

Selection Strategies for the Development of Maize Introgression Populations

Eva Herzog¹, Karen Christin Falke², Thomas Presterl³, Daniela Scheuermann³, Milena Ouzunova³, Matthias Frisch^{1*}

1 Institute of Agronomy and Plant Breeding II, Justus Liebig University, Giessen, Germany, **2** Institute for Evolution and Biodiversity, University of Münster, Münster, Germany, **3** KWS Saat AG, Einbeck, Germany

Abstract

Introgression libraries are valuable resources for QTL detection and breeding, but their development is costly and time-consuming. Selection strategies for the development of introgression populations with a limited number of individuals and high-throughput (HT) marker assays are required. The objectives of our simulation study were to design and compare selection strategies for the development of maize introgression populations of 100 lines with population sizes of 360–720 individuals per generation for different DH and S₂ crossing schemes. Pre-selection for complete donor chromosomes or donor chromosome halves reduced the number of simultaneous backcross programs. The investigated crossing and selection schemes differed considerably with respect to their suitability to create introgression populations with clearly separated, evenly distributed target donor chromosome segments. DH crossing schemes were superior to S₂ crossing schemes, mainly due to complete homozygosity, which greatly reduced the total number of disjunct genome segments in the introgression populations. The S₂ crossing schemes were more flexible with respect to selection and provided economic alternatives to DH crossing schemes. For the DH crossing schemes, increasing population sizes gradually over backcross generations was advantageous as it reduced the total number of required HT assays compared to constant population sizes. For the S₂ crossing schemes, large population sizes in the final backcross generation facilitated selection for the target segments in the final backcross generation and reduced fixation of large donor chromosome segments. The suggested crossing and selection schemes can help to make the genetic diversity of exotic germplasm available for enhancing the genetic variation of narrow-based breeding populations of crops.

Citation: Herzog E, Falke KC, Presterl T, Scheuermann D, Ouzunova M, et al. (2014) Selection Strategies for the Development of Maize Introgression Populations. PLoS ONE 9(3): e92429. doi:10.1371/journal.pone.0092429

Editor: Lewis Lukens, University of Guelph, Canada

Received: October 18, 2013; **Accepted:** February 21, 2014; **Published:** March 19, 2014

Copyright: © 2014 Herzog et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding from the German Federal Ministry of Education and Research (BMBF Grant 0315951) is gratefully acknowledged. [http://www.bmbf.de/en/index.php] The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have the following interests: Thomas Presterl, Daniela Scheuermann and Milena Ouzunova are employed by KWS Saat AG. There are no patents, products in development or marketed products to declare related to KWS SAAT AG and the subject matter of the publication. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

* E-mail: matthias.frisch@uni-giessen.de

Introduction

Introgression libraries are valuable resources for the identification of alleles of agricultural interest in exotic germplasm. They facilitate the introduction of new genetic variation into elite breeding germplasm by providing favorable chromosome segments from wild or exotic species in an adapted genetic background [1,2]. Ideally, an introgression library consists of a set of homozygous introgression lines (ILs) which carry short marker-defined chromosome segments from an exotic donor in a common genetic background. The concept was first described in tomato [3]. In the mean time, introgression libraries have been developed for the model species *Arabidopsis thaliana* [4,5], and in many agriculturally important crops, such as rice [6,7], barley [8,9], wheat [10,11], maize [12,13] and rye [14].

Introgression libraries are usually developed by marker-assisted backcrossing followed by selfing or production of double haploid (DH) lines. The backcross process for their development is costly and labor-intensive if complete coverage of the donor genome by short evenly distributed target chromosome segments is to be achieved. Often additional backcross programs have to be run for

the developed ILs in order to close gaps in donor genome coverage, or to shorten donor chromosome segments by additional recombination events [3,9]. In spite of the high resource requirements, only incomplete donor genome coverage has been achieved for most of the reported introgression libraries [9,14].

In previous simulation studies on introgression libraries, two generations of selfing were investigated for line development [15,16]. Recent genetic studies in maize were based on ILs that underwent two to five generations of selfing [17–19]. The use of DH technology has to our knowledge not yet been investigated in simulation studies on the development of introgression libraries. However, *in vivo* induction of maternal haploids is currently a routine method of DH production in commercial maize breeding programs. The main advantage of the DH technology is that complete homozygosity can be obtained after only two generations. In spite of this time-saving, the production of DH lines is still considerably more costly than conventional selfing [20]. Moreover, a current drawback of *in vivo* induction of maternal haploids in maize is that on average only one viable DH line can be derived from one backcross individual. It is therefore of economic interest to compare this method with S₂ crossing schemes which require

the same number of generations to evaluate the benefits of DH lines.

A possible approach to tackle the high costs required for the development of ideal introgression libraries would be to resort to introgression populations which are not perfect in appearance, but carry some additional donor segments outside the actual target segments. Such introgression populations could be developed with fewer individuals and marker assays. Complete coverage of the donor genome is desirable in order to capture the whole wealth of alleles of agricultural interest in the exotic donor. It is therefore one component of a minimum standard which introgression populations should meet. A second component are short, evenly distributed target donor chromosome segments in a clean adapted background, as they facilitate the use of the ILs in the following breeding process.

The design of the crossing scheme and the selection strategy are the most important factors that influence the distribution of donor chromosome segments in the introgression population. Falke et al. [16] suggested for the development of ideal introgression libraries that a chromosome-based selection strategy which pre-selects individuals carrying the donor alleles on complete chromosomes in generation BC_1 saves resources. Adapting and advancing this concept to crossing schemes with small population sizes might be an efficient approach to develop introgression populations with a limited number of marker assays.

The objectives of our simulation study were (1) to design selection strategies and crossing schemes for the development of maize introgression populations with limited resources, (2) to compare these selection strategies with respect to the distribution and length of donor chromosome segments and the required investments in terms of time, individuals and marker assays, (3) to give guidelines for the optimal experimental design for constructing introgression populations.

Materials and Methods

Software

All simulations were conducted in R version 3.0.0 [21] with the software package SelectionTools, which is available from <http://www.uni-giessen.de/population-genetics/downloads>.

Genetic Model

A genetic model of maize with 10 equally sized chromosomes of 200 cM length was used for the simulations. Genetic markers for selection were equally spaced. The distance between two adjacent marker loci was 1 cM. All markers were polymorphic between donor and recipient. It was assumed that markers were analyzed with high-throughput (HT) assays. One HT assay comprised genotyping one individual at all marker loci in the linkage map. Recombination was modelled assuming no interference in crossover formation [22]. Each simulation of an introgression population of 100 ILs was replicated 1,000 times in order to reduce sampling effects and to obtain results with high numerical accuracy and a small standard error.

