Theor Appl Genet (2009) 119:595–603
DOI 10.1007/s00122-009-1069-3

ORIGINAL PAPER

Selection strategies for the development of rye introgression libraries

Karen Christin Falke · Thomas Miedaner · Matthias Frisch

Received: 11 December 2008 / Accepted: 11 May 2009 / Published online: 31 May 2009

Abstract Computer simulations can be employed to find optimal procedures for developing introgression libraries in rye with marker-assisted backcrossing. Our objectives were to investigate the effects of the employed (1) breeding scheme, (2) selection strategy, and (3) population sizes on the donor genome coverage of the library, the number of introgression lines carrying additional donor chromosome segments outside the target regions, and the number of required marker data points. With respect to these target criteria, a BC_3S_2 breeding scheme and increasing population sizes from early to advanced generations were superior to a BC_2S_3 breeding scheme and constant population sizes. The smallest number of donor segments outside the target regions was reached with a three-stage selection strategy, which consists on selection for the target segment, selection for recombination at flanking markers and selection for recurrent parent alleles across the entire genome. Omitting the selection for flanking markers in generation BC_1 reduced considerably the number of required marker data points. A pre-selection of chromosomes consisting completely of donor genome in BC_1 was advantageous, if the effort in the breeding nursery should kept minimum. Adopting the described designs can help rye breeders to successfully develop introgression libraries.

Communicated by M. Bohn.

K. C. Falke · T. Miedaner
State Plant Breeding Institute, Universitaet Hohenheim,
70593 Stuttgart, Germany
e-mail: falke@pz.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

Introgression libraries consist of introgression lines carrying short marker-defined donor chromosome segments (target segments), which were introgressed into a common genetic background by marker-assisted backcrossing (Eshed and Zamir 1994). They were employed to detect and utilize favorable alleles from exotic accessions for enhancing the genetic variation of elite breeding material (Eshed et al. 1992) and to identify quantitative trait loci (QTL) underlying agronomically important traits (Tanksley et al. 1996). Gur and Zamir (2004) demonstrated in tomato that exploiting unexplored natural genetic variation via introgression libraries can lift yield barriers in variety development.

Two rye (*Secale cereale* L.) BC_2S_3 introgression libraries (Falke et al. 2008) were developed at the University of Hohenheim starting in 1999 from a cross of the elite inbred line L2053-N used as recurrent parent and the heterozygous Iranian primitive rye accession Altevogt 14160 used as donor parent. Amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR) markers were used to monitor in each backcross and selfing generation the target segments as well as additional donor chromosome segments in the introgression lines and the donor genome coverage of the introgression library. The mean number of additional donor segments per introgression line was five for library A and three for library B. Library A covered 74% of the donor genome and library B 59%. Favorable effects of the donor segments on per se (Falke et al. 2008, 2009a) and testcross (Falke et al. 2009b) performance for quality and agronomic traits as well as on pollen-fertility restoration (Falke et al. 2009c) have been detected.

In backcrossing programs, markers can be used for tracing the presence of the target gene or genome region to be introgressed (foreground selection) and for accelerating the

recovery of the recurrent parent genome (background selection). Main factors determining the efficiency of foreground and background selection are the breeding scheme, the selection strategy, and the population sizes (Hospital 2005; Frisch 2005). For introgression of individual target genes, simulation studies provided a useful tool to optimize marker-assisted backcrossing programs (Visscher et al. 1996; Frisch et al. 1999a; Prigge et al. 2008). In the context of introgression libraries, a first simulation study determined that in rye three backcross generations are more advantageous than two (Sušić 2005). Since high-throughput marker systems are still lacking in rye, the high costs for DNA extraction, PCR amplification, and conducting marker analyses are a limiting factor and only optimized breeding strategies allow the cost efficient development of introgression libraries. An in-depth study on such optimum designs is, however, not available yet.

The goal of our study was to find with computer simulations optimized procedures to develop introgression libraries in rye. In particular, our objectives were to (1) investigate different breeding schemes, (2) compare various selection strategies, and (3) examine the effects of increasing or decreasing population sizes from early to advanced generations with respect to the donor genome coverage of the library, additional non-target donor genome segments in introgression lines, and the number of marker data points required for developing the introgression library.

Simulations

Genetic map

We considered a genome model of rye consisting of seven chromosomes of 100 cM length. The genome was covered with equally spaced markers of 5 cM distance. The first marker on each chromosome had a distance of 2.5 cM from the telomere. All markers were polymorphic in the crossing parents from which the library was developed. The target donor genome segments were of 20 cM length. In consequence, the final introgression library consisted of 35 introgression lines.

