MINIMUM SAMPLE SIZE AND OPTIMAL POSITIONING OF FLANKING MARKERS IN MARKER-ASSISTED BACKCROSSING FOR TRANSFER OF A TARGET GENE

MATTHIAS FRISCH, MARTIN BOHN, AND ALBRECHT E. MELCHINGER*

ABSTRACT. In recurrent backcrossing designed for introgression of a target allele from a donor into the genetic background of a recurrent parent (RP), molecular markers can accelerate recovery of the recurrent parent genome (RPG). The objectives of this study were to determine in marker-assisted backcrossing (MAB) (i) the optimum positioning of flanking markers (d_1, d_2) and (ii) the minimum number of individuals (n) required for obtaining with a certain probability a given number of individuals that carry the donor allele at the target locus and have a minimum proportion of donor genome on the carrier chromosome. Analytical solutions and tabulated results are given for relevant parameters (d_1, d_2, n) required to obtain, with a specified probability of success, at least one desired individual. They depend on the length of the carrier chromosome, the chromosomal position of the target locus, its distance to the flanking marker loci, and the number of individuals evaluated. Our approach can increase the efficiency of MAB by reducing the number of individuals and marker data points required.

Recurrent backcrossing is a breeding method commonly employed to transfer alleles at one or more loci from a donor to a recurrent parent (Allard, 1960). Examples include the transfer of resistance alleles from a wild or unimproved form into elite breeding materials and cultivars or the transfer of a target allele introduced by genetic transformation into a line that is easy to handle in tissue culture but otherwise of no agronomic value (Ragot et al., 1995). Besides transfer of the target allele(s), the main goal is to recover the RPG as completely and as quickly as possible.

Molecular markers are used in recurrent backcrossing for two purposes: (i) as a diagnostic tool for tracing the presence of a target allele, for which direct selection is difficult or impossible (e.g., recessive alleles expressed at a late stage in plant development or quantitative trait loci) and/or (ii) for identifying individuals with a low proportion of the undesirable genome from the donor parent. Adopting the terminology of Hospital and Charcosset (1997), we refer to the first approach as "foreground selection" (for review see Melchinger, 1990) and to the second approach as "background selection" (for review see Visscher et al., 1996). As demonstrated by Tanksley et al. (1989) with computer simulations, use of molecular markers for background selection can accelerate recovery of the RPG by two or three generations.

Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany. Received 31 March 1998. *Corrsponding Author (melchinger@uni-hohenheim.de).

Published in Crop Sci. 39:967–975 (1999), Erratum Crop Sci. 39:1903. This version of the paper is correct and does not contain the errors listed in the Erratum.

Background selection has two goals: (i) reduction of the proportion of the donor genome on the carrier chromosome of the target allele and (ii) reduction of the donor genome on the non-carrier chromosomes. The length of the chromosome segment from the donor that is linked to the target allele ("linkage drag") is reduced by selecting individuals that carry the target allele and are homozygous for the RP alleles at tightly linked marker loci. In practical implementations of MAB, two questions of crucial importance are: How should the flanking markers be positioned? How many individuals must be generated and genotyped with molecular markers in order to reduce the undesirable donor genome below a certain threshold?

Hospital et al. (1992) determined optimum distances d_1 and d_2 between the target locus and the flanking marker loci in order to recover a maximum amount of the RPG on the carrier chromosome by applying equation

$$d_1 = d_2 = \frac{1}{2} \ln \left(1 + 2\sqrt{s}\right), \tag{1}$$

where s is the proportion of selected BC1 individuals. This approach is based upon the assumption of an infinite population size. However, in practical breeding programs the number of individuals available for testing and genotyping is limited.

Hospital and Charcosset (1997) investigated marker-assisted introgression of quantitative trait loci (QTL) combining foreground and background selection. They presented recurrence equations and pointed out these could be used to calculate the minimum population size $n^{(t)}$ needed in each backcross generation BCt. Values for $n^{(t)}$ are determined numerically before starting the breeding program.

The present study focuses on background selection for flanking markers in combination with selection for a target allele based on phenotypic evaluation. Our objectives were to determine (i) the optimum positioning of flanking markers and (ii) the minimum number of individuals that have to be genotyped in order to obtain at least m desired individuals, which carry the target allele and have a maximum proportion of RPG on the carrier chromosome. We provide closed analytical solutions for important situations in backcrossing that can be easily applied by breeders.

Methods

We consider a chromosome of length L. Positions on the chromosome are represented by a scale (in Morgan units) ranging from 0 to L. The target locus is located at position x and two flanking markers at positions y_l and y_r (Fig. 1). Let d_1 and d_2 denote the lengths of the intervals between the flanking markers and the target locus ($]y_l, x[$ and $]x, y_r[$, respectively), and l_1 and l_2 the lengths of the intervals between the target locus and the ends of the chromosome ($]0, x[$ and $]x, L[$, respectively). Without loss of generality, we assume $d_1 \leq d_2$. Note that in all subsequent equations d_i and l_i are in Morgan units while specifications in the text and tables are in cM for the sake of convenience. Adopting the terminology of Hospital and Charcosset (1997), we denote by z^- the genotype of an individual homozygous for the RP allele and by z^+ the genotype of an individual heterozygous for the RP at the locus at position z.

Abbreviations: BC, backcross; BCt, t-th backcross generation; cM, centimorgan; MAB, marker-assisted backcrossing; NRP, non-recurrent parent; QTL, quantitative trait loci; RP, recurrent parent; RPG, recurrent parent genome; RFLP, restriction fragment length polymorphism.

Under the assumptions (a) the average number of crossovers formed on a chromatid is equal to its length in Morgan units and (b) the locations of crossovers are uniformly and independently distributed on the chromatid, the random variable K, counting the number k of crossovers formed on a chromatid or a chromatid segment of length l, follows a Poisson distribution with parameter l (Libermann and Karlin, 1984):

$$P\left(K=k\right)=\frac{l^k}{k!}\,e^{-l}\;. \tag{2}$$

Assumptions (a) and (b), which also underlie Haldane's (1919) mapping function, imply that neither chiasma interference nor chromatid interference (Stam, 1979) occurs. This assures the stochastic independance of crossover formation in adjacent chromosome segments.