Crossing Schemes

Four crossing schemes were investigated: BC_2DH , BC_3DH , BC_2S_2 , BC_3S_2 . Each crossing scheme started with the cross of a homozygous donor and a homozygous recipient to create one F_1 individual. The F_1 individual was backcrossed to the recipient to create a BC_1 population of size n_{BC_1} . From the BC_1 population, the best individuals with the highest values of selection indices for the respective selection strategy were selected. Each of the selected BC_1 individuals was backcrossed to the recipient to create BC_2

sub-populations of size n_{BC_2} . From these BC_2 sub-populations, the best individuals with the highest values of the respective selection indices were selected. For the DH crossing schemes, *in vivo* induction of maternal haploids was assumed with a success rate of one viable DH line per backcross individual. For the BC_2DH schemes, one DH line was thus created from each of the selected BC_2 individuals. For the BC_2S_2 crossing schemes, the selected BC_2 individuals were selfed to create a fixed number of S_1 individuals. Each of the S_1 individuals was selfed again and one S_2 individual was created. For the BC_3 crossing schemes, each of the selected BC_2 individuals was backcrossed to the recipient to create BC_3 sub-populations of size n_{BC_3} . From these BC_3 sub-populations, the best individuals with the highest values of the respective selection indices were selected. The generations S_1 , S_2 or DH of the BC_3 crossing schemes were carried out as described for the BC_2 crossing schemes.

Evaluation of Selection Candidates

The final introgression populations should consist of 100 ILs which guarantee an acceptable resolution of QTL detection in maize, and which can be immediately used in further breeding steps. Each IL should ideally carry a 20 cM chromosome segment from the donor to provide a complete and even coverage of the donor genome without overlap. The 20 cM chromosome segments are hereafter simply referred to as “target segments”. To determine the selection index for an individual with respect to a given target segment, we denote with t_c the donor genome proportion of the chromosome on which the target segment is located, with t_h the donor genome proportion of the chromosome half on which the target segment is located and with t_s the donor genome proportion of the target segment itself. The values for the genetic background b_c , b_h , b_s correspond to t_c , t_h , t_s and denote the recipient genome proportion outside the respective chromosome region. Depending on the selection strategy, t and b are used to define selection indices.

Selection Strategies

We considered generations $g = \{BC_1, BC_2, BC_3, DH, S_1, S_2\}$ for selection. Generation DH was the generation in which homozygous diploid DH lines were available for selection. In each generation g , the genome was divided into selection regions that could either be 10 complete chromosomes, 20 chromosome halves or 100 target segments. For selection for complete donor chromosomes, a fixed number n_{sel} of best individuals for each of the chromosomes $c = 1, 2, \dots, 10$ with the highest values for selection index $i = t_c + b_c$ were selected. For selection for donor chromosome halves, a fixed number n_{sel} of best individuals for each of the chromosome halves $h = 1, 2, \dots, 20$ with the highest values for selection index $i = t_h + b_h$ were selected. For selection for donor target segments, a fixed number n_{sel} of best individuals for each of the target segments $s = 1, 2, \dots, 100$ with the highest values for selection index $i = t_s + b_s$ were selected.

Selection for complete donor chromosomes, donor chromosome halves and donor target segments were combined to form different selection strategies. Selection for complete donor chromosomes in a backcross generation is denoted by a C in the strategy name, selection for donor chromosome halves is denoted by an H, and selection for donor target segments is denoted by an S. For example, for strategy CH, selection for complete donor chromosomes was conducted in generation BC_1 while selection for donor chromosome halves was conducted in generation BC_2 . An overview of the investigated selection strategies is presented in Table 1. The investigated combinations of crossing scheme and selection strategy are listed in the first column of Table 2. For all

Table 1. Definition of the selection index i in generations BC₁, BC₂, BC₃, DH, S₁, S₂ for different selection strategies for developing introgression populations.

Strategy	Generation				
	BC ₁	BC ₂	BC ₃	S ₁	DH/S ₂
C	$t_c + b_c$	–	–	–	$t_s + b_s$
H	$t_h + b_h$	–	–	–	$t_s + b_s$
CC	$t_c + b_c$	$t_c + b_c$	–	–	$t_s + b_s$
HH	$t_h + b_h$	$t_h + b_h$	–	–	$t_s + b_s$
CH	$t_c + b_c$	$t_h + b_h$	–	–	$t_s + b_s$
CCC	$t_c + b_c$	$t_c + b_c$	$t_c + b_c$	–	$t_s + b_s$
HHH	$t_h + b_h$	$t_h + b_h$	$t_h + b_h$	–	$t_s + b_s$
CHH	$t_c + b_c$	$t_h + b_h$	$t_h + b_h$	–	$t_s + b_s$
HHS	$t_h + b_h$	$t_h + b_h$	$t_s + b_s$	–	$t_s + b_s$

Selection for complete donor chromosomes (C), selection for donor chromosome halves (H) and selection for donor target segments (S) were combined to form different selection strategies (left column). t_c , t_h and t_s denote the donor genome proportions of the chromosome on which the target segment is located, of the chromosome half on which the target segment is located and of the target segment itself. b_c , b_h and b_s correspond to t_c , t_h , t_s and denote the recipient genome proportion outside the respective chromosome region.

doi:10.1371/journal.pone.0092429.t001

selection strategies, the best 100 ILS for selection index $i = t_s + b_s$ were selected in generation DH or S₂, depending on the crossing scheme.

Population Sizes and Simulation Series

We investigated population sizes of $n_{tot} = 360 - 720$ individuals per backcross generation. This should be within a range which can be realized in practical maize breeding programs. Variations in population size were investigated to determine both the effect on preserving the target segments up to line development as well as on recovering the genotype of the recipient outside the target segments.

In the first series of simulations, basic crossing schemes were investigated. Selection was carried out in generation BC₁ for basic crossing schemes with two backcross generations, and in generations BC₁ and BC₂ for basic crossing schemes with three backcross generations. The total population size per generation was kept constant at $n_{tot} = 360$ individuals in every generation g .

In the second series of simulation, crossing schemes with high selection intensity were investigated. Population size was doubled compared to the basic crossing schemes ($n_{tot} = 720$) in every generation g , while the number of selected individuals was the same as for the basic crossing schemes. The crossing schemes with high selection intensity are denoted by BC₃–CC', BC₃–HH' and BC₃–CH' (Table 2). In the first and second series of simulations, all backcross individuals generated in the final backcross generation were used for line development for both DH and S₂ crossing schemes. One IL was derived from one backcross individual.

In the third series of simulations, crossing schemes with selection in the final backcross generation were investigated. n_{tot} was doubled to 720 individuals in the final backcross generation for the DH crossing schemes BC₂DH–CC, BC₂DH–HH, BC₂DH–CH, BC₃DH–CCC, BC₃DH–HHH, BC₃DH–CHH. This increase in population size was necessary to enable selection and to keep n_{tot} at 360 individuals in generation DH. For the corresponding S₂ schemes, n_{tot} was kept at 360 individuals also in the final backcross generation.

In the fourth series of simulations, crossing schemes with increasing population sizes were investigated. Selection was

conducted in the final backcross generation. The crossing schemes with increasing population sizes are denoted by BC₃–HHH* and BC₃–HHS*. The details concerning the total population size n_{tot} and population sizes in the sub-populations n_g for all investigated combinations of crossing scheme and selection strategy are summarized in Table 2. Schematic representations of the crossing schemes BC₃DH–HHH* and BC₃S₂–HHS* are given in Figure 1 and Figure 2 for illustration.