To model recombination along the chromosomes, no interference in crossover formation (Stam 1979) was assumed and recombination frequencies were related to the corresponding map distances with Haldane's (1919) mapping function. Software Plabsoft (Maurer et al. 2008) was used for the simulations.

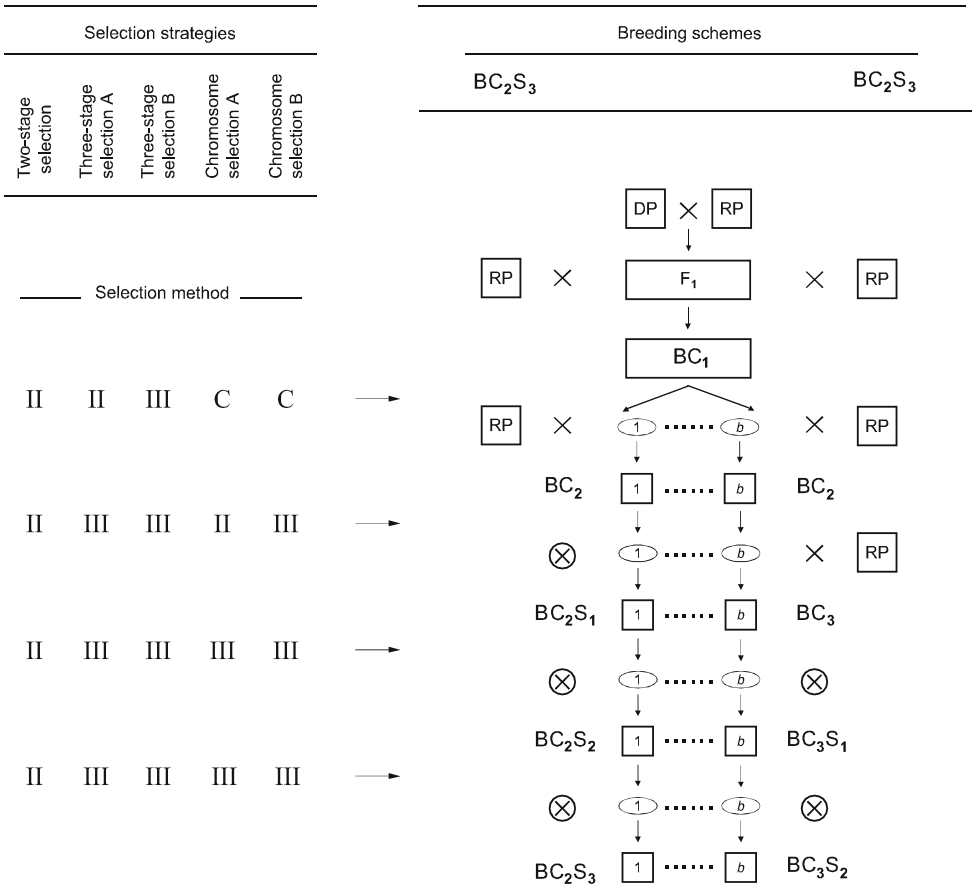
Breeding schemes

We simulated a BC_2S_3 breeding scheme, as carried out experimentally by Falke et al. (2008), and in addition a BC_3S_2 breeding scheme (Fig. 1). For both breeding schemes, the simulations started with a cross of the homozygous elite inbred line as recurrent parent and the heterozygous exotic donor parent to generate the F_1 base population. Random F_1 plants were individually backcrossed to the recurrent parent to create the BC_1 generation. BC_1 plants were selected according to the selection strategies described below and backcrossed to the recurrent parent for producing the BC_2 generation. Generation BC_2 was followed by three selfing generations (BC_2S_3 breeding scheme) or a third backcross generation and two selfing generations (BC_3S_2 breeding scheme). The simulations were repeated 5,000 times to reduce sampling effects and obtain results with a high numerical accuracy.

Selection strategies

For both breeding schemes, we investigated five selection strategies, which differ in the selection pressure applied to carrier versus non-carrier chromosomes of the target segments (Fig. 1).

1. *Two-stage selection*. In the first selection stage, foreground selection is carried out such that only plants carrying donor alleles at all markers on the target segment are pre-selected. In the second stage, selection is based on an index counting the number of RP alleles at the entire set of markers outside the target segment. Two-stage selection is employed in all generations.
2. *Three-stage selection A*. In generation BC_1, two-stage selection is carried out. Starting with generation BC_2, the following three-stage strategy is employed. After foreground selection in the first selection stage, a selection index is determined based on recombination events between the target segment and its flanking markers. The index is 2 for recombination between the target segment and both flanking markers, 1 for recombination on one side of the target segment, and 0 for no recombination. All plants with the largest value of the index are preselected. After this second selection stage, the third selection stage follows, based on the entire set of markers outside the target segment.
3. *Three-stage selection B*. Three-stage selection is employed in all generations.
4. *Chromosome selection A*. In generation BC_1, foreground selection is carried out such that plants carrying the donor alleles on entire chromosomes are pre-selected. This is followed by background selection based on the markers on the remaining chromosomes. In generation BC_2, two-stage selection is employed and in subsequent generations three-stage selection.