The probability that an odd number of crossovers occur (i.e., recombination occurs) in an interval is the recombination frequency r related to the map distance by Haldane's (1919) mapping function:

$$p=r=\sum_{\nu=0}^{\infty}P\left(K_i=2\nu+1\right)=\sinh\left(d_i\right)e^{-d_i}=\frac{1}{2}\left(1-e^{-2d_i}\right). \tag{3}$$

We define the events A: No crossover occurs in $]0,y_l[$; B: Recombination occurs in $]y_l,x[$; C: Recombination occurs in $]x,y_r[$; and D: No crossover occurs in $]y_r,L[$. Applying Eq. (2) and (3), the respective probabilities are

$$p_A = e^{-(l_1-d_1)} \tag{4}$$
$$p_B = \sinh\left(d_1\right)e^{-d_1}=(1-e^{-2d_1})/2 \tag{5}$$
$$p_C = \sinh\left(d_2\right)e^{-d_2}=(1-e^{-2d_2})/2 \tag{6}$$
$$p_D = e^{-(l_2-d_2)}. \tag{7}$$

On the basis of the genotype at the target locus and the two flanking marker loci, different types of individuals are defined (Table 1):

Type 1: An individual is heterozygous for the donor allele at the target locus and homozygous for the RP allele at both flanking markers.

Type 2, 2L/2R: A Type 2 individual is heterozygous for the donor allele at the target locus and homozygous for the RP allele at one of the flanking markers. The second flanking marker is heterozygous for the RP allele. Depending on whether

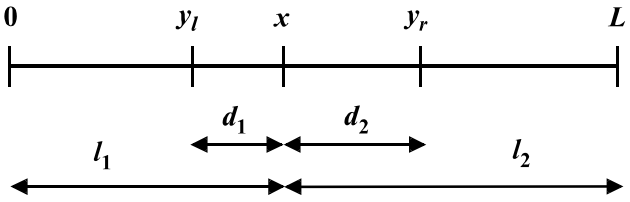

FIGURE 1. Chromosome of length L with target locus at position x and two flanking marker loci at positions y_l and y_r. l_1 and l_2 are the map distances between the target and the ends of the chromosomes, d_1 and d_2 are the map distances between the target locus and the flanking markers.

the flanking marker on the left or right hand side is fixed for the RP allele, we distinguish between individuals of Type 2L and Type 2R.

Type 3L/3R: An individual is heterozygous for the donor allele at the target locus and homozygous for the RP allele at one flanking marker, irrespective of the genotype at the other flanking marker. As before, we distinguish between Type 3L and Type 3R individuals, depending on which flanking marker is under consideration.

Type 4: An individual is heterozygous for the donor allele at the target locus and heterozygous for the RP allele at both flanking markers.

Type 5: An individual is homozygous for the RP allele at the target locus, i.e., it is not a carrier of the target allele.

Adding a further criterion, we define Type $1^\star$, $2^\star$, $2L^\star/2R^\star$, and $3L^\star/3R^\star$ as individuals of Type 1, 2, 2L/2R, and 3L/3R, respectively, which carry no chromosomal segments of the donor between the flanking marker(s) carrying homozygous the RP allele(s) and the respective end(s) of the chromosome (Table 1).

Probabilities $P(G|H)$ that a BC individual is of a certain Type (Event G), given that its non-recurrent parent (NRP) is of a specified genotype (Condition H) are presented in Table 1. While the conditional probabilities p_1, p_{2L}, p_{2R}, p_2, p_{3L}, p_{3R}, p_4, and p_5 are valid for any BC generation, $p_1^\star$, $p_{2L}^\star$, $p_{2R}^\star$, $p_2^\star$, $p_{3L}^\star$, and $p_{3R}^\star$ are valid only for generation BC1. In the following generations they are only exact under the condition that no recombination occurred on the carrier chromosome in previous BC generations. Otherwise, they are approximate and the exact probabilities are higher. The exact probabilities for generation BC$t+1$ in this case could be obtained by redefining p_A and p_D according to the observed genotype at additional markers on the carrier chromosome at the individual selected in generation BCt.

Concerning the carrier chromosome, an individual of Type $1^\star$ has by expectation the smallest proportion of donor genome; it can be regarded as the final product of a gene introgression program. As noticed by Young and Tanksley (1989), the probability that recombination occurs in small regions at both sides of the target locus in one BC generation is much lower than in two generations, when one recombination can occur in each. Therefore, the other types are defined in order to design breeding programs that reduce the donor genome on the carrier chromosome in successive BC generations.

If a particular genotype occurs with probability p, the number m of individuals of this type in a sample of size n is assumed to be binomially distributed:

$$P\left(Y=m\right)=\binom{n}{m}p^m\left(1-p\right)^{n-m}. \tag{8}$$

The probability q that in n genotypes (before performing any selection) at least one individual has the desired type is

$$q=P\left(Y>0\right)=1-P\left(Y=0\right)=1-\left(1-p\right)^n \tag{9}$$

and the probability q_m that there are at least m genotypes of the desired type is

$$\begin{aligned} q_m &= P\left(Y\geq m\right)=1-\sum_{i=0}^{m-1}P\left(Y=i\right) \\ &= \binom{n}{m}\left(1-p\right)^n\left(\frac{p}{1-p}\right)^m \operatorname{hg}\left([1,m-n],[1+m],\frac{p}{p-1}\right) \end{aligned} \tag{10}$$

TABLE 1. Definition of various Types of BC individuals depending on (i) the genotype at the target locus and flanking marker loci and (ii) on adjacent chromosome segments without crossovers. For a given genotype of the non-recurrent parent (NRP), conditional probabilities that BC individuals are of the specified Type are given.

	Event G:		Condition H:	Conditional
Type	Genotype	No cross-over in	NRP is of Genotype	probability $P(G\|H)$
1	$y_l^- x^+ y_r^-$	—	$y_l^+ x^+ y_r^+$	$p_1 = p_B p_C/2$
2L	$y_l^- x^+ y_r^+$	—	$y_l^+ x^+ y_r^+$	$p_{2L} = p_B(1-p_C)/2$
2R	$y_l^+ x^+ y_r^-$	—	$y_l^+ x^+ y_r^+$	$p_{2R} = (1-p_B)p_C/2$
2	2L or 2R			$p_2 = p_{2L} + p_{2R}$
3L	$y_l^- x^+$	—	$y_l^+ x^+$	$p_{3L} = p_B/2$
3R	$x^+ y_r^-$	—	$x^+ y_r^+$	$p_{3R} = p_C/2$
4	$y_l^+ x^+ y_r^+$	—	$y_l^+ x^+ y_r^+$	$p_4 = (1-p_B)(1-p_C)/2$
5	x^-	—	x^+	$p_5 = 1/2$
$1^\star$	$y_l^- x^+ y_r^-$	$]0, y_l[$ and $]y_r, L[$	$y_l^+ x^+ y_r^+$ †‡	$p_1^\star = p_A p_B p_C p_D/2$
$2L^\star$	$y_l^- x^+ y_r^+$	$]0, y_l[$	$y_l^+ x^+ y_r^+$ †	$p_{2L}^\star = p_A p_B(1-p_C p_D)/2$
$2R^\star$	$y_l^+ x^+ y_r^-$	$]y_r, L[$	$y_l^+ x^+ y_r^+$ ‡	$p_{2R}^\star = (1-p_A p_B)p_C p_D/2$
$2^\star$	$2L^\star$ or $2R^\star$			$p_2^\star = p_{2L}^\star + p_{2R}^\star$
$3L^\star$	$y_l^- x^+$	$]0, y_l[$	$y_l^+ x^+$ †	$p_{3L}^\star = p_A p_B/2$
$3R^\star$	$x^+ y_r^-$	$]y_r, L[$	$x^+ y_r^+$ ‡	$p_{3R}^\star = p_C p_D/2$

