Theor Appl Genet (2009) 119:23–32
DOI 10.1007/s00122-009-1013-6

ORIGINAL PAPER

Efficiency gain of marker-assisted backcrossing by sequentially increasing marker densities over generations

Vanessa Prigge · Albrecht E. Melchinger · Baldev S. Dhillon · Matthias Frisch

Received: 30 October 2008 / Accepted: 8 March 2009 / Published online: 29 April 2009

Abstract Expenses for marker assays are the major costs in marker-assisted backcrossing programs for the transfer of target genes from a donor into the genetic background of a recipient genotype. Our objectives were to (1) investigate the effect of employing sequentially increasing marker densities over backcross generations on the recurrent parent genome (RPG) recovery and the number of marker data points (MDP) required, and (2) determine optimum designs for attaining RPG thresholds of 93–98% with a minimum number of MDP. We simulated the introgression of one dominant target gene for genome models of sugar beet (*Beta vulgaris* L.) and maize (*Zea mays* L.) with varying marker distances of 5–80 cM and population sizes of 30–250 plants across BC_1 to BC_3 generations. Employing less dense maps in early backcross generations resulted in savings of over 50% in the number of required MDP compared with using a constant set of markers and was accompanied only by small reductions in the attained RPG values. The optimum designs were characterized by increasing marker densities and increasing population sizes in advanced generations for both genome models. We conclude that increasing simultaneously the marker density and the population size from early to advanced backcross generations results in gene introgression with a minimum number of required MDP.

Communicated by Q. Zhang.

V. Prigge · A. E. Melchinger (✉) · B. S. Dhillon
Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

M. Frisch
Institute of Agronomy and Plant Breeding II, Justus-Liebig-University, 35392 Giessen, Germany
e-mail: matthias.frisch@agrar.uni-giessen.de

Introduction

Marker-assisted backcrossing is widely used to introgress one or several target gene(s) from a donor into the genome of a recipient parent. The efficiency of a marker-assisted backcrossing program depends on the selection strategy, population size, distance of the markers flanking the target gene, and length of the chromosome segment attached to the target gene. Studies focusing on the optimum marker-assisted backcrossing design in relation to these factors have been reviewed by Frisch (2005) and Hospital (2005).

Marker-assisted selection for the recurrent parent alleles, also known as background selection, accelerates recovery of the recurrent parent genome (RPG) (Hospital et al. 1992; Ribaut and Hoisington 1998; Wang et al. 2007). The effect of the marker density on the recovery of the RPG in background selection was first investigated by Hospital et al. (1992). They studied six generations of backcrossing and concluded that two generations can be saved by conducting marker-assisted background selection. They added that two to three markers per 100 cM were optimal for controlling the genetic background of the recurrent parent. Visscher et al. (1996) suggested a marker distance of 10–20 cM to be appropriate. Considering a chromosome length of 100 cM, Visscher (1996) proposed to use two markers at a distance of 28 cM from the telomeres, while Servin and Hospital (2002) proposed to place two markers at 20 cM from the telomeres.

These studies applied simulation and theoretical approaches with varying marker densities, but in all

studies, a constant set of markers was used across generations of the marker-assisted backcrossing program. Hospital et al. (1992) suggested that the use of more than two markers per 100 cM was of small benefit in early generations, whereas in advanced backcross generations higher marker densities are relatively more efficient. However, a systematic study of the effect of employing more dense marker maps in advanced generations of a marker-assisted backcrossing program is lacking. Furthermore, in all these studies the target variable was only the proportion of the RPG recovered. Yet from a practical point of view, the number of required marker data points (MDP) is another key variable which determines the costs and, therefore, the applicability of marker-assisted backcrossing programs. Frisch et al. (1999a) assessed the number of required MDP when comparing various selection strategies in marker-assisted backcrossing for the introgression of one target gene, but they used a published map of maize (*Zea mays* L.) with a constant average marker density of 20 cM including two large gaps.

In this study, we examined the effects of constant and sequentially increasing marker densities across the first three backcross generations in marker-assisted backcrossing using the genome models of sugar beet (*Beta vulgaris* L.) and maize. We considered the introgression of one dominant target gene and phenotypic selection for the same.

The objectives of the study were to (1) evaluate the effect of employing maps with sequentially increasing marker densities across BC_1 to BC_3 generations on the RPG recovery and the MDP requirement, and (2) determine the optimum designs for attaining RPG thresholds of 93–98% with a minimum input of MDP and time.

Simulations

We conducted simulations on background selection in marker-assisted backcrossing with genome models similar to those of sugar beet and maize, and these models are hereafter referred to as sugar beet and maize, respectively. For sugar beet, we assumed nine chromosomes of length 100 cM and for maize ten chromosomes of length 160 cM. Furthermore, we assumed homozygous parents and polymorphism at all loci.

One dominant gene located at the center of chromosome 1, with a distance of 50 cM (sugar beet) or 80 cM (maize) from the telomere, was the target gene to be transferred. We assumed equidistant spacing of the markers in the genome with known map positions. This assumption is close to reality when using SSR and AFLP markers in commercial maize and sugar beet hybrid breeding programs. Two markers were located at the telomeres of each chromosome and the distance between two adjacent markers was d_i, where the subscript i stands for backcross generation BC_i. The simulations were carried out with Plabsoft (Maurer et al. 2008), assuming no interference in crossover formation. The simulation of each backcross program was repeated 10,000 times in order to reduce sampling effects and obtain results with high numerical accuracy and a small standard error. This assures that each observed difference between the RPG values of two simulated marker-assisted backcross programs is significant.

A two-stage selection procedure was employed for marker-assisted backcrossing. In the first selection step, plants carrying the target gene were preselected. In the second selection step, a selection index $i = \sum_m x_m$ was constructed, where summation is over markers and x_m is the number of recurrent parent alleles at the mth marker. This index was used to select one plant per backcross population as non-recurrent parent for the next backcross.