Selection index based on → Characteristics of the selected plants	Selection method II	III	C
	Selection stage		
DP alleles on the target segment → Selection of plants carrying the target segment	1	1	-
DP alleles on entire chromosomes → Selection of plants carrying donor chromosomes	-	-	1
RP alleles at markers flanking the target segment → Selection of plants with recombinations	-	2	-
RP alleles at markers across the entire genome → Selection of one plant with the maximum number of RP alleles	2	3	2
No. of backcrossing programs carried out simultaneously	$b = 35$	$b = 35$	$b = 7$

Fig. 1 Overview of two breeding schemes and five selection strategies employed for the development of rye introgression libraries starting with a cross of the recurrent parent (RP) and the donor parent (DP)

5. *Chromosome selection B*. In generation BC_1, selection is based on entire chromosomes, as in chromosome selection A. Three-stage selection is employed in all subsequent generations, including generation BC_2.

Success criteria

To quantify and compare the success of the different breeding schemes and selection strategies with respect to the genomic composition of the resulting introgression libraries

and the required costs, we determined for each scenario the following success criteria: (1) Donor genome coverage of the introgression library, assessed at markers on the target segments. (2) Proportion of the recurrent parent alleles at markers outside the target segments (RPM). (3) Average number of donor alleles at markers outside the target segments per introgression line (ADM). (4) Number of introgression lines within an introgression library, which carry in addition to the target segment further donor alleles at markers outside the target segments (LDM). (5) Number of marker data points (MDP) required to develop the introgression library. (To shorten notation, we abbreviate thousand MDP with kMDP.)

The development of an introgression library was considered as successful, if the following two criteria were met: (1) The donor genome coverage was 100%, i.e., each target segment is carried by one line of the library. (2) LDM < 1, i.e., the expected number of introgression lines within an introgression library, which are carrying donor alleles at markers outside the target segments, is smaller than one.

Population sizes

The breeding schemes and selection strategies were compared with respect to the success criteria employing constant population sizes n of 20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 plants throughout all backcross and selfing generations. For the BC_3S_2 breeding scheme, we additionally compared scenarios with decreasing and increasing populations sizes from early to advanced generations of the breeding scheme. These scenarios were characterized by differences $\delta = -50, \ldots, 50$ (in steps of 10) in the population sizes of two subsequent generations (Table 1).

Table 1 Decreasing and increasing population sizes from early to advanced generations employed in the simulations for the BC_3S_2 breeding scheme

δ	Population sizes (n)				
	BC_1	BC_2	BC_3	BC_3S_1	BC_3S_2
−50	220	170	120	70	20
−40	200	160	120	80	40
−30	180	150	120	90	60
−20	160	140	120	100	80
−10	140	130	120	110	100
0	120	120	120	120	120
10	100	110	120	130	140
20	80	100	120	140	160
30	60	90	120	150	180
40	40	80	120	160	200
50	20	70	120	170	220

δ is the difference in population size of two subsequent generations

Results

For both breeding schemes, a donor genome coverage of 100% was reached irrespective of the employed selection strategy and population sizes. The BC_3S_2 breeding scheme outperformed the BC_2S_3 breeding scheme with respect to genomic composition of the resulting introgression libraries. It reached greater values for RPM and smaller values for ADM and LDM for any combination of selection strategy and population sizes (Table 2). The BC_2S_3 breeding scheme required more MDP than the BC_3S_2 breeding scheme only for small population sizes ($n = 20$), but less for population sizes of $n = 40$ and larger.

For the BC_2S_3 breeding scheme, the success criterion that on expectation less than one introgression line within the final introgression library carries additional donor alleles at markers outside the target segment (LDM < 1) could not be achieved with any simulated scenario (Table 2). For the BC_3S_2 breeding scheme and constant population sizes across generations, LDM < 1 was reached for three-stage selection B with population sizes of $n \geq 100$, and for the other selection strategies with population sizes between $n \geq 120$ and 200.

Two-stage selection resulted in greater values for LDM than the other selection strategies with the exception of chromosome selection B (for $n = 20$; Table 2). This poor performance with respect to the genomic composition was accompanied with large numbers of required MDP, only three-stage selection B required more MDP (for $n \geq 100$).