Measures

To evaluate and compare introgression populations originating from different crossing and selection schemes, the following measures were determined: (a) the genome coverage of the donor O in percent, which is defined as the proportion of the donor genome which is covered by the introgression population, irrespective of whether by the target segments or other donor segments in the genetic background, (b) the depth of donor genome coverage T , which is defined as the average number of ILS in which each donor allele appears in the introgression population, (c) the number of disjunct genome segments in the introgression population S , (d) the resolution of the introgression population R in cM, which is defined as the total genome length of the genetic model in cM divided by S , (e) the average number of donor segments per IL N , (f) the average length of donor segments per IL L in cM, (g) the average total donor genome proportion of the introgression population D_t in percent, (h) the average donor genome proportion of the chromosomes carrying the respective target segments D_c in percent, (i) the average donor genome proportion of the target segments D_s in percent.

Results

High values for the donor genome coverage O around 99% were observed for all crossing schemes (Table 3). However, the resulting introgression populations differed substantially in the values for the number of disjunct genome segments S , the total donor genome proportion D_t , the donor genome proportion of the carrier chromosomes D_c and the donor genome proportion of the target segments D_s . BC₃ crossing schemes resulted in 2–3% lower values for D_t than BC₂ crossing schemes, even if the number of

Table 2. Subdivision of the total population sizes n_{tot} into sub-population sizes n_g in generations $g = BC_1, BC_2, BC_3, S_1, DH, S_2$ for different crossing and selection schemes for developing introgression populations.

Scheme	Generation				
	BC ₁	BC ₂	BC ₃	S ₁	DH/S ₂
Basic crossing schemes					
BC ₂ DH–C	1 × 1 × 360	1 × 10 × 36	–	–	10 × 36 × 1
BC ₂ DH–H	1 × 1 × 360	1 × 20 × 18	–	–	20 × 18 × 1
BC ₃ DH–CC	1 × 1 × 360	1 × 10 × 36	10 × 1 × 36	–	10 × 36 × 1
BC ₃ DH–HH	1 × 1 × 360	1 × 20 × 18	20 × 1 × 18	–	20 × 18 × 1
BC ₃ DH–CH	1 × 1 × 360	1 × 10 × 36	10 × 2 × 18	–	20 × 18 × 1
Crossing schemes with high selection intensity					
BC ₃ DH–CC'	1 × 1 × 720	1 × 10 × 72	10 × 1 × 72	–	10 × 72 × 1
BC ₃ DH–HH'	1 × 1 × 720	1 × 20 × 36	20 × 1 × 36	–	20 × 36 × 1
BC ₃ DH–CH'	1 × 1 × 720	1 × 10 × 72	10 × 2 × 36	–	20 × 36 × 1
Crossing schemes with selection in the final BC generation					
BC ₂ DH–CC	1 × 1 × 360	1 × 10 × 72	–	–	10 × 36 × 1
BC ₂ DH–HH	1 × 1 × 360	1 × 20 × 36	–	–	20 × 18 × 1
BC ₂ DH–CH	1 × 1 × 360	1 × 10 × 72	–	–	10 × (2 × 18) × 1
BC ₃ DH–CCC	1 × 1 × 360	1 × 10 × 36	10 × 1 × 72	–	10 × 36 × 1
BC ₃ DH–HHH	1 × 1 × 360	1 × 20 × 18	20 × 1 × 36	–	20 × 18 × 1
BC ₃ DH–CHH	1 × 1 × 360	1 × 10 × 36	10 × 2 × 36	–	20 × 18 × 1
Crossing schemes with increasing population sizes					
BC ₃ DH–HHH*	1 × 1 × 180	1 × 20 × 18	20 × 1 × 27	–	20 × 18 × 1
BC ₃ DH–HHS*	1 × 1 × 180	1 × 20 × 18	20 × 1 × 30	–	20 × (5 × 3) × 1
Basic crossing schemes					
BC ₂ S ₂ –C	1 × 1 × 360	1 × 10 × 36	–	10 × 36 × 1	10 × 36 × 1
BC ₂ S ₂ –H	1 × 1 × 360	1 × 20 × 18	–	20 × 18 × 1	20 × 18 × 1
BC ₃ S ₂ –CC	1 × 1 × 360	1 × 10 × 36	10 × 1 × 36	10 × 36 × 1	10 × 36 × 1
BC ₃ S ₂ –HH	1 × 1 × 360	1 × 20 × 18	20 × 1 × 18	20 × 18 × 1	20 × 18 × 1
BC ₃ S ₂ –CH	1 × 1 × 360	1 × 10 × 36	10 × 2 × 18	20 × 18 × 1	20 × 18 × 1
Crossing schemes with high selection intensity					
BC ₃ S ₂ –CC'	1 × 1 × 720	1 × 10 × 72	10 × 1 × 72	10 × 72 × 1	10 × 72 × 1
BC ₃ S ₂ –HH'	1 × 1 × 720	1 × 20 × 36	20 × 1 × 36	20 × 36 × 1	20 × 36 × 1
BC ₃ S ₂ –CH'	1 × 1 × 720	1 × 10 × 72	10 × 2 × 36	20 × 36 × 1	20 × 36 × 1
Crossing schemes with selection in the final BC generation					
BC ₂ S ₂ –CC	1 × 1 × 360	1 × 10 × 36	–	10 × 1 × 36	10 × 36 × 1
BC ₂ S ₂ –HH	1 × 1 × 360	1 × 20 × 18	–	20 × 1 × 18	20 × 18 × 1
BC ₂ S ₂ –CH	1 × 1 × 360	1 × 10 × 36	–	10 × 2 × 18	20 × 18 × 1
BC ₃ S ₂ –CCC	1 × 1 × 360	1 × 10 × 36	10 × 1 × 36	10 × 1 × 36	10 × 36 × 1
BC ₃ S ₂ –HHH	1 × 1 × 360	1 × 20 × 18	20 × 1 × 18	20 × 1 × 18	20 × 18 × 1
BC ₃ S ₂ –CHH	1 × 1 × 360	1 × 10 × 36	10 × 2 × 18	20 × 1 × 18	20 × 18 × 1
Crossing schemes with increasing population sizes					
BC ₃ S ₂ –HHH*	1 × 1 × 180	1 × 20 × 18	20 × 1 × 27	20 × 1 × 18	20 × 18 × 1
BC ₃ S ₂ –HHS*	1 × 1 × 180	1 × 20 × 18	20 × 1 × 23	20 × 5 × 4	100 × 1 × 4

The total population size in generation g is defined as $n_{tot} = n_{pop} \times n_{sel} \times n_g$. n_{pop} : number of sub-populations in generation $g - 1$; n_{sel} : number of individuals selected from the sub-populations in generation $g - 1$; n_g : population size per sub-population in generation g .
doi:10.1371/journal.pone.0092429.t002

generations of selection was the same. For example, the basic crossing scheme BC₃DH–CC resulted in a D_t of only 5.0%, while crossing scheme BC₂DH–CC with selection in the final backcross generation resulted in a D_t of 7.8%. An additional generation of selection in BC₂ schemes only resulted in minor

improvements of D_t of 0.4–1.4% compared to the basic crossing schemes without selection. For example, scheme BC₂DH–CC improved D_t only by 0.5% compared to scheme BC₂DH–C.