† and no crossovers in $]0, y_l[$ in previous BC generations.
‡ and no crossovers in $]y_r, L[$ in previous BC generations.

where 'hg' is the generalized hypergeometric function implemented in common calculation software such as Maple (Char et al., 1991).

In order to determine the dimensioning of generation BC$t+1$ based on the marker genotype of the selected individual in generation BCt, a combination of the conditional probabilities (Table 1) with Eq. (9) and (10) can be used. The parameters d_1, d_2, or n can be determined in such a way that at least one individual of a given Type is generated with probability q by combining Eq. (9) with the corresponding conditional probabilities. A more general result is obtained by inserting the corresponding conditional probability instead of p into Eq. (10). This can be used to design a MAB breeding program such that at least m individuals of a certain type are produced with probability q_m, but the equations for determining the parameters d_1, d_2, or n must be solved numerically for $m > 1$.

Optimum Positioning of Flanking Marker Loci

The minimum marker distance d_i required to obtain under Condition H (Table 1) with probability q at least one BC individual of Type 1, 2, or 3L/3R in a sample of n is derived by combining the respective conditional probabilities (Table 1) with

TABLE 2. Minimum marker distance [cM] to obtain with probability $q = 0.90$, 0.95, and 0.99 at least one individual of Type 1 or 3L (for definition see text), if n individuals are assayed in generation BC$t+1$. Assumptions: The non-recurrent parent has genotype $y_l^+x^+y_r^+$ for Type 1 individuals and genotype $y_l^+x^+$ for Type 3L individuals; Type 1: $d_1 = d_2$.

	Number n of BC individuals								
Type	50	75	100	125	150	175	200	300	500
	Minimum marker distance d_1 [cM]								
Probability $q = 0.90$*:*									
1	46	34	28	25	22	20	19	15	11
3L	10	7	5	4	4	3	3	2	1
Probability $q = 0.95$*:*									
1	58	42	34	29	26	23	22	17	13
3L	14	9	7	5	5	4	4	3	2
Probability $q = 0.99$*:*									
1	92	59	46	39	34	31	28	22	16
3L	22	14	10	8	7	6	5	4	2

Eq. (9). Solving for d_1 yields for Type 1 (assuming $d_1 = d_2$)

$$d_1 = -\frac{1}{2}\ln\left(1 - 2\sqrt{2 - 2\sqrt[n]{1-q}}\right), \tag{11}$$

for Type 2 (assuming $d_1 = d_2$)

$$d_1 = -\frac{1}{4}\ln\left(4\sqrt[n]{1-q} - 3\right), \tag{12}$$

and for Type 3L ($i = 1$) and Type 3R ($i = 2$)

$$d_i = -\frac{1}{2}\ln\left(4\sqrt[n]{1-q} - 3\right). \tag{13}$$

On the basis of Eq. (11) and (13), the minimum marker distances were calculated for population sizes n ranging from 50 to 500 and probabilities $q = 0.90$, 0.95, and 0.99 (Table 2). The values for Type 2 individuals (Eq. (12)) were not tabulated because they are half the values for Type 3L individuals. The minimum marker distance decreases with increasing number of genotyped individuals. In addition, the marker distance increases with increasing probability of q for any of the three types of individuals. The values range from tightly linked markers (2 cM to obtain with probability $q = 0.99$ at least one Type 3L individual in a backcross population of 500 individuals) to nearly unlinked flanking markers (46 cM to obtain with probability $q = 0.99$ at least one Type 1 individual in a sample of 100 BC1 individuals).

The minimum distance d_i required to warrant with probability q the occurrence of at least one individual of Type $1^\star$, $2^\star$, or $3L^\star/3R^\star$ in a sample of n individuals is derived by combining the respective conditional probabilities (Table 1) with Eq. (9).

TABLE 3. Minimum marker distance [cM] to obtain with probability $q = 0.99$ at least one individual of Type $2^\star$ or $3L^\star$ (for definition see text), if n individuals are assayed in generation BC$t+1$. Assumptions: No recombination event must have occurred on the carrier chromosome in the non-recurrent parent up to generation BCt; Type $2^\star$: the target locus is positioned in the center of the chromosome ($l_1 = l_2$) and the marker bracket is symmetric ($d_1 = d_2$).

l_1	Number n of BC individuals								
[cM]	50	75	100	125	150	175	200	300	500
	Minimum marker distance d_1 [cM]								
Type 2*:									
25	13	9	7	5	5	4	3	2	2
50	17	11	8	7	6	5	4	3	2
75	21	14	10	8	7	6	5	4	2
100	27	18	13	11	9	8	7	5	3
125	34	23	17	14	11	10	9	6	4
150	43	29	22	17	14	12	11	7	5
175	54	36	27	22	18	16	14	9	6
200	67	46	35	28	23	20	18	12	7
Type 3L*:									
25	23	16	12	10	8	7	6	4	3
50	29	20	15	12	10	9	8	6	4
75	37	25	19	16	13	11	10	7	4
100	47	32	25	20	17	15	13	9	5
125	59	41	31	25	21	19	16	11	7
150	73	52	40	32	27	24	21	14	9
175	90	65	50	41	34	30	26	18	11
200	108	80	63	52	44	38	34	23	14

Solving for d_i yields for Type $1^\star$ (assuming $d_1 = d_2$)

$$d_1 = \text{Arsinh}\left(e^{\frac{l}{2}}\sqrt{2 - 2\sqrt[n]{1-q}}\right), \tag{14}$$

for Type $2^\star$ (assuming $d_1 = d_2$)

$$d_1 = \text{Arsinh}\,\frac{1}{4}\left(e^{l_1} + e^{l_2} - \sqrt{e^{2l_1} + e^{2l_2} + e^{l_1+l_2}\left(16\sqrt[n]{1-q} - 14\right)}\right), \tag{15}$$

which simplifies for $l_1 = l_2$ to

$$d_1 = \text{Arsinh}\left(\frac{1}{2}e^{l_1}\left[1 - \sqrt{4\sqrt[n]{1-q} - 3}\right]\right), \tag{16}$$

and for Type $3L^\star$ ($i = 1$) and Type $3R^\star$ ($i = 2$)

$$d_i = \text{Arsinh}\left(2e^{l_i}\left[1 - \sqrt[n]{1-q}\right]\right), \tag{17}$$

where 'Arsinh' is the inverse of the hyperbolic sine. Note that for $l_1 = d_1$ and $l_2 = d_2$, Eq. (14) to (17) simplify to Eq. (11) to (13) as expected.