Three-stage selection A reached smaller LDM values than the other selection strategies with the exception of three-stage selection B (Table 2). This good performance was accompanied by few required MDP, only chromosome selection B in combination with small population sizes ($n \leq 40$) required less MDP.

Three-stage selection B reached the smallest LDM values of all selection strategies (Table 2). However, it required large numbers of MDP, only two-stage selection with small population sizes ($n \leq 60$) required more MDP. The selection pressure affects the carrier chromosome of the target segments since the first BC generation in three-stage selection B. Thus, linkage drag is here reduced most efficiently among all selection strategies.

Chromosome selection A reached lower LDM values than chromosome selection B but required more MDP (Table 2). The chromosome selection strategies did not reach LDM values as small as those of three-stage selection B. They also did not reach as small numbers of required MDP as those of three-stage selection A, with the exception of chromosomes selection B and small population sizes $n \leq 40$.

In the scenarios with decreasing and increasing population sizes from early to advanced generations of the BC_3S_2 breeding scheme (Table 1), only three-stage selection A and B reached LDM values <1 (Fig. 2). For all differences δ between

Table 2 Proportion of the recurrent parent alleles at markers outside the target segments (RPM), average number of donor alleles at markers outside the target segments per introgression line (ADM), number of introgression lines carrying donor alleles at markers outside the target segment (LDM), and total number of required marker data points in thousands (kMDP) for the BC_2S_3 and BC_3S_2 breeding schemes with five selection strategies and constant population sizes n in all generations

Selection strategy	Population size (n)									
	20[a]	40	60	80	100	120	140	160	180	200
BC_2S_3 breeding scheme										
Two-stage selection										
RPM	97.66	98.28	98.80	99.09	99.27	99.40	99.49	99.56	99.62	99.66
ADM	3.78	3.03	2.42	2.07	1.87	1.72	1.62	1.54	1.47	1.42
LDM	29.41	26.94	23.62	20.90	18.61	16.62	14.92	13.57	12.33	11.38
kMDP	36.45	48.09	66.67	84.49	101.72	118.53	134.98	151.20	166.89	182.77
Three-stage selection A										
RPM	97.43	98.19	98.79	99.15	99.37	99.53	99.64	99.72	99.78	99.82
ADM	4.48	3.71	3.18	2.88	2.66	2.47	2.32	2.17	2.06	1.96
LDM	27.04	23.20	17.98	14.06	11.26	9.07	7.41	6.15	5.09	4.30
kMDP	30.52	35.63	45.62	54.87	64.22	73.41	82.84	92.16	101.62	111.09
Three-stage selection B										
RPM	97.63	98.45	99.07	99.39	99.57	99.69	99.76	99.82	99.86	99.88
ADM	4.45	3.66	3.04	2.69	2.41	2.20	2.04	1.94	1.82	1.75
LDM	25.41	20.13	14.54	10.94	8.46	6.73	5.47	4.50	3.76	3.19
kMDP	35.95	49.33	68.43	86.02	102.60	118.45	134.05	149.44	165.40	181.16
Chromosome selection A										
RPM	96.30	97.31	98.13	98.63	98.96	99.20	99.38	99.49	99.59	99.66
ADM	5.87	4.79	4.02	3.57	3.26	2.99	2.77	2.59	2.44	2.30
LDM	29.97	26.77	22.13	18.31	15.21	12.77	10.80	9.28	7.93	6.96
kMDP	32.50	39.42	53.20	65.96	77.84	89.53	100.81	112.05	122.88	134.16
Chromosome selection B										
RPM	96.40	97.40	98.16	98.60	98.89	99.10	99.25	99.36	99.44	99.51
ADM	5.57	4.32	3.48	2.99	2.67	2.44	2.26	2.13	2.03	1.95
LDM	30.74	28.68	25.22	22.21	19.77	17.61	15.85	14.30	13.07	11.94
kMDP	28.46	34.19	46.36	58.26	69.98	81.70	93.17	104.50	115.80	127.00
BC_3S_2 breeding scheme										
Two-stage selection										
RPM	98.99	99.40	99.68	99.81	99.88	99.92	99.95	99.97	99.98	99.98
ADM	2.20	1.70	1.38	1.21	1.13	1.09	1.06	1.04	1.02	1.01
LDM	21.82	16.60	11.19	7.43	4.98	3.38	2.31	1.59	1.10	0.77
kMDP	35.53	53.37	74.70	94.92	114.57	133.82	152.87	171.27	189.68	207.86
Three-stage selection A										
RPM	99.16	99.62	99.85	99.93	99.97	99.98	99.99	99.99	100.00	100.00
ADM	2.35	1.84	1.53	1.37	1.32	1.25	1.23	1.20	1.16	1.21
LDM	17.06	9.88	4.65	2.24	1.16	0.62	0.34	0.21	0.12	0.07
kMDP	29.41	39.49	52.05	64.20	76.22	88.32	100.67	112.93	125.40	137.87
Three-stage selection B										
RPM	99.32	99.76	99.93	99.97	99.99	99.99	100.00	100.00	100.00	100.00
ADM	2.28	1.76	1.48	1.33	1.28	1.21	1.15	1.11	1.11	1.08
LDM	14.35	6.51	2.37	1.01	0.47	0.22	0.13	0.08	0.05	0.03
kMDP	33.39	52.15	74.10	95.24	115.60	135.59	155.25	174.24	193.51	213.00
Chromosome selection A										
RPM	98.80	99.43	99.76	99.89	99.94	99.97	99.98	99.99	99.99	100.00