The DH crossing schemes had in most cases better values for T , D_t , D_c , D_s and especially S than the S₂ crossing schemes (Table 3).

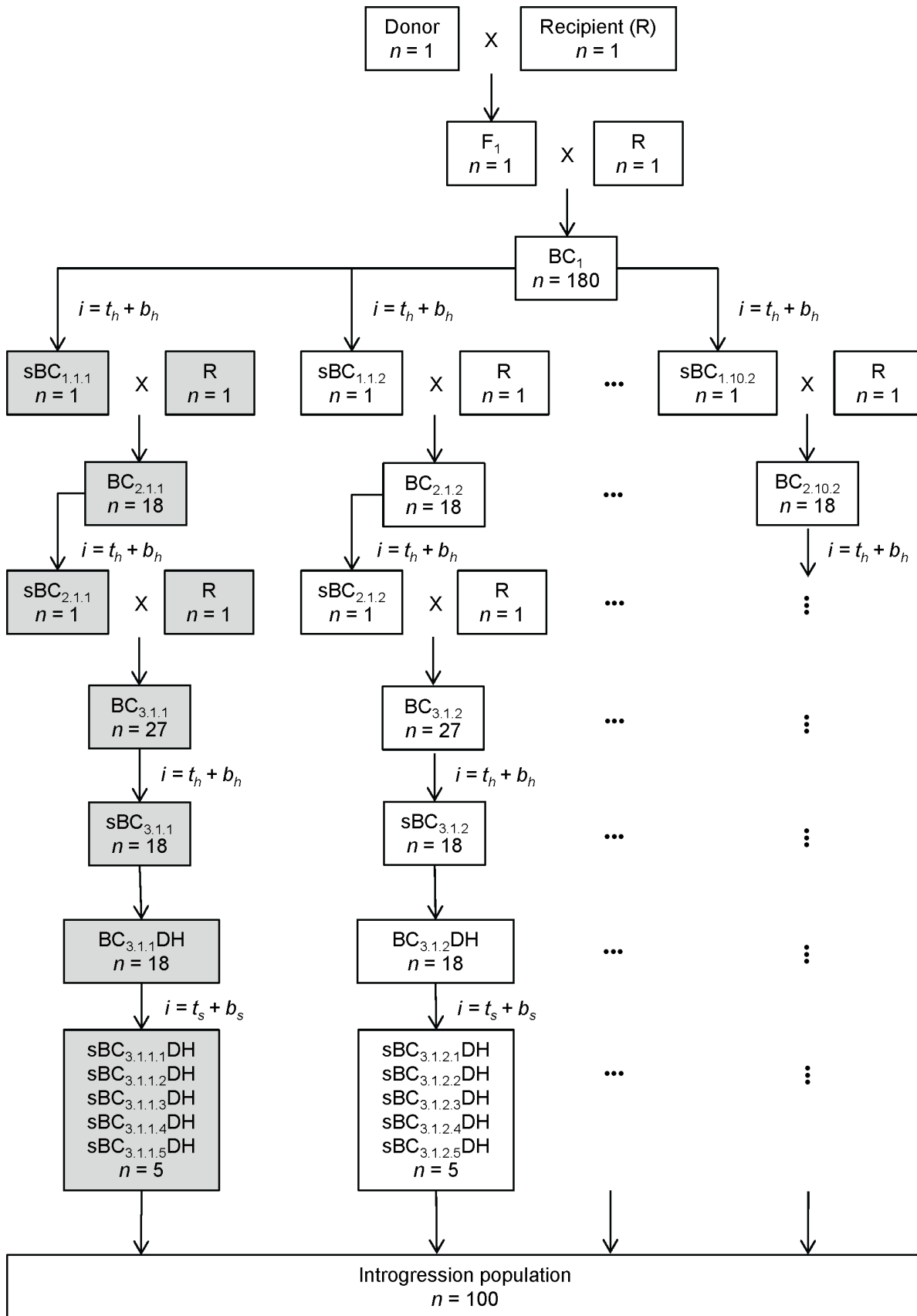


Figure 1. Schematic representation of crossing scheme $BC_3DH-HHH^*$. Crossing scheme $BC_3DH-HHH^*$ is characterized by increasing population sizes in the backcross generations and selection for donor chromosome halves in the final backcross generation. The parts highlighted in gray represent one branch of the crossing scheme. Sub-populations are indexed by BC_g , $BC_{g,c,h}$ and $BC_{g,c,h,s}$, where g is the respective backcross generation, c is the respective chromosome, h is the respective chromosome half, s is the respective target segment; $sBC_{g,c,h}$ and $sBC_{g,c,h,s}$ denote individuals selected for the respective selection regions.
doi:10.1371/journal.pone.0092429.g001

Very high values of $S > 1000$ segments were observed for the basic crossing schemes BC_2S_2-C and BC_2S_2-H . These crossing schemes had on average $N=1$ additional donor segment per IL compared to the corresponding DH crossing schemes. However, they were also characterized by incomplete homozygosity (Figure 3B). The S_2 crossing schemes with selection in the final backcross generation required 360 individuals and HT assays less than the corresponding DH crossing schemes (Tables 2 and 3). Nevertheless, the differences between DH and S_2 crossing schemes then diminished. For example, scheme BC_3S_2-HHH resulted in similar values for most measures as the corresponding scheme $BC_3DH-HHH$ (Table 3).

The differences in the total donor genome proportion D_t between selection for complete donor chromosomes and selection for donor chromosome halves ranged only between 0.1–0.7% for the same number of backcross generations and generations of selection. However, substantial differences were observed for the donor genome proportion of the carrier chromosomes D_c and the donor genome proportion of the target segments D_s . For selection for complete donor chromosomes, high values for D_c of up to 48% were observed. They were clearly visible in the graphical genotypes for schemes BC_2S_2-CC and BC_3DH-CC (Figure 3B and C). For selection for donor chromosome halves, the values for D_c were much lower and did not exceed 42% (Table 3). Without selection in the final backcross generation, selection for donor chromosome halves resulted in substantially reduced values for D_s . For example, the basic crossing schemes BC_3DH-HH and BC_3S_2-HH resulted in values for D_s of only 94% and 90%. Moreover, the ranges for D_s for these crossing schemes were substantially greater (Figure 4 for S_2 crossing schemes, for DH data not shown).

The basic crossing schemes BC_3DH-CH and BC_3S_2-CH which combined selection for complete donor chromosomes and selection for donor chromosome halves resulted in similarly low values for D_s of 93.7% and 89.9% as selection for donor chromosome halves only (Table 3). In addition, the combined strategies CH and CHH resulted in high values for D_c of up to 45.9%. The low values for D_s and the high values for D_c were reflected in the graphical genotype of scheme BC_3DH-CH , e.g. in ILs 47 and 54 (Figure 3A).