The upper 10% quantile (Q10) of the proportion of RPG recovered in the selected plants was employed to measure the success of a marker-assisted backcrossing program with respect to restoring the properties of the recipient parent. The Q10 values have the advantage that they contain not only information about the location of a distribution, as do the mean or median, but also include information about the variation of the distribution. They can be interpreted as "with a probability of 90% the marker-assisted backcrossing program reaches an RPG value that is greater than the Q10 value.", i.e. they not only describe what can be expected on average, but rather give a lower bound for the worst case.

The Q10 values describe the actual genetic composition of the genome of backcross plants. They were determined from the probability distribution of the RPG in the simulated genotypes including all genes in the genome. Hence, the values are the true RPG values, which are (1) independent of the marker map, and (2) not affected by an estimation bias (cf. Frisch and Melchinger 2006).

The number of required MDP was employed to measure the economic efficiency of a marker-assisted backcrossing program. From generation BC_2 onwards, only those markers were assumed to be analyzed (and, hence, contribute to the number of required MDP), which did not carry the recurrent parent allele in homozygous state in the previous generation.

Three series of simulations were carried out. In the first series, we determined the proportion of RPG recovered in typical backcrossing programs in sugar beet and maize, assuming only phenotypic selection for the target gene but no marker-assisted background selection. Since a typical backcrossing program usually consists of six generations (Allard 1960), Q10 values of RPG recovered in generation BC_6 can serve as threshold values for comparing

marker-assisted backcrossing and conventional backcrossing programs.

In the second series of simulations, we compared constant d_i in all generations of a marker-assisted backcrossing program with reduced d_i in advanced backcross generations. We investigated three scenarios: (1) constant marker densities in generations BC_1 to BC_3 ($d_1 = d_2 = d_3$) (2) higher marker densities in generations BC_2 and BC_3 than in generation BC_1 ($d_1 > d_2 = d_3$), and (3) sequential increase of marker densities from generations BC_1 to BC_3 ($d_1 > d_2 > d_3$) (Table 1). For both genome models, the simulations for the three scenarios were carried out assuming population sizes n of 40, 60, 80, 100, 125, 150, 175, 200, and 250 individuals, respectively, and n was held constant across backcross generations.

In the third series of simulations, we determined the optimum design of marker-assisted backcrossing programs intending to attain the Q10 thresholds of 93–96% in two backcross generations, and the thresholds of 96–98% in three backcross generations for both crop models. This was motivated by practical considerations because plant breeders often aim at predefined target values for the RPG recovery across various crops. As a starting point, we identified in the results of the second series of simulations the combinations of $d_i = (d_1, d_2, d_3)$ and n, which reached or surpassed a given threshold value and required the least number of MDP. For each set of starting parameters, n was altered in the backcross generations in steps of ten individuals and the attained Q10 values and required MDP were determined. Among these investigated parameter sets, the one that equaled or surpassed the corresponding threshold value and simultaneously required the least number of MDP was regarded as the optimum design. We did not investigate scenarios with decreasing population sizes from early to advanced backcross generations, because decreasing population sizes require more MDP that constant or increasing population sizes without resulting in a greater RPG recovery (Frisch and Melchinger 1999b).

Table 1 Distances d_i between adjacent markers investigated across backcross generations using variable marker densities in three different scenarios: (1) constant marker densities in generations BC_1 to BC_3 ($d_1 = d_2 = d_3$), (2) higher marker density in generations BC_2 and BC_3 than in generation BC_1 ($d_1 < d_2 = d_3$), and (3) sequential increase of marker density from generation BC_1 to BC_3 ($d_1 < d_2 < d_3$)

Scenario	Sugar beet			Maize		
	d_1	d_2	d_3	d_1	d_2	d_3
1	33	33	33	40	40	40
	20	20	20	20	20	20
	10	10	10	10	10	10
	5	5	5	5	5	5
2	50	33	33	80	40	40
	33	20	20	40	20	20
	20	10	10	20	10	10
	10	5	5	10	5	5
3	50	33	20	80	40	20
	33	20	10	40	20	10
	20	10	5	20	10	5

Values are in cM

Results

Phenotypic selection for the target allele without any selection for RPG, i.e. random selection among individuals carrying the target allele, resulted in lower RPG recovery than achieved with no selection (Table 2). The differences in mean RPG recovery using phenotypic versus no selection increased from generation BC_1 to BC_3 but declined thereafter, and the difference in generation BC_6 was similar to that in generation BC_1. Comparing the two genome models, the mean proportion of RPG recovery was higher in maize than in sugar beet by about 0.8% across generations BC_1 to BC_6. The Q10 values obtained in generation BC_6 in sugar beet and maize were 95.3 and 96.5%, respectively.

When employing marker-assisted background selection, in most cases, we found larger Q10 values for sugar beet than for maize considering the two marker-assisted backcrossing programs with identical d_1 and n (Tables 3, 4). In BC_1, the differences between the two models increased with increasing population size from 0.5 to 1.3%, but these differences diminished in generation BC_2 (0.3–0.5%) and had no relationship with n. In generation BC_3, these differences were generally 0–0.2%. The total genome length of the two crop models was reflected in the requirement of MDP, i.e. the number of MDP required for maize was approximately 1.8 times the MDP required for sugar beet.

