Institute of
Plant Breeding, Seed Science, and Population Genetics
University of Hohenheim
Chair of Population Genetics
Prof. Dr. Dr. h. c. H. H. Geiger

Experimental and simulation studies on introgressing genomic segments from exotic into elite germplasm of rye (*Secale cereale* L.) by marker-assisted backcrossing

Dissertation
submitted in fulfilment of the requirements for the degree
"Doktor der Agrarwissenschaften"
(Dr. sc. agr. / *Ph. D.* in Agricultural Sciences)
to the
Faculty of Agricultural Sciences

presented by
M. Sc. in Agricultural Sciences
Zoran Sušić
from Smederevska Palanka (Serbia and Montenegro)

Stuttgart – Hohenheim
2005

This thesis was accepted as a doctoral dissertation in fulfilment of the requirements for the degree “Doktor der Agrarwissenschaften (Dr. sc. agr. / *Ph. D.* in Agricultural Sciences)“ by the Faculty of Agricultural Sciences at the University of Hohenheim, on 11.10.2005.

Day of oral examination:	30.08.2005.

Examination Committee

Vice-Dean and Head of the Committee:	Prof. Dr. K. Stahr
Supervisor and Reviewer:	Prof. Dr. Dr. h. c. H. H. Geiger
Co-reviewer:	Prof. Dr. H.-P. Piepho
Additional examiner:	Prof. Dr. A. Schaller

Table of Contents

List of Abbreviations

BC - backcross
DC - donor chromosome
DG - donor genome
IL - introgression line
MAS - marker-assisted selection
MDP - marker data points
RPG - recurrent parent genome
S - selfing

1 General Introduction

Crop improvement is highly dependent upon finding and using genetic variation. Nevertheless, years of cultivation and selection by both farmers and breeders inevitably lead to reduction in genetic variation in elite germplasm (Hawks 1977, Goodman 1997). To increase genetic diversity and thus ensure long-term selection gain of elite breeding materials, the introgression of exotic germplasm is a promising approach (Tanksley and Nelson 1996). Primitive races and related species contain a wide range of variation for many traits, and for some traits (usually resistance traits) they may represent the only source of desirable genes (Hawks 1977, Goodman et al. 1987, Stalker 1980).

The potential value of exotic germplasm is well-known to plant breeders. However, it has not been intensively utilized in modern plant breeding due to a number of reasons. (1) Exotic germplasm lacks environmental adaptation that is of utmost importance for variety improvement. (2) There is a significant difference in performance between elite and exotic germplasm for polygenic traits. (3) Exotic germplasm of cross-fertilizers is lacking inbreeding tolerance and generally is not assigned to known heterotic groups, the two key-issues of hybrid breeding. (4) Genetic problems such as pleiotropy, coupling phase linkage between desired and undesired alleles, as well as epistasis may hinder a direct utilization of plant genetic resources in modern breeding (Haussmann et al. 2004).

Despite their agronomically inferior phenotypes, exotic germplasm may contain genomic segments that can improve oligo- and polygenically inherited traits even in highly-selected breeding populations (Frey et al. 1981, de Vicente and Tanksley 1993). This has already been shown in a number of studies involving wild tomato species *Lycopersicon pennellii* (Eshed and Zamir 1994, 1995, Eshed et al. 1996, Ronen et al. 2000), *L. hirsutum* (Bernacchi et al. 1998a, 1998b, Monforte and Tanksley 2000, Monforte et al. 2001) and *L. peruvianum* (Fulton et al. 1997). Agriculturally unadapted (exotic) sources were also used for the broadening of genetic diversity in breeding populations of maize (Ragot et al. 1995), sorghum (Tuinstra et al. 1998), rice (Xiao et al. 1998, Lin et al. 2000, Yan et al. 2002), and barley (Pillen et al. 2003, von Korff et al. 2004).

To expand the variability in hybrid rye breeding populations, East European cultivars, landraces from Europe, Asia and South America, as well as primitive populations from the Near East could be used as genetic resource. These resources have so far been used only for

extracting monogenically inherited traits such as self-fertility (Ossent 1938), resistance to powdery mildew or leaf rust (Rollwitz 1985). Furthermore, exotic germplasm was indispensable for establishing a hybridizing mechanism in rye, as the CMS-inducing Pampa cytoplasm was derived from an Argentinian landrace (Geiger and Schnell 1970) and effective restorer genes originated from Iranian and South American collections (Miedaner et al. 2000). A proper management of new variability in hybrid rye breeding can additionally enhance the genetic distance between the seed-parent and pollinator gene pools and thus contribute to an increase in heterosis (Geiger and Miedaner 1999). Unfortunately, non-adapted (exotic) germplasm is difficult to use in hybrid rye breeding, particularly for improving quantitative traits, because of its: i) low performance level, ii) high mutational load, and iii) unknown genetic distance to established heterotic pools.

Broadening the genetic base of elite breeding materials by introgressing exotic germplasm requires techniques that would minimize reduction in productivity by interfering genetic interactions between recipient and donor. This appears achievable by an introgression library approach in which introgression is restricted to one or a few short donor chromosome (DC) segments (Eshed et al. 1992). An introgression library consists of a set of lines, each carrying a single marker-defined DC segment introgressed from an agriculturally unadapted source into the background of an elite variety (Zamir 2001). Ideally, the introgressed DC segments are evenly distributed over the whole recipient genome and the total genome of the exotic donor is comprised in the established set of near-isogenic lines (NILs).

General breeding scheme for the establishment of an introgression library

The procedure for the establishment of an introgression library implies the systematic transfer of DC segments from a genetic resource accession (donor) into an elite line (recipient, recurrent parent) by marker-assisted backcrossing.

A breeding scheme for the establishment of an introgression library always starts from a cross between an elite line and exotic donor (Fig. 1). Once an F_1 generation is created, various numbers of backcross (BC) generations are produced in order to increase the proportion of the recurrent parent genome (RPG). Simultaneously, the proportion of the donor genome (DG) and the number of DC segments per introgression line (IL) are reduced. At the end of the BC procedure, the introgressed short DC segments are present in the heterozygous state (Fig. 1).

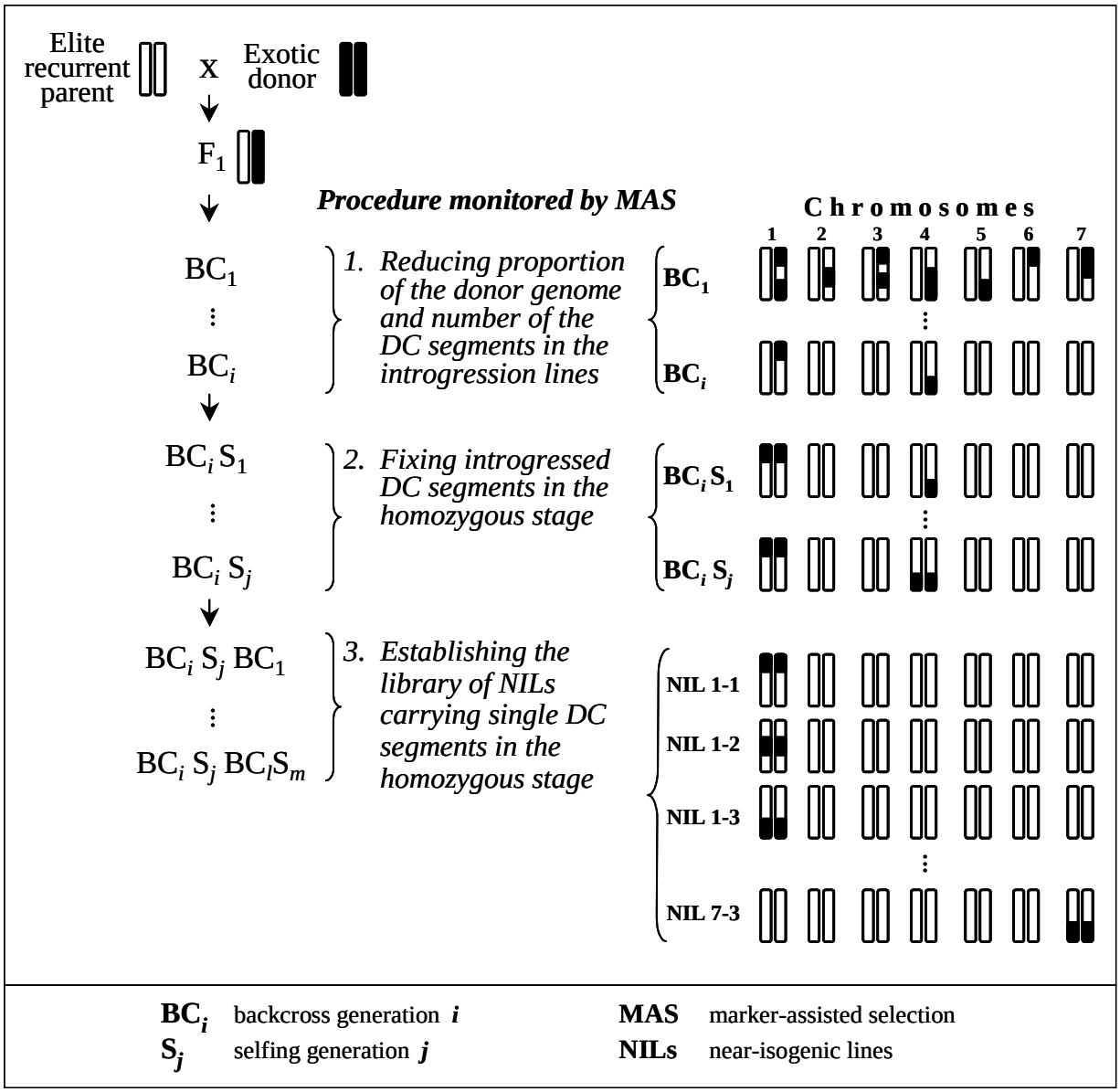


Figure 1. General breeding scheme for the establishment of an introgression library. Reduction in the donor genome proportion and the number of the donor chromosome (DC) segments during the introgression procedure is illustrated on the right-hand side. Each pair of bars represents one homologuous chromosome. The whole rye genome consists of seven chromosomes. The resulting introgression library (covering the complete donor genome) is shown for the first three NILs covering chromosome 1 (NIL 1-1, NIL 1-2, and NIL 1-3) as well as for the last NIL covering the proximal region of the chromosome 7 (NIL 7-3).

Additional selfing (S) generation(s) and marker-assisted selection (MAS) result in a library consisting of ILs with homozygous DC segments and covering the complete donor genome. Nevertheless, if the established introgression library comprises lines carrying more than one DC segment, additional backcrossing and selfing, in conjunction with MAS, have to be performed to obtain NIL harbouring single homozygous DC segments (Fig. 1).

Objectives

The main goals of this study were:

- Marker-assisted introgression of chromosome segments from a primitive rye population (donor) into an elite inbred line (recipient),
- Application of computer simulations to establish an effective and cost-efficient marker-assisted introgression strategy for creating an introgression library in rye.

Correspondingly, the thesis is organized in two sections (A and B, respectively).

2 Section A
Establishment of two rye introgression libraries by marker-assisted backcrossing

2.1 Introduction

For the purpose of genetic analysis, a set of NILs representing the complete DG has several advantages over other types of segregating populations, such as F_2, F_3, BC, or recombinant inbred line populations. Lines in the library differ from the elite line by only a small, defined chromosomal segment, and phenotypic differences between a line in the library and the elite line can be associated with the genes located in a single DC segment (Eshed et al. 1996). This represents the major advantage of the NIL approach, which is the elimination of genetic variation not associated with the introgressed DC segment (Eshed et al. 1996). Furthermore, the permanent genetic constitution of the homozygous NILs in the library enables to test the effects of the DC segments in different environments, with a high reproducibility of gene effects harboured in introgressed DC segments (Eshed et al. 1996, Paterson et al. 1990). Finally, an important advantage of the NIL approach is that the genetic analysis is performed on an elite genetic background being directly usable for the development of new varieties.

In respect of the simultaneous development of plant materials and the identification of genome regions controlling qualitative and quantitative traits, the introgression library approach has been suggested as an efficient method (Kaeppler et al. 1993, Kaeppler 1997, Stuber et al. 1999). Mapping quantitative trait loci (QTL), based on segregating populations, provides only rough positions of QTL (for review see Kearsey 2002). Since the QTL approach is impaired by the great genetic variation in the genetic background, NILs have been utilized as a resource to improve QTL analysis (Kaeppler 1997, Kearsey 2002). However, the approach of establishing an introgression library is based on systematic development of potentially superior ILs without any prior identification of the effects of possible QTL carried by the introgressed DC segment. The identification and mapping of QTL is an extra benefit at no additional costs, which can be obtained when the established ILs are evaluated (Stuber 1999). The usefulness of NILs in mapping QTL has already been demonstrated in various crops, such as tomato (Eshed and Zamir 1995, Eshed et al. 1996, Monforte et al. 2001, Brouwer and St Clair 2004), maize (Koester et al. 1993, Salvi et al. 2002), rice (Lin et al.

2000, Ioannidou et al. 2003), barley (Han et al. 1999, Matus et al. 2003), and bread wheat (Prasad et al. 2003).

In addition, NILs facilitate fine-mapping of valuable DC segments and/or QTL (Paterson et al. 1990) and may lead towards gene discovery (Zamir 2001). Fine-mapping techniques allow to reduce the confidence interval of a QTL position and thus focus on a smaller sub-set of possible candidate genes that might be responsible for the trait in question (Kearsey 2002). Having identified a small number of potential candidate genes, the final stage of the procedure involves various approaches to identify the particular candidate gene that is responsible for the polymorphism identified by the QTL (Albert and Tanksley 1996).

On the other hand, the introgression library approach also has some shortcomings. The main drawback is the length of time and amount of work required for the development of introgression libraries (Zamir and Eshed 1998, Kearsey 2002). This particularly applies to libraries covering the complete donor genome and thus comprising a large number of ILs. In addition, experimental design (number of locations and repetitions) needs to be considered in experiments comparing ILs for differences due to short DC regions. Such experiments need to ensure that any differences among tested ILs are truly genetic and not due to environmental effects. The smaller the DC segment effects to be studied, the more important these precautions become.

Although a great research effort is required to generate introgression libraries, plant breeders express increasing interest in this approach (Mank et al. 2003). Introgression libraries have been developed for tomato (*Lycopersicon esculentum*) and its wild relatives. These include *L. pennellii* (Eshed et al. 1992, Eshed and Zamir 1994), *L. hirsutum* (Monforte and Tanksley 2000), and *Solanum lycopersicoides* (Chetelat and Meglic 2000). In lettuce (*Lactuca sativa*), Jeuken and Lindhout (2004) established a backcross inbred line library from a cross between *L. sativa* (recurrent parent) and *L. saligne* (wild lettuce, donor). Introgression library has also been developed in barley (*Hordeum vulgare*), using an exotic barley accession (*H. vulgare* ssp. *spontaneum*) as a donor (von Korff et al. 2004).

A somewhat similar approach integrating variety development and identification of genome regions affecting quantitative traits is the advanced backcross-QTL analysis (AB-QTL) proposed by Tanksley and Nelson (1996). In contrast to the introgression library approach the AB-QTL scheme applies negative (phenotypic) selection to reduce frequency of deleterious donor alleles through the BC_2 and BC_3 generation. The QTL are then mapped in an advanced backcross generation (BC_2 or BC_3) and ILs specific for the detected QTL are

subsequently developed. Because the AB-QTL scheme uses backcrossing based on phenotypic selection, the resulting lines depend on a scholastic approach, and the probability of collectively encompassing all of the genetic material from the donor (*i.e*. having each of the DC segments represented in at least one IL) is very low (Beckmann and Soller 1986) unless large number of lines is generated and evaluated (Stuber et al. 1999). In addition, if the goal is to improve a trait such as yield, phenotypic selection in early BC generations could be counter-productive. Favourable yield genes may be linked to the genes associated with the deleterious or undesirable traits, and would be eliminated during the backcrossing process (Stuber et al. 1999). Nevertheless, the AB-QTL approach was successfully used in tomato (Bernacchi et al. 1998a, 1998b, Fulton et al. 2000) and barley (Pillen et al. 2003) to introgress QTL from the exotic into cultivated varieties.

Objectives

The systematic development of an introgression library in rye, providing a complete coverage of the DG, has not been described yet. The main objective was to establish two rye introgression libraries by marker-assisted backcrossing. Each IL should carry one to three DC segments, and DC segments carried by the established ILs should jointly represent most of the donor genome.

2.2 Materials and Methods

Plant material

A cross between a homozygous rye inbred line L2053-N (recurrent parent) and a heterozygous Iranian primitive rye population Altevogt 14160 (donor) was used to generate two rye populations (F and G). The recurrent parent L2053-N is an elite inbred line already used as a parent in several registered hybrid rye varieties (H. Wortmann, personal communication). The donor population, provided by the Polish Botanical Garden at Warsaw, represents a primitive rye population found as a weed in wheat and barley fields in the Near East. The population is characterized by early heading, non-shattering ears, high susceptibility to lodging, low kernel weight and grain yield.

Repeated backcrossing and subsequent selfing till BC_2S_3 was chosen as introgression method (Fig. 2).

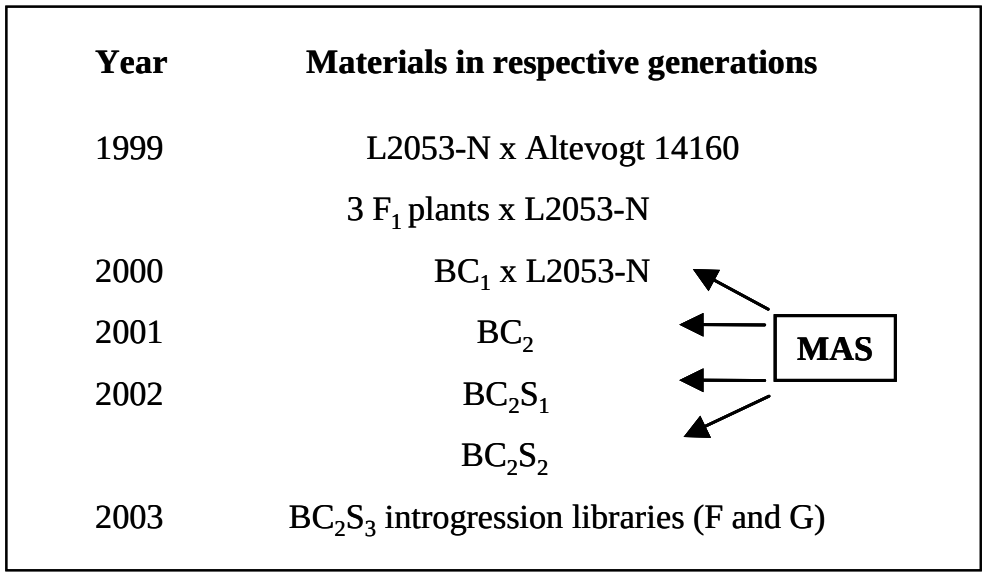


Figure 2. Breeding scheme for the build-up of genetic materials. For abbreviations see Fig. 1.

The backcrossing procedure was started with three F_1 plants. After two cycles of backcrossing and three cycles of selfing, the progeny of one F_1 plant was used to develop introgression library F, and the remaining two to develop a combined introgression library G.

Mapping analyses

Due to the lack of sufficient SSR markers in rye at the beginning of the study, initially constructed genetic maps in generation BC_1 of both populations were based on AFLP marker data only. The genetic maps were based on 90 randomly chosen BC_1 individuals in the library F, and 88 individuals in the library G. To identify AFLP primer combinations (PCs) being polymorphic in populations F and G, 48 PCs were screened. Since the donor population was heterozygous, BC_0 plant samples were used for the screening. Based on the number and scorability of DNA fragments, as well as their distribution over the fingerprint, 13 PCs were selected to construct the genetic map of population F, and 10 PCs for population G.

After the first rye SSR markers had become available (Hackauf and Wehling 2002a, b), these were stepwise included into both maps. Combined (AFLP+SSR) genetic maps were based on 76 BC_1 individuals in population F and 88 BC_1 individuals in population G. The software package JoinMap 3.0 (Van Ooijen and Voorrips 2001) was applied to calculate genetic linkage maps of both populations. To assign markers to linkage groups, pairwise comparisons and grouping of markers were performed applying a LOD value of 3.0. Using the Kosambi mapping function (Kosambi 1944), recombination frequencies among marker loci were converted into centiMorgan (cM) units.

The AFLP marker assays (Vos et al. 1995) using co-dominant scoring were conducted by Keygene N. V. (Wageningen, The Netherlands). The SSR marker data (Oetting et al. 1995) were assessed by the Federal Centre for Breeding Research on Cultivated Plants (BAZ, Groß Lüsewitz, Germany), PLANTA GmbH (Einbeck, Germany) and SAATEN-UNION Resistance Laboratory GmbH (Leopoldshöhe, Germany).

Marker-assisted selection

Genetic maps based on AFLP markers were used to identify and monitor DC segments, and to select the best-suited progenies in generations BC_1 and BC_2. Afterwards, in generations BC_2S_1 and BC_2S_2, monitoring and selection were done using SSR markers along with AFLP markers. Monitoring of DC segments throughout the introgression program was carried out by at least two flanking markers. For DC segments longer than 10 cM, additional markers were used within the DC segment.

Graphical genotypes of the analyzed individuals were generated applying the software packages Genome Typer (proprietary software package of Keygene N. V.) and GGT (van

Berloo 1999). Graphical genotypes allowed the monitoring of DC segments over the introgression generations and facilitated the detection of too frequent double crossovers.

Graphical genotypes do not show the exact location of crossovers. As the most likely position of a crossover, half the distance between the two closest markers is used by both software. Interpretation of crossover events should therefore be done cautiously, particularly if crossover occurs between markers positioned at large distances.

The following criteria were used to select parent plants:

i) chromosomal localization of DC segments: to cover most of the DG ILs were selected if they harboured a particular target DC segment.

ii) number of DC segments per IL: to develop final ILs harbouring not more than three DC segments, an IL was selected if it carried not more than five DC segments in generation BC_2, and not more than four DC segments in generation BC_2S_1 (maximum two of them in the heterozygous stage).

iii) proportion of RPG per IL: to facilitate a rapid restoration of recurrent parent genetic background, among ILs carrying identical target DC segments those with higher proportion of RPG were preferred.

2.3 Results

2.3.1 Genetic maps

Introgression library F

The first genetic map of population F comprised 137 AFLP markers, with a total map length of 572.8 cM (Appendix, Fig. A1). The markers were assigned to seven major linkage groups corresponding to the seven rye chromosomes, and two minor uncoupled linkage groups (5R-a and A). Even though satisfactory marker density was achieved (approx. one marker every 4 cM), several gaps remained within the genome, particularly on chromosomes 1R, 4R, 6R, and 7R.

The combined genetic map of population F contained 137 AFLP and 59 SSR markers (Appendix, Fig. A2). The SSR markers bridged some of the gaps between AFLP marker positions (chromosomes 4R and 7R) and extended the map towards the distal regions of the chromosomes (particularly of chromosomes 2R, 3R, and 6R). Moreover, three SSR markers (SCM76, SCM174, and SCM179) on the distal region of the chromosome 5R allowed to join the previously uncoupled linkage group 5R-a to chromosome 5R. Thus, the length of the final, combined genetic map of population F was 683.1 cM.

Introgression library G

The AFLP genetic map of population G comprised 194 markers, with a total map length of 615.4 cM (Appendix, Fig. A3). The AFLP markers were classified into seven major linkage groups corresponding to seven rye chromosomes, and one minor uncoupled linkage group 1R-a. Although a reasonable marker density was achieved (average distance between markers was 2.5 cM), an unequal marker distribution resulted in several gaps within the genome, especially on chromosomes 2R, 5R, and 6R.

The combined genetic map of population G comprised 194 AFLP and 56 SSR markers (Appendix, Fig. A4). Similarly to the genetic map used for establishing library F, SSR markers bridged large gaps between AFLP marker positions (on chromosomes 1R, 5R, and 6R) and stretched the map towards the distal regions of the chromosomes (5R and 6R). New SSR markers in the distal region of the chromosome 1R (SCM127a_01, FvLRMS 66, and SCM127b_01) allowed to join the previously uncoupled linkage group 1R-a to chromosome 1R, extending it for additional 46 cM. Consequently, the length of the combined map of

population G reached 685.2 cM, which is approx. 10% longer than the map initially obtained by AFLP markers only (Appendix, Fig. A4).