Doubling population sizes n_{tot} from 360 to 720 individuals in the crossing schemes with high selection intensity reduced the total donor genome proportion D_t from 5.0–5.1% to 3.6–3.8% compared to the basic DH crossing schemes, and from 5.3–5.7% to 4.3–4.4% compared to the basic S_2 crossing schemes (Table 3). The donor genome proportion of the carrier chromosomes D_c was reduced by about 4.2–7.5% for the DH crossing schemes, and by about 0.9–6.9% for the S_2 crossing schemes. The reduction of the donor genome proportion of the target segments D_s in combination with increased ranges that was observed with selection for donor chromosome halves in the basic crossing schemes was not observed in the crossing schemes with high selection intensity (Table 3 and Figure 4). D_s was increased by 5.2% for crossing scheme BC_3DH-HH' and by 8.6% for crossing scheme BC_3S_2-HH' compared to the basic crossing schemes BC_3DH-HH and BC_3S_2-HH . However, these

improvements were only achieved with 2160 HT assays compared to 1080 HT assays in the basic crossing schemes (Table 3).

The crossing schemes with selection in the final backcross generation resulted in values for D_t that were 1.1–1.2% higher for the DH crossing schemes and 0.6–1.4% higher for the S_2 crossing schemes compared to the crossing schemes with high selection intensity. The ranges of D_s for selection for donor chromosome halves were about the same size as for the crossing schemes with high selection intensity (Figure 4). The average values for D_s were 0.5% lower for scheme $BC_3DH-HHH$ and 0.3% lower for scheme BC_3S_2-HHH (Table 3). The number of required HT assays was reduced by 360 for the DH crossing schemes and by 720 for S_2 crossing schemes compared to the crossing schemes with high selection intensity. For the crossing schemes with selection in the final backcross generation, selection for donor chromosome halves was the most advantageous selection strategy with respect to the genetic background and to the target segments. Most notably, the crossing schemes $BC_3DH-HHH$ and BC_3S_2-HHH resulted in the lowest values for the donor genome proportion of the carrier chromosomes D_c . Compared to the most efficient basic crossing schemes BC_2DH-C and BC_3DH-CC , the crossing schemes BC_2DH-HH and $BC_3DH-HHH$ resulted in small improvements of both the genetic background and D_s . However, in both cases 720 additional HT assays had to be invested. For the S_2 crossing schemes with selection in the final backcross generation, high values of D_c of 38.1–48.3% were observed. Large donor chromosome segments on the carrier chromosomes were also visible in the graphical genotypes for schemes BC_2S_2-CC and BC_3S_2-HHH (Figure 3B and D). The high values for D_c were associated with a considerable reduction of the number of disjunct genome segments S of >200 segments for the BC_2 crossing schemes and of 100–200 segments for the BC_3 crossing schemes compared to the basic S_2 crossing schemes (Table 3).

The crossing schemes with increasing population sizes reduced the number of required HT assays for DH crossing schemes by 360 in comparison to the crossing schemes with selection in the final backcross generation and constant population sizes. The crossing schemes $BC_3DH-HHH^*$ and $BC_3DH-HHS^*$ resulted in similar values for most measures as the crossing scheme $BC_3DH-HHH$. However, D_c and D_s were slightly reduced for crossing scheme $BC_3DH-HHS^*$. Compared to the most efficient basic crossing scheme BC_3DH-CC , crossing scheme $BC_3DH-HHH^*$ required 360 additional HT assays, but reduced D_c by 1.9% and increased D_s by 0.6%. The crossing scheme $BC_3S_2-HHH^*$ resulted with 38.0% in a much higher D_c than the crossing scheme $BC_3S_2-HHS^*$ with 30.4%. For crossing scheme $BC_3S_2-HHS^*$, the average D_s was only 96.2% and the range for D_s was higher than for the crossing schemes $BC_3S_2-HHH^*$ and BC_3S_2-HHH (Figure 4). However, D_t and D_c were the lowest for all investigated crossing schemes, with the exception of the crossing schemes with high selection intensity and $n_{tot}=720$ (Table 3). The clear-cut separation of the target segments is also visible in the graphical genotype (Figure 3F).