In sugar beet, the threshold Q10 value of 95.3% could be attained in generation BC_2 for all constant marker densities with population sizes between 250 ($d = 33$) and 100 ($d = 5$; Table 3). This indicates a saving of four backcross generations. For these d, the least MDP (5,090) were required for $d = 20$ cM and $n = 150$. Considering the parameter settings with increasing marker density in generation BC_2, i.e. decreasing d, Q10 values of at least 95.3% were attained in BC_2 in all cases when $d \leq 33$ cM with population sizes of 100 and 150. The MDP requirement reached a minimum of 3,830 for marker distances of 33 and 20 cM in generations BC_1 and BC_2 [$d_i = (33, 20)$ cM] and $n = 150$ individuals. This requirement was substantially lower than the smallest value obtained with constant d. In generation BC_3, Q10 values were larger than 95.3%

Table 2 Simulation results for the mean and the 10% quantile (Q10) of the distribution of the recurrent parent genome in generation BC_t with random choice of individuals carrying the target allele in sugar beet and maize as well as expected values for the mean without selection

Generation	No selection	Random selection among individuals carrying the target allele			
		Sugar beet model		Maize model	
	Mean	Mean	Q10	Mean	Q10
BC_1	75.00	73.2	65.4	73.8	67.2
BC_2	87.50	85.1	78.5	85.9	80.4
BC_3	93.75	91.3	86.2	92.1	88.1
BC_4	96.88	94.6	90.9	95.4	92.5
BC_5	98.43	96.3	93.6	97.1	95.0
BC_6	99.21	97.3	95.3	98.1	96.5
BC_7	99.61	97.9	96.2	98.6	97.4
BC_8	99.80	98.3	96.8	98.9	97.9
BC_9	99.90	98.6	97.2	99.1	98.3
BC_{10}	99.95	98.7	97.6	99.2	98.5

Values are in %

for all parameter sets evaluated. The least dense map with $d_i = (50, 33, 33)$ cM and the smallest population size of $n = 40$ individuals were sufficient to obtain a Q10 value of 96.5% and also required the least number of MDP (870).

In maize, the Q10 threshold value of 96.5% could not be attained in generation BC_2 (Table 4). However, in the BC_3 generation this threshold was attained for all parameter sets with a population size of $n = 40$, except when using the least dense marker map [$d_i = (80, 40, 40)$ cM], where $n = 60$ was needed to attain the threshold. The Q10 values in generation BC_3 indicated a saving of three backcross generations. For constant d, the combination of $d = 40$ cM and $n = 40$ individuals, having a Q10 value of 96.6%, required the least number of MDP (1,440). Among the parameter settings with sequential increase in marker density, the minimum number of MDP (1,200) were required with $d_i = (80, 40, 20)$ cM and $n = 40$ individuals resulting in a Q10 value of 97.0%.

For two-generation marker-assisted backcrossing programs in sugar beet, the optimum marker density was $d_i = (33, 20)$ for all investigated Q10 threshold values (Table 5). Population sizes between $n_i = (40, 40)$ and $n_i = (210, 250)$ and MDP between 1,120 and 5,540 were required to reach Q10 threshold values between 93 and 96%, respectively. In three-generation marker-assisted backcrossing programs with sugar beet, the combination of increasing population sizes and increasing marker density across generations resulted in optimum designs requiring a minimum of 630 [$n_i = (30, 30, 40)$ and $d_i = (50, 50, 33)$ cM] to 1,950 [$n_i = (60, 60, 80)$ and $d_i = (33, 20, 10)$ cM] MDP to achieve Q10 values between 96 and 98%, respectively.

For two-generation marker-assisted backcrossing programs in maize, the optimum marker densities were $d_i = (40, 20)$ cM for Q10 threshold values between 93 and 95%, while a marker density of $d_i = (20, 10)$ cM was optimal for reaching the Q10 threshold of 96% (Table 5). The optimum population sizes ranged from $n_i = (40, 60)$ to $n_i = (200, 250)$ and required between 1,970 and 15,030 MDP to reach Q10 values between 93 and 96%, respectively. Optimum three-generation marker-assisted backcrossing programs with maize were also characterized by increasing population sizes and increasing marker densities across backcross generations. These optimum designs required between 860 [$n_i = (30, 30, 40)$ and $d_i = (80, 40, 40)$ cM] and 2,780 [$n_i = (50, 60, 70)$ and $d_i = (40, 20, 10)$ cM] MDP to obtain Q10 threshold values between 96 and 98%, respectively.

The Q10 value of 96% is included in the simulations for both two- and three-generation marker-assisted backcrossing programs. This value was attained with a minimum of 30–40 individuals in three-generation programs, whereas 200–250 individuals were necessary in two-generation programs for both crop models. Three-generation programs required 630 (sugar beet) to 860 (maize) MDP, whereas in two-generation programs about 8 (sugar beet) to 16 (maize) times more MDP were required to attain a Q10 value of 96%.

Discussion

We compared various combinations of d_1 and n to attain a specific Q10 value for the recovery of RPG with the least number of MDP in order to minimize the input of financial resources in marker-assisted backcrossing programs. This is a simplification because other factors such as number of crosses and DNA extractions to be performed or number of

Table 3 Sugar beet: simulation results for the 10% quantile (Q10) of the distribution of the recurrent parent genome and the total number of marker data points (MDP) required in a backcross program to introgress one target allele, using constant population sizes n and different marker densities d in backcrossing generations BC_1 to BC_3