Minimum marker distances to obtain at least one individual of Type $2^\star$ or Type $3L^\star$ with probability $q = 0.99$ were calculated for chromosome segments l_1 of varying length (between 25 and 200 cM) and for population sizes n ranging from 50 to 500 individuals (Table 3). For Type $2^\star$ individuals, a symmetric marker bracket ($d_1 = d_2$) and a target locus in the middle of the chromosome ($l_1 = l_2$) was assumed. The minimum marker distances increase with increasing length of the respective chromosome arm and the values vary from tightly linked markers to nearly unlinked markers. For Type $1^\star$ individuals, large numbers of individuals are required (data not shown). For example, for a chromosome of length 125 cM and $n = 300$, the distance of the flanking markers $d_1 = d_2$ has to be larger than 33 cM in order to obtain with probability $q = 0.99$ at least one individual of Type $1^\star$.

Minimum Sample Size Required in MAB

Suppose flanking markers are positioned at distances d_1 and d_2 from the target locus and at least one individual of a certain Type should be recovered with probability q. Given the NRP satisfies Condition H (Table 1), the minimum sample size n needed to achieve this goal with probability q is $n = \ln(1-q)/\ln(1-p)$, were p equals the corresponding conditional probability $P(G|H)$ given in Table 1.

Hence, for getting at least one desired BC individual with probability q, the minimum sample size is for Type 1

$$n = \frac{\ln(1-q)}{\ln\left(1 - \frac{1}{8}\left[1 - e^{-2d_1}\right]\left[1 - e^{-2d_2}\right]\right)}\,, \tag{18}$$

for Type 2

$$n = \frac{\ln(1-q)}{\ln\left(\frac{3}{4} + \frac{1}{4}e^{-2(d_1+d_2)}\right)}\,, \tag{19}$$

and for Type 3L ($i = 1$) and Type 3R ($i = 2$)

$$n = \frac{\ln(1-q)}{\ln\left(\frac{3}{4} + \frac{1}{4}e^{-2d_i}\right)}\,. \tag{20}$$

Minimum sample sizes were calculated according to Eq. (18) to (20) for marker distances d_1 and d_2 ranging from 5 to 50 cM, and probability $q = 0.99$ (Table 4). Within marker brackets of 5 to 50 cM, the required n for Type 1 individuals ranges from 4066 to 90. The required n for Type 2 individuals varies between 100 and 20 and that for Type 3L/3R individuals ranges between 192 and 27 individuals. For tightly linked flanking markers, n for Type 2 individuals is about half the sample size required for Type 3L individuals. For large marker brackets, the values of n for Type 2 and Type 3L individuals are almost equal.

The number of individuals required to reduce the length of the chromosome segment linked to the target allele below a fixed threshold ($g = d_1 + d_2$) for Type 1 individuals is minimized, if d_1 and d_2 are of equal size (Table 4). For example for a marker bracket of length $g = 40$ cM, $n = 359$ individuals are required if $d_1 = 15$ cM and $d_2 = 25$ cM, whereas $n = 337$ individuals are required if $d_1 = d_2 = 20$ cM. A

TABLE 4. Minimum number of individuals n required to obtain with probability $q = 0.99$ at least one individual of Type 1, 2, or 3L (for definition see text), if n individuals are assayed in generation BC$t + 1$. Assumptions: The non-recurrent parent has genotype $y_l^+ x^+ y_r^+$ for Type 1 and 2 individuals and genotype $y_l^+ x^+$ for Type 3L individuals.

d_2	d_1 [cM]								
[cM]	5	10	15	20	25	30	35	40	50
	Minimum number of individuals n								
Type 1:									
5	4066	2134	1492	1172	982	856	767	701	611
10		1119	782	615	515	449	402	367	320
15			547	429	359	313	281	256	233
20				337	282	246	220	201	175
25					236	206	184	168	146
30						179	160	146	127
35							144	131	114
40								120	104
50									90
Type 2:									
5	100	69	54	45	39	35	32	29	26
10		54	45	39	35	32	29	27	24
15			39	35	32	29	27	26	23
20				32	29	27	26	24	23
25					27	26	24	23	22
30						24	23	23	21
45							23	23	21
40								21	20
50									20
Type 3L:									
	192	100	69	54	45	39	35	32	27

Type 2 individual occurs if a cross-over is formed between the two flanking marker loci, no matter whether it is formed in $]y_l, x[$ or $]x, y_r[$. Hence, in contrast to Type 1 individuals, a symmetric placement of the markers has no influence on the required minimum sample size for Type 2 individuals.

The minimum number of individuals needed to obtain with probability q at least one BC individual is for Type 1*

$$n = \frac{\ln(1-q)}{\ln\left(1 - \frac{1}{2}\sinh(d_1)\sinh(d_2)\,e^{-l}\right)}\,, \tag{21}$$

for Type 2*

$$n = \frac{\ln(1-q)}{\ln\left(1 - \frac{1}{2}\sinh(d_1)\,e^{-l_1} - \frac{1}{2}\sinh(d_2)\,e^{-l_2} + \sinh(d_1)\sinh(d_2)\,e^{-l}\right)}\,, \tag{22}$$

TABLE 5. Minimum number of individuals required to obtain with probability $q = 0.99$ at least one individual of Type $2^\star$ or $3L^\star$ (for definition see text), if n individuals in generation BC$t+1$ are assayed. Assumptions: No recombination event occurred on the carrier chromosome in the non-recurrent parent up to generation BCt; Type $2^\star$: the target locus is positioned in the center of the chromosome ($l_1 = l_2$) and the marker bracket is symmetric ($d_1 = d_2$).