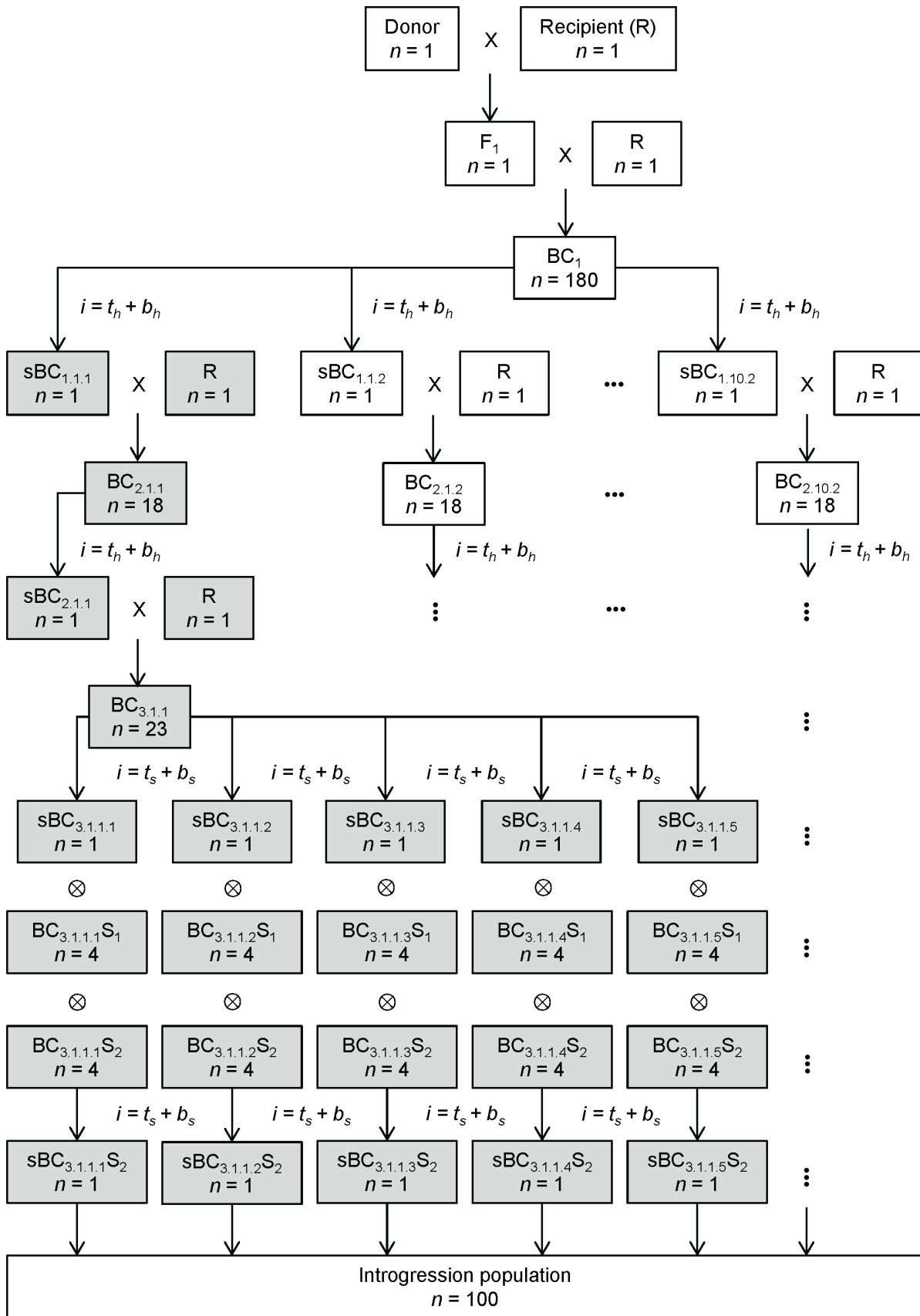


Figure 2. Schematic representation of crossing scheme $BC_3S_2-HHS^*$. Crossing scheme $BC_3S_2-HHS^*$ is characterized by increasing population sizes in the backcross generations and selection for target segments in the final backcross generation. The parts highlighted in gray represent one branch of the crossing scheme. Sub-populations are indexed by BC_g , $BC_{g,c,h}$ and $BC_{g,c,h,s}$ where g is the respective backcross generation, c is the respective chromosome, h is the respective chromosome half, s is the respective target segment; $sBC_{g,c,h}$ and $sBC_{g,c,h,s}$ denote individuals selected for the respective selection regions.
doi:10.1371/journal.pone.0092429.g002

Discussion

Measures for Characterizing Introgression Populations

Measures for the description of introgression populations should allow to distinguish between introgression populations of different structure. Complete donor genome coverage O is desirable in order to make the complete genetic variation of the donor available for the breeding process. However, high values for O can also be caused by donor segments outside the target segments which could not be removed from the genetic background. O is therefore only informative if interpreted in relation to measures which reflect the distribution of the donor genome in the introgression population. A distinctive description of introgression populations is possible with the total donor genome proportion D_t , the donor genome proportion on the carrier chromosomes D_c and the donor genome proportion of the actual target segments D_s .

A high total donor genome proportion D_t is often associated with a high number of disjunct genome segments S . S determines the resolution R , which is an important parameter for the accuracy of QTL detection. However, if S is greater than the number of ILs, the problem of overparameterization arises with classical linear model approaches. This issue has only in part been resolved by using statistical methods which pre-select a reduced number of ILs for the linear model [23].

High values for the donor genome proportion on the carrier chromosomes D_c and the depth of donor genome coverage T reflect undesired donor segments attached to the actual target segments. Such large donor segments which overlap between ILs have been reported to increase the risk of false-positive effects in QTL detection and reduce the power of QTL detection [24]. This is mainly a problem if linkage maps with large distances between adjacent markers of 10 cM or more are employed, because QTLs located between the last marker of the target segment and the next marker outside the target segment are incorrectly assigned to the target segments. With dense marker maps which are now available this problem should be overcome. However, large donor segments also increase the risk of linkage drag in the breeding process and often require further steps of separation [24].

Low values for the donor genome proportion of the target segments D_s indicate a loss of target segments and potentially useful alleles. This is a problem that arises with small population sizes as were investigated in the present study [16]. Even if the missing target segments are present in the genetic background of other ILs, this might impair QTL detection and the further use of the ILs for the breeding progress.

We therefore argue that short non-overlapping target segments in a clean recipient background are advantageous also with dense marker maps. For 20 cM target segments and a genomic model of 10 equally sized chromosomes of 200 cM length, this corresponds to $D_t = 1\%$, $D_c = 10\%$ and $D_s = 100\%$ in the ideal case. The effort and time required for developing introgression populations with such characteristics is beyond the scope of most breeding programs. With the limited population sizes and number of HT assays investigated in this study, these ideal values could not be achieved with two or three backcross generations (Table 3). We therefore considered those crossing and selection schemes as efficient which with a given limited resource input resulted in the

highest coverage of target segments D_s in combination with low overlap of target segments reflected in D_c and T and a low total donor genome proportion D_t .

With respect to QTL detection, it can be expected that the optimal values for the suggested measures will depend on the statistical method and the genetic architecture of the trait. They could be determined for a given statistical method by including QTLs of different number and effect in future simulation studies. We plan further investigations in this area of research.

Crossing Schemes

BC_3 crossing schemes had 2–3% lower values for the total donor genome proportion D_t than BC_2 crossing schemes (Table 3), even if no selection for the genetic background was conducted in generation BC_3 . Selection in generation BC_2 , as was investigated with the crossing schemes BC_2-CC , BC_2-HH and BC_2-CH , only resulted in a reduction of D_t of 0.4–1.4% compared to the basic crossing schemes BC_2-C and BC_2-H (Table 3). An explanation for this comparatively small reduction is that the limiting factor for the reduction of D_t is the number of recombinations during meiosis. Hence, even though BC_2 crossing schemes have a time advantage, the effect of a third backcross generation cannot be compensated by investing in additional marker analyses. We therefore conclude that BC_3 crossing schemes result in introgression populations with an improved structure, and that the time investment in the additional backcross generation is worthwhile.

DH crossing schemes were for most measures superior to the corresponding S_2 crossing schemes. The differences were most pronounced in the number of disjunct genome segments S . Even though the S_2 schemes on average had a slightly higher number of donor segments per IL N , it seems that the very high values for S that were observed especially in the BC_2-S_2 crossing schemes mainly had to be attributed to incomplete homozygosity (Figure 3B). It can be expected that introgression populations with $S > 1000$ segments in 100 ILs (Table 3) are not suitable for effective QTL detection. We therefore conclude that the DH method is essential for short crossing schemes with only two backcross generations.

A drawback of the DH method is that with current protocols of *in vivo* DH induction of maternal haploids, only a very limited number of viable DH lines can be derived from one backcross individual. We expect that our assumption of one DH line per backcross individual is a conservative, but realistic estimate. In contrast, with selfing, many progenies can be derived from one selected backcross individual. In the S_2 crossing schemes, it is consequently comparatively cheap and easy to conduct selection in the final backcross generation. For the DH schemes, selection in the final backcross generation could only be conducted if population size in this generation was higher than the desired number of final DH lines. As a result, the S_2 crossing schemes with selection in the final backcross generation required 360 HT assays less than the corresponding DH schemes (Table 3). Moreover, the selected fractions of best backcross individuals were much greater for the DH than for the S_2 crossing schemes (Table 2). This resulted in a lower selection intensity for both the selection region of the final backcross generation and the genetic background in

Table 3. Measures evaluated for introgression populations resulting from different crossing and selection schemes.