2.3.2 Marker-assisted backcrossing program

The number of analysed and selected individuals and primers used to identify and monitor DC segments varied in each introgression generation of both libraries (Table 1). During the introgression process, the total number of individuals analysed by molecular markers was increased and the number of primer combinations per line was reduced (Table 1).

Table 1. Number of selected individuals and primers used to characterize each generation of introgression libraries F and G.

Generation	Process	Number of plants		Number of primer combinations	
		Analysed	Selected	AFLP	SSR
	---------- Introgression library F ----------				
	Screening of parents (Recurr. par. / Exotic donor)	2		48	
BC_1	Mapping (AFLP)	90		13	-
BC_1	Mapping (AFLP+SSR)	76		13	59
BC_1	Selection	68	9	13	-
BC_2	Selection	154	19	5-8 / line	-
BC_2S_1	Selection	190	17	1-5 / line	39
BC_2S_2	Selection	256	40	1 / line	20
	---------- Introgression library G ----------				
	Screening of parents (Recurr. par. / Exot. donor)	2		48	-
BC_1	Mapping (AFLP)	88		13	-
BC_1	Mapping (AFLP+SSR)	88		13	56
BC_1	Selection	69	9	10	-
BC_2	Selection	196	18	5-7 / line	-
BC_2S_1	Selection	133	22	2-5 / line	36
BC_2S_2	Selection	267	40	1-2 / line	32

Frequency distributions for the (*i*) number of DC segments (Library F, Fig. 3; Library G, Fig. 4), (*ii*) total DC segment length (Library F, Fig. 5; Library G, Fig. 6), and (*iii*) recovered RPG proportion (Library F, Fig. 7; Library G, Fig. 8), observed in the complete set of analysed BC_1 plants prior to selection were compared with respective values in the finally selected set of BC_2S_2 plants (lines). As desired, marker-assisted selection resulted in a low number of DC segments, a low total DC segment length , and a high RPG proportion.

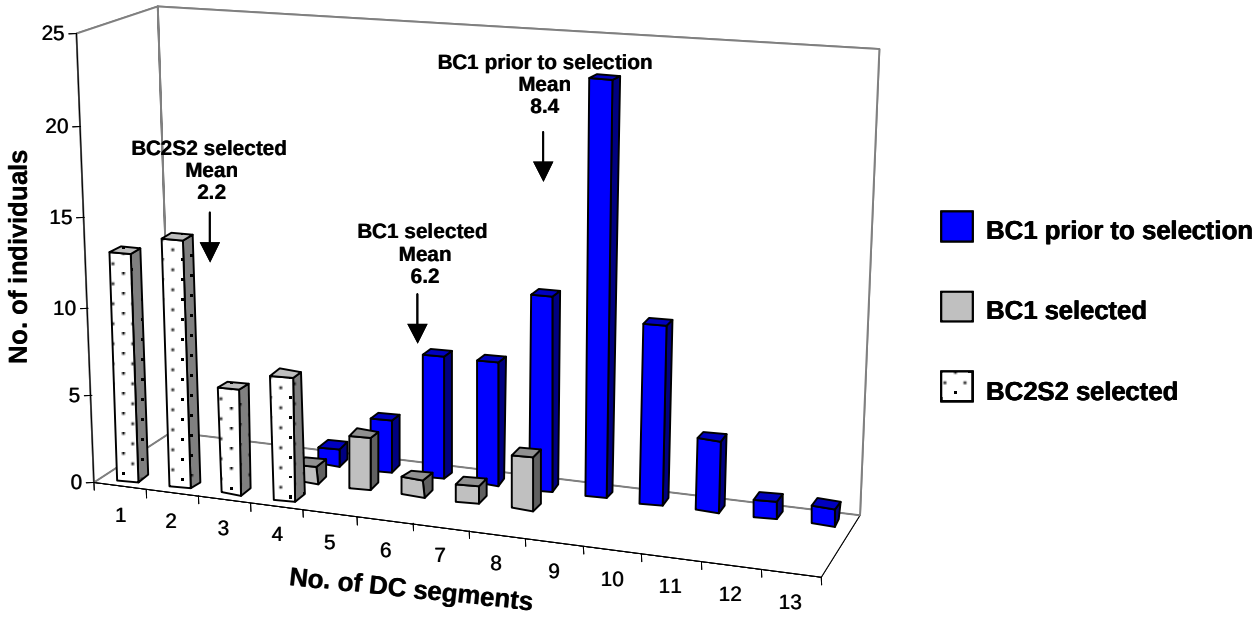


Figure 3. Library F: Frequency distribution of the number of donor chromosome (DC) segments per individual in the complete set of analysed BC_1 individuals prior to selection (68), the selected fraction of 9 BC_1 individuals, and in the finally selected set of 40 BC_2S_2 individuals (lines).

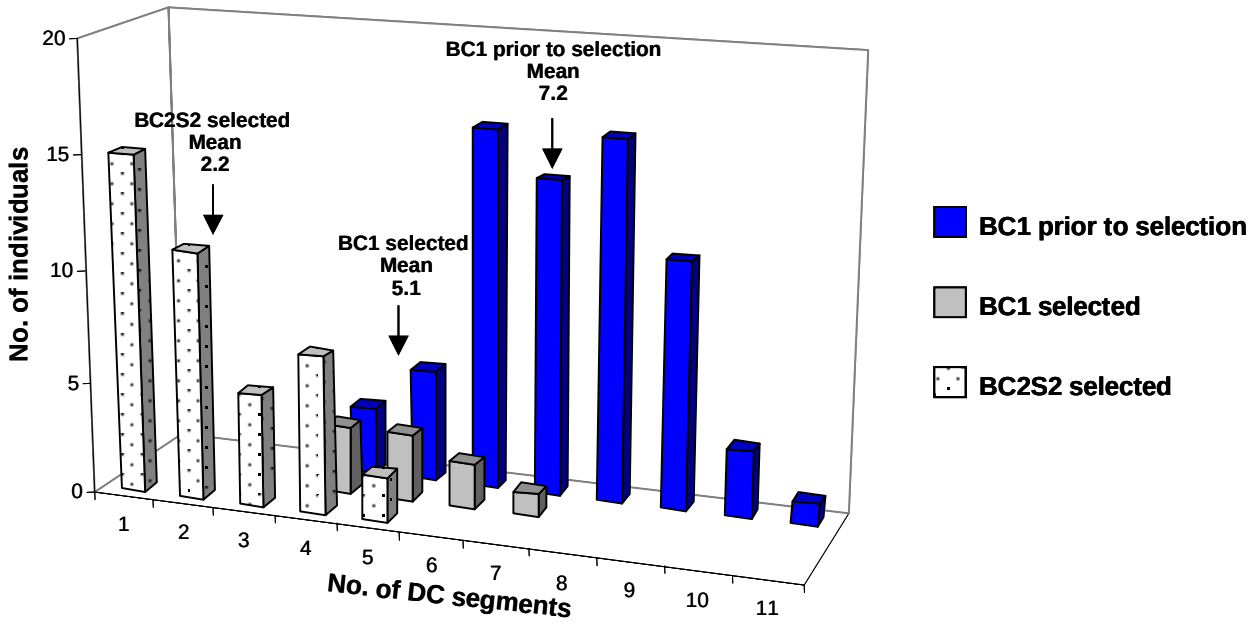


Figure 4. Library G: Frequency distribution of the number of donor chromosome (DC) segments per individual in the complete set of analysed BC_1 individuals prior to selection (69), the selected fraction of 9 BC_1 individuals, and in the finally selected set of 40 BC_2S_2 individuals (lines).

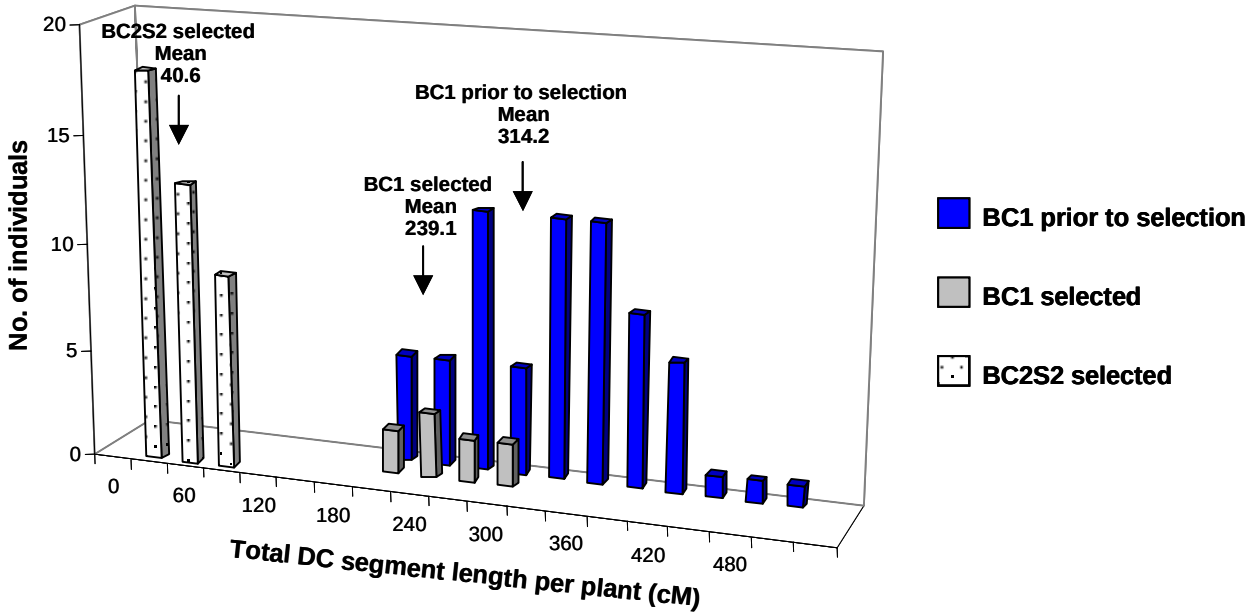


Figure 5. Library F: Frequency distribution of the total donor chromosome (DC) segment length per individual in the complete set of analysed BC_1 individuals prior to selection (68), the selected fraction of 9 BC_1 individuals, and in the finally selected set of 40 BC_2S_2 individuals (lines).

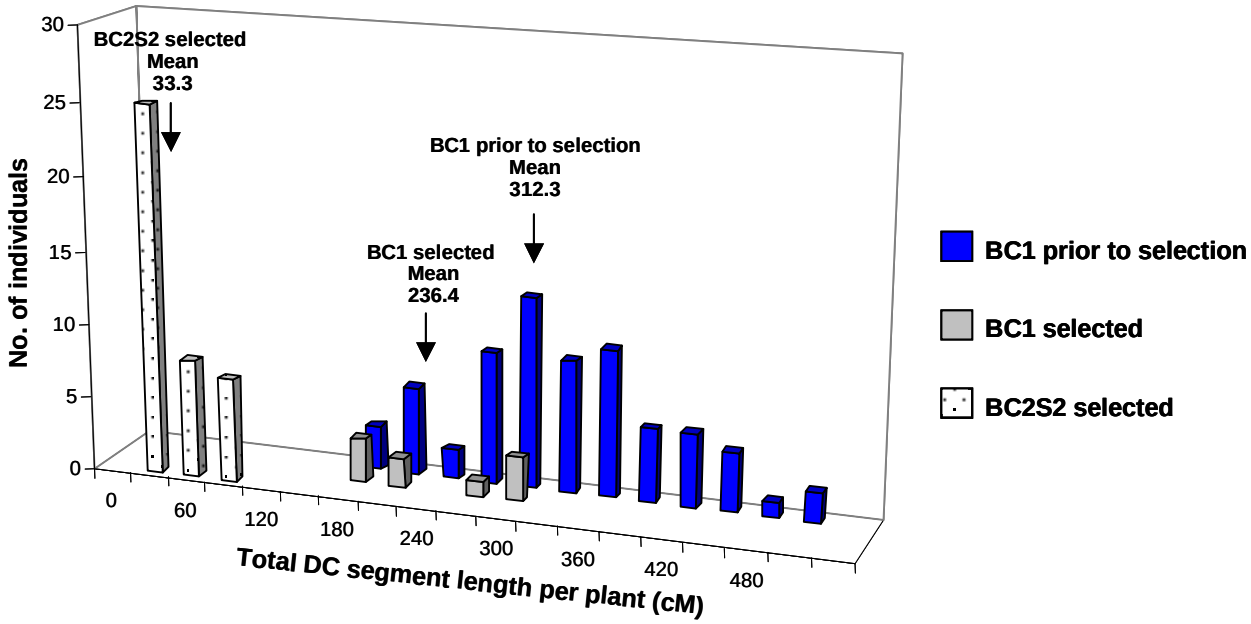


Figure 6. Library G: Frequency distribution of the total donor chromosome (DC) segment length per individual in the complete set of analysed BC_1 individuals prior to selection (69), the selected fraction of 9 BC_1 individuals, and in the finally selected set of 40 BC_2S_2 individuals (lines).

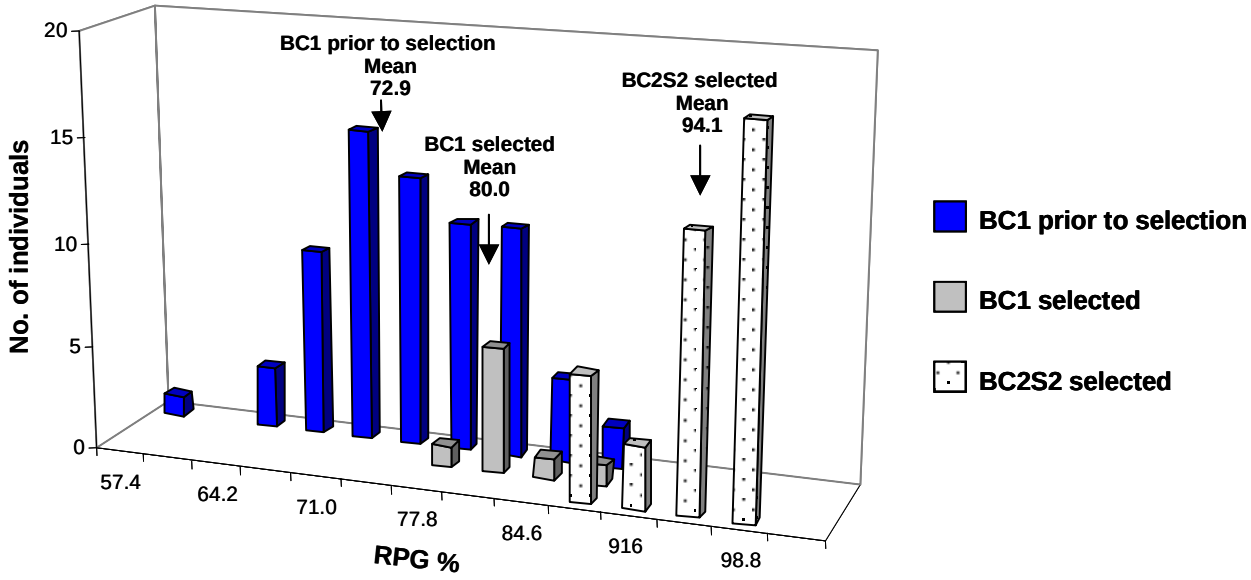


Figure 7. Library F: Frequency distribution of the recurrent parent genome (RPG) proportion in the complete set of analysed BC_1 individuals prior to selection (68), the selected fraction of 9 BC_1 individuals, and in the finally selected set of 40 BC_2S_2 individuals (lines).

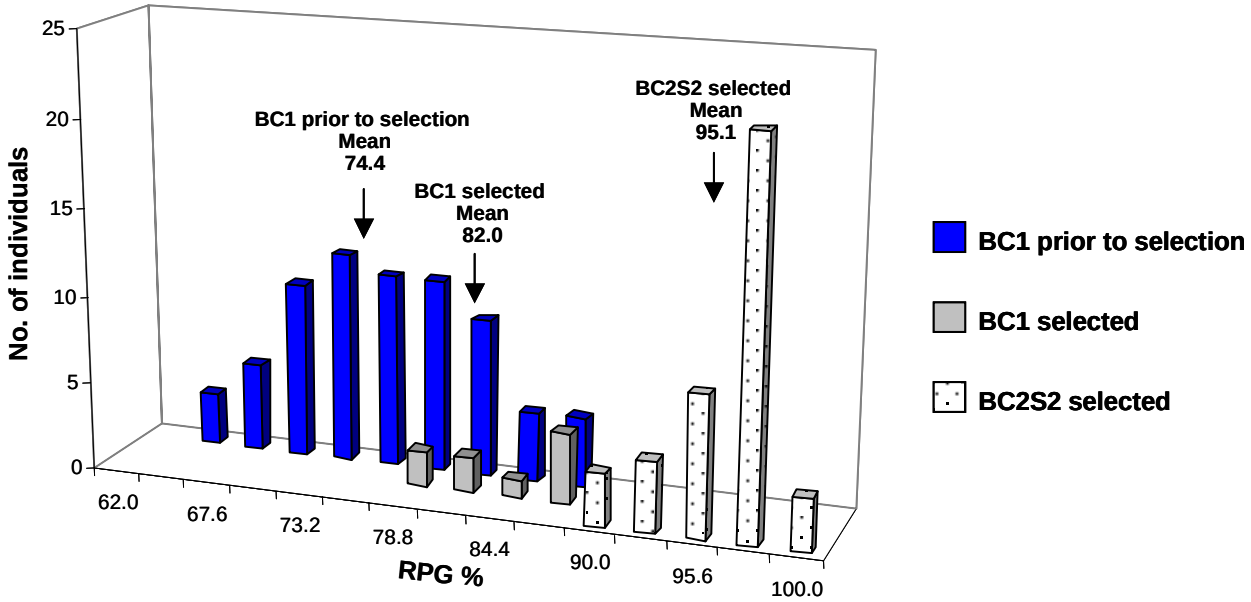


Figure 8. Library G: Frequency distribution of the recurrent parent genome (RPG) proportion in the complete set of analysed BC_1 individuals prior to selection (69), the selected fraction of 9 BC_1 individuals, and in the finally selected set of 40 BC_2S_2 individuals (lines).

The criteria used for MAS did not affect the proportion of short and long DC segments harboured in the complete and selected sets of BC_1 plants, neither in library F nor in G. Thus, the range and mean values for the length of individual DC segments were almost equal in the complete and selected set of BC_1 plants (data not shown).

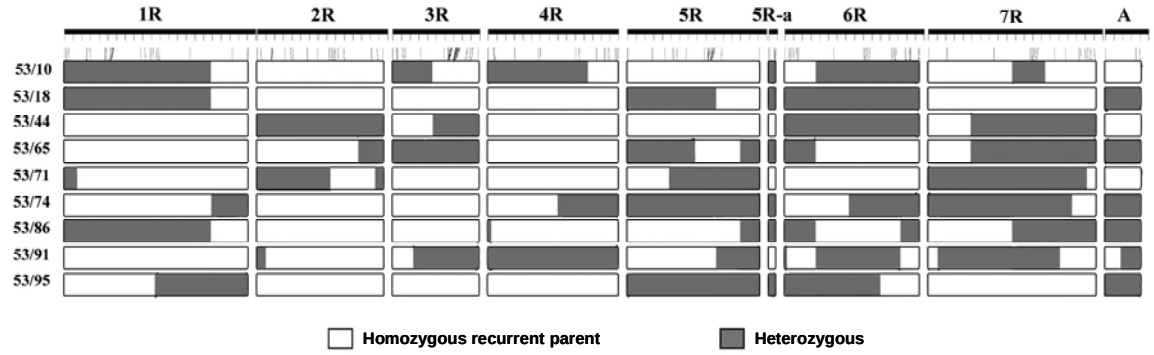
Table 2. Means and ranges for the number of donor chromosome (DC) segments per introgression line (IL), length of individual DC segment, total length of DC segments per IL, proportion of the recurrent parent genome (RPG), and donor genome coverage of the plants selected in BC_1 to BC_2S_2

Gene-ration	Number of DC segments per IL (cM)		Length of individual DC segment (cM)		Total length of DC segments per IL (cM)		Proportion of RPG (%)		Donor genome coverage (%)
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	
	---------- Introgression library F ----------								
BC_1	6.2	4-8	43.5	2.0-93.0	239.1	197.0-283.0	80.0	76.1-85.5	100
BC_2	2.5	1-4	29.8	5.0-72.0	64.5	13.5-139.0	89.8	82.0-95.1	90
BC_2S_1	2.3	1-4	21.6	2.0-72.0	42.5	10.0-115.0	92.9	86.0-98.0	72
BC_2S_2	2.2	1-4	18.3	2.0-72.0	40.6	10.0-87.5	94.1	87.5-98.5	72
	---------- Introgression library G ----------								
BC_1	5.1	4-7	44.6	4.5-117.2	236.4	145.0-289.2	82.0	76.8-85.5	100
BC_2	2.8	2-5	28.5	2.0-70.0	40.6	7.0-106.0	89.3	82.0-94.0	70
BC_2S_1	2.6	1-5	19.9	2.0-55.0	34.5	4.0-95.0	94.2	88.0-98.0	63
BC_2S_2	2.2	1-5	14.8	1.5-45.5	33.3	7.0-73.0	95.1	88.5-99.0	63

During the marker-assisted introgression process, in both libraries, an apparent decrease in the (*i*) number of DC segments per IL, (*ii*) individual DC segment length, and (*iii*) total length of DC segments per IL was observed among the selected sets of individuals in the respective generations (Table 2). Consequently, with each BC and S introgression generation the average proportion of the RPG increased, indicating a progressive elimination of the donor marker alleles (Table 2).

Four generations of MAS in library F resulted in 40 BC_2S_2 ILs, carrying on average 2.2 DC segments, with a mean length of 18.3 cM. The RPG proportion in library F ranged from 87.5% to 98.5%, with a mean of 94.1% (Table 2). In library G, 40 BC_2S_2 ILs harboured on average 2.2 DC segments, with a mean length of 14.8 cM. The finally selected set of ILs contained 88.5% to 99.0% of the RPG, with a mean of 95.1%.

In library F, the set of BC_1 plants selected to constitute generation BC_2 provided at least a double coverage of the donor genome for most of the chromosome regions (Fig. 9).



Each horizontal line in the figure represents an individual plant and each vertical block represents a chromosome. The black bars above the drawing designate marker positions.

Figure 9. Graphical genotypes representing the donor genome coverage in the selected set of 9 BC_1 individuals of introgression library F

In generation BC_2, library F covered approx. 90% of the total DG (Table 2). Larger gaps in DG coverage remained on chromosomes 1R (25 cM) and 4R (15 cM), and a minor gap of 8 cM was detected on chromosome 2R (Fig. 10).