Table 2 continued

Selection strategy	Population size (n)									
	20[a]	40	60	80	100	120	140	160	180	200
ADM	2.83	2.15	1.75	1.57	1.48	1.39	1.33	1.26	1.26	1.17
LDM	20.19	12.54	6.42	3.39	1.88	1.07	0.66	0.40	0.25	0.17
kMDP	29.55	42.23	58.23	73.14	87.94	102.12	116.41	130.57	144.25	158.25
Chromosome selection B										
RPM	98.62	99.24	99.62	99.79	99.88	99.93	99.95	99.97	99.98	99.99
ADM	2.89	2.19	1.75	1.54	1.41	1.32	1.26	1.21	1.19	1.16
LDM	22.55	16.53	10.21	6.42	4.11	2.64	1.76	1.15	0.80	0.56
kMDP	27.82	38.52	53.04	67.37	81.34	95.25	109.21	122.93	136.69	150.30

A donor genome coverage of 100% was reached for both breeding schemes irrespective of the selection strategy and the population size

[a] $n = 40$ in selfing generations

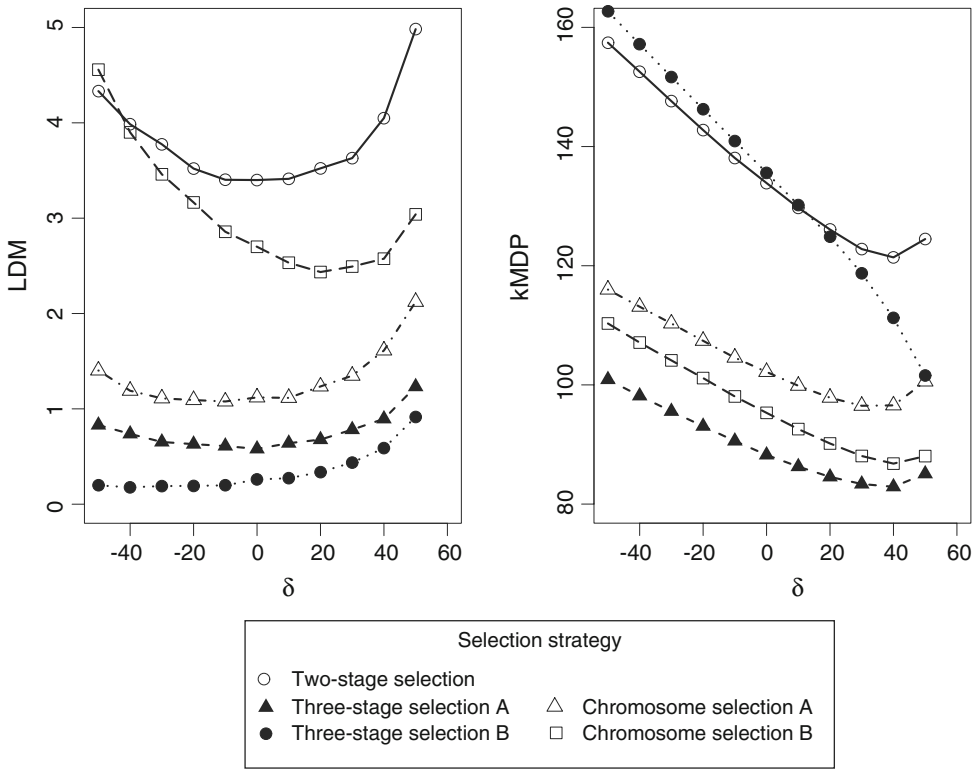


Fig. 2 Number of introgression lines carrying donor alleles at markers outside the target segment (LDM) and total number of required marker data points in thousands (kMDP) for the BC_3S_2 breeding scheme with decreasing and increasing population sizes from early to advanced generations and five selection strategies. δ denotes the difference in population size between two subsequent generations

the population sizes of subsequent generations, the lowest LDM values were reached by three-stage selection B and the lowest number of MDP were required by three-stage selection A. Three-stage selection A in combination with increasing population sizes ($\delta = 40$) reached an LDM value smaller than one with the minimum number of required MDP (≈80,000).