Generation	Q10 (%)/MDP								
	$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$	$n = 250$
Constant d (cM) across generations ($d_1 = d_2 = d_3$)									
$d = 33$									
BC_1	79.6/720	80.7/1,080	81.6/1,440	82.1/1,800	82.7/2,250	83.1/2,700	83.5/3,150	83.7/3,600	84.3/4,500
BC_2	92.4/950	93.3/1,400	93.8/1,850	94.2/2,290	94.5/2,840	94.7/3,390	94.9/3,930	95.1/4,480	95.3/5,560
BC_3	96.6/1,000	97.0/1,470	97.2/1,920	97.4/2,380	97.5/2,940	97.5/3,490	97.6/4,040	97.6/4,590	97.7/5,690
$d = 20$									
BC_1	80.2/1,080	81.5/1,620	82.4/2,160	82.9/2,700	83.5/3,370	83.9/4,050	84.3/4,720	84.6/5,400	85.1/6,750
BC_2	93.1/1,420	94.0/2,100	94.6/2,770	94.9/3,440	95.2/4,260	95.5/5,090	95.7/5,900	95.9/6,730	96.2/8,350
BC_3	97.1/1,510	97.6/2,220	97.9/2,900	98.1/3,590	98.2/4,430	98.3/5,270	98.4/6,100	98.4/6,940	98.5/8,590
$d = 10$									
BC_1	80.4/1,970	81.8/2,970	82.7/3,960	83.3/4,950	83.9/6,190	84.4/7,430	84.7/8,670	85.0/9,900	85.5/12,380
BC_2	93.5/2,600	94.4/3,850	94.9/5,070	95.3/6,300	95.7/7,820	96.0/9,320	96.2/10,830	96.3/12,320	96.6/15,310
BC_3	97.5/2,780	98.0/4,070	98.3/5,340	98.4/6,600	98.6/8,160	98.7/9,700	98.8/11,240	98.8/12,770	98.9/15,820
$d = 5$									
BC_1	80.6/3,780	81.9/5,670	82.8/7,550	83.4/9,450	84.0/11,820	84.5/14,180	84.9/16,540	85.1/18,900	85.7/23,640
BC_2	93.6/4,970	94.5/7,340	95.1/9,680	95.5/12,020	95.8/14,920	96.1/17,790	96.3/20,650	96.5/23,500	96.7/29,190
BC_3	97.6/5,300	98.2/7,760	98.5/10,180	98.6/12,580	98.8/15,570	98.9/18,510	99.0/21,440	99.0/24,350	99.1/30,150
Variable d_i (cM) across generations ($d_1 > d_2 = d_3$)									
$d_i = (50, 33, 33)$									
BC_1	78.6/580	79.6/840	80.4/1,120	80.9/1,390	81.5/1,720	81.8/2,060	82.2/2,400	82.4/2,740	82.8/3,410
BC_2	92.1/810	93.0/1,180	93.5/1,540	93.9/1,900	94.2/2,350	94.4/2,790	94.6/3,230	94.8/3,680	95.1/4,550
BC_3	96.5/870	96.9/1,250	97.1/1,630	97.3/2,000	97.4/2,460	97.5/2,910	97.6/3,360	97.6/3,810	97.7/4,690
$d_i = (33, 20, 20)$									
BC_1	79.6/770	80.7/1,130	81.6/1,500	82.1/1,850	82.7/2,300	83.1/2,760	83.5/3,200	83.7/3,660	84.3/4,550
BC_2	93.0/1,120	93.9/1,630	94.4/2,120	94.8/2,620	95.2/3,230	95.4/3,830	95.6/4,430	95.8/5,030	96.1/6,220
BC_3	97.1/1,210	97.6/1,740	97.9/2,260	98.1/2,770	98.2/3,400	98.3/4,020	98.4/4,640	98.4/5,250	98.5/6,460
$d_i = (20, 10, 10)$									
BC_1	80.2/1,120	81.5/1,670	82.4/2,200	82.9/2,750	83.5/3,420	83.9/4,100	84.3/4,770	84.6/5,440	85.1/6,800
BC_2	93.4/1,750	94.3/2,550	94.9/3,330	95.3/4,110	95.6/5,060	95.9/6,020	96.1/6,960	96.3/7,900	96.6/9,760
BC_3	97.5/1,930	98.0/2,770	98.3/3,590	98.4/4,410	98.6/5,410	98.7/6,400	98.8/7,370	98.8/8,350	98.9/10,270
$d_i = (10, 5, 5)$									
BC_1	80.4/2,070	81.8/3,060	82.7/4,050	83.3/5,040	83.9/6,280	84.4/7,520	84.7/8,750	85.0/9,990	85.5/12,470
BC_2	93.6/3,260	94.5/4,730	95.1/6,180	95.4/7,610	95.8/9,380	96.1/11,130	96.3/12,870	96.5/14,600	96.7/18,040
BC_3	97.6/3,590	98.2/5,160	98.5/6,680	98.6/8,180	98.8/10,030	98.9/11,850	99.0/13,660	99.0/15,450	99.1/19,010
Variable d_i (cM) across generations ($d_1 > d_2 > d_3$)									
$d_i = (50, 33, 20)$									
BC_1	78.6/580	79.6/840	80.4/1,120	80.9/1390	81.5/1,720	81.8/2,060	82.1/2,400	82.4/2,740	82.8/3,410
BC_2	92.1/830	93.0/1,200	93.5/1,560	93.9/1,920	94.2/2,360	94.4/2,810	94.6/3,250	94.8/3,690	95.1/4,560
BC_3	96.9/930	97.4/1,330	97.7/1,720	97.9/2,110	98.0/2,580	98.2/3,060	98.3/3,520	98.3/3,990	98.4/4,910
$d_i = (33, 20, 10)$									
BC_1	79.6/770	80.7/1,130	81.6/1,500	82.1/1,850	82.7/2,300	83.1/2,760	83.5/3,200	83.7/3,660	84.3/4,550
BC_2	93.0/1,140	93.9/1,640	94.4/2,130	94.8/2,630	95.2/3,240	95.4/3,840	95.6/4,450	95.8/5,040	96.1/6,240
BC_3	97.4/1,320	97.9/1,880	98.2/2,410	98.4/2,950	98.5/3,610	98.7/4,260	98.7/4,900	98.8/5,530	98.9/6,800
$d_i = (20, 10, 5)$									
BC_1	80.2/1,120	81.5/1,670	82.4/2,200	82.9/2,750	83.5/3,420	83.9/4,100	84.3/4,770	84.6/5,440	85.1/6,800
BC_2	93.4/1,780	94.3/2,580	94.9/3,360	95.3/4,130	95.6/5,090	95.9/6,040	96.1/6,990	96.3/7,920	96.6/9,790
BC_3	97.6/2,110	98.1/3,000	98.4/3,870	98.6/4,710	98.7/5,760	98.8/6,780	99.0/7,800	99.0/8,800	99.1/10,780