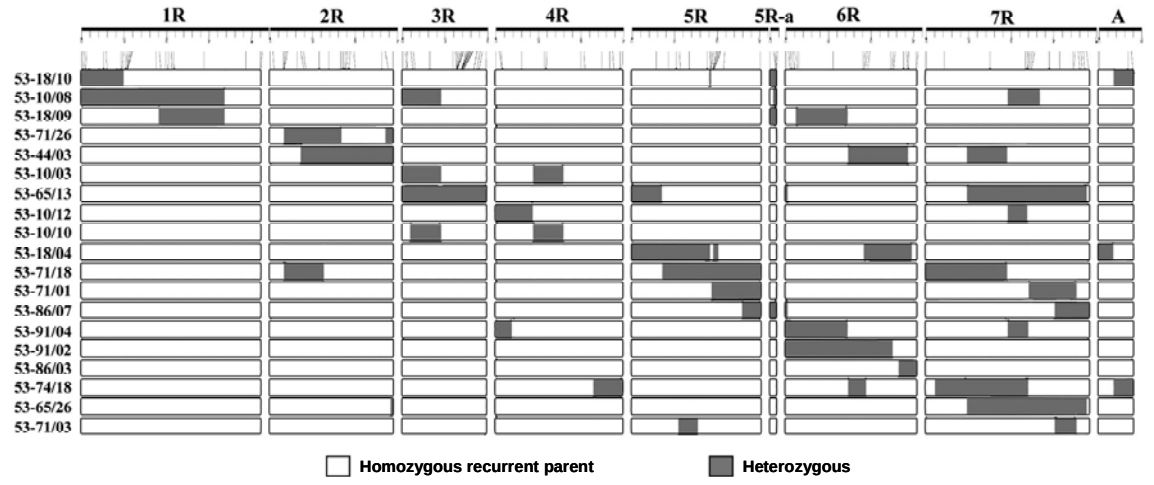


Figure 10. Graphical genotypes representing the donor genome coverage in the selected set of 19 BC_2 individuals of introgression library F

Additional large gaps were detected in generation BC_2S_1 on chromosomes 2R (30 cM), 4R (38 cM), and 7R (35 cM), whereas, new minor gaps appeared on chromosomes 5R and 6R (Fig. 11). Thus, in generation BC_2S_1 library F covered approx. 72% of the donor genome (Table 2, Fig. 11). Several short genome regions on chromosomes 1R, 4R, 5R, and 6R carried DC segments in the heterozygous state. Moreover, several segments on these chromosomes were characterized as "heterozygous or homozygous" due to uncertainty caused by the dominant scoring of several AFLP markers (Fig. 11). Since 13 BC_2S_1 individuals carried DC segments in heterozygous or uncertain state, an additional S generation was required to complete the introgression library.

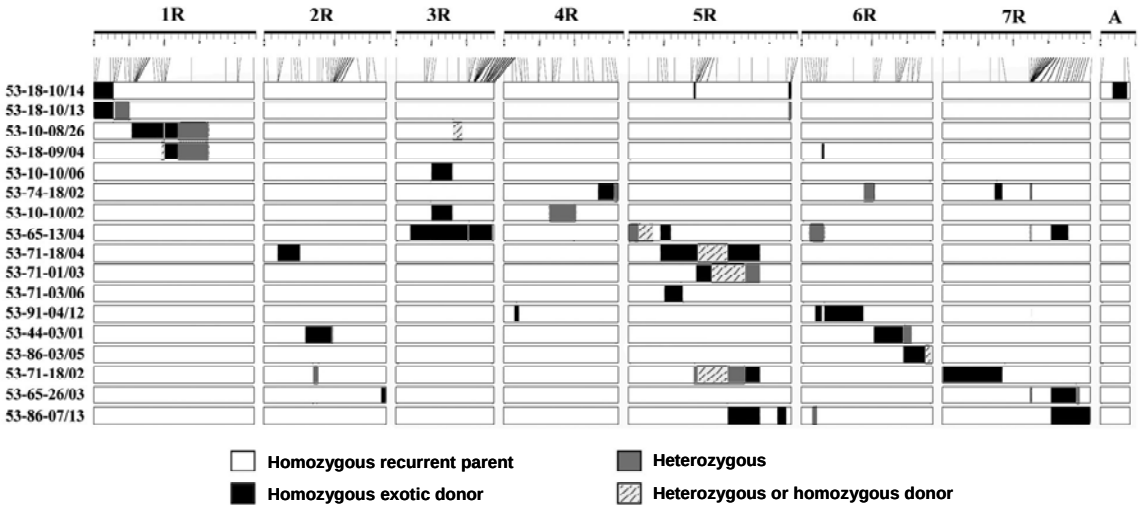


Figure 11. Graphical genotypes representing the donor genome coverage in the selected set of 17 BC_2S_1 individuals of introgression library F

In the final introgression generation (BC_2S_2), library F covered approx. 72% of the donor genome (Table 2). Thus, no new gaps showed up from BC_2S_1 to BC_2S_2 (Fig. 12).

All finally selected ILs carried only homozygous DC segments. Exceptions were two ILs harbouring short heterozygous DC segments on chromosome 4R, represented by single markers (point introgressions), as well as four ILs harbouring point introgressions of unknown state on chromosomes 1R and 5R (Fig. 12).

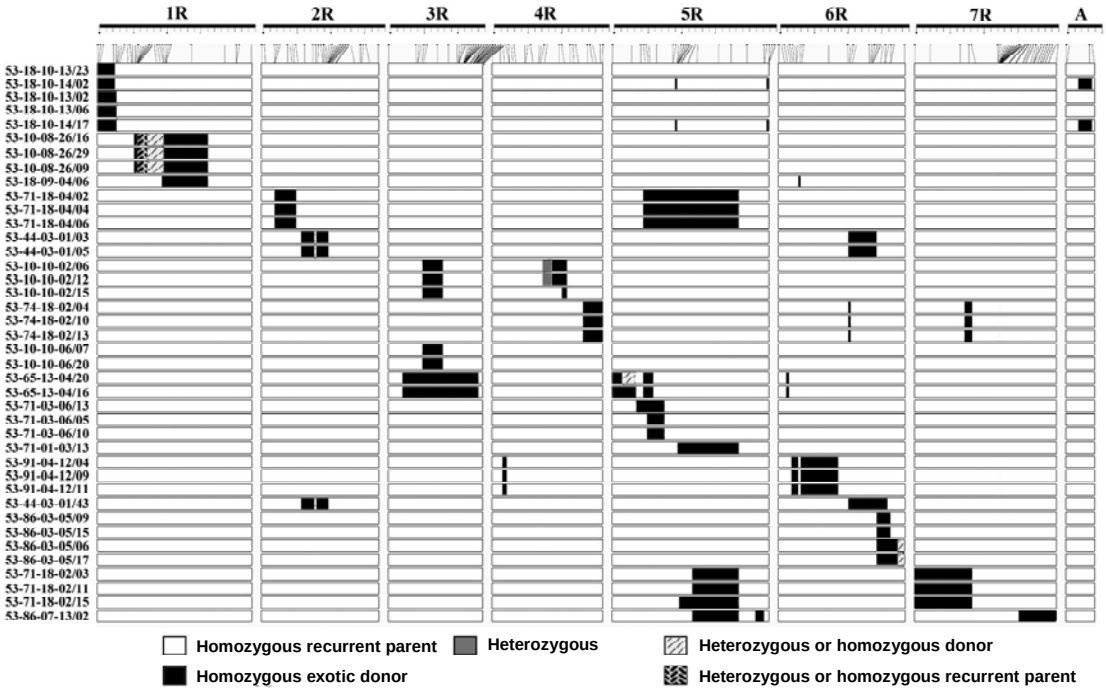


Figure 12. Graphical genotypes representing the coverage of the total donor genome in a selected set of 40 BC_2S_2 individuals of the introgression library F

In library G, the selected set of BC_1 plants provided at least a double coverage of the donor genome for most of the chromosome regions (Fig. 13). The only exception was the proximal region of chromosome 5R, that was covered by a single DC segment (Fig. 13).

In generation BC_2, introgression library G covered approx. 70% of the total DG (Table 2). The largest gaps appeared on chromosomes 2R (two gaps, 25 cM and 30 cM) and 5R (40 cM), whereas minor gaps were present throughout the genome (1R - 10 cM, 4R - 10 cM, 5R - 5 cM, 6R - 15 cM, and 7R - 10 cM) (Fig. 14).

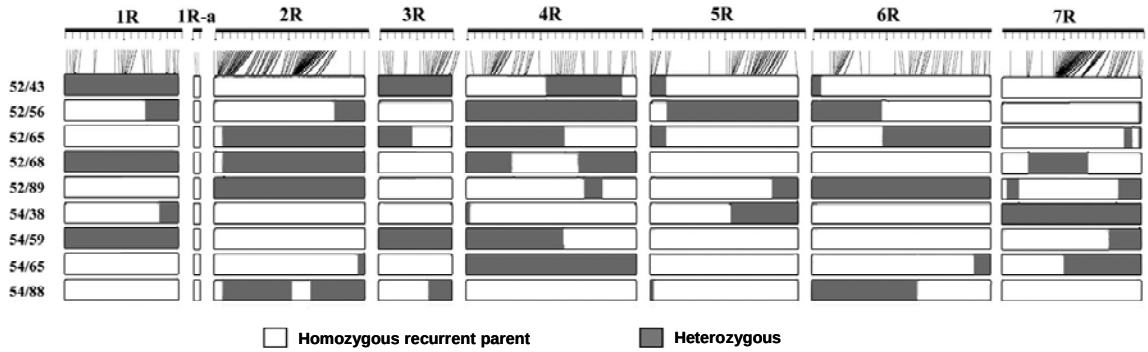


Figure 13. Graphical genotypes representing the donor genome coverage in the selected set of 9 BC_1 individuals of introgression library G

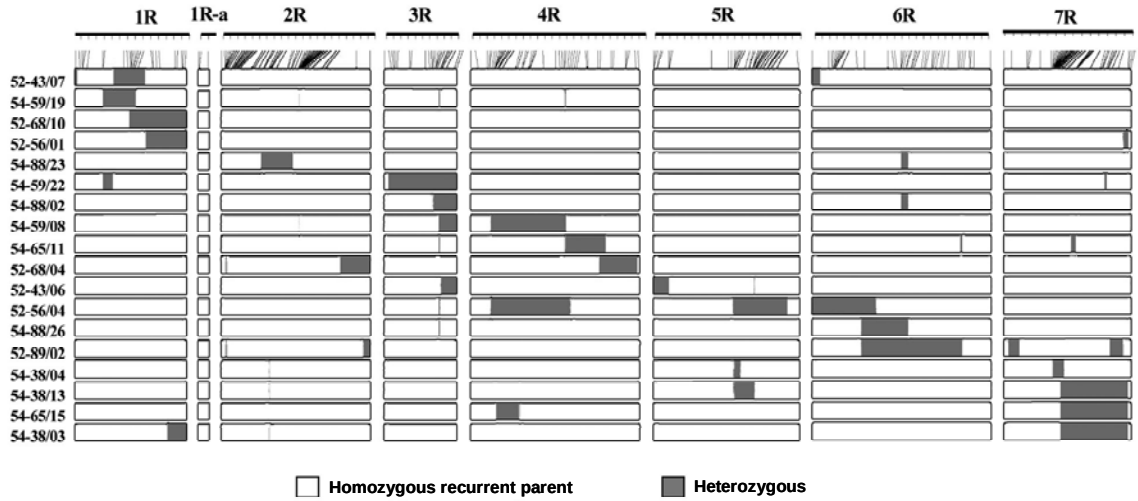


Figure 14. Graphical genotypes representing the donor genome coverage in the selected set of 18 BC_2 individuals of introgression library G

In comparison with generation BC_2, two additional gaps were detected in generation BC_2S_1 on chromosome 6R (10 cM and 20 cM), whereas gaps already present on chromosome 2R increased further (Fig. 15). As a result, in generation BC_2S_1 library G covered approx. 63% of the total DG (Table 2). Some genome regions in generation BC_2S_1 were covered either by heterozygous or DC segments in uncertain (heterozygous or homozygous) state (Fig. 15). To obtain these segments in the homozygous state and to complete the introgression library G, 18 BC_2S_1 individuals were selected for an additional generation of selfing.

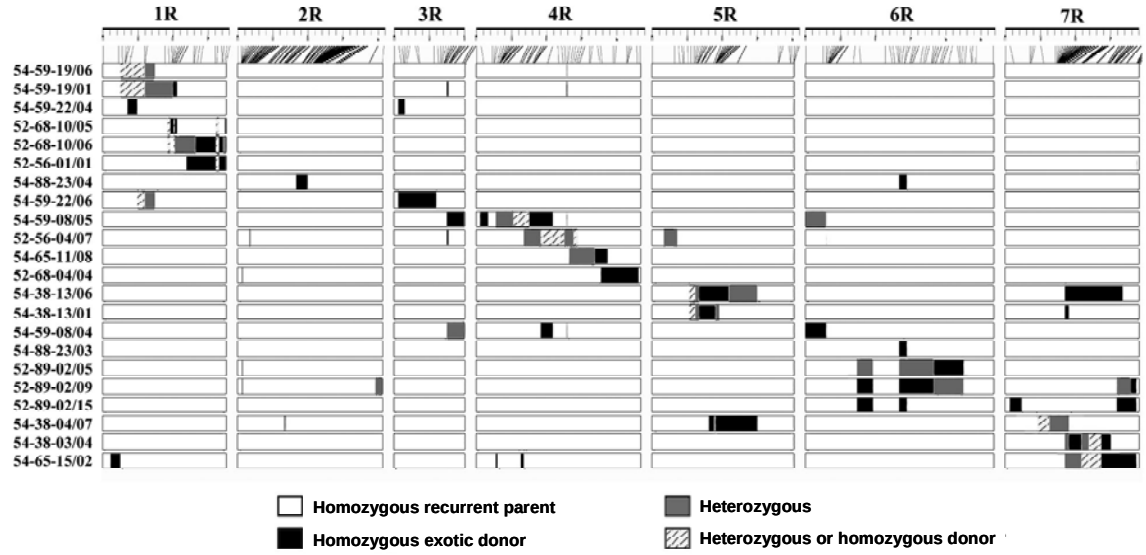


Figure 15. Graphical genotypes representing the donor genome coverage in the selected set of 22 BC_2S_1 individuals of introgression library G

From generation BC_2S_1 to BC_2S_2 no new gaps showed up (Fig. 16). Thus the finally selected set of 40 BC_2S_2 ILs in library G covered approx. 63% of the donor genome (Table 2). The majority of the introgressed DC segments was fixed in homozygous state. Exceptions

were four short heterozygous DC segments on chromosome 4R and several DC segments of uncertain state (chromosomes 1R and 5R), represented by point introgressions (Fig. 16).

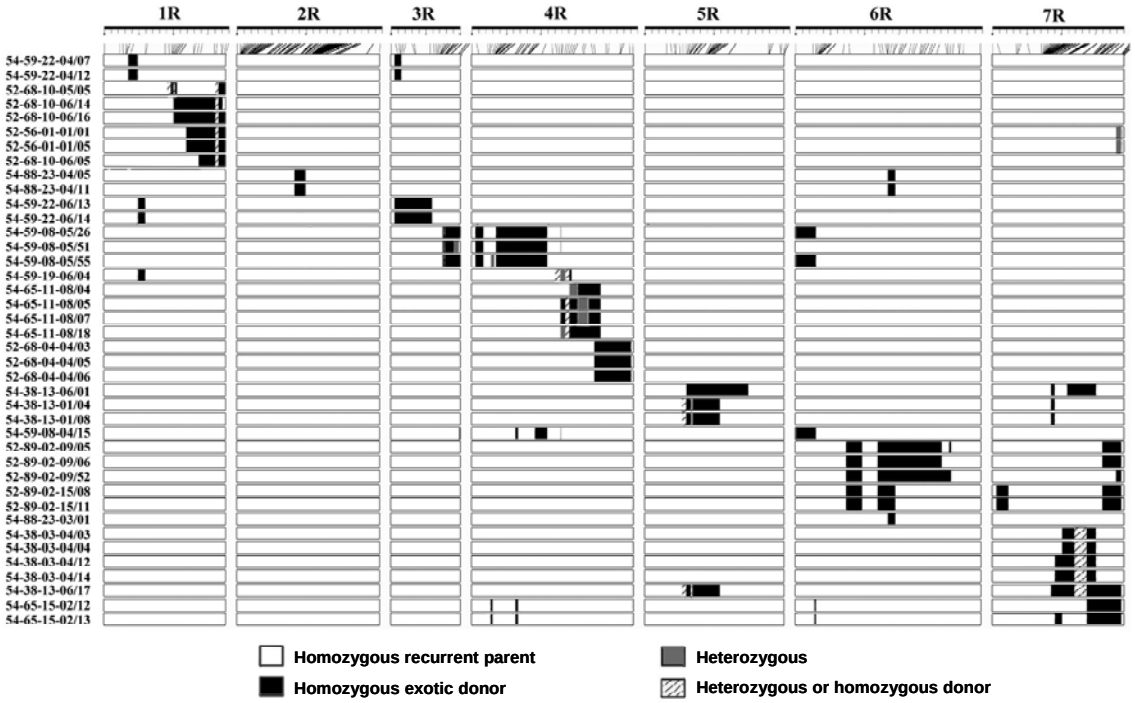


Figure 16. Graphical genotypes representing the donor genome coverage in the selected set of 40 BC_2S_2 individuals of introgression library G

2.4 Discussion

2.4.1 Comparisons of rye introgression libraries with existing libraries in other crops

Several studies reported about the establishment of introgression libraries employing different strategies. The majority of them were based on MAS only (Eshed and Zamir 1994, Chetelat and Meglic 2000, Monforte and Tanksley 2000, von Korff et al. 2004). Others employed either a combination of phenotypic selection and MAS (Eshed et al. 1992) or random backcrossing followed by MAS of ILs (Jeuken and Lindhout 2004). For the development of these introgression libraries various numbers of BC and S generations were employed. For example, libraries consisted of BC_1S_6 ILs (Eshed et al. 1992), BC_3S_1 ILs (Mank et al. 2003), BC_3S_1 ILs and BC_3S_2 recombinant backcross lines (Monforte and Tanksley 2000), $BC_{(1-2)}F_{(2-6)}$ ILs (Chetelat and Meglic 2000), BC_5S_1 ILs (Jeuken and Lindhout 2004), or BC_2 double-haploid lines (von Korff et al. 2004). By subsequent backcrossing and selfing some of the introgression libraries were advanced to NILs harbouring one single short DC segment, (*e.g.*, $BC_1S_6BC_3S_1$ and $BC_1S_2BC_3S_1$ NILs, Eshed and Zamir 1994).

In the present study, two rye introgression libraries jointly consisting of 80 BC_2S_3 ILs were successfully established after four generations of MAS, which may be considered as a relatively short procedure. Generally, approaches based on MAS only are the most effective due to several reasons. First, they are considerably shorter because of the rapid reconstitution of the recurrent parent genome (Howell et al. 1996). Second, in random backcrossing, there is a high probability that some DC segments will be lost (Beckmann and Soller 1986), unless large numbers of individuals are generated and evaluated (Stuber et al. 1999). Missing introgressions would not be noticed immediately, but only after MAS has been applied. Third, approaches combining phenotypic selection and MAS result in ILs harbouring higher DC segment numbers than if only MAS is applied. This may lead to “cancelling effects” at the phenotypic level, when a favourable and unfavourable segment occur within the same IL. To avoid this obstacle, higher population sizes are required to detect ILs in which favourable segments are pyramidized.

Combining two types of molecular markers (AFLP and SSR) proved powerful for the establishment of the present introgression libraries. In generation BC_1, AFLP primer combinations provided on average 11 markers/PC in library F and 19 markers/PC in library G. Hence, few PCs were sufficient to quickly and efficiently locate DC segments across the whole genome. This was also true for the generation BC_2, when a relatively large proportion

of the genome (25%) was still segregating. However, in the subsequent S generations, particularly in the final generation (BC_2S_2) when the proportion of segregating DC segments to be monitored declined, the application of SSR markers became much more cost-efficient. The use of AFLPs in the later generations would have lead to numerous useless monomorphic markers throughout the RPG and therefore would have been more expensive. Another shortcoming of AFLPs was the difficulty in the codominant scoring of several markers, especially in the selfing generations. A rather even distribution of the mapped SSR markers within both libraries F and G allowed to substitute AFLPs in the advanced introgression generations, and to use them only in cases where no SSR marker was available for monitoring a particular DC segment.

In both libraries, the average proportions of the RPG in the selected set of individuals in generations BC_1 and BC_2 were higher than theoretically expected (75% and 87.5%, respectively). The average proportion of the RPG further increased with each S generation, approaching in the finally selected sets of BC_2S_2 ILs 94.1% in library F and 95.1% in library G (Table 2). This suggests that efficient selection criteria were applied in rye marker-assisted backcrossing program to increase the RPG proportion.

Considering rye introgression libraries F and G together, 35% of the developed ILs (28 of 80) harboured single DC segments, whereas an additional 30% (24 ILs) carried two DC segments. Only 14 ILs (17.5%) did not match the objective of the study, and carried more than three DC introgressions (Figs. 3 and 6). That was opposite to tomato introgression libraries where the complete set (Eshed and Zamir 1994) or the majority of the developed ILs (Monforte and Tanksley 2000) carried a single DC segment. High numbers of DC segments per IL in the present rye libraries were due solely to point introgressions. This could have arisen from a double crossover between two closely positioned recurrent parent markers. However, the probability of a double crossover within a very small mapping distance is extremely low. Erroneous marker scorings can be excluded as an explanation, because the short introgressions were observed in two subsequent introgression generations (BC_2S_1 and BC_2S_2). As the most probable reason for detecting point introgressions, inaccurate map positions of the markers for these short DC segments could be suggested. Extension of the present F and G rye genetic maps with additional markers surrounding point introgression sites may elucidate the true cause of the observed short DC segments.

The mean length of individual DC segments decreased during the introgression process. Observed values in generations BC_1 and BC_2 (Table 2) were in agreement with the experimental results on other crops. Two sets of lines (BC_1 and BC_2) were developed in tomato, harbouring DC segments of 47 cM and 31 cM, respectively (Fulton et al. 1997). Von Korff et al. (2004) established two barley IL libraries having on average 46 cM and 48.1 cM long introgressions in generation BC_1. Furthermore, the mean length of individual DC segments in libraries F and G (generations BC_1 and BC_2) were in accordance with the predictions of Fulton et al. (1997), who calculated a mean segment length of 50 cM in generation BC_1 and 34 cM in generation BC_2 of an unselected population, using computer simulations. In the present rye libraries, the mean individual DC segment length in the finally selected set of 40 BC_2S_2 ILs was reduced to 18.3 cM in library F and 14.8 cM in library G (Table 2). These values were much lower than those presented for three tomato introgression libraries, which were 47 cM (Eshed et al. 1992), 33 cM (Eshed and Zamir 1994), and 25.2 cM (Chetelat and Meglic 2000), as well as for an introgression library of lettuce (33 cM; Jeuken and Lindhout 2004). The short mean DC segment length observed in the present rye libraries should have a greater potential for mapping gene effects located in the introgressed DC segments.

In comparison with the introgression libraries of tomato (Eshed and Zamir 1994), lettuce (Jeuken and Lindhout 2004), and barley (von Korff et al. 2004), which covered almost the complete DG, the two rye libraries display a lower DG coverage (library F 72%, library G 63%, jointly approx. 80%). Minor gaps appeared across the whole genome in both rye libraries, with large ones occurring on chromosome 2R in both libraries and on chromosome 4R in library F. In generation BC_1, no distorted segregation ratio favouring the recurrent parent was observed for the loci in gap regions. Hence, the main reason for the appearance of such gaps in DG coverage might be insufficient progeny sizes per IL during the introgression process. Small progeny sizes especially occurred in generation BC_2 since high sterility of BC_2 plants, caused by extremely high temperatures in the greenhouse during the flowering period, led to many failures. Thus, in library G the selection resulted in only three BC_2 plants carrying DC segments on chromosome 2R (Fig. 14). Consequently, the coverage of chromosome 2R in the final BC_2S_2 generation was only 10%, with two large gaps appearing in the regions from 0 cM to 40 cM and from 50 cM to 102 cM (Fig. 16). Another reason for the appearance of gaps in DG coverage might be a low marker density, which may have resulted in undiscovered DC segments (*e.g.*, proximal regions of chromosomes 1R and 7R in library F, Fig. A2; proximal regions of the chromosomes 5R and 6R in library G, Fig. A4). Absence of a certain portion of

the DG in the final set of ILs was also observed in introgression libraries in tomato (Chetelat and Meglic 2000, Monforte and Tanksley 2000) and lettuce (Jeuken and Lindhout 2004), and was attributed to unfavourable combinations of DC segments, as well as to insufficient population sizes.

The absence of a homozygous donor parent in constituting the two rye introgression libraries was a disadvantage for the applied marker-assisted backcrossing procedure, which in some cases resulted in difficulties to identify the origin (donor or recurrent parent) of the scored marker bands. This particularly applied to AFLP bands. Therefore, the presence and exact length of some DC segments were not unambiguous in all cases. This inaccurate monitoring of a few DC segments may have additionally reduced the DG coverage.

Even though the genetic maps used for the marker-assisted establishment of the present libraries had high average marker densities, small DC segments could have remained undetected during the introgression procedure, particularly within gap regions of the genetic maps.