Discussion

Earlier simulation studies on introgression libraries

In a simulation study, Sušić (2005) investigated the development of introgression libraries in rye and determined that three generations of backcrossing are superior over two. His approach differs from our study in (1) the use of the recurrent parent genome content as success criterion, (2) the investigated population sizes, and (3) the employed selection strategies.

The recurrent parent genome content reached after six backcross generations with random selection of plants carrying the target gene was suggested as success criterion in simulations studies on marker-assisted backcrossing (Frisch et al. 1999a). The main focus in gene introgression programs is the phenotypic congruency of the recurrent parent and the converted line, which is highly correlated with the recurrent parent genome content of the converted line. Therefore, this criterion is suitable for marker-assisted

gene introgression programs. However, the main focus of introgression libraries is on obtaining a defined genomic composition of the introgression lines. This genomic composition is described in detail with the success criteria RPM, ADM, and LDM employed in our study.

The population sizes investigated by Sušić (2005) were held constant across all generations. However, increasing population sizes from early to advanced generations require less MDP than constant population sizes (Frisch et al. 1999a). Therefore, we investigated in our study also scenarios with increasing population sizes.

Population sizes of $n < 20$ in selfing generations bear the risk of reducing severely the donor genome coverage of the introgression library. Consider a backcross plant carrying a target segment of length $d = 0.2$ M. The probability that, after selfing, a progeny carries the target segment homozygous is $p = (1 - r)^2/4$ with $r = (1 - e^{-2d})/2$ when assuming no interference. To obtain with the probability of success q at least one plant carrying the complete target segment homozygous, population sizes of at least $n = \ln(1 - q)/\ln(1 - p)$ are required (cf. Frisch et al. 1999b). Thus, for $q = 99.9\%$, population sizes $n > 36$ are required. Therefore, we employed a minimum population size of $n = 40$ to avoid a loss of donor genome coverage through recombination in the target segments during the selfing generations.

The two-stage selection strategy adopted by Sušić (2005) requires more MDP than alternative selection strategies (Frisch et al. 1999a). However, the high costs for marker analyses are the limiting factor in the development of introgression libraries in rye. Therefore, we devised in our study new selection strategies developed especially for developing introgression libraries and compared their efficiency with respect to the MDP and LDM values.

Breeding scheme

As a starting point for our simulations, we choose a BC_2S_3 breeding scheme as carried out experimentally by Falke et al. (2008). However, LDM values smaller than one were not reached, even with population sizes of $n = 200$ (Table 2). In additional simulations (data not shown), we found that at least population sizes of $n = 600$ and 300 kMDP were necessary to reach LDM < 1 with the BC_2S_3 breeding scheme. Replacing one selfing generation by a backcross generation requires the same time for the development of the introgression library, but with the BC_3S_2 breeding scheme LDM values <1 were reached for many scenarios (Table 2). Consequently, the BC_3S_2 breeding scheme is better suited to develop introgression libraries in rye than the BC_2S_3 breeding scheme, because it efficiently reduces the donor genome outside the target segments.

Breeding schemes with four, five or six generations of backcrossing were investigated in further series of simulations. The increasing number of backcrossing generations reduced the necessary population sizes at the expense of time and number of MDP required (Table 3). Thus, breeding

Table 3 Proportion of the recurrent parent alleles at markers outside the target segments (RPM), average number of donor alleles at markers outside the target segments per introgression line (ADM), number of introgression lines carrying donor alleles at markers outside the target segment (LDM), and total number of required marker data points in thousands (kMDP) for BC_4S_2, BC_5S_2, and BC_6S_2 breeding schemes with three-stage selection A and constant population sizes n in all generations