l_1	d_1 [cM]								
[cM]	5	10	15	20	25	30	35	40	50
	Minimum number of individuals n								
Type 2*:									
25	121	62	43	33	-†	-	-	-	-
50	155	79	54	41	34	29	25	23	-
75	198	100	68	52	42	36	31	28	23
100	253	128	86	65	53	44	38	34	28
125	324	163	110	83	67	56	48	43	34
150	415	209	140	106	85	71	61	53	43
175	533	267	179	135	108	90	77	68	54
200	683	343	229	172	138	115	98	86	68
Type 3L*:									
25	235	116	77	57	-†	-	-	-	-
50	302	150	99	74	58	48	41	35	-
75	388	193	128	95	75	62	53	46	36
100	499	248	164	123	97	80	68	59	46
125	641	319	212	158	125	104	88	76	60
150	823	410	272	203	162	134	114	99	77
175	1058	527	350	261	208	172	147	127	100
200	1359	678	450	336	268	222	189	164	129

† Not possible because $d_1 \leq l_1$.

and for Type $3L^\star$ ($i = 1$) and Type $3R^\star$ ($i = 2$)

$$n = \frac{\ln(1-q)}{\ln\left(1 - \frac{1}{2}\sinh(d_i)\, e^{-l_i}\right)}\,. \tag{23}$$

The minimum sample sizes for getting at least one individual of Type $2^\star$ or $3L^\star$ with probability $q = 0.99$ were calculated for values of l_1 and l_2 varying from 25 to 200 cM and flanking marker distances d_1 and d_2 from 5 to 50 cM (Table 5). For Type $2^\star$ individuals, equal marker distances ($d_1 = d_2$) and a target locus in the middle of the chromosome ($l_1 = l_2$) were assumed. The sample sizes for a marker distance of 10 cM range from 62 to 343 individuals for Type $2^\star$ and from 116 to 678 individuals for Type $3L^\star$. In all cases tabulated, the n values for Type $2^\star$ individuals were about half the corresponding n values required for Type $3L^\star$ individuals. The n values required for obtaining at least one individual of Type $1^\star$ with probability

q are generally large (data not shown). For example, for a chromosome of length 100 cM at least 616 individuals are required if $d_1 = d_2 = 20$ cM.

Design of a Two-Generation MAB Program

We consider here dimensioning a two-generation MAB program that has the goal to produce with probability $q_1^{(2)}$ at least one individual of Type 1 in generation BC2. Flanking markers are assumed to be located at distances d_1 and d_2 with $d_1 \leq d_2$.

From generation BC1, one individual of the most desirable Type is selected in the given order: Type 1 $\succ$ Type 2L $\succ$ Type 2R $\succ$ Type 4. If in generation BC1 more than one individual satisfying the strongest condition is found, selection between them can be performed based on analysis of further marker loci (located either on the carrier or on non-carrier chromosomes) in order to determine the most desirable individual for producing generation BC2. Similar selection schemes were proposed by various authors (e.g., Tanksley et al., 1989; Hospital and Charcosset, 1997), but in the later study no distinction was made between Type 2L and Type 2R individuals, even if $d_1 \neq d_2$. Provided the (very unlikely) case that none of the BC1 individuals carries the target allele, the BC program failed in BC1.

Assume we used $n^{(1)}$ individuals in BC1 and selected one as described above, then the probabilities $z_i^{(1)}$ that the selected BC1 individual is of Type i ($i \in \{1, 2L, 2R, 4\}$) are:

$$z_1^{(1)} = 1 - (1 - p_1)^{n^{(1)}} \tag{24}$$

$$z_{2L}^{(1)} = (1 - p_1)^{n^{(1)}} - (1 - p_1 - p_{2L})^{n^{(1)}} \tag{25}$$

$$z_{2R}^{(1)} = (1 - p_1 - p_{2L})^{n^{(1)}} - (1 - p_1 - p_2)^{n^{(1)}} \tag{26}$$

$$z_4^{(1)} = (1 - p_1 - p_2)^{n^{(1)}} - 1/2^{n^{(1)}} \tag{27}$$

The probability that the target allele is lost in BC1 ($i = 5$) is:

$$z_5^{(1)} = 1/2^{n^{(1)}} \tag{28}$$

The probabilities (24) to (28) sum up to 1 and cover all possible results for generation BC1.

Depending on the Type i of the selected individual in BC1, the population size $n_i^{(2)}$ is chosen such that at least one BC2 individual of Type 1 is generated with probability $q_1^{(2)}$. The respective mimimum population sizes are (from Eq. (9)):

$$n_1^{(2)} = \ln(1 - q_1^{(2)}) / \ln(1/2) \tag{29}$$

$$n_{2L}^{(2)} = \ln(1 - q_1^{(2)}) / \ln(1 - p_{3R}) \tag{30}$$

$$n_{2R}^{(2)} = \ln(1 - q_1^{(2)}) / \ln(1 - p_{3L}) \tag{31}$$

$$n_4^{(2)} = \ln(1 - q_1^{(2)}) / \ln(1 - p_1) \tag{32}$$

Irrespective of the Type i of the selected BC1 individual, choice of $n_i^{(2)}$ according to Eq. (29) to (32) assures that with probability $q_1^{(2)}$ at least one individual of Type 1 is produced in generation BC2. Hence, the probability of success of the entire MAB program is $(1 - z_5^{(1)})q_1^{(2)}$. Note that $z_5^{(1)} \approx 0$ for values of $n^{(1)}$ typically used in a BC program.

For choosing a value for $n^{(1)}$ we propose two methods. First, $n^{(1)}$ can be determined such that at least one Type 2 individual is generated in BC1 with a given probability, e.g., $q_2^{(1)} = 0.99$. This procedure has the advantage of a simple calculation by using Eq. (19) but is not optimizing the procedure.