Values for MDP are rounded to multiples of ten

Table 4 Maize: simulation results for the 10% quantile (Q10) of the distribution of the recurrent parent genome and the total number of marker data points (MDP) required in a backcross program to introgress one target allele, using constant population sizes n and different marker densities d in backcrossing generations BC_1 to BC_3

Generation	Q10 (%)/MDP								
	$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$	$n = 250$
Constant d (cM) across generations ($d_1 = d_2 = d_3$)									
$d = 40$									
BC_1	79.0/1,000	80.0/1,500	80.6/2,000	81.1/2,500	81.6/3,130	81.9/3,750	82.2/4,370	82.5/5,000	82.9/6,250
BC_2	92.0/1,340	92.7/1,990	93.2/2,630	93.5/3,260	93.9/4,060	94.1/4,840	94.3/5,620	94.5/6,400	94.7/7,960
BC_3	96.6/1,440	96.9/2,120	97.1/2,780	97.2/3,440	97.2/4,260	97.3/5,070	97.3/5,880	97.3/6,680	97.3/8,280
$d = 20$									
BC_1	79.7/1,800	80.7/2,700	81.5/3,590	82.0/4,500	82.5/5,620	82.8/6,750	83.2/7,880	83.4/9,000	83.9/11,250
BC_2	92.8/2,410	93.6/3,570	94.2/4,720	94.5/5,870	94.9/7,290	95.2/8,700	95.4/10,120	95.5/11,530	95.8/14,320
BC_3	97.4/2,590	97.8/3,810	98.1/5,000	98.2/6,200	98.3/7,670	98.5/9,130	98.5/10,590	98.6/12,040	98.6/14,920
$d = 10$									
BC_1	79.9/3,400	81.0/5,100	81.7/6,790	82.3/8,510	82.8/10,610	83.2/12,750	83.5/14,880	83.7/17,000	84.2/21,250
BC_2	93.1/4,560	93.9/6,750	94.5/8,910	94.9/11,070	95.2/13,740	95.5/16,410	95.7/19,080	95.9/21,730	96.1/27,010
BC_3	97.7/4,900	98.1/7,190	98.4/9,440	98.6/11,690	98.7/14,460	98.8/17,220	98.9/19,980	99.0/22,700	99.0/28,140
$d = 5$									
BC_1	80.0/6,590	81.1/9,900	81.8/13,210	82.3/16,500	82.8/20,630	83.2/24,730	83.6/28,860	83.9/33,020	84.3/41,240
BC_2	93.1/8,820	94.0/13,070	94.5/17,300	95.0/21,470	95.3/26,680	95.6/31,830	95.8/36,980	96.0/42,140	96.3/52,350
BC_3	97.8/9,480	98.2/13,930	98.5/18,340	98.6/22,660	98.8/28,060	98.9/33,390	99.0/38,700	99.1/44,030	99.2/54,520
Variable d_i (cM) across generations ($d_1 > d_2 = d_3$)									
$d_i = (80, 40, 40)$									
BC_1	77.0/620	77.9/920	78.4/1,220	78.8/1,520	79.3/1,890	79.5/2,270	79.8/2,640	79.9/3,020	80.3/3,770
BC_2	91.4/980	92.1/1,430	92.5/1,880	92.9/2,320	93.3/2,880	93.5/3,420	93.7/3,970	93.9/4,520	94.1/5,600
BC_3	96.4/1,080	96.7/1,560	96.9/2,050	97.1/2,520	97.1/3,100	97.1/3,680	97.2/4,250	97.2/4,820	97.2/5,960
$d_i = (40, 20, 20)$									
BC_1	79.0/1,040	80.0/1,540	80.6/2,040	81.1/2,540	81.6/3,170	81.9/3,790	82.2/4,410	82.5/5,040	82.9/6,290
BC_2	92.6/1,660	93.4/2,430	93.9/3,190	94.3/3,930	94.7/4,860	94.9/5,790	95.1/6,710	95.3/7,620	95.6/9,440
BC_3	97.3/1,840	97.8/2,670	98.0/3,480	98.2/4,270	98.3/5,250	98.4/6,230	98.5/7,200	98.5/8,150	98.6/10,060
$d_i = (20, 10, 10)$									
BC_1	79.7/1,880	80.7/2,780	81.5/3,680	82.0/4,580	82.5/5,700	82.8/6,830	83.2/7,950	83.4/9,070	83.9/11,330
BC_2	93.0/3,040	93.8/4,430	94.4/5,810	94.8/7,160	95.1/8,850	95.4/10,530	95.6/12,190	95.8/13,830	96.1/17,130
BC_3	97.7/3,380	98.1/4,880	98.4/6,350	98.5/7,790	98.7/9,580	98.8/11,350	98.9/13,090	98.9/14,820	99.0/18,270
$d_i = (10, 5, 5)$									
BC_1	79.9/3,570	81.0/5,260	81.7/6,960	82.3/8,660	82.8/10,790	83.2/12,910	83.5/15,040	83.7/17,170	84.2/21,410
BC_2	93.2/5,800	94.0/8,440	94.5/11,050	94.9/13,650	95.3/16,840	95.5/20,020	95.8/23,180	96.0/26,300	96.2/32,560
BC_3	97.7/6,450	98.2/9,300	98.5/12,090	98.7/14,850	98.8/18,230	98.9/21,580	99.0/24,900	99.1/28,180	99.2/34,740
Variable d_i (cM) across generations ($d_1 > d_2 > d_3$)									
$d_i = (80, 40, 20)$									
BC_1	77.0/620	77.9/920	78.4/1,220	78.8/1,520	79.3/1,890	79.5/2,270	79.8/2,640	79.9/3,020	80.3/3,770
BC_2	91.4/990	92.1/1,440	92.6/1,890	92.9/2,340	93.3/2,890	93.5/3,440	93.7/3,990	93.9/4,530	94.1/5,620
BC_3	97.0/1,200	97.4/1,720	97.7/2,230	97.9/2,740	98.0/3,360	98.1/3,970	98.2/4,590	98.3/5,190	98.3/6,400
$d_i = (40, 20, 10)$									
BC_1	79.0/1,040	80.0/1,540	80.6/2,040	81.1/2,540	81.6/3,170	81.9/3,790	82.2/4,410	82.5/5,040	82.9/6,290
BC_2	92.6/1,690	93.4/2,450	93.9/3,210	94.3/3,960	94.7/4,890	94.9/5,820	95.1/6,720	95.3/7,650	95.6/9,460
BC_3	97.5/2,050	98.0/2,920	98.2/3,780	98.4/4,620	98.6/5,670	98.7/6,690	98.8/7,690	98.8/8,710	98.9/10,700
$d_i = (20, 10, 5)$									
BC_1	79.7/1,880	80.7/2,780	81.5/3,680	82.0/4,580	82.5/5,700	82.8/6,830	83.2/7,950	83.4/9,070	83.9/11,330
BC_2	93.0/3,090	93.8/4,480	94.4/5,850	94.8/7,220	95.1/8,910	95.4/10,570	95.6/12,230	95.8/13,880	96.1/17,170
BC_3	97.7/3,760	98.2/5,360	98.4/6,900	98.6/8,440	98.8/10,320	98.9/12,160	99.0/14,000	99.0/15,800	99.1/19,400