Breeding scheme	Population sizes (n)								
	20[a]	30[a]	40	50	60	70	80	90	100
BC_4S_2									
RPM	99.67	99.83	99.91	99.96	99.98	99.99	99.99	100.00	100.00
ADM	1.63	1.41	1.29	1.18	1.12	1.08	1.05	1.02	1.01
LDM	9.53	5.66	3.48	1.77	0.91	0.48	0.24	0.12	0.07
kMDP	31.33	38.33	44.88	52.67	60.26	67.81	75.32	82.79	90.27
BC_5S_2									
RPM	99.83	99.93	99.97	99.99	100.00	100.00	100.00	100.00	100.00
ADM	1.34	1.19	1.11	1.06	1.03	1.02	1.01	1.01	1.00
LDM	5.98	2.83	1.37	0.57	0.22	0.10	0.04	0.02	0.01
kMDP	33.91	42.39	50.47	59.69	68.67	77.60	86.47	95.43	104.30
BC_6S_2									
RPM	99.90	99.97	99.99	100.00	100.00	100.00	100.00	100.00	100.00
ADM	1.20	1.09	1.04	1.02	1.00	1.00	1.00	1.00	1.00
LDM	3.89	1.51	0.59	0.20	0.06	0.02	0.01	0.00	0.00
kMDP	36.66	46.61	56.08	66.59	77.04	87.37	97.73	107.96	118.25

[a] $n = 40$ in selfing generations

schemes with four or more backcrossing generations have only a limited utility.

In conclusion, our results suggest that the BC_3S_2 breeding scheme is the most efficient for developing rye introgression libraries.

Selection strategies

The investigated selection strategies differ in (1) the selection pressure applied to the carrier versus non-carrier chromosomes of the target segments in backcross and selfing generations and (2) the number of simultaneously conducted backcross programs b (Fig. 1). These two factors affect both the genomic composition of the introgression library and the number of required MDP.

Three-stage selection differs from two-stage selection in the pre-selection of plants with recombination between the target segments and flanking markers, which is carried out between the foreground and background selection stages. This efficiently reduces the length of the donor genome (linkage drag) attached to the target segments (Frisch and Melchinger 2001). Since linkage drag is a main source of donor genome in converted lines when no selection for flanking markers is carried out, this reduction is the reason for the smaller LDM values of three-stage selection compared with two-stage selection. When three-stage selection is employed in generation BC_2 or later, only the few pre-selected recombinant plants are analyzed for the entire set of markers outside the target region. This explains the smaller number of MDP required for three-stage selection A compared with two-stage selection.

In three-stage selection A, a high selection pressure is imposed on the carrier chromosomes of the target segments starting with generation BC_2, whereas in three-stage selection B it is imposed starting with generation BC_1. This results in the most efficient reduction of the linkage drag and the lowest LDM values of three-stage selection B among all selection strategies. When three-stage selection is employed in generation BC_1, it does not result in a reduction of the required MDP, because analysis of the entire set of markers is required for all BC_1 plants to conduct foreground selection for the target segments. However, the higher selection pressure on the carrier chromosome of the target segments is accompanied by a smaller selection pressure on the remaining chromosomes. Thus, in generation BC_2, a higher proportion of heterozygous markers need to be analyzed with three-stage selection B than with three-stage selection A. This explains the greater number of MDP required with three-stage selection B.

Chromosome selection A and B differ from two- and three-stage selection in the selection of plants carrying entire donor chromosomes in generation BC_1. Thus, chromosome selection A and B require in BC_2 only as many simultaneously conducted backcross programs b as the species has chromosomes, i.e., seven in rye. In contrast, for two- and three-stage selection, separate backcross programs are needed for each target segment. A segment length of 20 cM and a genome length of 7 M result hence in $b = 35$ simultaneously conducted backcross programs. Consequently, the chromosome selection strategies have the advantage of saving money and time by reducing the effort in the breeding nursery. This aspect is of particular importance for developing introgression libraries with short target segments, because the development of such libraries needs otherwise a large number of simultaneously conducted backcross programs in generation BC_1.

In conclusion, the suitability of the investigated selection strategies depends on the goals during the development of the introgression library: (1) Three-stage selection B is superior to the other strategies if the lowest possible LDM values are desired. (2) Three-stage selection A is superior to the other strategies if the number of required MDP should be minimum by simultaneously obtaining low LDM values. (3) Chromosome selection A is superior to other selection strategies if the effort in the breeding nursery should be minimum.

Population sizes

The employed population sizes are, besides the breeding scheme and the selection strategy, an important factor affecting the genomic composition of introgression lines and the number of MDP required for developing introgression libraries.

Decreasing population sizes from early to advanced generations ($\delta < 0$) resulted in larger LDM values than constant or increasing population sizes ($\delta \geq 0$; Fig. 2). Further, they require more MDP, because a large numbers of markers, which are not yet fixed for the RP allele, are analyzed in the large populations of early generations. Consequently, decreasing population sizes show no advantages with any of the investigated selection strategies.