Scheme	<i>O</i>	<i>T</i>	<i>S</i>	<i>R</i>	<i>N</i>	<i>L</i>	<i>D_t</i>	<i>D_c</i>	<i>D_s</i>	HT
Basic crossing schemes										
BC ₂ DH–C	99.9	8.3	691	2.9	6.2	28.7	8.3	38.9	98.7	720
BC ₂ DH–H	100.0	8.9	751	2.7	6.4	29.2	8.9	37.3	94.1	720
BC ₃ DH–CC	99.2	5.1	457	4.4	3.8	29.8	5.0	35.2	97.8	1080
BC ₃ DH–HH	99.8	5.2	487	4.1	3.9	29.8	5.1	33.3	94.2	1080
BC ₃ DH–CH	99.6	5.2	469	4.3	3.8	30.5	5.1	35.1	93.7	1080
Crossing schemes with high selection intensity										
BC ₃ DH–CC'	99.3	3.6	389	5.1	3.0	27.8	3.6	27.7	98.8	2160
BC ₃ DH–HH'	99.9	3.8	406	4.9	3.0	29.7	3.8	29.1	99.4	2160
BC ₃ DH–CH'	99.6	3.9	399	5.0	3.0	30.0	3.8	29.9	99.0	2160
Crossing schemes with selection in the final BC generation										
BC ₂ DH–CC	99.9	7.8	676	3.0	5.9	28.7	7.8	41.0	98.9	1440
BC ₂ DH–HH	100.0	8.1	716	2.8	6.0	29.0	8.1	38.9	99.3	1440
BC ₂ DH–CH	99.9	8.5	679	2.9	6.0	30.4	8.5	43.7	98.8	1440
BC ₃ DH–CCC	99.1	4.9	457	4.4	3.7	30.2	4.8	35.8	98.0	1800
BC ₃ DH–HHH	99.9	4.8	464	4.3	3.6	31.0	4.7	33.9	98.9	1800
BC ₃ DH–CHH	99.6	4.8	450	4.4	3.5	31.4	4.7	35.0	98.5	1800
Crossing schemes with increasing population sizes										
BC ₃ DH–HHH*	99.9	5.0	492	4.1	3.8	29.7	5.0	33.3	98.4	1440
BC ₃ DH–HHS*	99.8	4.9	484	4.1	3.8	29.4	4.8	32.5	97.5	1440
Basic crossing schemes										
BC ₂ S ₂ –C	100.0	11.4	1021	2.0	7.3	26.4	9.3	41.4	97.6	720
BC ₂ S ₂ –H	100.0	11.4	1073	1.9	7.4	25.9	9.3	36.3	90.4	720
BC ₃ S ₂ –CC	99.3	6.9	684	2.9	4.5	27.7	5.7	39.2	97.0	1080
BC ₃ S ₂ –HH	99.9	6.3	702	2.8	4.4	26.3	5.3	33.0	90.3	1080
BC ₃ S ₂ –CH	99.7	6.4	681	2.9	4.3	26.9	5.3	35.0	89.9	1080
Crossing schemes with high selection intensity										
BC ₃ S ₂ –CC'	99.5	5.1	585	3.4	3.6	26.6	4.3	32.3	98.8	2160
BC ₃ S ₂ –HH'	99.9	5.1	591	3.4	3.5	27.8	4.4	32.1	98.9	2160
BC ₃ S ₂ –CH'	99.7	5.2	581	3.4	3.5	28.2	4.4	33.4	98.4	2160
Crossing schemes with selection in the final BC generation										
BC ₂ S ₂ –CC	99.3	10.5	795	2.5	6.5	27.2	8.3	48.3	98.0	1080
BC ₂ S ₂ –HH	99.9	9.7	785	2.5	6.2	27.0	7.9	42.8	98.8	1080
BC ₂ S ₂ –CH	99.8	9.7	761	2.6	6.1	27.6	7.9	45.9	98.4	1080
BC ₃ S ₂ –CCC	98.3	7.1	588	3.4	4.2	30.5	5.7	44.6	97.5	1440
BC ₃ S ₂ –HHH	99.7	5.9	510	3.9	3.7	31.0	5.0	38.1	98.6	1440
BC ₃ S ₂ –CHH	99.3	6.0	509	3.9	3.7	31.3	5.0	39.5	98.2	1440
Crossing schemes with increasing population sizes										
BC ₃ S ₂ –HHH*	99.7	5.8	508	3.9	3.8	30.7	4.9	38.0	98.7	1440
BC ₃ S ₂ –HHS*	99.8	5.1	596	3.4	3.7	26.1	4.3	30.4	96.2	1400

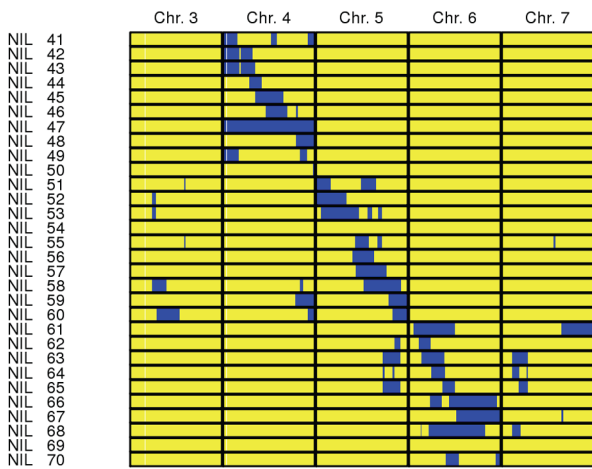
O: donor genome coverage in percent; *T*: depth of donor genome coverage; *S*: number of disjunct genome segments; *R*: resolution; *N*: number of donor segments per IL; *L*: length of donor segments per IL in cM; *D_t*: total donor genome proportion in percent; *D_c*: donor genome proportion of carrier chromosomes in percent; *D_s*: donor genome proportion of target segments in percent; HT: the required number of HT assays. Measures are arithmetic means over 1,000 replications. doi:10.1371/journal.pone.0092429.t003

the DH crossing schemes. We therefore suggest that a comparison of DH and S₂ crossing schemes should take the distinctive features of both methods into account. The evaluation of efficiency should also be based on the number of required HT assays. Considering this, S₂ crossing schemes which exploit their selection advantages represent economic and easy-to-handle alternatives to DH crossing schemes.