Values for MDP are rounded to multiples of ten

Table 5 Simulation results for the optimum design of marker-assisted backcrossing programs in sugar beet and maize with the goal of recovering a certain proportion of recurrent parent genome in a two- or three-generation approach, and the number of marker data points (MDP) required with this approach

Q10 (%)	Sugar beet model			Maize model		
	d_i (cM)	n_i	MDP	d_i (cM)	n_i	MDP
Two-generation marker-assisted backcrossing programs						
93	33, 20	40, 40	1,120	40, 20	40, 60	1,970
94	33, 20	60, 70	1,710	40, 20	80, 80	3,210
95	33, 20	100, 125	2,810	40, 20	150, 175	6,120
96	33, 20	210, 250	5,540	20, 10	200, 25	15,030
Three-generation marker-assisted backcrossing programs						
96	50, 50, 33	30, 30, 40	630	80, 40, 40	30, 30, 40	860
97	50, 33, 20	40, 40, 60	990	80, 40, 20	30, 40, 50	1,120
98	33, 20, 10	60, 60, 80	1,950	40, 20, 10	50, 60, 70	2,780

n population size and d marker density employed in backcrossing generations BC_1 to BC_3

plants to be raised also contribute to the costs of a marker-assisted backcrossing program. However, the marker analyses are the major cost factor by far and, therefore, we focused specifically on the possibilities of minimizing the number of required MDP.

Threshold values

The proportion of RPG recovered and, consequently, the Q10 values were lower for sugar beet than for maize, the difference being 1.2–1.9% in BC_1 to BC_6 generations (Table 2). Hence, if the Q10 values obtained after generation BC_6 are used as threshold, then the actual similarity of the introgressed progeny and the recurrent parent will vary depending on the crop under consideration. For example, introgressed progenies in sugar beet are expected to be less similar to the recurrent parent than those in maize. From a practical point of view, however, there seems no need to set different Q10 values, i.e. different degrees of similarity between the introgressed progeny and the recurrent parent in marker-assisted backcrossing programs for different crops. A more general strategy is to define Q10 thresholds which depend on the actual goal of the introgression program as well as the performance level of the donor but are independent of the crop under consideration.

In backcross programs, the recurrent parent usually shows excellent agronomic performance but lacks one trait which is present in the donor and is to be transferred. If the donor parent is adapted or related germplasm with acceptable agronomic performance, it seems sufficient to recover at least 95% of the RPG. Without background selection, a mean RPG of at least 95% requires Q10 values of about 93% (Table 2). If the donor parent is unadapted and shows poor agronomic performance, the recovery of a higher proportion of the RPG is advisable to minimize the risk of undesirable effects of such a donor genome on the phenotype of the introgressed progeny. Here, a mean RPG recovery of at least 98% may be a reasonable guideline, and this corresponds to Q10 values of at least 96% without background selection. Plant breeders always endeavor to rapidly develop improved cultivars. The time factor has become more important in the present era of intellectual property rights, and particularly in transgenics' development. Hence, we addressed Q10 values of 93–96% and 96–98% in two- and three-generation marker-assisted backcrossing programs, respectively, when identifying optimum designs (Table 5).

Population size

The optimum designs were mostly characterized by larger population sizes in advanced backcross generations. The optimum designs for three-generation marker-assisted backcrossing programs required a maximum of 80 individuals in generation BC_3 (Table 5). These results are in accordance with earlier studies of Frisch et al. (1999a) and Frisch and Melchinger (2001a, b). For two-generation marker-assisted backcrossing programs, however, population sizes up to 250 individuals still resulted in considerable increases of the Q10 values, which is also reflected in the optimum designs. Our study identified the optimum population size to be 30–80 individuals for three-generation marker-assisted backcrossing programs, whereas for two-generation marker-assisted backcrossing programs 40–250 individuals are needed.