The second method is to choose $n^{(1)}$ such that at the expected number of individuals required for the two-generation program is minimized. The number of individuals required in BC2 depends on the choice of $n^{(1)}$. Its expectation is

$$E(n^{(2)}) = \sum_{i \in \{1,2L,2R,4\}} z_i^{(1)}\, n_i^{(2)} \tag{33}$$

and the expectation for the total number of individuals required in BC1 and BC2 is

$$E(n^{(1)} + n^{(2)}) = n^{(1)} + E(n^{(2)}) \tag{34}$$

Partial differentiation with regard to $n^{(1)}$ yields

$$\begin{aligned}
\frac{\partial E(n^{(1)} + n^{(2)})}{\partial n^{(1)}} =& 1 + \frac{\ln(1 - q_1^{(2)})\,(1 - p_1)^{n^{(1)}} \ln(1 - p_1)}{\ln(2)} \\
&+ \frac{\ln(1 - q_1^{(2)})\,(1 - p_1)^{n^{(1)}} \ln(1 - p_1)}{\ln(1 - p_{3R})} \\
&- \frac{\ln(1 - q_1^{(2)})\,(1 - p_1 - p_{2L})^{n^{(1)}} \ln(1 - p_1 - p_{2L})}{\ln(1 - p_{3R})} \\
&+ \frac{\ln(1 - q_1^{(2)})\,(1 - p_1 - p_{2L})^{n^{(1)}} \ln(1 - p_1 - p_{2L})}{\ln(1 - p_{3L})} \\
&- \frac{\ln(1 - q_1^{(2)})\,(1 - p_1 - p_2)^{n^{(1)}} \ln(1 - p_1 - p_2)}{\ln(1 - p_{3L})} \\
&+ \frac{\ln(1 - q_1^{(2)})\,(1 - p_1 - p_2)^{n^{(1)}} \ln(1 - p_1 - p_2)}{\ln(1 - p_1)} \\
&+ \frac{\ln(1 - q_1^{(2)})\,(1/2)^{n^{(1)}} \ln(2)}{\ln(1 - p_1)}
\end{aligned} \tag{35}$$

Equating Eq. (35) to zero and solving for $n^{(1)}$ yields the extrema of Eq. (34). By considering the second partial derivation or by calculating $E(n^{(1)} + n^{(2)})$ at the extrema, a value for $n^{(1)}$ that minimizes the expected total number of individuals can be found.

A third possibility would be to choose $n^{(1)}$ in order to minimize the required number of marker analyses in BC1 and BC2 $E(a^{(1)} + a^{(2)})$ by partial differentiation of

$$E(a^{(1)} + a^{(2)}) = 2n^{(1)} + z_{2L}^{(1)}\, n_{2L}^{(2)} + z_{2R}^{(1)}\, n_{2R}^{(2)} + 2z_4^{(1)}\, n_4^{(2)} \tag{36}$$

with regard to $n^{(1)}$ and equating the result to zero. In this calculation, only marker analyses for marker loci flanking the target locus are taken into account.

Generalizations

We now consider the general description of a MAB program in which selection of individuals is performed based on the ranking described above: Type 1 $\succ$

Type 2L $\succ$ Type 2R $\succ$ Type 4. If the target allele is lost in any generation, the BC programm is considered as not successful. Let the vector

$$\mathbf{z}^{(t+1)} = \left[z_1^{(t+1)}, z_{2L}^{(t+1)}, z_{2R}^{(t+1)}, z_4^{(t+1)}, z_5^{(t+1)}\right]' \tag{37}$$

consist of the probabilities that an individual of Type 1 to 4 is selected in generation $t+1$ and the probability that the BC program failed in generation BC$t+1$ due loss of the target allele. The probabilities $\mathbf{z}^{(t+1)}$ can be caluclated as

$$\mathbf{z}^{(t+1)} = \mathbf{P}\mathbf{z}^{(t)}\,, \tag{38}$$

where the transition matrix $\mathbf{P}$ is defined as

$$\mathbf{P} = \begin{bmatrix} 1-(\frac{1}{2})^{n_1} & 1-(1-p_{3R})^{n_{2L}} & 1-(1-p_{3L})^{n_{2R}} & 1-(1-p_1)^{n_4} & 0 \\ 0 & (1-p_{3R})^{n_{2L}} - (\frac{1}{2})^{n_{2L}} & 0 & (1-p_1)^{n_4} - (1-p_1-p_{2L})^{n_4} & 0 \\ 0 & 0 & (1-p_{3L})^{n_{2R}} - (\frac{1}{2})^{n_{2R}} & (1-p_1-p_{2L})^{n_4} - (1-p_1-p_2)^{n_4} & 0 \\ 0 & 0 & 0 & (1-p_1-p_2)^{n_4} - (\frac{1}{2})^{n_4} & 0 \\ (\frac{1}{2})^{n_1} & (\frac{1}{2})^{n_{2L}} & (\frac{1}{2})^{n_{2R}} & (\frac{1}{2})^{n_4} & 1 \end{bmatrix} \tag{39}$$

The values for n_1 to n_4 for the final BC generation can be derived from Eq. (9), as shown in Eq. (29) to (32). For generation BC0 (=F_1), the initial values are $\mathbf{z}^{(0)} = (0,0,0,1,0)'$.

Modifications of $\mathbf{P}$ can be applied to alternative situations. For example, if the target allele got lost in generation BC$t+1$, the breeder usually backs up one generation by using either remnant seed or another Type i individual to produce generation BC$t+1$ anew with n_{i^*} individuals. This procedure corresponds to substituting column five by the same expressions as in the column with power n_i but using population size n_{i^*}.

Given that n_4 determined by Eq. (32) is generally a high number, a MAB program over $t+1$ generations must be regarded as failed, if only a Type 4 individual can be selected in generaton BCt. A transition matrix taking this into account is obtained by replacing the first element of the fourth column by 0 and by replacing in the other elements of this column n_4 by the maximal number of individuals ($n_{\max}$) that can be handled.

In conclusion, applying Eq. (38) with the original or a modified transition matrix allows the dimensioning of a wide range of MAB programs consisting of one or more generations. In particular, previously described results (Eq. (24) to (28)) can be obtained as special cases of Eq. (38).

When comparing Eq. (38) with the transition probabilities presented by Hospital and Charcosset (1997) in their Eq. (A.16) to (A.19) for QTL introgression, two main differences are: (1) We recommend to use variable population sizes $n_1^{(t)}, \ldots, n_4^{(t)}$ that depend on the Type of the selected individual in generation BCt. In contrast, Hospital and Charcosset (1997) suggest a fixed population size $n_1^{(t)} = n_{2L}^{(t)} = n_{2R}^{(t)} = n_4^{(t)}$.

(2) We prefer Type 2L over Type 2R individuals, while Hospital and Charcosset take one at random if no Type 1 but several Type 2 individuals are present (even if $d_1 < d_2$).

Discussion

The distances d_1 and d_2 between the flanking markers and the target locus are key parameters in MAB. Hospital et al. (1992) determined d_1 and d_2 in order to recover a maximum amount of the RPG on the carrier chromosome. Their calculations were based on the assumptions of a target locus located in the middle of a chromosome and an infinite sample size. However, if markers are spaced according to the rule (see Eq. (1)) described by these authors, MAB with sample sizes typically used in practical breeding programs has little chance of success in one BC generation. Considering $n = 100$ individuals from which one has to be selected ($s = 0.01$), the optimum distance of two flanking markers is $d_1 = d_2 = 9$ cM, based on Eq. (1). For this marker distance the probability that in 100 BC1 individuals, there is at least one of Type 1, is only $q = 0.29$ (Eq. (9)). Instead of determining d_1 and d_2 to maximize the percentage of the RPG on the carrier chromosome, we propose to choose d_1, d_2, and n according to our Eq. (11) to (17) such that with probability q, there will be at least one individual of the desired Type.