Introgression lines can be applied successfully for a rapid identification of DNA markers specific for a certain genome region (Martin et al. 1991, Eshed and Zamir 1994). Thus, the established rye ILs could be used to add new markers to the current gap regions in the genetic maps of libraries F and G (unless the gaps are not identical), and would allow to identify potentially undiscovered DC segments. Currently, the majority of the introgressed DC segments in the two libraries is carried by more than one BC_2S_3 IL. These ILs represent internal repetitions, aimed to increase the precision of phenotypic evaluation and to identify small DC segments that could have escaped detection.

2.4.2 Possible applications of rye introgression libraries

The starting point in the application of the two introgression libraries in rye breeding and genomics would be the phenotypic evaluation of the marker-characterized DC segments by evaluating the rye ILs in large-scale field experiments. The differently performing ILs would be presumed to have received DC segments that contain phenotypically detectable donor loci (Stuber et al. 1999) that may be favourable or detrimental. Data assessment may include agronomically important qualitative and quantitative traits, such as date of heading, plant height, lodging resistance, grain yield, 1000-kernel weight, test weight, and falling number. Additionally, particular quality traits, important for the use of rye for baking, feeding, or as an energy resource, can be analysed. As rye ILs carry only a small fraction of the exotic DG, the

number of unfavourable or even deleterious genes affecting vigour and fertility is greatly reduced and yield-associated traits can be measured with higher accuracy. Compared with trait-specific sets of NILs that are exclusively used for the mapping of particular genes (Koester et al. 1993, van Berloo et al. 2001, Shen et al. 2001), the main advantage of introgression libraries is that they can be used as a comprehensive mapping resource for a simultaneous analysis of genome regions underlying various agronomically important traits (Peleman and Rouppe van der Voort 2003). Immediate applicability of ILs in breeding is another important advantage of the introgression library approach. If the recipient is an elite line or a registered variety (like the rye inbred line L2053), a derived IL harbouring a favourable DC segment may represent a significant improvement, thus leading to a new variety (Eshed and Zamir 1995).

Particularly important for hybrid rye breeding is that introgression libraries can be used to map loci contributing to heterosis (Zamir 2001, Peleman and Rouppe van der Voort 2003). For this purpose, ILs could be crossed to cytoplasmic-male sterile (CMS) testers to determine their combining ability. This would create an "F_1 introgression library" which would then be phenotyped to detect heterotic effects caused by introgressed DC segments. This approach may provide an efficient way to improve the combining ability of lines, by broadening their genetic base with exotic DC segments harbouring new favourable genes. Furthermore, genes for new quality traits may be found that would, for the first time, allow the initiation of well-directed marker-assisted breeding programs for special uses of rye.

Some of the developed rye ILs harboured more than a single DC segment (Figs. 12 and 16). Since no sufficient overlap for the identified DC segments exists in the two libraries, it is not always possible to determine which of the segments was responsible for an observed phenotypic effect. In this case, additional backcrossing to the recurrent parent are required to generate segregating families in which the introgressed DC segments will be dispersed by recombination events, allowing the breeder to develop a sub-library of ILs carrying the respective DC segments individually or in various combinations (Eshed and Zamir 1994).

Introgression libraries combined with recent advantages in genomics allow not only to broaden the genetic base of crop populations and to achieve faster and more predictable progress from selection, but also to isolate and functionally characterize individual genes. However, the identification of genes underlying favourable DC segment effects is a highly demanding task even with the most advanced molecular genetic tools (Stuber et al. 1999).

Once a favourable single DC segment carrying desirable genes is identified a set of recombinant backcross lines can be developed for this DC segment to allow its high-

resolution (fine) mapping (Paterson et al. 1990). Potential association among genes affecting different traits could thus be broken. As the starting point for fine-mapping of favourable DC segments, rye ILs carrying beneficial DC segments should be further backcrossed to create NILs. This kind of molecular characterisation of favourable DC segments harboured in rye ILs could provide an insight into the function and structure of rye DC segments and pave the way for mapping and verifying candidate genes, in particular those related to hybrid vigour and abiotic stress (cold tolerance). Moreover, since open-pollinated genetic resources are expected to carry a high mutational load, comparisons among ILs or sub-ILs would allow to identify and eliminate unfavourable and/or sub-lethal mutations in introgression programs.

The rye introgression libraries developed in this study can be regarded as a valuable tool to proceed towards genomic studies of not only rye but also other cereals. The impact of rye DC segments should be the greatest for triticale since triticale carries a complete rye genome and no translocations are necessary to transfer favourable rye DC segments. Furthermore, rye is the only hybrid crop among the small grains and could be used as a model plant for breeding hybrids in triticale in the future. Wheat ILs carrying translocations on the short arm of chromosome 1R possess improved agronomic performance (Moonen and Zeven 1984, Lukaszewski 1990, Carver and Rayburn 1994, Villareal et al. 1996, Lukaszewski 2001, Mago et al. 2002), particularly when the source of the rye donor chromatin was selected carefully (Kim et al. 2004). Thus, it can be assumed that well-characterized rye DC segments will also have an important impact on wheat improvement. It is therefore particularly promising to analyse the molecular architecture of hybrid vigour, as the basis of the yield advantage of hybrids over their parental inbred lines. In addition, rye is well adapted to abiotic stress conditions - in comparison with other *Triticeae*, rye is the most cold tolerant cereal, which is an upcoming issue for the production of competitive rye hybrids in Eastern Europe.

In summary, introgression libraries can make a wide array of previously unexplored genetic variation rapidly available to plant breeders and geneticists. In this respect, the rye introgression libraries established in this study represent a dynamic new resource that could substantially foster rye breeding programs in years to come.

3 Section B
Optimisation of the establishment of rye introgression libraries: Simulation study

3.1 Introduction

Marker-assisted backcrossing

With the advent of DNA marker technology, new perspectives were offered to plant breeders. In this respect, the application of molecular marker technology in backcross breeding (marker-assisted backcrossing) offers great possibilities of improving efficiency and effectiveness in the selection of plant phenotypes with desired combination of traits.

Marker-assisted backcrossing represents a powerful tool for the manipulation of oligogenic traits under various conditions (for review see Melchinger 1990). Many studies focused on the optimization aspects of application of molecular markers in BC breeding programs, as well as on various models and strategies for improving the efficiency of introgression of a single gene from a donor into recipient genotype (Hospital et al. 1992, Frisch et al. 1999a, 2000, Frisch and Melchinger 2001b, Reyes-Valdés 2000, Servin and Hospital 2002, Ribaut et al. 2002).

In marker-assisted backcrossing, molecular markers are applied for two purposes: *i*) tracing the presence of a target allele, and *ii*) identifying individuals with a high proportion of the RPG (Frisch et al. 1999b). Adopting the terminology of Hospital and Charcosset (1997), the former application is defined as “foreground selection”, and the latter as “background selection” (for review see Visscher et al. 1996).

Marker-assisted foreground selection is particularly effective for the introgression of genes for which direct selection is difficult or impossible (*e.g.*, recessive alleles expressed at a late stage in plant development) (Frisch et al. 1999b). For a successful foreground selection a close marker/trait association is necessary. This requires tight linkages between the respective loci.

The goal of marker-assisted background selection is to recover the RPG as rapidly and completely as possible. Originally proposed by Tanksley et al. (1989), marker-assisted background selection surveys the parental origin of alleles at marker loci throughout the entire genome. In combination with foreground selection, background selection allows the selection

of individuals that do not only carry the target gene but also are homozygous for the recurrent parent alleles at a large proportion of loci. Hence, marker-assisted background selection can reduce the number of BC generations required for gene introgression (Frisch et al. 1999a, Stam 2003).

Marker-assisted backcrossing is also suited for the introgression of a DC segment harbouring favourable alleles at quantitative trait loci (QTL). The length of DC segment to be introgressed depends on an accurate identification of QTL affecting the trait of interest. Thus, reliable QTL identification before starting marker-assisted backcrossing introgression is of essential importance (Gallais et al. 2000). However, it is very difficult to obtain reliable, unbiased QTL estimates (*e.g.*, Hyne et al. 1995, Beavis 1998, Melchinger et al. 1998, Utz et al. 2000, Kearsey 2000), which is significant obstacle for successful introgression of QTL using marker-assisted backcrossing. Several authors considered the optimization aspects of multiple gene introgression (Frisch and Melchinger 2001a) or QTL introgression (Visscher et al. 1996, Hospital and Charcosset 1997, Hospital et al. 2000, Hospital 2002) by repeated marker-assisted backcrossing.

In summary, marker-assisted backcrossing efficiently moderates the shortcomings of the backcrossing based on phenotypic selection, by *i*) minimizing the length of the linkage drag around the target gene on carrier chromosome, and *ii*) accelerating the recovery of the RPG on non-carrier chromosomes.

Marker-assisted backcrossing for introgression library development

An approach related to gene introgression includes the development of NILs harbouring a single DC segment (Lynch and Walsh 1998). It implies a systematic transfer of DC segments from a donor into recurrent parent (elite line) by marker-assisted backcrossing. Each NIL in the library differs from the elite line by only a small, unique, marker-defined chromosomal segment, containing alleles that may cause phenotypic differences between the elite line and the NIL. An important advantage of the NIL approach for QTL identification is that the use of NILs does not require prior assessment and knowledge of QTL. Instead, QTL introgression and QTL detection are accomplished simultaneously (Bernardo 2002).

Even though the NIL approach has been used as a tool for an accurate QTL mapping of various agronomically important traits (Eshed and Zamir 1995, Ramsay et al. 1996, Monforte and Tanksley 2000, Zamir 2001, Kearsey 2002, Peleman and Rouppe van der Voort 2003)

there has been no theoretical study aiming to optimize the procedure for creating an introgression library.

In plant breeding and population genetics, computer simulations are a powerful tool to investigate problems for which no analytical solutions are available. However, none of the available simulation software packages has so far been applied to optimize strategies for the introgression of genomic segments from exotic germplasm into elite breeding materials.

Objectives

The goal of the simulation study was to establish an effective marker-assisted introgression strategy for creating an introgression library in rye. Each IL of the library should carry a single marker-defined DC segment, and jointly they should cover most of the donor genome.

In particular, the objectives were to:

1) investigate the influence of the number of BC and S generations,

2) examine the effects of the different progeny sizes per IL from generation BC_2 onwards, and

3) study the influence of DC segment length and marker density

on the: *i*) population and progeny sizes in individual BC and S generations and *ii*) total number of marker data points (MDP) required to reach the RPG and DG threshold restrictions.

3.2 Materials and Methods

To establish an optimal procedure for the creation of a rye introgression library by marker-assisted backcrossing, various introgression strategies were simulated using the software PLABSIM version 2 (Frisch et al. 2000). The software simulates the recombination process during meiosis by a random walk algorithm (Crosby 1973) on the basis of Haldane's (1919) mapping function, and assumes the absence of interference in crossover formation.

Genetic map and target DC segments

Simulations were based on map-length estimates obtained from genotyping the BC_1 generation of rye mapping population F (introgresion library F, Fig. 4). The map comprised seven linkage groups representing seven rye chromosomes, with a total length of 665 cM. To accurately monitor the proportion of RPG on each chromosome, equally spaced hypothetical marker loci were assumed and used for selection. Regarding the density of hypothetical marker loci, two genetic maps were applied. The first map comprised 140 loci spaced at 5 cM distance. The second map consisted of 43 marker loci, spaced at 20 cM distance.

Strategies for the establishment of the introgression libraries comprising ILs carrying target DC segments of 20 cM and 40 cM length were simulated. To fulfill the requirements concerning the DG coverage and the number of target DC segments per IL, a set of ILs carrying single target DC segments on different genome regions was defined before the start of simulations. Those target DC segments were defined as the uninterrupted sequences of target loci, identical in length and number of loci. Target DC segments were aligned with no overlapping loci and, taken together, covered completely the genome of the donor parent. A hypothetical example of such a set of ILs comprising an introgression library is illustrated in Figure 17.

For the simulation of an introgression library carrying DC segments of 20 cM length, a genetic map of 5 cM distance among hypothetical marker loci was used. In this case, a particular DC segment was defined as an uninterrupted sequence of four target loci. The total length of the target DC segments was counted as the length between two flanking target markers plus two times half distance to the first adjacent non target markers (the place of recombination was assumed in the middle of the distance between a flanking target marker and the first adjacent non target marker). Analogously, in simulations aiming to establish a library carrying DC segments of 40 cM length, two approaches were used to define DC segments: *i*)

eight target loci when a genetic map with marker loci spaced at 5 cM distance was applied and *ii*) two target loci when a genetic map with marker loci spaced at 20 cM distance was used.

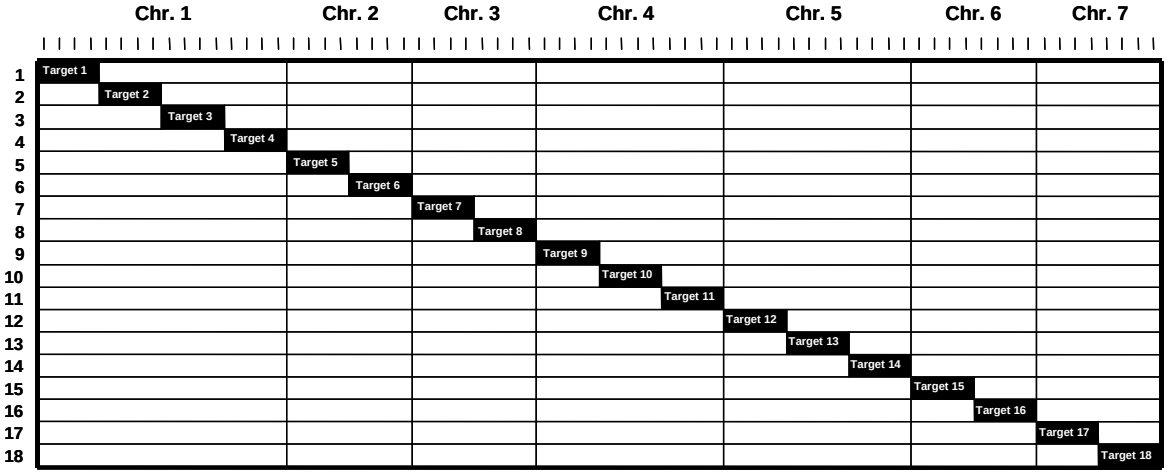


Figure 17. Example of a set of graphical genotypes showing the coverage of the donor genome in a simulated introgression library comprising 18 introgression lines (ILs) carrying single target donor chromosome segments of identical length (horizontal black rectangles). Each horizontal line represents the total genome of one IL and each vertical block refers to one chromosome. The bars above the graphical genotypes represent equally spaced hypothetical marker loci used for selection.

To cover the total DG of 665 cM, introgression libraries carrying DC segments of 20 cM comprised 36 ILs, whereas libraries bearing DC segments of 40 cM consisted of 18 ILs. Due to the fact that map-length estimates of the particular linkage groups were predefined (genetic map of library F), some DC segments covering distal regions of the linkage groups were shorter than specified (20 cM or 40 cM).

Recurrent parent genome and donor genome threshold values

The minimum proportion of recurrent parent alleles to be recovered in a simulated introgression program was designated as the RPG threshold. Its value was defined as the mean RPG value reached in generation BC_6 by applying a random selection among all individuals carrying target DC segments. Simulations assuming selection for the presence of the target alleles only (foreground selection) and no marker-assisted background selection for recurrent parent loci (Frisch et al. 1999a) were conducted to define the RPG threshold value. For these simulations, genetic map with marker loci spaced at 5 cM distance was used. Target DC segment of 20 cM length was defined as an uninterrupted sequence of four target loci. In simulations for 40 cM DC segment length the same proportion of non-target donor alleles was used and no double cross-over between target loci within DC segments was assumed.

Population size per generation equalled 1000 individuals. Each simulation was repeated 1000 times.

To assure a high coverage of the DG in simulated introgression scenarios, a deviation of only 1% from the maximum possible DG coverage was allowed. Thus, a DG threshold of 99% was defined.

Population size in generation BC_1 and progeny size per IL in generation BC_2 and onwards

To define population size in generation BC_1, preliminary simulation runs were performed. Sizes ranging from 5 to 1000 were tested. Selection criteria were based on the output estimates for the DG coverage, proportion of the RPG recovered, and number of MDP employed in generation BC_1. For further simulations, only those population sizes were selected that *i*) provided complete DG coverage and *ii*) resulted in an increased proportion of the RPG recovered, without simultaneous substantial increase in the number of MDP.

From generation BC_2 onwards, progeny size per IL was *i*) kept constant, *ii*) increased, or *iii*) decreased. For each BC_1 population size, various number of simulations was performed to determine the progeny size per IL in subsequent generations, necessary to meet the RPG and DG threshold restrictions.

Selection strategy

A two-stage selection strategy (Frisch et al. 1999a) was employed to simulate the marker-assisted introgression procedure. First, all individuals of a BC_t or BC_tS_x line, carrying a single pre-defined target DC segment, were selected (foreground selection), thus ensuring a high DG coverage. Second, among selected individuals, those with the highest RPG proportion were chosen (background selection). Thus, the selected set of ILs displayed the maximum DG coverage and maximum RPG recovery that were possible to obtain with a particular combination of population size in BC_1 and progeny size per IL.

Simulation runs

Each simulation run started from an F_1 cross between recurrent parent and donor. Both parents were assumed to be homozygous and to carry different alleles at all marker loci. A selected BC_1 plant generally served as a common parent for several BC_2 lines since BC_1 plants mostly harbour more than one DC segment.

From generation BC_2 onwards, selection was practiced only within lines. The best plant of each line was used for the establishment of the next generation. The number of backcrosses varied from two to four, with one or two additional generations of selfing to complete the introgression process.

Simulation of each introgression strategy was repeated 100 times. This way, the standard errors (SE) given as a percentage of the mean values were below 1% for all of the observed output parameters, except for the average number of non-target DC segments per IL.

Simulation outputs

Simulation outputs contained mean values (across simulation runs) and standard deviations of the *i*) total number of donor alleles at the target segments, *ii*) total number of RPG alleles, *iii*) total number of MDP employed, and *iv*) total number of uninterrupted sequences of donor alleles in the introgression library (further referred to as number of blocks).

Coverage of the DG (expressed as percentage) was calculated by dividing the number of the donor alleles at the target loci carried by the complete set of ILs with a total number of target alleles. Proportion of the RPG recovered in the introgression library (expressed as percentage) was determined by dividing the number of RPG alleles carried by the complete set of ILs with the total number of alleles monitored (both on target and background loci). Each analysis of a marker locus in a BC or S individual was counted as an MDP. In generation BC_1, the total set of markers was analyzed, whereas in the following generations only those markers were used that were not yet fixed for the recurrent parent or donor allele in the respective individuals. The total number of MDP required in each generation was summed over the whole introgression program.

Empirical SE of mean values across simulation runs were calculated for the *i*) number of MDP employed, *ii*) DG coverage, and *iii*) average number of non-target DC segments per IL.

Evaluation criteria

For the evaluation of introgression strategies, the following criteria were employed:

i) Number of MDP and

ii) Number of introgression generations

required to reach the RPG and DG threshold restrictions.

The estimated costs of raising, crossing, and selfing the plants varied only slightly among the introgression scenarios and were negligible in comparison with the costs for marker analyses. Thus, the total cost of the introgression scenarios was estimated only on the basis of the number of the MDP employed.

For each simulated introgression strategy, one optimum scenario (a particular combination of population size in BC_1 and progeny size per IL) was defined regarding the minimum number of MDP employed.

Influence of the number of BC and S generations

Six strategies consisting of consisting of two to four BC and one to two S generations: BC_2S_1, BC_2S_2, BC_3S_1, BC_3S_2, BC_4S_1, and BC_4S_2 were simulated. Simulated introgression strategies were based on constant progeny size per IL for a particular BC_1 population size.

Comparison of introgression variants with increasing and decreasing progeny size per IL

The BC_3S_1 strategy was used to evaluate the effects of increasing and decreasing progeny size per IL from generation BC_2 onwards. Five scenarios were simulated within the variant with increasing progeny size (ratios 1:1.5:2, 1:1:2, 1:2:3, 1:1:3, and 1:3:9). Analogously, five scenarios were simulated within the variant of decreasing progeny size (ratios 2:1.5:1, 2:1:1, 3:2:1, 3:1:1, and 9:3:1).

Influence of the DC segment length and marker density

To investigate the effects of DC segment length and marker density on the introgression procedure, the BC_2S_1 and BC_3S_1 strategies were used, with constant progeny size per IL for each particular BC_1 population size.

Three combinations of DC segment length and marker density were tested:

i) 20 cM DC segment length / 5 cM marker density (further referred to as variant 20/5),

ii) 40 cM DC segment length / 5 cM marker density (further referred to as variant 40/5), and

iii) 40 cM DC segment length / 20 cM marker density (further referred to as variant 40/20).

3.3 Results

3.3.1 Preliminary simulations for determining the RPG threshold value and the suitable population sizes in generation BC_1

When only foreground selection was performed, the mean RPG recovered was about 3-4% below the values expected in backcrossing without selection (Table 3). After six generations of backcrossing with foreground selection, the mean RPG recovered reached 95.60%.

Table 3. Simulation results for the mean proportion of the recurrent parent genome in various backcross (BC) generations with foreground selection of individuals carrying the target donor chromosome segments of 20 cM length and expected values for backcrossing without selection.

Generation	With foreground selection	Without selection
	---------- % ----------	
BC_1	72.30	75.00
BC_2	83.80	87.50
BC_3	89.42	93.75
BC_4	92.68	96.88
BC_5	94.53	98.43
BC_6	**95.60**	99.22
BC_7	96.29	99.61
BC_8	96.71	99.82
BC_9	96.98	99.91
BC_{10}	97.22	100.00

The RPG threshol value allowed for a small proportion of non-target donor alleles (1.62%). Any further increase in the proportion of RPG towards complete RPG recovery would have required considerable increase in progeny size per IL from generation BC_2 to BC_3S_1 and a corresponding raise in the number of MDP employed (Table 4). Therefore, this deviation from an "ideal" introgression library was accepted, and the value of 95.60% was subsequently used as the RPG threshold for strategies with 20 cM DC segment length.

Standard errors of the means across simulation runs for DG coverage, proportion of the RPG recovered and number of the MDP employed as a function of various progeny sizes per IL from generation BC_2 to BC_3S_1 are given in Appendix, Table A1.

To compare introgression strategies for DC segments of 20 cM with those of 40 cM length, an equal proportion of non-target DG (1.62%) was used. As the maximum proportion of the target DG in introgression libraries carrying DC segments of 40 cM length equalled

5.55%, the remaining proportion of the total genome was 92.83%. The latter value was therefore used for the establishment of an introgression library with DC segments of 40 cM length.

Table 4. Coverage of the donor genome (DG), percentage of the recurrent parent genome (RPG) recovered and the number of marker data points (MDP) employed in different BC_3S_1 scenarios as a function of progeny sizes per introgression line (IL) from generation BC_2 to BC_3S_1. Simulations were based on constant progeny size per IL from generation BC_2 onwards and a BC_1 population size of 100 individuals.

Progeny size per IL from generation BC_2 to BC_3S_1	Coverage of the DG (%)	RPG recovered (%)	Number of MDP
19	99.92	95.60	52 700
50	100.00	96.52	107 500
100	100.00	96.89	192 900
250	100.00	97.10	440 500
500	100.00	97.11	846 300
1000	100.00	97.20	1 659 300
1500	100.00	97.21	2 475 000
2000	100.00	97.22	3 290 300

[a] Values for the number of MDP are rounded to multiples of hundred.

The minimum population size providing total DG coverage in generation BC_1 was 17. (Table 5).

Table 5. Coverage of the donor genome (DG), percentage of the recurrent parent genome (RPG) recovered, and the number of marker data points (MDP) needed in generation BC_1, as a function of the population size in generation BC_1.

Population size in BC_1[a]	Coverage of the DG (%)	RPG recovered (%)	Number of MDP
5	97.29	75.11	700
10	99.80	78.20	1 400
17	100.00	80.37	2 380
25	100.00	81.62	3 500
40	100.00	83.25	5 600
60	100.00	84.47	8 400
80	100.00	85.31	11 200
100	100.00	85.85	14 000
130	100.00	86.56	18 200
160	100.00	86.97	22 400
200	100.00	87.38	28 000
300	100.00	88.14	42 000
400	100.00	88.68	56 000
500	100.00	89.15	70 000
750	100.00	90.00	105 000
1000	100.00	90.39	140 000

[a] Bold letters designate population sizes in BC_1 selected for further simulations.