Marker positions

In the present study, we used hypothetical marker maps with evenly spaced markers of which the first and the last were located at the telomeres of a chromosome. This is in

contrast to earlier studies (Frisch et al. 1999a; Frisch and Melchinger 2001a), wherein published maps were used because evenly spaced maps were often not available at that time. With the advances made in technologies, a large number of markers have now become available which facilitate the use of maps with evenly spaced markers in most of the economically important crops.

The relevance of even versus uneven marker spacing is demonstrated when comparing the results of Frisch et al. (1999a) and the present study. Using an unevenly spaced map of maize for two-stage selection with constant marker density (average $d = 20$) and constant population size of $n = 40, 60, 80$ (Frisch et al. 1999a) resulted in Q10 values, which were about 1% lower than those obtained in the present study with the same parameter settings in generation BC_3 (Table 4). However, the increased proportion of RPG recovered in the present study also required almost 300 MDP more than reported by Frisch et al. (1999a).

The advantage of using an evenly spaced marker map becomes obvious, however, when comparisons are made for a given Q10 threshold. For example, Frisch et al. (1999a) attained the Q10 of 97.3% in generation BC_3 by employing two-stage selection, a map with unevenly spaced markers (average $d = 20$ cM) and a constant $n = 80$, which required 4,390 MDP. Using an evenly spaced map ($d = 20$ cM), we obtained a Q10 value of 97.4% with less individuals ($n = 40$) which required only 2,590 MDP (Table 4). Similar results were obtained for other Q10 values. Thus, we conclude that the use of evenly spaced marker maps leads to substantial savings in MDP required to attain a given Q10 threshold as compared to the map with unevenly spaced markers.

We assumed two markers to be located at the telomeres of each chromosome (e.g. on a 100 cM chromosome with $d_i = 50$, three markers were located at positions 0, 50, and 100 cM, respectively). An alternative would be to position the first and the last markers with distance $d/2$ from the telomere as suggested by Wang and Bernardo (2000), or to use optimized distances between the telomere and the first marker as suggested by Visscher (1996) and Servin and Hospital (2002).

When using increasing marker densities in advanced backcross generations, our model of marker positioning has the advantage that when doubling the marker density from generation i to generation $i + 1$ (i.e. $d_{i+1} = 2d_i$), the markers analyzed in generation i are a subset of the markers analyzed in generation $i + 1$. This is not the case for the above alternative models of marker positioning. They require in generation $i + 1$ a map consisting of entirely new markers which have not been previously analyzed in generation i. This requires additional marker analyses and a larger number of available marker loci than our approach. Furthermore, additional simulations (results not shown) suggested that there is only a marginal increase in RPG recovered as indicated in a slightly higher Q10 value when employing markers located $d/2$ distant from the telomere, which does not justify extra effort. Therefore, we suggest to use maps with markers located at the telomeres for marker-assisted backcrossing.

In our maps, we assumed the locus of the target allele to be located at the center of chromosome 1. In generation BC_3, the number of MDP required and the proportion of RPG recovered depend largely on the distance of the markers flanking the target locus because most of the markers are expected to carry the recurrent parent allele in homozygous state in this generation. Hence, the location of the target locus per se plays only a minor role and we expect the general trends presented here to hold true irrespective of the map position of the target locus.

Marker density

Employing linkage maps with higher marker densities, i.e. smaller marker distances ($d_1 > d_2 > d_3$), in advanced as compared to early backcross generations resulted in higher Q10 values than using maps with constant marker densities ($d_1 = d_2 = d_3$) and also in higher number of required MDP (Tables 3, 4). The increase in the number of required MDP is only moderate, because many of the additionally investigated markers are already fixed for the recurrent parent allele in earlier backcross generations. The higher Q10 values, through the use of maps with sequentially increasing marker densities across generations, can be explained by the fact that on average one crossover per meiosis occurs on a chromosome segment of 100 cM length. Consequently, in early backcross generations the chromosomes are expected to have relatively long segments originating from the parental genomes. These segments can be efficiently monitored with low marker densities. In advanced backcross generations each chromosome is expected to consist of shorter segments of the parental genomes and, thus, a higher marker density results in a greater selection intensity.

All optimum designs (Table 5) are characterized by increasing marker densities (decreasing d) across backcross generations. In a two-generation marker-assisted backcrossing program, with one exception, the optimal d_i was 33 and 40 cM in generation BC_1 and 20 cM in generation BC_2 in both crop models. In a three-generation marker-assisted backcrossing program, a sequential increase of marker density across generations was generally optimal and the range of d_i covered all marker densities except for 5 cM. We conclude that in marker-assisted backcrossing programs an increase of the marker density in advanced generations substantially reduces the number of MDP required when compared with constant marker densities. Furthermore, the effects of increasing both marker density

and population size in advanced generations result in a synergistic effect with respect to saving MDP in a marker-assisted backcrossing program.

Genome length

The variance of the parental genome contribution to BC individuals in the maize genome is smaller than in the sugar beet genome (Frisch and Melchinger 2007). This explains the lower response to marker-assisted background selection in generations BC_1 and BC_2 observed in maize when compared with sugar beet. In generation BC_3, most markers are expected to carry the recurrent parent alleles in homozygous form and the main source of donor genome is expected to be the donor chromosome segment attached to the target gene. The length of this segment depends on the map distance between the target gene and the flanking markers (Frisch and Melchinger 2001c). Hence, assuming the same marker distances, the length is expected to be identical in both crop models. However, due to the smaller total genome length of sugar beet, the donor chromosome segment attached to the target gene corresponds to a higher percentage of the entire genome. This explains the greater Q10 values attained for maize when compared to sugar beet in various backcross generations. We conclude that smaller genomes require markers more tightly linked to the target gene than larger genomes to attain comparable Q10 values with the same number of generations.