Hospital and Charcosset (1997) presented an approach for determining the minimum sample size in QTL introgression programs under both marker-assisted foreground and background selection over an arbitrary number of BC generations. They gave recurrence equations in terms of recombination frequencies that can be used to obtain numerical solutions for the population sizes $n^{(t)}$ required in BCt in order to select with probability $z_i^{(w)}$ an individual of Type i ($i \in \{1, 2L, 2R, 4\}$) in generation BCw. Considering the special case of one QTL with known map position, their Eq. (A.16) to (A.18) could be applied to determine population sizes $n^{(1)}, \ldots, n^{(w)}$ for obtaining at least one Type 1 individual in generation BCw. Since these values are determinded 'a priori' (before starting the breeding program), $n^{(t+1)}$ is applied in generation BC$t+1$ irrespective of the observed marker genotype of the selected individual in generation BCt. While this procedure assures that with probability $z_1^{(w)}$ a least one individual of Type 1 is generated up to generation BCw, it has the following consequences, illustrated here for $w = 2$: (1) If there neither a Type 1 individual nor a Type 2 individual is found in BC1, the actual probability of success $q_{1,4}^{(2)}$ with $n^{(2)}$ in BC2 is lower than $z_1^{(2)}$ ($q_{i,j}^{(t)}$ is the probability that at least one individual of Type i is produced in generation t under the condition that the NRP is of Type j; $i, j \in \{1, 2L, 2R, 4\}$). (2) If the selected individual in BC1 is either of Type 1 or Type 2, then using $n^{(2)}$ results in a probability of success $q_{1,1}^{(2)}$, $q_{1,2L}^{(2)}$ or $q_{1,2R}^{(2)}$ in BC2 higher than $z_1^{(2)}$ (and consequently $n^{(2)}$ is higher than required for having success with probability $q_1^{(2)}$). This is demonstrated numerically in Example 3. A general proof of this proposition follows directly from the Theorem of Total Probability.

In contrast, we recommend for MAB over several BC generations a sequential approach, in which calculation of $n^{(t+1)}$ depends on the observed marker genotype of the individual selected in generation BCt by using Eq. (29) to (32). For each possible Type i of the selected individual, $n_i^{(t+1)}$ is calculated such that at least one individual of Type j ($i, j \in \{1, 2L, 2R, 4\}$) is generated with probability $q_j^{(t+1)}$

in generation BC$t+1$. This ensures that (i) the actual probability of success in BC$t+1$ is always $q_j^{(t+1)}$, even if in BCt no flanking marker was fixed, and (ii) only the number of individuals actually required to reach a given $q_j^{(t+1)}$ in BC$t+1$ are generated, if already one or two flanking markers are fixed in BCt.

In order to determine the dimensioning of a MAB program, we propose first to check whether, depending on the markers and the population size available, an individual of Type 1 or $1^\star$ can be generated in BC1 with a given probability $q_1^{(1)}$ or $q_{1^\star}^{(1)}$, respectively. Provided this is possible and economical with the available resources, BC1 is dimensioned accordingly. If not, a two-generation BC program is designed as described above. The proposed procedure is illustrated by the following three Examples.

Example 1: With probability $q_1^{(1)} = 0.99$ at least one Type 1 individual should be generated in BC1. Due to practical limitations, the maximum population size per generation that can be handled by the breeder is $n^{(1)} = 150$. According to Table 2, the minimum flanking markers distances are $d_1 = d_2 = 34$ cM. Note that the expected value of the donor segments in the interval $[y_r, y_l]$ is $(d_1 + d_2)/2$ and $d_1 + d_2$ is only the maximum.

Example 2: With probability $q_1^{(1)} = 0.99$ at least one Type 1 individual should be generated in BC1. On the basis of prior linkage information, it is known that the map distances between the target locus and its two flanking markers are $d_1 = 20$ cM and $d_2 = 25$ cM. According to Table 4, at least $n^{(1)} = 282$ BC1 individuals have to be produced.

Example 3: The length of the linkage drag $d_1 + d_2$ should be reduced with probability $q_1^{(2)} = 0.99$ below a threshold of 10 cM ($d_1 = d_2 = 5$ cM) in a two-generation BC program. Numerical results for three alternative strategies are given in Table 6. Strategy A was proposed by Hospital and Charcosset (1997) and employs a fixed population size in generation BC$t+1$, irrespective of the type of the selected individual in generation BCt. According to their Table 6 (for $m = 1$, $S = 10$, $S^\star = 10$), choosing $n^{(1)} = 118$ and $n^{(2)} = 200$ assures $z_1^{(2)} = 0.99$ for a Type 1 individual. In strategy B, $n^{(1)} = 100$ is chosen to generate at least one BC1 individual of Type 2 with probability $q_2^{(1)} = 0.99$ (Eq. (19)). Depending on the genotype of the selected individual in BC1, $n_1^{(2)} = 7$, $n_{2L}^{(2)} = n_{2R}^{(2)} = 192$ or $n_4^{(2)} = 4066$ is chosen in BC2 (based on Eq. (29) to (32)), to warrant that a Type 1 individual in generation BC2 is found with probability $q_1^{(2)} = 0.99$. In Strategy C, $n^{(1)} = 114$ is chosen to minimize the expected total number of individual assayed in BC1 and BC2 by equating Eq. (35) to zero. For generation BC2, $n_1^{(2)}, \ldots, n_4^{(2)}$ are chosen as in Strategy B in order to assure that a Type 1 individual is found with probability $q_1^{(2)} = 0.99$. The results demonstrate that taking into consideration the genotype of the selected individual in BCt for dimensioning BC$t+1$ is superior over Strategy A with regard to (a) the expected total number of indivduals in the breeding program and (b) warranting a constant probability of success for all possible genotypes that might be selected in BC1.

The difference between the required individuals for Strategies A and C increases substantially for asymmetric marker brackets. Assume that in the above example $d_1 = 5$ cM and $d_2 = 15$ cM. According to Eq. (35), choosing $n^{(1)} = 101$ minimizes

the expected total number of individuals and yields $E(n^{(2)}) = 76$ for Strategy C. For $n^{(1)} = 101$, Strategy A would require $n^{(2)} = 122$ individuals.