Larger population sizes (from 17 to 300) increased the proportion of the RPG recovered (from 80.37% to 88.68%) and the number of MDP needed (from 2380 to 42000). On the other hand, sizes higher than 300 caused only a slight increase in the proportion of the RPG recovered (approximately 1.7%), but required a substantial increase in the number of MDP. Therefore, BC_1 population sizes higher than 300 were not taken into further consideration.

Standard errors of the means across simulation runs for DG coverage and proportion of the RPG recovered, as a function of various BC_1 population sizes are given in Appendix, Table A2.

3.3.2 Influence of the number of BC and S generations

Progeny size per IL required to reach the RPG threshold

Considerable variation in progeny size per IL required to reach the RPG threshold was observed among and within simulated introgression strategies. The highest progeny size was observed in the BC_2S_1 introgression strategy. Increasing the number of BC or S generations resulted in a steady reduction in the required progeny size. Consequently, the BC_4S_2 strategy required the smallest progeny size (Table 6).

The effect of the BC_1 population size on the progeny size per IL was the stongest in the BC_2S_1 strategy, where an increase in the BC_1 population size from 17 to 300 individuals resulted in a decrease in progeny size from 180 to 62. Higher numbers of introgression generations significantly reduced the effect of the BC_1 population size on the progeny size per IL. For example, in the BC_4S_2 strategy an increase in BC_1 population size from 17 to 300 individuals resulted in progeny sizes of 9 to 6 (Table 6).

In the BC_2S_1 introgression strategy, the BC_1 population size of 200 and progeny size per IL of 65 individuals resulted in achieving the RPG threshold with the minimum number of MDP. This scenario was therefore considered as optimum for the BC_2S_1 strategy (Table 6, bold letters). In the optimum scenario for the BC_2S_2 strategy, a considerable reduction in progeny size (30) was observed for the same BC_1 population size of 200 individuals. Higher numbers of BC or S generations resulted in a continual decrease in the BC_1 population size and progeny size per IL in the optimum scenarios (Table 6).

Table 6. Progeny size per introgression line (IL) from generation BC_2 onwards, the number of marker data points (MDP) required to reach the recurrent parent genome (RPG) threshold of 95.60%, donor genome (DG) coverage and average number of non-target donor chromosome (DC) segments per IL, as a function of the number of backcross and selfing generations, based on constant progeny size from generation BC_2 onwards. Bold letters designate optimum scenarios for each introgression strategy.

Pop. size in generation BC_1	Introgression strategies					
	BC_2S_1	BC_2S_2	BC_3S_1	BC_3S_2	BC_4S_1	BC_4S_2
	---------- Progeny size per IL ----------					
17	180	57	27	18	11	9
40	126	45	23	16	10	9
60	106	38	21	15	**10**	**8**
80	95	37	20	14	10	8
100	90	35	**19**	**13**	9	8
130	84	33	19	13	9	8
160	76	33	18	13	8	7
200	**65**	**30**	17	12	8	7
300	62	27	16	12	8	6
	---------- MDP [a] ----------					
17	397 900	144 800	74 800	54 200	37 600	33 000
40	244 300	100 000	59 100	45 500	34 300	32 000
60	195 800	84 000	54 300	44 000	**33 800**	**30 500**
80	170 800	80 300	53 500	43 200	36 400	32 500
100	158 700	76 900	**52 700**	**42 200**	36 900	34 000
130	149 500	75 800	55 300	45 400	39 400	38 000
160	137 000	75 100	56 100	48 600	41 000	39 500
200	**122 000**	**73 400**	59 300	51 500	46 000	44 700
300	126 800	83 100	69 700	63 900	59 100	55 700
	---------- DG coverage (%) ----------					
17	100.00	100.00	99.94	99.96	97.22	98.69
40	100.00	100.00	99.84	99.98	96.69	98.77
60	100.00	100.00	99.78	99.95	**96.16**	**97.93**
80	100.00	100.00	99.66	99.92	96.08	97.69
100	100.00	100.00	**99.62**	**99.85**	94.78	97.85
130	100.00	100.00	99.62	99.81	94.46	97.65
160	100.00	100.00	99.55	99.90	92.99	95.88
200	**100.00**	**100.00**	99.27	99.83	92.53	95.71
300	100.00	100.00	99.30	99.76	92.93	92.81
	---------- Average number of non-target DC segments per IL ----------					
Optimum	**0.37**	**0.40**	**0.30**	**0.32**	**0.24**	**0.27**
Suboptimum value						
Highest	0.42	0.48	0.33	0.36	0.29	0.29
Lowest	0.36	0.40	0.27	0.28	0.21	0.21

[a] Values for MDP are rounded to multiples of hundred.

An increase in the BC_1 population size beyond optimum resulted only in a slight reduction in the progeny size per IL in all introgression strategies, even when the BC_1 population size was substantially higher (Table 6).

Comparing the strategies that contained the identical numbers of introgression generations (BC_2S_2 with BC_3S_1, and BC_3S_2 with BC_4S_1), those having a higher number of BC and a single S generation resulted in the lower BC_1 population size and progeny size per IL. For example, the optimum scenario for the BC_2S_2 strategy required 200 BC_1 individuals and progeny size of 30 plants, whereas the optimum scenario in the BC_3S_1 strategy required 100 BC_1 individuals and the progeny size of 19 individuals (Table 6).

Number of MDP required to reach the RPG threshold

A substantial range of variation in MDP required to reach the RPG threshold was observed among and within six simulated introgression strategies (Tables 6). In general, longer strategies required less MDP. Thus, the highest number of MDP was requred for the BC_2S_1 introgression strategy, and the lowest for the BC_4S_2 (Figure 18).

The effect of BC_1 population size on the number of MDP was considerably reduced with the increase in the number of introgression generations. The influence of the BC_1 population size was therefore the strongest in the BC_2S_1 and the lowest in the BC_4S_2 strategy (Table 6, Figure 18).

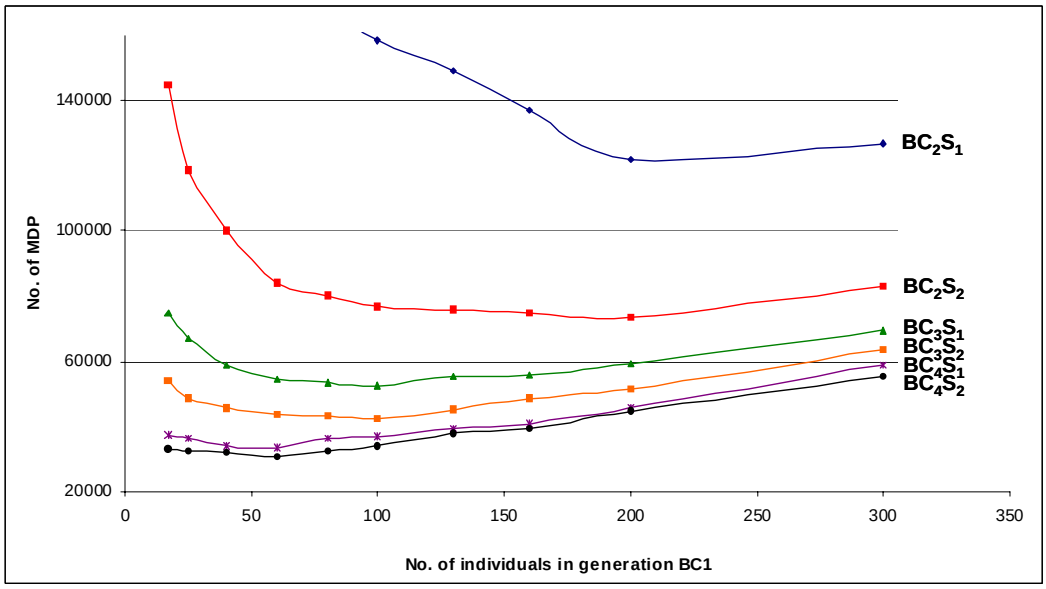


Figure 18. Number of marker data points (MDP) required to reach the recurrent parent genome threshold restriction in the BC_2S_1, BC_2S_2, BC_3S_1, BC_3S_2, BC_4S_1, and BC_4S_2 introgression strategies as a function of population size in the generation BC_1. Simulation was based on constant progeny size per introgression line from generation BC_2 onwards.

The number of MDP required for an optimally dimensioned introgression scenario was the highest (122 000 MDP) in the shortest procedure (BC_2S_2 strategy) and the lowest (33 500 MDP) in the longest one (BC_4S_2). The strategies that were equally long but with different number of BC and S generations (*e.g.* BC_2S_2 *vs.* BC_3S_1) also differed in the number of MDP. The strategies with a higher number of BC and lower number of S generations were always the most efficient ones, *e.g.* required lesss MDP (Table 6).

DG coverage

A relatively high coverage of the DG (92.81% - 100%) was achieved in all 54 simulated introgression scenarios (Tables 6). In the BC_2S_1 and BC_2S_2 strategies, a complete DG coverage (100%) was achieved with all applied BC_1 population sizes, whereas in the BC_3S_1 and BC_3S_2 strategies the coverage varied between 99 and 100%. Only the latter two strategies did not meet the targeted DG coverage of 99% when the RPG proportion of at least 95.60% was considered as the threshold restriction.

Average number of non-target DC segments per IL

The average number of non-target DC segments per IL varied between 0.21 and 0.48, with a trend towards lower values in the longer introgression strategies (Tables 6). Within each strategy, a higher number of individuals in generation BC_1 resulted in a lower average number of non-target DC segments per IL (data not shown).

Standard errors of the means across simulation runs for MDP number, DG coverage and average number of non-target DC segments per IL as a fuction of the number of BC and S generations are given in Appendix, Table A3.

Progeny size per IL in the BC_4S_1 and BC_4S_2 strategies satisfying the RPG threshold alone or both the RPG and DG thresholds

In the BC_4S_1 and BC_4S_2 strategies, the progeny sizes per IL were sufficient to meet the RPG threshold only (Table 6). Therefore, additional simulations were performed to define progeny sizes needed to satisfy both the RPG and DG threshold restrictions.

For the same BC_1 population size, progeny sizes required to meet both thresholds were considerably higher than those meeting the RPG threshold alone (Table 7).

Table 7. Progeny size per introgression line (IL) from generation BC_2 onwards, the number of marker data points (MDP), donor genome (DG) coverage, percentage of the recurrent parent genome (RPG) recovered, and number of non-target donor chromosome (DC) segments per IL when the RPG threshold alone is reached, compared to those when both the RPG and DG thresholds are reached. Simulations were based on constant progeny sizes per IL from generation BC_2 onwards. Bold letters designate optimum scenarios for each introgression strategy.

Population size in generation BC_1	BC_4S_1 strategy		BC_4S_2 strategy	
	RPG threshold alone	RPG and DG threshold	RPG threshold alone	RPG and DG threshold
	---------- Progeny size per IL ----------			
17	11	15	9	10
40	10	15	9	10
60	**10**	**15**	**8**	**10**
80	10	15	8	10
300	8	15	6	10
	---------- MDP [a] ----------			
17	37 600	48 600	32 900	35 100
40	34 300	45 500	32 000	34 700
60	**33 800**	**44 900**	**30 500**	**34 400**
80	36 400	46 900	32 500	36 800
300	59 100	71 200	55 700	62 900
	---------- DG coverage (%) ----------			
17	97.22	99.02	98.69	99.31
40	96.69	99.05	98.77	99.25
60	**96.16**	**99.00**	**97.93**	**99.20**
80	96.08	99.11	97.69	99.37
300	92.93	99.11	92.81	99.31
	---------- RPG recovered (%) ----------			
17	95.60	95.95	95.60	95.70
40	95.60	96.00	95.67	95.81
60	**95.63**	**96.08**	**95.61**	**95.84**
80	95.68	96.12	95.67	95.91
300	95.63	96.26	95.60	96.04
	---------- Average number of non-target DC segments per IL ----------			
Optimum	0.24	0.19	0.27	0.23
Suboptimum value				
Highest	0.29	0.21	0.29	0.27
Lowest	0.21	0.15	0.21	0.18

Different BC_1 population sizes in the strategies that satisfied both thresholds did not cause any changes in the progeny size per IL, neither in BC_4S_1 nor in BC_4S_2 introgression strategy. The required values for progeny size were 15 in the BC_4S_1 strategy, and 10 in the BC_4S_2 strategy (Table 7).

The requirement to meet (not only the RPG but also) the DG threshold caused the increase in the MDP number, as well as the surpass of the RPG threshold (Table 7).

Increase in the BC_1 population size from 17 to 300 individuals induced a continuous reduction in the average number of non-target DC segments per IL in both strategies (data not shown). Moreover, scenarios that satisfied both thresholds resulted in a lower average number of non-target DC segments per IL (Table 7).

Standard errors of the means across simulation runs for number of MDP, DG coverage, percentage of the RPG recovered, and average number of non-target DC segments per IL when RPG threshold alone or both thresholds were reached are given in Appendix, Table A4.

3.3.3 Comparison of introgression variants with increasing and decreasing progeny size per IL

As BC_3S_1 strategy was optimal when a constant progeny size per IL was applied, it was additionally employed to compare introgression variants with increasing and decreasing progeny size per IL, starting with a BC_1 population size of 100 individuals (Tables 8 and 9).

Introgression variant with increasing progeny size per IL

The total number of individuals needed to reach the RGP threshold varied from 2224 (scenario with ratio 1:1.5:2) to 2908 (scenario with ratio 1:3:9). All scenarios with increasing ratios required a higher total number of individuals than the one with the constant ratio 1:1:1 (Table 8).

The increase in progeny size ratio hardly influenced the required number of MDP, which varied from 48500 (in the 1:1.5:2 scenario) to 54000 (in the 1:1:3 scenario). Regarding the number of MDP, most of the scenarios within the variant with increasing progeny size required somewhat less MDP than the scenario with the constant ratio. A single exception was the ratio 1:1:3 (Table 8).

All simulated scenarios with increasing progeny size ratios, except the scenario with the ratio 1:3:9, resulted in the DG coverage higher than 99% threshold value (Table 8).

Table 8. Progeny size per introgression line (IL) from generation BC_2 onwards, total number of individuals, and number marker data points (MDP) required to reach recurrent parent genome threshold of 95.60%, coverage of the donor genome (DG), and the average number of non-target donor chromosome (DC) segments per IL in BC_3S_1 strategy, as a function of increasing progeny size from generation BC_2 to BC_3S_1. Scenarios were simulated for a BC_1 population size of 100 individuals.

		Increasing progeny size ratio (BC_2 – BC_3S_1)				
Generation	**1:1:1**	**1:1.5:2**	**1:1:2**	**1:2:3**	**1:1:3**	**1:3:9**
	---------- Progeny size per IL ----------					
BC_2	19	**13**	16	11	15	6
BC_3	19	**20**	16	22	15	18
BC_3S_1	19	**26**	32	33	45	54
	---------- Total number of individuals ----------					
	2152	**2224**	2332	2476	2800	2908
	---------- MDP [a] ----------					
	52300	**48500**	51900	50000	54000	50700
	---------- DG coverage (%) ----------					
	99.62	**99.90**	99.60	99.84	99.97	98.52
	---------- Average number of non-target DC segments per IL ----------					
	0.30	**0.30**	0.30	0.29	0.27	0.34

[a] Values for MDP are rounded to multiples of hundred.

In comparison with the scenario with the constant progeny size ratio, all scenarios (except the one with the ratio 1:3:9) resulted in either equal or slightly lower average number of non-target DC segments per IL (Table 8).

Standard errors of the means across simulation runs for the MDP number, DG coverage, and average number of non-target DC segments per IL, as a function of increasing progeny size are given in Appendix, Table A5.

Introgression variant with decreasing progeny size per IL

Among five scenarios with decreasing progeny sizes, a relatively low range of variation was observed in the total number of individuals required to reach the RGP threshold (from 1972 in the ratio 9:3:1 to 2440 in the ratio 3:1:1). In comparison with the scenario with the constant progeny size ratio of 1:1:1, only scenario with the ratio 9:3:1 required lower total number of individuals (Table 9).

Table 9. Progeny size per introgression line (IL) from generation BC_2 onwards, total number of individuals and number of marker data points (MDP) required to reach recurrent parent genome threshold of 95.60%, coverage of the donor genome (DG) and average number of non-target donor chromosome (DC) segments per IL in BC_3S_1 strategy, as a function of decreasing progeny size from generation BC_2 to BC_3S_1. Scenarios were simulated for a BC_1 population size of 100 individuals.

		Decreasing progeny size ratio (BC_2 – BC_3S_1)				
Generation	**1:1:1**	**2:1.5:1**	**2:1:1**	**3:2:1**	**3:1:1**	**9:3:1**
	---------- Progeny size per IL ----------					
BC_2	19	25	30	29	39	36
BC_3	19	19	15	19	13	12
BC_3S_1	19	12	15	9	13	4
	---------- Total number of individuals ----------					
	2152	2162	2260	2152	2440	1972
	---------- MDP [a] ----------					
	52 300	57 600	62 300	61 500	72 700	66 400
	---------- DG coverage (%) ----------					
	99.62	97.92	99.06	95.30	98.30	83.25
	---------- Average number of non-target DC segments per IL ----------					
	0.30	0.30	0.29	0.28	0.28	0.22

[a] Values for MDP are rounded to multiples of hundred.

Decrease in the progeny size ratio increased the number of MDP from 57600 (scenario with the ratio 2:1.5:1) to 72700 (scenario with the ratio 3:1:1). Regarding the number of MDP, all scenarios within the variant with decreasing progeny size required more MDP than the scenario with the constant ratio (Table 9).

The resulting coverage of the DG was considerably affected by decreasing progeny size ratios (Table 9). Scenario with the ratio 9:3:1 resulted in the DG coverage of only 83.25%, whereas the highest DG coverage (99.06%) was observed in the scenario with the ratio 2:1:1. The scenario with the ratio 2:1:1 was the only one meeting the DG threshold restriction of 99%. However, even this scenario did not perform better in the DG coverage than the scenario with the constant ratio (99.62%).

The average number of non-target DC segments per IL was hardly influenced by decreasing progeny size ratios (Table 9). In comparison with the scenario with the constant ratio, all scenarios resulted in either equal or slightly lower average number of non-target DC segments per IL (Table 9).

Standard errors of the means across simulation runs for the number of MDP, DG coverage, and average number of non-target DC segments per IL, as a function of decreasing progeny size are given in Appendix, Table A6.

3.3.4 Influence of the DC segment length and marker density

Donor chromosome segment length

Increase in the DC segment length from 20 cM to 40 cM, while maintaining the same marker density of 5 cM, reduced the optimal BC_1 population size in the BC_3S_1 introgression strategy from 100 to 60 (bold letters, Table 10). Moreover, a large reduction in the progeny size per IL was observed for all BC_1 population sizes, in both the BC_2S_1 and BC_3S_1 strategies (variant 20/5 *vs.* variant 40/5, Table 10).

In BC_2S_1 and BC_3S_1 strategies a substantial reduction of roughly 50% was revealed in the total number of individuals required to reach the RPG threshold when DC segment length was increased from 20 cM to 40 cM, in the BC_2S_1 and BC_3S_1 strategies (Table 10). Furthermore, the number of MDP required to reach the RPG threshold was approximately halved with the increased DC segment length, in both strategies (Table 10).

Increase in DC segment length in the BC_2S_1 introgression strategy did not cause any changes in the DG coverage (Table 11). In both 20/5 and 40/5 variants, the DG coverage reached maximum value of 100% for all BC_1 population sizes. In the BC_3S_1 strategy, however, increase in DC segment length reduced the DG coverage, especially in scenarios with larger population sizes in generation BC_1. Thus, in the BC_3S_1 strategy variant 20/5, all simulated scenarios surpassed the 99% threshold value, whereas in the variant 40/5 scenarios beginning with BC_1 population size larger than 40 individuals did not reach this restriction (Table 11).

Increase in DC segment length in both strategies slightly increased the average number of non-target DC segments per IL (Table 11).

Standard errors of the means across simulation runs for the number of MDP, DG coverage and average number of non-target DC segments per IL, as a function of DC segment length are given in Appendix, Table A7.

Table 10. Progeny size per introgression line (IL) from generation BC_2 onwards, total number of individuals and number of marker data points (MDP) required to reach recurrent parent genome threshold of 95.60%, as a function of donor chromosome (DC) segment length and marker density. Bold letters designate optimum scenarios for each introgression strategy.

Population size in generation	Combinations of DC segment length and marker density					
	BC_2S_1 strategy			BC_3S_1 strategy		
BC_1	20/5[a]	40/5[b]	40/20[c]	20/5[a]	40/5[b]	40/20[c]
	Progeny size per IL					
17	180	140	52	27	23	11
25	148	118	43	25	22	10
40	126	105	36	23	20	**9**
60	106	85	33	21	**18**	8
80	95	74	30	20	17	8
100	90	71	**28**	**19**	17	7
130	84	66	26	19	16	7
160	76	61	24	18	15	7
200	**65**	**54**	22	17	15	7
300	62	50	20	16	14	6
	Total number of individuals					
17	12 977	5 057	1 889	2 933	1 259	611
25	10 681	4 273	1 573	2 725	1 213	565
40	9 112	3 820	1 336	2 524	1 120	**526**
60	7 692	3 120	1 248	2 328	**1 032**	492
80	6 920	2 744	1 160	2 240	998	512
100	6 580	2 656	**1 108**	**2 152**	1 018	478
130	6 178	2 506	1 066	2 182	994	508
160	5 632	2 356	1 024	2 104	970	538
200	**4 880**	**2 144**	992	2 036	1 010	578
300	4 764	2 100	1 020	2 028	1 056	624
	MDP [d]					
17	379 900	176 800	20 300	74 800	38 600	6 000
25	309 100	140 100	16 200	67 100	35 800	5 500
40	244 300	118 200	13 400	59 100	33 200	**5 400**
60	195 800	93 700	12 600	54 300	**31 500**	5 800
80	170 800	82 000	12 300	53 500	32 500	6 500
100	158 700	80 100	**12 200**	**52 700**	34 300	6 900
130	149 500	76 700	12 400	55 300	36 700	8 100
160	137 000	75 300	13 200	56 100	39 500	9 400
200	**122 000**	**73 000**	14 300	59 300	44 600	11 000
300	126 800	81 000	17 700	69 700	56 600	14 900

[a] 20/5 - variant with 20 cM DC segment length and 5 cM marker density (standard scenario).
[b] 40/5 - variant with 40 cM DC segment length and 5 cM marker density.
[c] 40/20 - variant with 40 cM DC segment length and 20 cM marker density.
[d] Values for number of MDP are rounded to multiples of hundred.

Table 11. Coverage of the donor genome (DG) and average number of non-target DC segments per IL when recurrent parent genome (RPG) threshold is reached as a function of donor chromosome (DC) segment length and marker density. Bold letters designate optimum scenarios for each introgression strategy.

Population size in generation	Combinations of DC segment length and marker density					
	BC_2S_1 strategy			BC_3S_1 strategy		
BC_1	20/5[a]	40/5[b]	40/20[c]	20/5[a]	40/5[b]	40/20[c]
	---------- DG coverage (%) ----------					
17	100.00	100.00	99.95	99.94	99.38	95.93
25	100.00	100.00	99.99	99.92	99.34	95.20
40	100.00	100.00	99.98	99.84	99.23	**93.74**
60	100.00	100.00	99.91	99.78	**98.84**	91.38
80	100.00	100.00	99.85	99.66	98.62	91.14
100	100.00	100.00	**99.83**	**99.62**	98.62	89.64
130	100.00	100.00	99.71	99.62	98.48	89.51
160	100.00	100.00	99.67	99.55	97.99	89.21
200	**100.00**	**100.00**	99.56	99.27	97.92	87.85
300	100.00	100.00	99.27	99.30	97.67	85.35
	---------- Average number of non-target DC segments per IL ----------					
Optimum	**0.37**	**0.45**	**0.75**	**0.30**	**0.38**	**0.66**
Suboptimum						
Highest value	0.44	0.51	0.83	0.34	0.42	0.70
Lowest value	0.36	0.42	0.69	0.27	0.34	0.52

[a]20/5 - variant with 20 cM DC segment length and 5 cM marker density (standard scenario).
[b] 40/5 - variant with 40 cM DC segment length and 5 cM marker density.
[c] 40/20 - variant with 40 cM DC segment length and 20 cM marker density.