With increasing population sizes ($\delta > 0$) from early to advanced generations, a minimum of the required MDP was observed at $\delta \approx 40$ for all selection strategies except three-stage selection B (Fig. 2). At this minimum, three-stage selection A required the least MDP of all selection strategies and simultaneously reached LDM values smaller than one. We, therefore, conclude that the optimal design to develop a rye introgression library carrying on expectation less than one line with donor alleles outside the target segment is a BC_3S_2 breeding scheme with three-stage selection A and increasing population sizes.

Comparison with experimental data

The rye introgression libraries of Falke et al. (2008) were developed with a BC_2S_3 breeding scheme and a sequential selection strategy consisting of three stages. (1) Foreground selection for the target segments. (2) Pre-selection based on a low number of additional donor chromosome segments outside the target segments. (3) Selection of the final plants for backcrossing/selfing on basis of the RP alleles at markers outside the target segments. This procedure can be regarded as a refinement of the two-stage selection strategy described in this study and, therefore, the results of Falke et al. (2008) can be compared with the presented simulations. Each library consists of 40 BC_2S_3 introgression lines, which contained on average five (introgression library A) and three (introgression library B) target segments and covered 74 and 59% of the donor genome. Thus, the average number of target segments per line is in good agreement with our ADM values for the BC_2S_3 breeding scheme using two-stage selection and small population sizes. In contrast, the donor genome coverage and the number of required MDP of Falke et al. (2008) were lower than in our simulations. From a retrospective point of view, our simulations suggest that employing the BC_3S_2 breeding scheme, three-stage selection A, and larger population sizes in the selfings could be regarded as an optimized design of this experiment.

The results of our study showed that the breeding scheme, the selection strategy, and the population sizes considerably influence the genomic composition of the introgression lines and the effort for developing introgression libraries. Consequently, an optimal design is of crucial importance for efficiently developing introgression libraries. The guidelines developed in this study may help plant breeders and geneticists to successfully develop introgression libraries in particular in rye, but also in other crops with a similar genome size, such as barley. Moreover, the fine tuning of breeding programs for the development of introgression libraries can be investigated with individual simulations depending on the specific goals in developing the library.

Acknowledgments The present study was financially supported by a grant from the German Federal Ministry of Food, Agriculture and Consumer Protection (BMELV) via the Federal Agency for Agriculture and Food (BLE) and the "Gemeinschaft zur Förderung der privaten deutschen Pflanzenzüchtung" (GFP), Grant No. PGI-06.01-28-1-43.017-07.

References

Eshed Y, Zamir D (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. Euphytica 79:175–179

Eshed Y, Abu-Abied M, Saranga Y, Zamir D (1992) *Lycopersicon esculentum* lines containing small overlapping introgressions from *L. pennellii*. Theor Appl Genet 83:1027–1034

Falke KC, Sušić Z, Hackauf B, Korzun V, Schondelmaier J, Wilde P, Wehling P, Wortmann H, Mank J, Rouppe van der Voort J, Maurer H, Miedaner T, Geiger HH (2008) Establishment of introgression libraries in hybrid rye (*Secale cereale* L.) from an Iranian primitive rye accession as a new tool for rye breeding and genomics. Theor Appl Genet 117:641–652

Falke KC, Wilde P, Wortmann H, Geiger HH, Miedaner T (2009a) Identification of genomic regions carrying QTL for agronomic and quality traits in rye (*Secale cereale* L.) introgression libraries. Plant Breed (in press)

Falke KC, Sušić Z, Wilde P, Wortmann H, Möhring J, Piepho H-P, Geiger HH, Miedaner T (2009b) Testcross performance of rye introgression lines developed by marker-based backcrossing using an Iranian accession as donor. Theor Appl Genet 118:1225–1238

Falke KC, Wilde P, Miedaner T (2009c) Rye introgression lines as source of alleles for pollen-fertility restoration in Pampa CMS. Plant Breed. doi:10.1111/j.1439-0523.2008.01589.x

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 (2001) Marker-assisted backcrossing for simultaneous introgression of two genes. Crop Sci 41:1716–1725

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

Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. PLoS Biol 2:1610–1615

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Hospital F (2005) Selection in backcross programmes. Phil Trans R Soc B 360:1503–1511

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

Prigge V, Maurer HP, Mackill DJ, Melchinger AE, Frisch M (2008) Comparison of the observed with the simulated distributions of the parental genome contribution in two marker-assisted backcross programs in rice. Theor Appl Genet 116:739–744

Stam P (1979) Interference in genetic crossing over and chromosome mapping. Genetics 92:573–594

Sušić Z (2005) Experimental and simulation studies on introgressing genomic segments from exotic into elite germplasm of rye (*Secale cereale* L.) by marker-assisted backcrossing. PhD thesis. Universitaet of Hohenheim, Stuttgart, Germany

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92:213–224

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