Our approach neglects the effects of interference. In the case of negative chiasma interference (Stam, 1979), multiple crossovers in a chromosome segment occur less frequently than expected in the absence of interference. This results in an overestimation of the probabilities for the occurrence of Type 1 and Type $1^\star$ individuals. Hence, the values for the minimum sample sizes and minimum marker distances obtained by our equations and presented in the tables are underestimated when negative interference exists. For positive chiasma interference the situation is reverse. To obtain individuals of Type 2, $2^\star$, 3L/3R and $3L^\star/3R^\star$ only one crossover is required in a small chromosome region adjacent to the target locus. Hence, in these cases interference has only a minor effect on the calculated parameters.

Modelling interference by relating recombination frequencies and map distance via a mapping function that takes interference into account is not possible. Calculation of probabilities of joint events as products of probabilites of simple events

TABLE 6. Numerical results of three alternative approaches for dimensioning a two-generation BC program designed to obtain at least one Type 1 individual ($d_1 = d_2 = 5$ cM) in generation BC2 with probability $q_1^{(2)} = 0.99$. For definition of symbols see text.

Parameter	Strategy A: Hospital and Charcosset (1997)	Strategy B: $q_2^{(1)} \geq 0.99$	Strategy C: Minimize $E(n^{(1)} + n^{(2)})$
$n^{(1)}$	118	100	114
$z_1^{(1)}$	0.1251	0.1071	0.1211
$z_{2L}^{(1)}$	0.4356	0.8029	0.8146
$z_{2R}^{(1)}$	0.4356	0.0814	0.0598
$z_4^{(1)}$	0.0037	0.0086	0.0044
$z_5^{(1)}$	2^{-118}	2^{-100}	2^{-114}
$n_1^{(2)}$	200	7	7
$n_{2L}^{(2)}$	200	192	192
$n_{2R}^{(2)}$	200	192	192
$n_4^{(2)}$	200	4066	4066
$q_{1,1}^{(2)}$	$1 - 2^{-200}$	0.9900	0.9900
$q_{1,2R}^{(2)}$	0.9999	0.9900	0.9900
$q_{1,2L}^{(2)}$	0.9999	0.9900	0.9900
$q_{1,4}^{(2)}$	0.2027	0.9900	0.9900
$E(n^{(1)} + n^{(2)})$	318	306	300

(as needed for all defined Types of individuals except those of Type 3L/3R, 4, and 5) requires the stochastic independence of recombination events in two adjacent intervals. This is ensured by Haldane's mapping function but not by other commonly employed mapping functions, which take interference into account (such as Kosambi, Morgan, Carter-Falconer, or Felsenstein).

Our approach emphasizes recovery of the RPG on the carrier chromosome. This is required when a genotype should be completely converted like in the examples described above. A total conversion is only necessary when the agricultural properties of the donor are poor (Tanksley et al., 1989). If the target allele is already introduced into elite lines, a partially converted genotype may be superior to a totally converted one, because the donor segments may contain favorable alleles (Lee, 1995) and a strict reduction of the linkage drag may not be nessecary. Background selection on the non-carrier chromosomes is not considered in this treatise. Results on the dimensioning of BC programs, comparing the effect of various selection strategies on the recovery of the RPG on the carrier and non-carrier chromosomes are presented in a companion paper (Frisch et al., 1999).

Comparison with a practical example. Ragot et al. (1995) demonstrated that MAB can be efficiently used for introgressing a transgene construct, containing the *Bt*-gene, from a transformed parent into an elite maize inbred. On the basis of the RFLP genotype, plants showing a maximum number of RP marker alleles were selected. A total of 61 RFLP markers were used for genotyping each of the BC1 individuals. Following the approach proposed in this study, out of the 15 markers evaluated on the carrier chromosome, only two flanking markers could have been used for the detection of Type 1 individuals. Selection among the detected Type 1 individuals could have been perfomed using the additional markers in order to detect putative Type $1^\star$ individuals and select among them. This illustrates that our approach can substantially increase the efficiency of MAB by reducing the total number of marker data points required.

Acknowledgments. The financial support from fellowships by Pioneer Hi-Bred Intl. Inc., USA, and KWS Kleinwanzlebener Saatzucht AG, Germany, to M. Frisch ixs gratefully acknowledged. We greatly appreciate the helpful suggestions of three anonymous reviewers for improving the quality of this paper.

References

Allard, R.W. 1960. Principles of Plant Breeding. John Wiley and Sons, Inc., New York.

Char, B.W., K.O. Geddes, G.H. Gonnet, B.L. Leong, M.B. Monogan, S.M. Watt. 1991. Maple V Library Reference Manual. Springer-Verlag, Berlin.

Haldane, J.B.S. 1919. The combination of linkage values and the calculation of distances between linked factors. J. Genet. 8:299-309.

Hospital, F., and A. Charcosset. 1997. Marker assisted introgression of quantitative trait loci. Genetics 147:1469-1485.

Hospital, F., C. Chevalet, and P. Mulsant. 1992. Using markers in gene introgression breeding programs. Genetics 132:1199-1210.

Frisch, M., M. Bohn, and A.E. Melchinger. 1999. Comparison of selection strategies for marker-assisted selection in transfer of a target gene by recurrent backcrossing. Crop Sci. 39:1295–1301.

Libermann, U., and S. Karlin. 1984. Theoretical models of genetic map functions. Theor. Population Biol. 25:331-346.

Lee, M. 1995. DNA markers on plant breeding programs. Adv. Agr. 55:265-344.

Melchinger, A.E. 1990. Use of molecular markers in breeding for oligogenic disease resistance. Plant Breeding 104:1-19.

Ragot, M., M. Biasiolli, M.F. Delbut, A. Dell'Orco, L. Malgarini, P. Thevenin, J. Vernoy, J. Vivant, R. Zimmermann, and G. Gay. 1995. Marker-assisted backcrossing: a practical example. In: Techniques et utilisations des marqueurs moleculaires. Montepellier, France. March 29-31, 1994. INRA, Paris.

Stam, P. 1979. Interference in genetic crossing over and chromosome mapping. Genetics 92:873-594.

Tanksley, S.D., N.D. Young, A.H. Patterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: New tools for an old science. Bio/Technology 7:257-263.

Visscher, P.M., C.S. Haley, and R. Thompson. 1996. Marker-assisted introgression in backcross breeding programs. Genetics 144:1923-1932.

Young, N.D., and S.D. Tanksley. 1989. RFLP analysis of the size of chromosomal segments retained around the *tm*-2 locus of tomato during backcross breeding. Theor. Appl. Genet. 77:353-259.