Marker density

Lower marker density (variant 40/5 *v.s.* variant 40/20) reduced of the optimal BC_1 population size (from 200 to 100 in the strategy BC_2S_1 and from 60 to 40 in the strategy BC_3S_1) and caused a substantial reduction in the progeny size per IL for all BC_1 population sizes (Table 10).

Total number of individuals required to reach the RPG threshold was more than halved with lower marker density (Table 10). Furthermore, lower marker density resulted in large reduction in the MDP number, in both strategies (Table 10). Thus, the variant 40/20 required approximately four to eight times less MDP (Table 10).

In the BC_2S_1 strategy, decrease in marker density from 5 cM to 20 cM slightly reduced the complete DG coverage (Table 11). Nevertheless, all scenarios with lower density (variant 40/20) surpassed the DG threshold restriction of 99%. However, in BC_3S_1 strategy, larger reduction in DG coverage was observed with decrease in marker density, especially for

scenarios beginning with larger BC_1 population sizes. Thus, in the variant 40/5 threshold for the DG coverage was reached only when BC_1 population size was not larger than 40 individuals, whereas in all scenarios within the variant 40/20 the DG coverage was considerably below the threshold value of 99% (Table 11).

Lower marker density in both BC_2S_1 and BC_3S_1 introgression strategies increased the average number of non-target DC segments per IL (Table 11).

Standard errors of the means across simulation runs for the number of MDP, DG coverage and average number of non-target DC segments per IL, as a function of marker density are given in Appendix, Table A7.

3.4 Discussion

3.4.1 Optimal dimensioning of the BC_1 population size

Population size in generationBC_1 should be large enough to provide a sufficiently high probability of finding parent individuals that will jointly cover the complete DG and, in addition, recover a high proportion of the RPG. The first of these requirements can easily be achieved with a relatively small population size (N≈20) (Table 5). However, introgression scenarios starting with a small BC_1 population size allow only mild background selection against non-target DC segments. Consequently, larger progeny sizes per IL and more MDP are required in subsequent BC and S generations to meet the RPG threshold restriction (Table 6). On the other hand, increasing the BC_1 population and decreasing progeny size per IL within a particular introgression strategy is reasonable only until a reduction in the total number of MDP is observed (Table 6). Further increase in BC_1 population increases MDP number even though applied progeny size per IL is lower. Generation BC_1 is the most demanding concerning the number of MDP needed per individual. Thus, marker assays in generation BC_1 affect significantly the final cost of a certain introgression scenario, and an optimal dimensioning of the BC_1 population size is an important prerequisite for an effective marker-assisted introgression procedure.

Optimizing the BC_1 population size is more important in the short introgression strategies than in the longer ones since progeny size from BC_2 onwards and consequently the number of MDP needed to meet the RPG threshold react more strongly to the BC_1 population size in the short strategies (Table 6).

In the two longest introgression strategies (BC_4S_1 and BC_4S_2), even a very large BC_1 population size does not meet the DG coverage threshold because the progeny size per IL needed in subsequent generations to meet the RPG threshold are too small (Table 6).

3.4.2 Influence of the number of BC and S generations

Short introgression strategies require larger progeny sizes per IL and, consequently, more MDP than long strategies, because the probability of finding individuals carrying the target DC segment and meeting the RPG threshold level is smaller in early than in late introgression generations (Table 6).

Lower coverage of the DG was the limiting factor for the establishment of the introgression libraries using longer (BC_4S_1 and BC_4S_2) strategies (Table 6). To determine the progeny size per IL sufficient to meet both RPG and DG threshold restrictions in the BC_4S_1 and BC_4S_2 strategies, additional simulations were performed. Results show that the double restriction considerably (up to 50%) increases the required progeny size per IL and, consequently, the number of MDP (Table 7). This increase did not only enhance the proportion of the DG coverage but also that of the RPG recovery. As a consequence, the number of non-target DC segments decreased (Table 7). These two facts are of high importance for obtaining "real" NILs, carrying from the donor nothing else but a single target segment.

In conclusion, the optimal dimensioning of the progeny size per IL in introgression strategies is of decisive importance for meeting the DG threshold restriction, particullary in longer introgression strategies.

The number of non-target DC segments per IL was not regarded as a restriction in any introgression strategy. Results demonstrate that this parameter does not vary considerably among the optimized introgression strategies (Table 6).

An additional BC instead of one S generation makes the establishment of an introgression library by marker-assisted backcrossing more efficient (Table 6). If strategies that containe identical numbers of introgression generations are compared (*e.g*, BC_2S_2 with. BC_3S_1, and BC_3S_2 with. BC_4S_1), those having a higher number of BC and a single S generation require lower progeny sizes per IL and less MDP (Table 6). Moreover, optimum scenarios of the introgression strategies comprising more BC generations (*e.g*, BC_2S_1 *v.s.* BC_3S_1, and BC_2S_2 *v.s.* BC_3S_2) require smaller BC_1 population sizes to meet the RGP threshold restriction. The explanation lies in the fact that, in addition to the marker-assisted background selection, backcrossing itself increases the proportion of the RPG in successive BC generations. By expectation, the DG proportion is halved with each BC generation, irrespective of its amount present in the non-recurrent parent. A lower BC_1 population size was required for optimum scenarios of the introgression strategies comprising more BC generations because additional backcrosses accelerate the RPG recovery. In S generations, however, reducing the DG proportion is apparently lower and can be obtained only by selection among segregants for high RPG proportion. Thus, a parent individual carrying the target DC segment and a sufficiently high RPG proportion can be found only if sufficiently large progeny sizes are grown. In addition, progenies obtained by selfing carry a larger proportion of the DG than those obtained by backcrossing in the same introgression generation (*e.g.*, BC_2S_2 IL *v.s.* BC_3S_1

IL). This fact, together with a larger progeny size per IL, increases the total number of MDP required for marker-assisted background selection in shorter introgression strategies.

To designate which of the six simulated introgression strategies can be recommended, two efficiency criteria should be considered: *i*) the final cost and *ii*) the length of the introgression strategy. The total number of MDP may be considered as the major cost factor. Substantial reductions in the number of MDP are achievable by choosing long instead of short introgression strategies (Table 6). This particularly holds true when comparing strategies BC_2S_1, BC_2S_2 and BC_3S_1 with each other. Further increase in the number of BC and S generations (BC_3S_2, BC_4S_1 and BC_4S_2 strategies) results in declining cost savings. Reductions are even smaller if BC_4S_1 and BC_4S_2 strategies meeting both the RPG and DG threshold restrictions are taken into consideration (Table 7).

The length of a breeding cycle is of decisive importance in plant breeding, where shorter cycles are usually preferable. Thus, a medium long strategy, *e.g.* BC_3S_1 procedure, may be a good compromise between cost and speed of an introgression program.

3.4.3 Comparison of introgression variants with increasing and decreasing progeny size per IL

Introgression variant with increasing progeny size (especially the scenario with slightly increasing ratio 1:1.5:2, Table 8) is more recommendable for the establishment of a rye introgression library than the variants with constant or decreasing progeny size (Table 9), because a smaller number of MDP is needed to meet the RPG threshold without a reduction in DG coverage and without an increase in the average number of non-target DC segments per IL.

The number of MDP needed in one scenario depends not only on progeny size per IL but also on the number of DC segments (target and non-target). Thus, for reducing the MDP number, it is reasonable to increase progeny size in advanced introgression generations when the number of non-target DC segments is already reduced by the BC procedure. This way, the number of MDP needed for background selection in advanced introgression generations is lower. On the other hand, a small progeny size in the early BC generations allows only very mild background selection against non-target DC segments. This disadvantage, however, seems to be minor compared to the advantage of a strong background selection in more advanced generations.

Variant with decreasing progeny size (especially scenarios with broad ratios, 3:2:1 and 9:3:1) is particularly unfavourable because it requires larger number of MDP and, moreover, results in a DG coverage below the threshold restriction of 99% (Table 9). Larger number of MDP is the result of a strong background selection against non-target DC segments in early BC generations, when large proportion of the DG is present. In addition, DG coverage below the threshold restriction is attributable to the small progeny size in the final BC_3S_1 generation that was not sufficient to avoid losing a considerable portion of the target DC segments.

3.4.4 Influence of the DC segment length and marker density

Targeting longer DC segments and using genetic maps with lower marker density allow a remarkable reduction in input parameters (progeny size per IL, total number of individuals, and number of MDP) required to reach the RPG threshold restriction (Table 10). This is expected since *i*) fewer ILs are needed to cover the complete DG if the target DC segments are longer and *ii*) a lower marker density leads to a smaller number of characterised DC segments. Lower number of individuals to be genotyped and markers to be employed lead to a substantial reduction in the overall breeding and genotyping effort.

However, achieving DG coverage is hampered by longer DC segments and lower marker density in scenarios in which a reduced progeny size per IL is used. Reduced progeny size allows only mild foreground selection, resulting in DG coverage below the threshold restriction (BC_3S_1 strategy, variants 40/5 and 40/20) (Tables 10 and 11). Large reduction in the progeny size also reduces the effectiveness of background selection and substantially increases the average number of non-target DC segments per IL (Table 11).

Development of the ILs carrying a single target DC segment and, subsequently, the isolation of putative candidate genes are two main aims for using ILs in breeding. Larger DC segments, however, increase the probability of carrying unfavourable loci as well. Moreover, the longer DC segments are the more likely they carry more than one gene controlling the trait in question. If those genes are associated in the repulsion phase, favourable gene effects can be cancelled partially or completely. On the other hand, if two favourable genes are associated in the coupling phase, they reinforce each other simulating one strong ("ghost") gene. Furthermore, large DC segments harbour many additional loci affecting other traits. This may cause problems in the process of identification and isolation of genes controlling the trait of interest, if strong unfavourable interactions occur between the target and non-target genes.

The major disadvantage of the genetic maps with large marker distances is the unknown information about possible double cross-over within a particular marker interval. In case of an exotic donor, large segments increase the unfavourable linkage drag around the target genes, and thus may mask the positive effect of the introgressed DC segments. If the introgressed chromosome segments originate from an elite donor source, double cross-over within a DC segment may remain undetected and leads to misinterpretation of the effects of such a long DC segment. Moreover, if double cross-over remains undetected within already "recovered" RPG, it results in a new, not characterized DC segment that may have an influence on already characterised DC segments and their effects on line performance.

In conclusion, lower initial effort for the establishment of an introgression library will later on require additional efforts for an efficient utilisation of such ILs in breeding and genomics.

4 Final Remarks

4.1 Deviations of the experimental study from an optimum dimensioning

This study describes a new approach of utilizing exotic rye germplasm for broadening the genetic base of elite rye breeding materials by enhancing genetic diversity for economically important agronomic traits.

Since the results of the simulation study became available after the marker-assisted establishment of the two rye introgression libraries had been finished, dimensioninig of the experimental study deviated from that determined as optimum in the simulation study:

a) The BC_2S_2 introgression strategy was used in the experimental study, whereas the BC_3S_1 strategy proved most recommendable in the simulation study.

b) The BC_1 population sizes used for the establishment of libraries F and G (68 and 69, respectively) were far from the optimum value (200) observed in the simulation study for the chosen BC_2S_2 strategy.

c) The mean progeny sizes per IL from generation BC_2 onwards varied between 7 and 21 whereas the optimum progeny size would have been two to three times higher.

d) The total number of individuals employed (690 in library F, 684 in library G) was approximately five times lower than the optimum number as determined in the simulation study (3440).

As a consequence of those deviations, the DG coverage was incomplete (library F 72%, library G 63%, jointly approx. 80%), and most ILs harboured more than a single DC segment. The application of the results of the simulation study would have increased considerably the value of the developed ILs (higher DG coverage, lower number of DC segments per IL), even with limited resources. Anyway, the experimental study resulted in the successful establishment of the first two rye introgression libraries worldwide, comprising 40 BC_2S_3 ILs per library and covering large portion of the donor genome.

4.2 Outlook and final conclusion

Before applying the introgression libraries in practical rye breeding, the effects of the introgressed DC segments on agronomically important qualitative and quantitative traits need to be examined, by evaluating the ILs *per se* and their testcrosses in multi-environment field experiments. Introgression lines with beneficial DC segments may directly be used in practical hybrid rye breeding programs.

Moreover, selected ILs showing favourable phenotypic performances may further be backcrossed to create NILs, carrying a single marker-characterized short DC segment each. These NILs would be an ideal starting point for high-resolution mapping of candidate genes located in a DC segment, as well as for the isolation and functional characterisation of genes, of economic importance. Thus, the two rye introgression libraries and the results of the simulation study mark important milestones for the targeted exploitation of exotic rye germplasm and provide a promising opportunity to proceed towards functional genomics in rye.

5 Summary

The introgression of exotic germplasm is a promising approach to increase the genetic diversity of elite rye breeding materials. Even though exotic germplasm may contain genomic segments that can improve oligo- and polygenically inherited traits, it has not been intensively utilized in modern rye breeding due to its agronomically inferior phenotypes and low performance level. Introgression of exotic germplasm requires techniques that would minimize negative side effects attributable to genetic interactions between recipient and donor. This appears achievable by the introgression library approach involving the systematic transfer of donor chromosome (DC) segments from an agriculturally unadapted source (donor) into an elite line (recipient, recurrent parent). A set of introgression lines (ILs) is thus developed, in which introgression is restricted to one or a few short DC segments. Ideally, the introgressed DC segments are evenly distributed over the whole recipient genome and the total genome of the exotic donor is comprised in the established set of ILs.

The systematic development of an introgression library in rye has not been described yet. The main objectives of this study were to *i*) establish two rye introgression libraries by marker-assisted backcrossing, comprising of ILs each harbouring one to three DC segments and jointly covering most of the donor genome (DG), and *ii*) apply computer simulations to develop a highly effective and cost-efficient marker-assisted introgression strategy for the creation of introgression libraries in rye.

A cross between a homozygous elite rye inbred line L2053-N (recurrent parent) and a heterozygous Iranian primitive rye population Altevogt 14160 (donor) was used as base material to generate the two libraries (F and G). Repeated backcrossing (BC) and subsequent selfing (S) until generation BC_2S_3 were chosen as the introgression method. The AFLP and SSR markers were employed to select individuals carrying desired DC segments, starting from generation BC_1 to generation BC_2S_2. The chromosomal localization of DC segments, the number of DC segments per IL, and the proportion of recurrent parent genome were used as criteria to select parent individuals. This procedure resulted in the first two rye introgression libraries worldwide, comprising 40 BC_2S_3 ILs per library and covering 72% of the total DG in library F and 63% in library G (jointly approximately 80%). Most of the established ILs harboured one to three homozygous DC segments (on average 2.2 in both libraries), with a mean length of 18.3 cM in library F and 14.3 cM in library G.

Computer simulations were conducted using the software PLABSIM version 2 to evaluate and optimize strategies for developing an introgression library in rye. Simulations were based on map-length estimates obtained from genotyping the BC_1 generation of population F (7 chromosome pairs, genome size 665 cM). Six strategies differing in the number of BC and S generations were analysed, by setting the restrictions of sufficient DG coverage and RPG recovery. The medium-long BC_3S_1 strategy proved to be the most recommendable. It allows to achieve close to 100% DG coverage with moderate progeny sizes (19 individual per IL) in the individual generations and an acceptable total number of marker data points (52700), thus providing a good compromise between the cost and speed of an introgression procedure. Longer strategies are somewhat more cost-efficient but too time-demanding. The reverse is true for shorter strategies. An optimal allocation of resources is achieved by starting an introgression strategy with a small BC_1 population (between 60 and 200 individuals) and stepwise increasing the progeny size per IL from about 15 to about 25-35 individuals in the succeeding generations.

Targeting longer DC segments and using genetic maps with lower marker density allow a remarkable reduction in resources. This approach, however, possesses shortcomings when implementation in breeding is considered. The longer DC segments more likely carry i) unfavourable loci as well, ii) more than one gene controlling the trait in question, or iii) many additional loci affecting other traits. The major disadvantage of genetic maps with large marker distances is the unknown information about possible double cross-overs within marker intervals. All above-mentioned disadvantages may cause problems in the process of identification and isolation of genes controlling the trait of interest. Thus, a lower initial effort for the establishment of an introgression library will later on require additional efforts for using the ILs in breeding and genomics.

Since the results of the simulation study became available after the marker-assisted establishment of the two rye introgression libraries had been finished, the dimensioning of the experimental study deviated from the optimum dimensioning determined in the simulation study: *i*) The BC_2S_2 introgression strategy was used in the empirical approach, whereas the BC_3S_1 strategy proved to be most recommendable in the simulation study. *ii*) The BC_1 population sizes of libraries F and G (68 and 69, respectively) were far below the optimum value (200) determined in the simulation study for the chosen BC_2S_2 strategy. *iii*) The mean progeny sizes per IL from generation BC_2 onwards varied between 7 and 21, whereas the optimum progeny size would have been two to three times higher. *iv*) The total number of analysed individuals (690 in library F, 684 in library G) was considerably lower than the

optimum determined in the simulation study (3440). As a consequence, the coverage of the donor genome in the two libraries was incomplete and most ILs harboured more than a single DC segment. The potential application of the results of the simulation study would have increased the value of the developed ILs (higher DG coverage, lower number of DC segments per IL) considerably, despite limited resources.

The effects of the introgressed DC segments on agronomically important qualitative and quantitative traits still need to be examined in multi-environmental field experiments. Introgression lines with beneficial DC segments may directly be used in practical hybrid rye breeding programs. Moreover, such ILs may be further backcrossed to create near isogenic lines (NILs) each carrying a single marker-characterized short DC segment. These NILs are an ideal starting point for high-resolution mapping and for the isolation and functional characterisation of candidate genes.

The two rye introgression libraries and the results of the simulation study mark important milestones for the targeted exploitation of exotic rye germplasm and provide a promising opportunity to proceed towards functional genomics in rye.

6 Zusammenfassung

Durch die Einkreuzung genetischer Ressourcen kann die Diversität von Elitezuchtmaterial bei Roggen erhöht werden. Diese vielversprechende Möglichkeit wurde in der aktuellen Züchtung bisher kaum genutzt, obwohl exotisches Genmaterial einzelne günstige agronomische Eigenschaften besitzen kann. Der Hauptgrund dafür ist die geringe Anpassung solcher Populationen an unsere Produktions- und Umweltbedingungen. Durch die markergestützte Entwicklung von Introgressionsbibliotheken, die zur systematischen Übertragung von einzelnen Donorchromosom (DC)-Segmenten aus einer nicht-adaptierten Quelle (Donor) in Elitezuchtmaterial (Empfänger, Rekurrenter Elter) führen, könnten agronomisch ungünstige Nebeneffekte vermieden werden. Dabei werden Introgressionslinien (ILs) entwickelt, die nur ein oder wenige, kurze DC-Segmente enthalten. Die eingelagerten DC-Segmente sollten dabei gleichmäßig über das Empfängergenom verteilt sein und möglichst das Gesamtgenom des Donors repräsentieren.

Eine systematische Entwicklung einer Introgressionsbibliothek wurde bei Roggen bisher noch nicht beschrieben. Die wesentlichen Ziele der Arbeit waren deshalb (1) zwei Introgressionsbibliotheken des Roggens durch markergestützte Rückkreuzung zu entwickeln, deren ILs jeweils nur ein bis drei DC-Segmente beinhalten, die zusammen den größten Teil des Donorgenoms abdecken und (2) durch Computersimulationen kosten- und zeiteffiziente Strategien für den Aufbau von Introgressionsbibliotheken zu entwickeln.

Ausgangspunkt der Erstellung der Introgressionsbibliotheken (F bzw. G) war eine Kreuzung zwischen der homozygoten Elitelinie L2053-N (Rekurrenter Elter) und der heterozygoten iranischen Primitivroggenpopulation Altevogt 14160 (Donor). Durch zwei Rückkreuzungen und anschließende Selbstungen wurde das Material bis zur BC_2S_3-Generation entwickelt. AFLP- und SSR-Marker dienten in den ersten vier Generationen dazu, die Einzelpflanzen mit den erwünschten DC-Segmenten zu selektieren. Als Selektionskriterien dienten die chromosomale Lokalisation der DC-Segmente, die Anzahl DC-Segmente pro IL und der Anteil des rekurrenten Elterngenoms. Dadurch entstanden die weltweit ersten beiden Roggen-Introgressionsbibliotheken. Sie bestehen aus je 40 BC_2S_3-ILs und umfassen 72% (Bibliothek F) bzw. 63% (Bibliothek G) des gesamten Donorgenoms; zusammen wurden rund 80% erreicht. Die meisten ILs enthalten ein bis drei homozygote DC-Segmente (Mittel 2,2) mit einer mittleren Länge von 18.3 cM in Bibliothek F und 14.3 cM in Bibliothek G.

Die Computersimulationen wurden mit dem Programmpaket PLABSIM, Version 2, durchgeführt, um eine optimale Introgressionsstrategie für Roggen zu entwickeln. Sie basierten

auf sieben Chromosomen mit einer durchschnittlichen Kartenlänge der Introgressionsbibliothek F von 665 cM. Sechs unterschiedliche Strategien, die sich in der Anzahl der Rückkreuzungs- und Selbstungsgenerationen unterschieden, wurden untersucht, wobei als Kriterien eine ausreichende Donorgenomabdeckung und ein maximaler Anteil des Empfängergenoms dienten. Als am effizientesten erwies sich der Aufbau einer BC_3S_1- Bibliothek. Dabei können mit einer relativ geringen Anzahl an Nachkommen je IL (N=19) und einer angemessenen Anzahl an Markerdatenpunkte (52.700) nahezu 100% des Donorgenoms abgedeckt werden. Eine länger dauernde Introgression ist etwas kostengünstiger, aber natürlich zeitaufwändiger. Strategien kürzerer Dauer sind wesentlich teurer. Eine optimale Allokation an Ressourcen wird erreicht, wenn man die Introgression mit einer kleinen BC_1-Population (60 bis 200 Einzelpflanzen) beginnt und in den nachfolgenden Generationen die Nachkommenschaftsgröße je IL schrittweise erhöht (etwa 15 auf 25-35 Einzelpflanzen).

Die Kapazitäten können wesentlich vermindert werden, wenn längere DC-Segmente angestrebt bzw. genetische Karten mit geringerer Markerdichte verwendet werden. Dies hat jedoch für die Verwendung in der Pflanzenzüchtung erhebliche Nachteile. Längere DC-Segmente enthalten mit größerer Wahrscheinlichkeit (1) auch agronomisch ungünstige Loci, (2) mehr als ein Gen des Zielmerkmals, wenn es sich um polygenische Eigenschaften handelt, oder (3) zahlreiche Gene, die andere Eigenschaften betreffen. Der größte Nachteil von genetischen Karten mit großen Markerabständen ist die Unsicherheit bezüglich doppelter Crossover-Ereignisse innerhalb der Markerintervalle. All diese Nachteile verursachen Probleme bei der späteren Identifikation der Zielgene und der Genisolation. Eine zu starke Kostenreduktion am Beginn der Introgression muss daher mit erhöhten Kosten bei der Verwendung der IL in der praktischen Züchtung und bei genomischen Ansätzen erkauft werden.