The length of the genome also affects the number of MDP required as this number increased approximately proportionally to the total genome length (Table 5). We confirmed this trend in additional simulations with a model of the wheat genome consisting of 21 chromosomes of length 180 cM (results not shown). We conclude that the genome length is a key parameter in the costs of a marker-assisted backcrossing program with genome-wide marker-assisted background selection.

In our simulations, we assumed no reduction of the recombination frequency in the proximity of the introgressed gene. If resistance genes originating from exotic material are to be introgressed, such a reduction of the recombination frequency may result in an extension of the genome length, with the effect of requiring larger populations and more marker analyses for efficiently reducing the donor genome.

Design of marker-assisted backcrossing programs

The BC_1 generation of a marker-assisted backcrossing program is usually the most expensive because all background selection markers need to be analyzed in a large number of individuals. In subsequent backcross generations, more than half of the markers heterozygous in the preceeding generation are expected to be fixed for the recurrent parent allele and, therefore, need not be analyzed again. In order to reduce the number of required MDP in generation BC_1, Frisch et al. (1999a) proposed to conduct selection for recombinants on the carrier chromosome by initially analyzing the markers flanking the target locus. Frisch et al. (1999b) suggested to use small population sizes in generation BC_1 and larger populations in advanced BC generations to reduce the number of required MDP.

Our results show that the approach of increasing population sizes can be refined by additionally increasing marker densities sequentially across backcross generations. The advantage of using increasing marker density and population size is that only one laboratory cycle for marker analyses is required. In contrast, the three-stage selection strategy of Frisch et al. (1999a) requires initial analyses of the markers flanking the target locus and subsequent marker assays of the selected individuals with the remaining background selection markers. This consumes considerably more time and effort than the approach presented here.

By employing the optimum design (Table 5), the introgression of an allele from an adapted donor parent can be conducted in two-generation marker-assisted backcrossing programs with about 1,100–5,500 MDP in sugar beet and 2,000–15,000 MDP in maize (Q10 = 93–96%). The exploitation of an agronomically poor donor parent may require three generations of marker-assisted backcrossing and about 600–2,000 MDP in sugar beet and 900–2,800 MDP in maize (Q10 = 96–98%). Attaining a Q10 value of 96% requires considerably more MDP in both crop models when conducting two- instead of three-generation marker-assisted backcrossing programs (5,540 vs. 630 MDP in sugar beet and 15,030 vs. 860 MDP in maize). Consequently, reducing a marker-assisted backcrossing program from three to two generations substantially increases the cost of the program. However, if there is an urgency and the resources are available the task can be accomplished efficiently in two backcross generations. The main goal of using markers for background selection is to speed up the recovery of the RPG. Having a converted genotype available for variety development 1 year or season earlier is one of the major determinants of economic returns and success of the variety. Thus, the advantage in time may be well worth the additional costs associated with MDP analyses in two-generation marker-assisted backcrossing programs.

The results presented here were based on assumptions related to markers with known map position and rather laborious to handle such as SSR or AFLP markers. Novel high-throughput marker technologies, such as SNP arrays, providing cheap data points are promising to further extend the applicability of the marker-assisted backcrossing concept. We plan to conduct further simulation studies

addressing specifically the optimization of marker-assisted backcrossing programs employing high-throughput marker technologies and further genetic scenarios such as the introgression of recessive genes or the introgression of more than one target gene.

Acknowledgments The financial support from the Bundesministerium für Wirtschaftliche Zusammenarbeit (BMZ, Germany) is gratefully acknowledged. We thank two anonymous reviewers for their helpful comments and suggestions.

References

Allard RW (1960) Principles of plant breeding. Wiley, New York

Frisch M (2005) Optimum design of marker-assisted backcross programs. In: Loerz H, Wenzel G (eds) Biotechnology in agriculture and forestry, vol 55. Springer, Berlin, pp 319–334

Frisch M, Melchinger AE (2001a) Marker-assisted backcrossing for simultaneous introgression of two genes. Crop Sci 41:1716–1725

Frisch M, Melchinger AE (2001b) Marker-assisted backcrossing for introgression of a recessive gene. Crop Sci 41:1485–1494

Frisch M, Melchinger AE (2001c) The length of the intact donor chromosome segment around a target gene in marker-assisted backcrossing. Genetics 157:1343–1356

Frisch M, Melchinger AE (2006) Marker-based prediction of the parental genome contribution to inbred lines derived from biparental crosses. Genetics 174:795–803

Frisch M, Melchinger AE (2007) Variance of the parental genome contribution to inbred lines derived from biparental crosses. Genetics 176:477–488

Frisch M, Bohn M, Melchinger AE (1999a) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39:1295–1301

Frisch M, Bohn M, Melchinger AE (1999b) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39:967–975

Hospital F (2005) Selection in backcross programmes. Phil Trans R Soc B360:1503–1512

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

Maurer HP, Melchinger AE, Frisch M (2008) Population genetic simulation and data analysis with Plabsoft. Euphytica 161:133–139

Ribaut J-M, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends Plant Sci 3:236–239

Servin B, Hospital F (2002) Optimal positioning of markers to control genetic background in marker-assisted backcrossing. J Heredity 93(3):214–217

Visscher PM (1996) Proportion of the variation in genetic composition in backcrossing programs explained by genetic markers. J Heredity 87:136–138

Visscher PM, Haley CS, Thompson R (1996) Marker-assisted introgression in backcross breeding programs. Genetics 144:1923–1932

Wang J, Bernardo R (2000) Variance of marker estimates of parental contribution to F2 and BC1-derived inbreds. Crop Sci 40:659–665

Wang J, Chapman SC, Bonnett DG, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci 47:582–588