Die Ergebnisse der Simulationsstudien lagen erst vor als die experimentelle Erstellung der beiden Roggen-Introgressionsbibliotheken abgeschlossen war. Deshalb ergaben sich folgende Unterschiede zur optimalen Strategie: (1) Es wurde die BC_2S_2–Generation genutzt, obwohl die BC_3S_1–Generation in der Simulationsstudie besser abschnitt. (2) Die Populationsgröße der BC_1–Generation war für die Introgressionsbibliothek F (N=68) bzw. G (N=69) weit von der optimalen Größe (N=200) entfernt, (3) die mittlere Anzahl Nachkommen je IL ab der BC_2-Generation schwankte zwischen 7 und 21, während die zwei- bis dreifache Größe optimal gewesen wäre, (4) die Gesamtzahl bearbeiteter Einzelpflanzen (N=690 bzw. 684 für Introgressionsbibliothek F bzw. G) war deutlich geringer als aufgrund der Simulationsstudie erforderlich (N=3440). Dies führte zu einer unvollständigen Abdeckung des Donorgenoms in beiden Introgressionsbibliotheken und erklärt, warum die meisten IL mehr als ein DC-Segment enthal-

ten. Hätten die Ergebnisse der Simulationsstudie für die empirische Studie verwendet werden können, dann wäre der Wert der entstanden IL trotz begrenzter Kapazitäten bezüglich der genannten Merkmale höher gewesen.

Der Einfluss der nicht-adaptierten DC-Segmente auf agronomisch wichtige, qualitativ bzw. quantitativ vererbte Merkmale muss noch in umfangreichen Felduntersuchungen ermittelt werden. ILs mit günstigen DC-Segmenten können direkt in der praktischen Hybridroggenzüchtung verwendet werden. Sie können aber auch weiter rückgekreuzt werden, um naheisogenische Linien (NILs) zu erzeugen, von denen jede nur ein einzelnes, kurzes DC-Segment trägt, das durch Marker gezielt angesprochen werden kann. Diese NILs wären ein günstiger Ausgangspunkt für eine hochauflösende Kartierung und die Isolierung und funktionale Charakterisierung von Kandidatengenen.

Die Erstellung der beiden Introgressionsbibliotheken und die Ergebnisse der Simulationsstudie sind wichtige Meilensteine für die gezielte Nutzung genetischer Ressourcen und eröffnen vielversprechende Möglichkeiten für die funktionale Genomanalyse bei Roggen.

7 References

Albert, K.D., and S.D. Tanksley. 1996. High resolution mapping and isolation of a yeast artificial chromosome containing *fw2.2*: a major fruit weight quantitative trait locus in tomato. Proc. Natl. Acad. Soc. USA. 93:15503-15507.

Beavis, W.D. 1998. QTL analyses: power, precision, and accuracy. p. 145-162. *In* Paterson A.H. (ed.) Molecular Dissection of Complex Traits. CRC Press, New York.

Beckmann, J.S., and M. Soller. 1986. Restriction fragment length polymorphisms in plant genetic improvement, Oxford Surv. Plant Mol. Cell Biol. 3:197-250.

Bernacchi, D., T. Beck-Bunn, D. Emmatty, Y. Eshed, S. Inai, J. Lopez, V. Petiard, H. Sayama, J. Uhlig, D. Zamir, and S. Tanksley. 1998a. Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor. Appl. Genet. 97:170-180.

Bernacchi, D., T. Beck-Bunn, Y. Eshed, J. Lopez, V. Petiard, J. Uhlig, D. Zamir, and S. Tanksley. 1998b. Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. Theor. Appl. Genet. 97:381-397.

Bernardo, R. 2002. Breeding for Quantitative Traits in Plants. Stemma Press, Woodbury.

Brouwer, D.J., and D.A. St Clair. 2004. Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. Theor. Appl. Genet. 108:628-638.

Carver, B.F., and A.L. Rayburn. 1994. Comparison of related wheat stocks possessing 1b or 1rs.1bl chromosomes - agronomic performance. Crop Sci. 34:1505-1510.

Chetelat, R.T., and V. Meglic. 2000. Molecular mapping of chromosome segments introgressed from *Solanum lycopersicoides* into cultivated tomato (*Lycopersicon esculentum*). Theor. Appl. Genet. 100:232-241.

Crosby, J.L. 1973. Computer Simulations in Genetics. Wiley, New York.

de Vicente, M.C., and S.D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585-596.

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

Eshed, Y., and D. Zamir. 1995. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147-1162.

Eshed, Y., G. Gera, and D. Zamir. 1996. *Lycopersicon esculentum* lines containing small overlapping introgressions from *L. pennellii*. Theor. Appl. Genet. 83:1027-1034.

Eshed, Y., M. Abu-Abied, Y. Saranga, and D. Zamir. 1992. A genome-wide search for wild-species alleles that increase horticultural yield of processing tomatoes. Theor. Appl. Genet. 93:877-886.

Frey, K.J., T.S. Cox, D.M. Rodgers, and P. Bramel-Cox. 1981. Increasing cereal yields with genes from wild and weedy species. p. 51-68. Journal paper No. J-11254 of the Iowa Agric. and Home Econ. Exp. Stn., Ames, Iowa 50011, USA.

Frisch, M., and A.E. Melchinger. 2001a. Marker-assisted backcrossing for simultaneous introgression of two genes. Crop Sci. 41:1716-1725.

Frisch, M., and A.E. Melchinger. 2001b. Marker-assisted backcrossing for introgression of a recessive gene. Crop Sci. 41:1485-1494.

Frisch, M., M. Bohn, and A.E. Melchinger. 1999a. Comparison of selection strategies for marker-assisted backrossing of a gene. Crop Sci. 39:1295-1301.

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

Frisch, M., M. Bohn, and A.E. Melchinger. 2000. PLABSIM: Software for simulation of marker-assisted backrossing. J. Hered. 91:86-87.

Fulton, T.M., T. Beck-Bunn, D. Emmatty, Y. Eshed, J. Lopez, V. Petiard, J. Uhlig, D. Zamir, and S.D. Tanksley. 1997. QTL analysis in an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. Theor. Appl. Genet. 95: 881–894.

Fulton, T.M., S. Grandillo, T. Beck-Bunn, E. Fridman, A. Frampton, J. Lopez, V. Petiard, J. Uhlig, D. Zamir, S. D. Tanksley. 2000. Advanced backcross QTL analysis of a

Lycopersicon esculentum × Lycopersicon parviflorum cross. Theor. Appl. Genet. 100: 1025–1042.

Gallais, A., A. Charcosset, I. Goldinger, F. Hospital, and L. Moreau. 2000. Progress and prospects for marker-assisted selection. p. 183-197. *In* Quantitative genetics and breeding methods: the way ahead. Proceedings of the Eleventh Meeting of the EUCARPIA Section Biometrics in Plant Breeding. Paris, France.

Geiger, H.H., and F.W. Schnell. 1970. Cytoplasmatic male sterility in rye (*Secale cereale* L.). Crop Sci. 35:27-38.

Geiger, H.H., and T. Miedaner. 1999. Hybrid rye and heterosis. p. 439-450. *In* Coors, J.G., and S. Pandey (eds.) Genetics and Exploitation of Heterosis in Crops. Crop Sci. Soc. America, Madison, Wisconsin, USA.

Goodman, M.M. 1997. Broadening the genetic diversity in breeding by use of exotic germplasm. p. 58-59. *In* CIMMYT. Book of Abstracts. The Genetics and Exploitation of Heterosis in Crops. An Intern. Symp., Mexico, D.F., Mexico.

Goodman, R.M., H. Hauptil, A. Crossawy, and V.C. Knaut. 1987: Gene transfer in crop improvement. Science 236:48-54.

Hackauf, B., and P. Wehling. 2002a. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed. 121:17-25.

Hackauf B., and P. Wehling. 2002b. Development of microsatellite markers in rye: map construction. p. 333-340. *In* Proceedings of the EUCARPIA Rye Meeting, July 4-7, 2001, Radzików, Poland.

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

Han, F., S.E. Ullrich, J.A. Clancy, and I. Romagosa. 1999. Inheritance and fine mapping of a major barley seed dormancy QTL. Plant Sci. 143:113-118.

Haussmann, B.I.G., H.K. Parzies, T. Presterl, Z. Sušić, and T. Miedaner. 2004. Plant genetic resources in crop improvement. Plant Genetic Resources 2:3-21.

Hawks, J.G. 1977. The importance of wild germplasm in plant breeding. Euphytica 26:615-621.

Hospital, F. 2002. Marker-assisted backcross breeding: a case study in genotype building theory. p. 135-141. *In* M.S. Kang (ed.) Quantitative Genetics, Genomics and Plant Breeding. CAB International, New York and Oxon.

Hospital, F., and A. Charcosset. 1997. Marker-assisted introgresion 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.

Hospital, F., I. Goldringer, and S.J. Openshaw. 2000. Efficient marker-based recurrent selection for multiple quantitative trait loci. Genetical Research 75:357-368.

Howell, P.M., D.F. Marshall, and D.J. Lydiate. 1996. Towards developing intervarietal substitution lines in *Brassica napus* using marker-assisted selection. Genome 39:348-358.

Ioannidou, D., A. Pinel, C. Brugidou, L. Albar, N. Ahmadi, A. Ghesquiere, M. Nicole, and D. Fargette. 2003. Characterisation of the effects of a major QTL of the partial resistance to rice yellow mottle virus using a near-isogenic-line approach. Physiological and Molecular Plant Pathology 63:213-221.

Jeuken, M.J.W., and P. Lindhout. 2004. The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. Theor. Appl. Genet. 109:394-401.

Kaeppler, S.M. 1997. Quantitative trait locus mapping using sets of near-isogenic lines: relative power comparisons and technical considerations. Theor. Appl. Genet. 95:384-392.

Kaeppler, S.M., R.L. Phillips, and T.S. Kim. 1993. Use of near-isogenic lines derived by backcrossing or selfing to map qualitative traits. Theor. Appl. Genet. 87:233-237.

Kearsey, M.J. 2002. QTL analysis: Problems and (possible) solutions. p. 45-58. *In* Kang M. S. (ed.) Quantitative Genetics, Genomics and Plant Breeding. CAB International, New York and Oxon.

Kim, W., J.W. Johnson, R.A. Graybosch, and C.S. Gaines. 2004. The effect of T1dl.1rs wheat-rye chromosomal translocation on agronomic performance and end-use quality of soft wheat. Cereal Research Communications 31:301-308.

Koester, R.P., P.H. Sisco, and C.W. Stuber. 1993. Identification of quantitative trait loci controlling days to flowering and plant height in two near isogenic lines og maize. Crop Sci. 33:1209-1216.

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

Lin, H. X., T. Yamamoto, T. Saski, and M. Yano. 2000. Characterisation and detection of epistatic interactions of 3QTLs, *Hd1*, *Hd2* and *Hd3*, controlling heading date in rice using nearly isogenic lines. Theor. Appl. Genet. 101:1021-1028.

Lukaszewski, A.J. 1990. Frequency of 1RS-1AL and 1RS-1BL translocations in United States wheats. Crop Sci. 30:1151-1153.

Lukaszewski, A.J. 2001. Breeding behavior of the cytogenetically engineered wheat-rye translocation chromosomes 1rs.1bl. Crop Sci. 41:1062-1065.

Lynch, M., and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc. Sunderland.

Mago, R., W. Spielmeyer, G.J. Lawrence, E.S. Lagudah, J.G. Ellis, and A. Pryor. 2002. Identification and mapping of molecular markers linked to rust resistance genes located on chromosome 1rs of rye using wheat-rye translocation lines. Theor. Appl. Genet. 104:1317-1324.

Mank, R., H. Verbakel, H. Witsenboer, and J. Peleman. 2003. Marker assisted construction of a high resolution introgression line library in tomato using *Lycopersicon hirsutum*. p. 508. *In* Plant and Animal Genomes XI Conference. San Diego.

Martin, G.B., G.K. Williams, and S.D. Tanksley. 1991. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proc. Natl. Acad. Sci. USA. 88:2336-2340.

Matus, I., A. Corey, T. Filichkin, P.M. Hayes, M.I. Vales, J. Kling, O. Riera-Lizarazu, K. Sato, W. Powell, and R. Waugh. 2003. Development and characterization of recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in a *Hordeum vulgare* subsp. *vulgare* background. Genome 46:1010-1023.

Melchinger A.E., H.F. Utz, and C.C. Schoen. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149 (1): 383-403.

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

Miedaner, T., C. Glass, F. Dreyer, P. Wilde, H. Wortmann, and H.H. Geiger. 2000. Mapping of genes for male-fertility restoration in 'Pampa' CMS winter rye *(Secale cereale* L.). Theor. Appl. Genet. 101:1226-1233.

Monforte, A.J., and S.D. Tanksley. 2000. Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery. Genome 43: 803-813.

Monforte, A.J., E. Friedman, D. Zamir, and S.D. Tanksley. 2001. Comparison of a set of allelic QTL-NIL for chromosome 4 of tomato: Deductions about natural variation and implications for germplasm utilization. Theor. Appl. Genet. 102:572-590.

Moonen, J.H.E., and A.C. Zeven. 1984. SDS-PAGE of the high-molecular-weight subunits of wheat glutenin and characterisation of 1R (1B) substitution and 1BL/1RS translocation lines. Euphytica 33:3-8.

Oetting, W.S., H.K. Lee, D.J. Flanders, G.L. Wiesner, T.A. Sellers, and R.A. King. 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30, 450-458.

Ossent, H.P. 1938. Zehn Jahre Roggenzüchtung in Müncheberg. Züchter 10:255-261.

Paterson, A.H., J.W. DeVerna, B. Lanini, and S.D. Tanksley. 1990. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecific cross of tomato. Genetics 124:735-742.

Peleman, J.D., and J.R. van der Voort. 2003. Breeding by design. Trends in Plant Science. 8(7):330-334.

Pillen, K., A. Zacharias, and J. Léon. 2003. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 107:340-352.

Prasad, M., N. Kumar, P.L. Kulwal, M.S. Roder, H.S. Balyan, H.S. Dhaliwal, and P. Gupta. 2003. QTL analysis for grain protein content using SSR markers and validation studies

using NILs in bread wheat. Theor. Appl. Genet. 106:659-667.

Ragot, M., P.H. Sisco, D.A Hoisington, and C.W. Stuber. 1995. Molecular-marker-mediated characterization of favorable exotic alleles at quantitative trait loci in maize. Crop Sci. 35:1306-1315.

Ramsay L.D., D.E. Jennings, E.J.R. Bohuon, A.E. Arthur, D.J. Lydiate, M.J. Kearsey, and D.F. Marshal. 1996. The construction of a substitution library of recombinant backcross lines in *Brasica oleracea* for the precision mapping of quantitative trait loci. Genome 39: 558-567.

Reyes-Valdés, M.H. 2000. A model for marker-based selection in gene introgression breeding programs. Crop Sci. 40:91-98.

Ribaut, J.M., C. Jiang, and D. Hoisington. 2002. Simulation experiments on efficiencies of gene introgression by backcrossing. Crop. Sci. 42:557-565.

Rollwitz, W. 1985. Untersuchungen zur Bewertung von Roggenzuchtmaterial bezüglich Braunrostresistenz und Schaffung von Ausgangsmaterial für die Züchtung. Diss. Rostock, Deutschland.

Ronen, G., L. Carmel-Goren, D. Zamir, and J. Hirschberg. 2000. An alternative pathway to ß-carotene formation in plant chromoplasts discovered by map-based cloning of ß- and old-gold color mutations in tomato. Proc. Natl. Acad. Sci. USA 97.11102-11107.

Salvi, S., R. Tuberosa, E. Chiapparino, M. Maccaferri, S. Veillet, L. Van Beuningen, P. Isaac, K. Edwards, and R.L. Phillips. 2002. Toward positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. Plant Molecular Biology 48:601-613.

Servin, B., and F. Hospital. 2002. Optimal position of markers to control genetic background in marker-assisted backrossing. J. Hered. 93:214-217.

Shen, L., B. Courtois, K.L. McNally, S. Robin, and Z. Li. 2001. Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. Theor. Appl. Genet. 103: 75-83.

Stalker, H.T. 1980. Utilisation of wild species for crop improvement. Adv. Agron. 33:111-147.

Stam, P. 2003. Marker-assisted introgression: speed at any cost? p. 117-124. *In* Proceedings of the Eucarpia Meeting on Leafy Vegetables Genetics and Breeding. Noordwijkerhout, The Netherlands, 19-21 March 2003.

Stuber, C.W., M. Polacco, and M. Lynn. 1999. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. Crop Sci. 39:1571-1583.

Tanksley, S.D., and J.C. Nelson. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTL from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92: 191-203.

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

Tuinstra, M.R., G. Ejeta, and P. Goldsbrough. 1998. Evaluation of near-isogenic sorghum lines contrasting for QTL markers associated with drought tolerance. Crop. Sci. 38:835-842.

Utz, H.F., A.E. Melchinger, and C.C. Schön. 2000. Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839-1849.

van Berloo, R. 1999. Computer note. GGT: Software for the display of the graphical genotype. J. Hered. 90:328-329.

van Berloo, R., H. Aalbers, A. Werkman, and R.E. Niks 2001. Resistance QTL confirmed through development of QTL-NILs for barley leaf rust resistance. Mol. Breed. 8:187-195.

van Ooijen, J.W., and R.E. Voorrips. 2001. JoinMap version 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands.

Villareal, R.L., E. Deltoro, S. Rajaram, and A. Mujeebkazi. 1996. The effect of chromosome 1al/1rs translocation on agronomic performance of 85 F2-derived F6 lines from three *Triticum aestivum* L. crosses. Euphytica 89:363-369.

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

Von Korff, M., H. Wang, J. Léon, and K. Pillen. 2004. Development of candidate introgression lines using an exotic barley accession (*Hordeum vulgare* ssp. *spontaneum*) as donor. Theor. Appl. Genet. 109:1736-1745.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

Xiao, J., J. Li, S. Grandillo, S.N. Ahn, L. Yuan, S.D. Tanksley, and S.R. McCouch. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899-909.

Yan, H.H., G.Q. Liu, Z.K. Cheng, X.B. Li, G.Z. Liu, S.K. Min, and L.H. Zhu. 2002. A genome-specific repetative DNA sequence from *Oryza eichingeri*: characterization, localization, and inrtogression to *O. sativa*. Theor. Appl. Genet. 104:177-183.

Zamir, D. 2001. Improving plant breeding with exotic genetic libraries. Nature Rev. Genet. 2:983-989.

Zamir, D., and Y. Eshed. 1998. Case history in germplasm introgression: Tomato genetics and breeding using nearly isogenic introgression lines derived from wild species. p. 207-217. *In* Paterson, A.H. (ed.) Molecular Dissection of Complex Traits. CRC Press, New York.

8 Appendix

Table of Contents

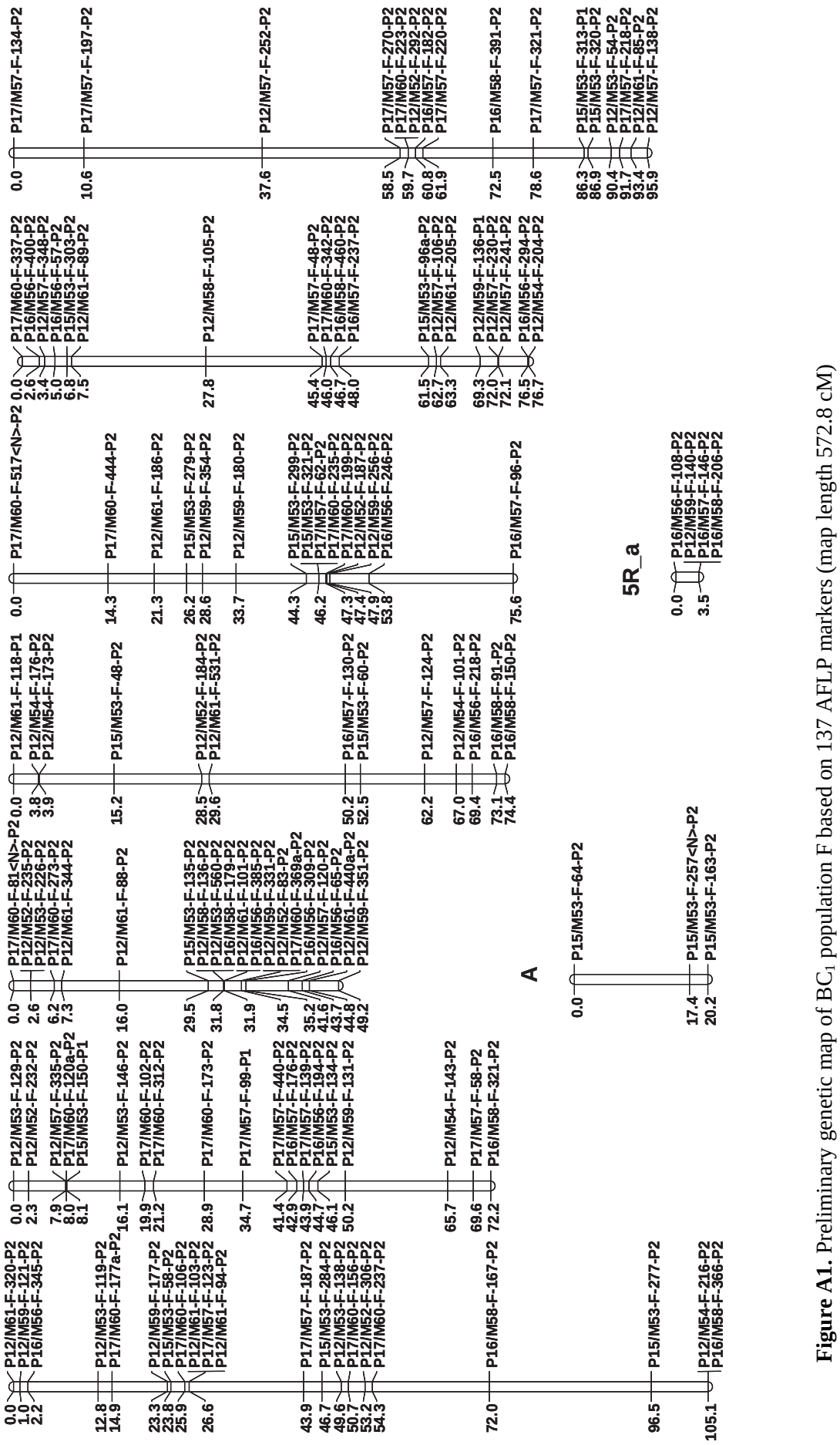


Figure A1. Preliminary genetic map of BC_1 population F based on 137 AFLP markers (map length 572.8 cM)

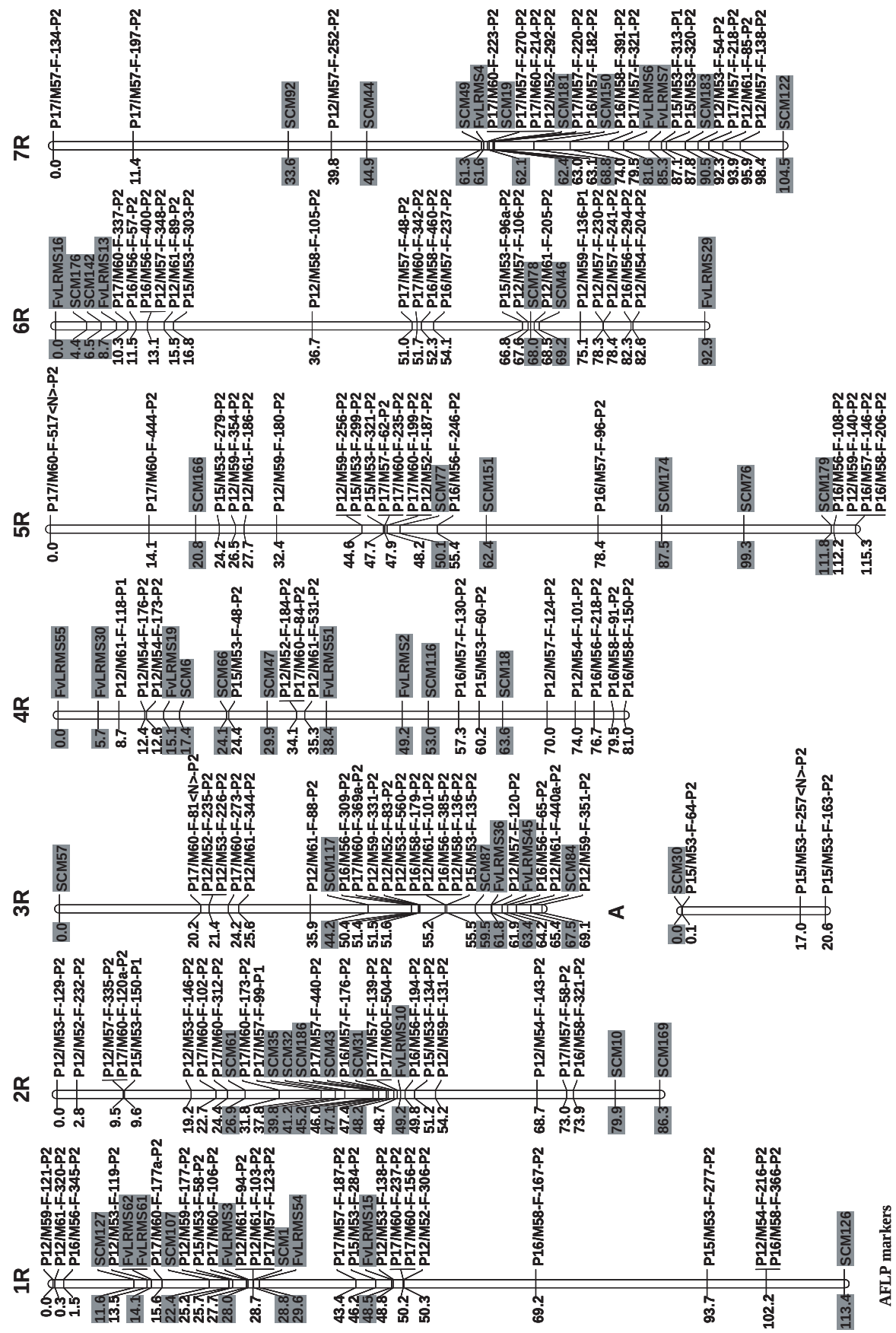


Figure A2. Extended genetic map of BC_1 population F based on 137 AFLP and 59 SSR markers (map length 683.1 cM)

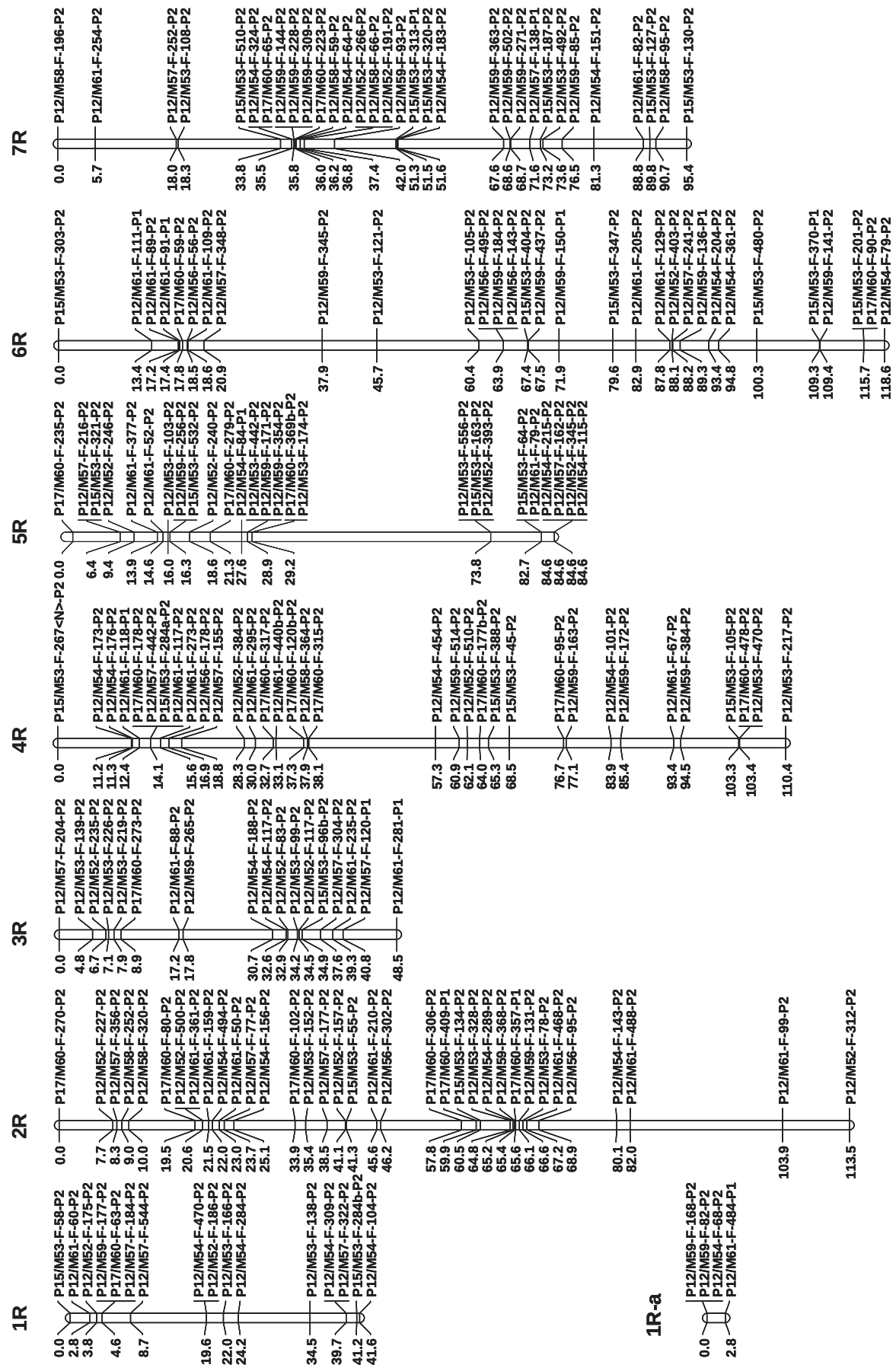


Figure A3. Preliminary genetic map of BC_1 population G based on 194 AFLP markers (map length 615.4 cM)

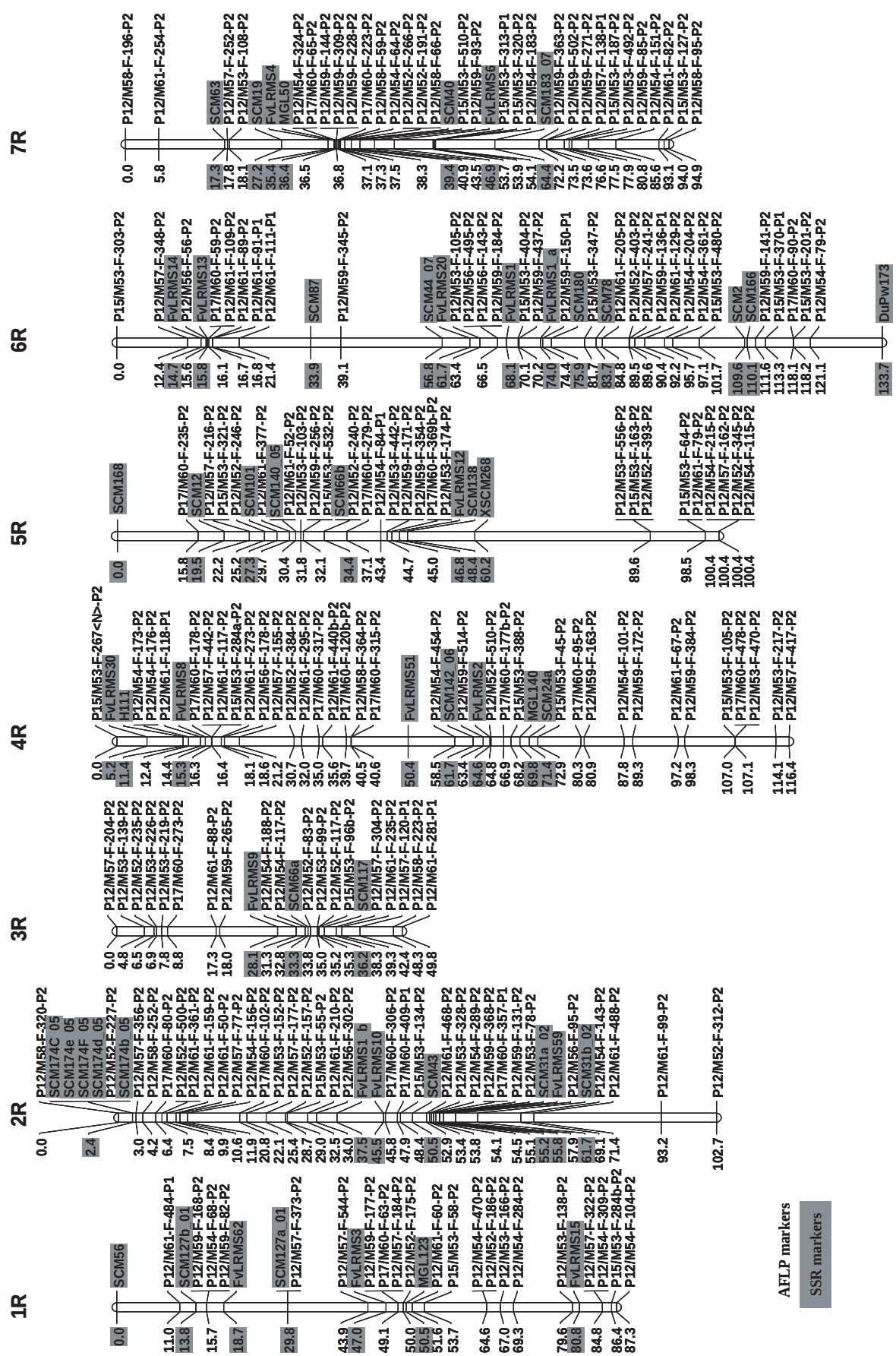


Figure A4. Extended genetic map of BC_1 population G based on 194 AFLP and 56 SSR markers (map length 685.2 cM)

Table A1. Standard errors (SE) given as a percentage of the mean values across 100 simulation runs for donor genome (DG) coverage, recurrent parent genome (RPG) recovered and number of required marker data points (MDP) in BC_3S_1 scenarios, as a function of progeny sizes per introgression line (IL) from generation BC_2 to BC_3S_1. Simulations were based on constant progeny size per IL from generation BC_2 onwards and a BC_1 population size of 100 individuals.

Progeny size per IL	DG coverage	RPG recovered	MDP number
	---------- SE (%) ----------		
19	0.05	0.24	0.52
50	0.00	0.12	0.54
100	0.00	0.09	0.56
250	0.00	0.004	0.69
500	0.00	0.002	0.64
1000	0.00	0.001	0.71
1500	0.00	0.001	0.62
2000	0.00	0.001	0.57

Table A2. Standard errors (SE) given as a percentage of the mean values across 100 simulation runs for donor genome (DG) coverage and recurrent parent genome (RPG) recovered in generation BC_1, as a function of the population size in generation BC_1.

Population size in BC_1	DG coverage	RPG recovered
	SE (%)	
5	0.004	0.246
10	0.001	0.178
17	0.000	0.164
25	0.000	0.131
40	0.000	0.113
60	0.000	0.118
80	0.000	0.095
100	0.000	0.097
130	0.000	0.080
160	0.000	0.076
200	0.000	0.074
300	0.000	0.064
400	0.000	0.068
500	0.000	0.060
750	0.000	0.050
1000	0.000	0.047

* Bold letters designate BC_1 population sizes used for further simulations.

Table A3. Standard errors (SE) given as a percentage of the mean values across 100 simulation runs for the number of marker data points (MDP) required to reach recurrent parent genome threshold of 95.60%, donor genome (DG) coverage and average number of non-target donor chromosome (DC) segments per introgression line (IL), as a function of the number of backcross and selfing generations, based on constant progeny size from generation BC_2 onwards. Bold letters designate optimum scenarios for each introgression strategy.

Population size in gen. BC_1	Introgression strategies					
	BC_2S_1	BC_2S_2	BC_3S_1	BC_3S_2	BC_4S_1	BC_4S_2
	---------- SE (%) ----------					
	MDP number					
17	0.74	0.89	0.77	0.74	0.66	0.66
40	0.65	0.67	0.63	0.73	0.57	0.54
60	0.73	0.69	0.57	0.60	**0.54**	**0.50**
80	0.65	0.64	0.49	0.53	0.48	0.40
100	0.69	0.55	**0.52**	**0.44**	0.41	0.38
130	0.55	0.53	0.43	0.36	0.33	0.29
160	0.51	0.51	0.33	0.32	0.26	0.26
200	**0.43**	**0.42**	0.34	0.31	0.21	0.22
300	0.39	0.31	0.23	0.23	0.15	0.15
	DG coverage					
17	0.00	0.00	0.02	0.05	0.15	0.21
40	0.00	0.00	0.03	0.08	0.18	0.23
60	0.00	0.00	0.04	0.10	**0.19**	**0.25**
80	0.00	0.00	0.05	0.11	0.21	0.29
100	0.00	0.00	**0.05**	**0.11**	0.23	0.25
130	0.00	0.00	0.05	0.11	0.25	0.27
160	0.00	0.00	0.07	0.12	0.29	0.23
200	**0.00**	**0.00**	0.07	0.13	0.30	0.28
300	0.00	0.00	0.07	0.14	0.30	0.37
	Average number of non-target DC segments per IL					
Optimum	**2.75**	**2.62**	**2.90**	**2.95**	**3.15**	**3.70**
Suboptimum						
Highest value	2.96	2.80	2.39	3.07	3.23	3.28
Lowest value	2.76	2.49	2.92	3.33	3.75	4.31

Table A4. Standard errors (SE) given as a percentage of the mean values across 100 simulation runs for the number of marker data points (MDP), donor genome (DG) coverage, percentage of the recurrent parent genome (RPG) recovered, and average number of non-target donor chromosome (DC) segments per introgression line (IL) when the RPG threshold alone is reached, compared to those when both the RPG and DG threshold are reached. Simulations were based on constant progeny sizes per IL from generation BC_2 onwards. Bold letters designate optimum scenarios for each introgression strategy.

Population size in generation BC_1	BC_4S_1 strategy		BC_4S_2 strategy	
	RPG threshold alone	RPG plus DG threshold	RPG threshold alone	RPG plus DG threshold
	---------- SE (%) ----------			
	MDP number			
17	0.66	0.72	0.66	0.75
40	0.57	0.58	0.54	0.51
60	**0.54**	**0.48**	**0.50**	**0.58**
80	0.48	0.38	0.40	0.46
100	0.41	0.40	0.38	0.40
130	0.33	0.37	0.29	0.37
160	0.26	0.35	0.26	0.30
200	0.21	0.30	0.22	0.28
300	0.15	0.21	0.15	0.18
	DG coverage			
17	0.15	0.09	0.21	0.08
40	0.18	0.09	0.23	0.11
60	**0.19**	**0.10**	**0.25**	**0.12**
80	0.21	0.09	0.29	0.07
100	0.23	0.07	0.25	0.08
130	0.25	0.10	0.27	0.09
160	0.29	0.10	0.23	0.09
200	0.30	0.08	0.28	0.08
300	0.30	0.10	0.37	0.10
	Percentage of the RPG recovered			
17	0.026	0.021	0.025	0.025
40	0.026	0.020	0.028	0.024
60	**0.026**	**0.015**	**0.024**	**0.023**
80	0.029	0.021	0.027	0.021
100	0.025	0.016	0.030	0.024
130	0.026	0.016	0.027	0.024
160	0.026	0.015	0.030	0.025
200	0.029	0.019	0.030	0.020
300	0.025	0.017	0.029	0.022
	Average number of non-target DC segments per IL			
Optimum	**3.15**	**3.70**	**3.70**	**3.23**
Suboptimum				
Highest value	3.23	3.43	3.28	3.09
Lowest value	3.75	4.31	4.31	3.56

Table A5. Standard errors (SE) given as a percentage of the mean values across 100 simulation runs for the number of marker data points (MDP) required to reach recurrent parent genome threshold of 95.60%, coverage of the donor genome (DG) and average number of non-target donor chromosome (DC) segments per introgression line (IL) in BC_3S_1 strategy, as a function of increasing progeny size from generation BC_2 to BC_3S_1. Scenarios were simulated for a BC_1 population size of 100 individuals.

Generation	Increasing progeny size ratio (BC_2 – BC_3S_1)					
	1:1:1	**1:1.5:2**	**1:1:2**	**1:2:3**	**1:1:3**	**1:3:9**
	---------- SE (%) ----------					
	MDP number					
BC_3S_1	0.52	**0.43**	0.46	0.45	0.41	0.50
	DG coverage					
	0.05	**0.03**	0.02	0.06	0.03	0.17
	Average number of non-target DC segments per IL					
	2.90	**2.80**	2.95	3.13	2.52	2.49

Table A6. Standard errors (SE) given as a percentage of the mean values across 100 simulation runs for the number of marker data points (MDP) required to reach recurrent parent genome threshold of 95.60%, coverage of the donor genome (DG) and average number of non-target donor chromosome (DC) segments per introgression line (IL) in BC_3S_1 strategy, as a function of decreasing progeny size from generation BC_2 to BC_3S_1. Scenarios were simulated for a BC_1 population size of 100 individuals.

Generation	Decreasing progeny size ratio (BC_2 – BC_3S_1)					
	1:1:1	**2:1.5:1**	**2:1:1**	**3:2:1**	**3:1:1**	**9:3:1**
	---------- SE (%) ----------					
	MDP number					
BC_3S_1	0.52	0.53	0.59	0.56	0.60	0.59
	DG coverage					
	0.05	0.14	0.10	0.22	0.12	0.48
	Average number of non-target DC segments per IL					
	2.90	2.83	2.77	3.44	3.07	4.22

Table A7. Standard errors (SE) given as a percentage of the mean values across 100 simulation runs for the number of marker data points (MDP), coverage of the donor genome (DG) and average number of non-target donor chromosome (DC) segments per introgression line (IL) when recurrent parent genome threshold of 95.60% is reached, as a function of DC segment length and marker density. Bold letters designate optimum scenarios for each introgression strategy.

Population size in generation	Ratios between DC segment lengths and marker densities					
	BC_2S_1 strategy			BC_3S_1 strategy		
BC_1	20/5[a]	40/5[b]	40/20[c]	20/5[a]	40/5[b]	40/20[c]
	---------- SE (%) ----------					
	MDP number					
17	0.74	0.77	0.81	0.77	0.75	0.60
25	0.65	0.67	0.73	0.79	0.59	0.66
40	0.65	0.68	0.64	0.63	0.51	**0.46**
60	0.73	0.54	0.56	0.57	**0.51**	0.33
80	0.65	0.59	0.50	0.49	0.36	0.30
100	0.69	0.54	**0.47**	**0.52**	0.33	0.29
130	0.55	0.39	0.40	0.43	0.28	0.22
160	0.51	0.41	0.35	0.33	0.22	0.20
200	**0.43**	**0.37**	0.25	0.34	0.19	0.13
300	0.39	0.28	0.15	0.23	0.12	0.08
	DG coverage					
17	0.00	0.00	0.03	0.02	0.08	0.23
25	0.00	0.00	0.01	0.02	0.08	0.31
40	0.00	0.00	0.02	0.03	0.09	**0.29**
60	0.00	0.00	0.03	0.04	**0.10**	0.40
80	0.00	0.00	0.04	0.05	0.11	0.40
100	0.00	0.00	**0.04**	**0.05**	0.11	0.47
130	0.00	0.00	0.05	0.05	0.12	0.43
160	0.00	0.00	0.06	0.07	0.16	0.47
200	**0.00**	**0.00**	0.07	0.07	0.15	0.48
300	0.00	0.00	0.10	0.07	0.16	0.56
	Average number of non-target DC segments per IL					
Optimum	**2.75**	**2.67**	**2.59**	**2.90**	**3.89**	**2.50**
Suboptimum						
Highest value	2.96	2.59	2.55	2.39	3.17	2.52
Lowest value	2.76	2.73	2.77	2.92	3.61	3.07

[a] 20/5 - variant with 20 cM DC segment length and 5 cM marker density (standard scenario)

[b] 40/5 - variant with 40 cM DC segment length and 5 cM marker density

[c] 40/20 - variant with 40 cM DC segment length and 20 cM marker density

Acknowledgements

I am very grateful to Prof. H. H. Geiger firstly for giving me the opportunity to work on this project and secondly for the authoritative scientific guidance and support during the creation of this PhD thesis.

I express my sincere appreciation to Dr. T. Miedaner for his professional support in the experimental part of the study, his constructive comments to the first draft of the PhD, and an outstanding tolerance.

I am grateful to Dr. M. Frisch for his brilliant ideas on how to reach the objectives of the theoretical study, the time he spent as my personal LINUX system admistrator, his comments to the first draft of the PhD, and encouragements in times I thought that simulations would last forever.

I am especially grateful to our collaborating partners in Keygene (Dr. J. R. van der Voort), Lochow-Petkus GmbH (Dr. V. Korzun), Saaten Union Resistenz Labor (Dr. J. Schondelmeier) and Federal Center for Breeding Research on Cultivated Plants, Institute of Agricultural Crops Groß Lüsewitz (Dr. B. Hackauf) for the marker analyses in the experimental part of this PhD, and for their scientific support and advice.

I particularly thank Ms B. Lieberherr for her outstanding skills in the technical support of the project, and to all the members of the Rye group for being available every time when any kind of help was required.

Of course, professional joy would be only half as pleasing without having friends to make an enjoyable environment. Thanks to Zeljko, Ina, Andrés, Katinka, Bettina, Heiko, Martin, Matthias, and to all football players for relaxing Tuesdays. I also gratefully appreciate the kindness and understanding of Ms S. Egelhaaf and colleagues at the Institute 350.

I dedicate special gratitude to my parents who trusted me and supported me by the time I decided to come to Hohenheim and take this way and, of course my immense gratitude to the dearest Jasmina for her never-ending love and encouragements. At the end, thanks to my beloved Sava for making me busy during the last several months.

The study was done within GABI project and was funded by the BMBF (*Bundesministerium für Bildung und Forschung*) under the number 0312289 B, and co-financed by the private breeding companies HYBRO GmbH & Co. KG, Schenkenberg and Lochow-Petkus GmbH, Bergen. The responsibility for the content of this publication lies with the author.

Curriculum Vitae

PERSONAL DETAILS

Name:	Zoran Sušić
Date of Birth:	04 October 1970
Place of Birth:	Smederevska Palanka, Serbia and Montenegro
Nationality:	Serbian
Marital Status:	since 25 June 2000 married to Jasmina Muminović, one child (son Sava)

ELEMENTARY AND HIGH SCHOOL EDUCATION

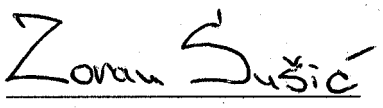
1977/78	–	1984/85	Elementary School Smederevska Palanka
1985/86	–	1988/89	High School of Mathematics and Natural Sciences Smederevska Palanka (Serbia and Montenegro)

UNIVERSITY EDUCATION

1989/90	–	1992/93	University of Belgrade (Serbia and Montenegro) Faculty of Agriculture, Department of Crop Sciences degree *B.Sc. in Agriculture* (20 September 1994)
1994/95	–	1995/96	University of Belgrade (Serbia and Montenegro) Faculty of Agriculture, Department of Genetics and Plant Breeding degree *M.Sc. in Agricultural Sciences* (16 January 1998)
2001/02	–	2003/04	University of Hohenheim (Germany) Faculty of Agriculture Institute of Plant Breeding, Seed Science and Population Genetics candidate for a degree *Ph.D. in Agricultural Sciences*

EMPLOYMENT RECORD

01.10.1996 – 31.01.2001	Researcher and Breeder Department of Genetics and Breeding of *Cucurbitaceae* Institute Serbia - Center for Vegetable Crops - Palanka, Smederevska Palanka (Serbia and Montenegro)
01.02.2001 – 30.11.2004	Research Assistant Institute of Plant Breeding, Seed Science and Population Genetics University of Hohenheim, Stuttgart (Germany)



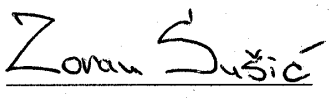
Zoran Sušić

Erklärung

Hiermit versichere ich an Eides statt, dass die vorliegende Arbeit von mir selbst verfasst und nur unter Zuhilfenahme von angegebenen Quellen und Hilfsmitteln angefertigt wurde.

Diese Arbeit wurde noch keiner anderen Prüfungsbehörde vorgelegt.

Stuttgart-Hohenheim,
den 17. Juni 2005



Zoran Sušić