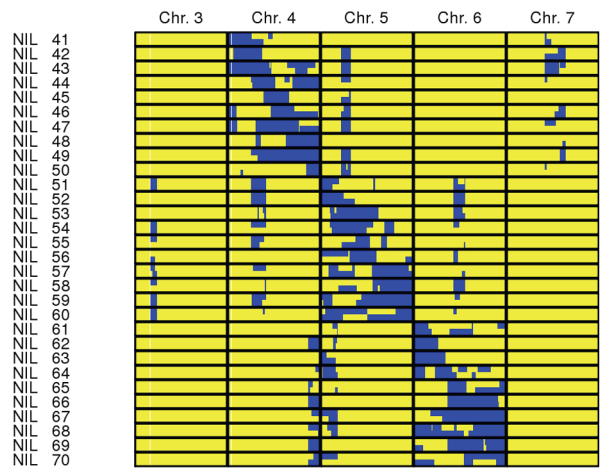
Selection Strategies for Small and Constant Population Sizes

For a given genetic model and crossing scheme, the selection strategy is the most important factor that influences the structure of the resulting introgression population. In the following paragraphs, different aspects such as the length of the selection

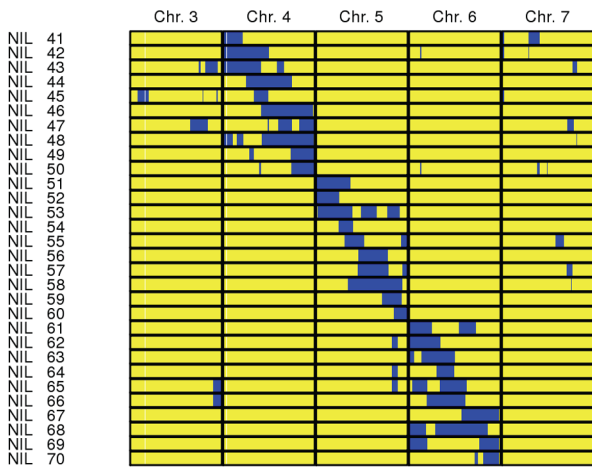
A



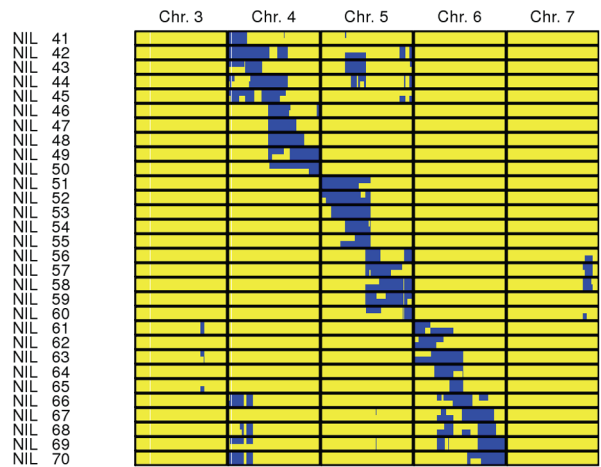
B



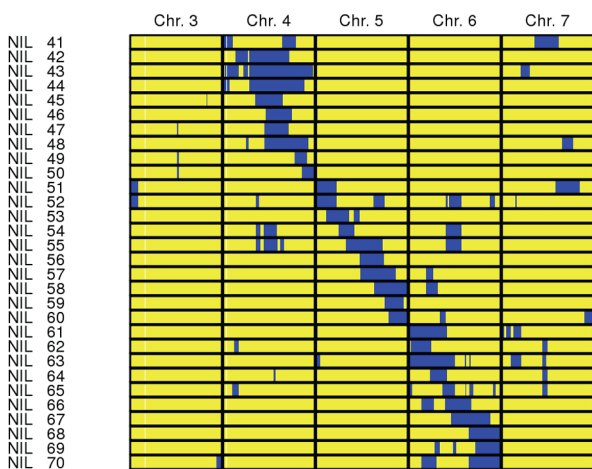
C



D



E



F

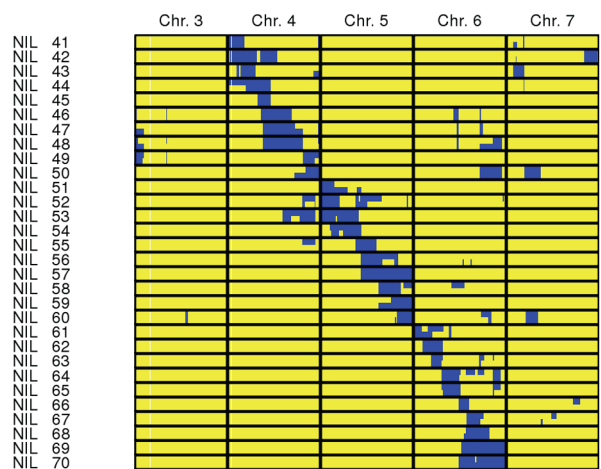


Figure 3. Graphical genotypes of introgression populations resulting from six different crossing schemes. A: BC₃DH–CH; B: BC₂S₂–CC; C: BC₃DH–CC; D: BC₃S₂–HHH; E: BC₃DH–HHH^{*}; F: BC₃S₂–HHS^{*}. The graphical genotypes display the chromosomes 3 to 7 of ILS 41–70 and are examples from one simulation run. Chromosome segments which stem from the donor are displayed in blue, whereas chromosome segments which stem from the recipient are displayed in yellow. The graphical genotypes illustrate the differences between the alternative crossing schemes with respect to their suitability to create introgression populations with complete donor genome coverage and clearly separated, evenly distributed target donor chromosome segments. doi:10.1371/journal.pone.0092429.g003

regions, the number of generations of selection and the required population sizes for effective selection are discussed.

Selection strategies which pre-select individuals carrying complete donor chromosomes reduce the number of simultaneous backcross programs to the number of chromosomes [16]. They are therefore suitable for breeding programs with limited resources. However, for long chromosomes of 200 cM length, selection for complete donor chromosomes preserved large donor chromosome segments on the carrier chromosomes up to line development (Figure 3C and B). This was reflected in high values for the proportion of donor genome on the carrier chromosomes D_c of up to 48% (Table 3). The selection regions for selection in the backcross generations were therefore reduced to donor chromosome halves for selection strategies H, HH and HHH. In all four series of simulations, selection for donor chromosome halves resulted in the desired reduction of D_c compared to selection for complete donor chromosomes (Table 3). Other measures for the genetic background were approximately equivalent. We therefore conclude that for crop species with long chromosomes such as maize, wheat or rapeseed, selection for donor chromosome halves reduces the length of the donor segments attached to the actual target segments and the risk of linkage drag.

However, for crossing schemes without selection in the final backcross generation and constant population sizes of $n_{tot} = 360$ individuals, selection for donor chromosome halves resulted in a considerable reduction of the donor genome proportion of the target segments D_s of up to 7%. Moreover, the estimated values for D_s were less reliable for these crossing schemes, e.g., in schemes BC₂S₂–H and BC₃S₂–HH (Figure 4). These findings have to be attributed to the small population sizes n_g in the sub-populations and the structure of the selection index i . In generation DH or S₂, population sizes were reduced to $n_g = 18$ individuals with selection for donor chromosome halves (Table 2). Without selection in the final backcross generation, around 50% of the ILS developed from the backcross individuals are expected to carry no donor allele at a given locus within the respective target segment. The probability to find five ILS with complete donor target segments for the introgression population was therefore even further reduced. As the selection index $i = t_s + b_s$ weighed the target segments and the genetic background equally, a clean genetic background sometimes outweighed a reduced D_s and led to the observed loss of target segments in these small sub-populations. We therefore conclude that a sufficiently large population size is the crucial

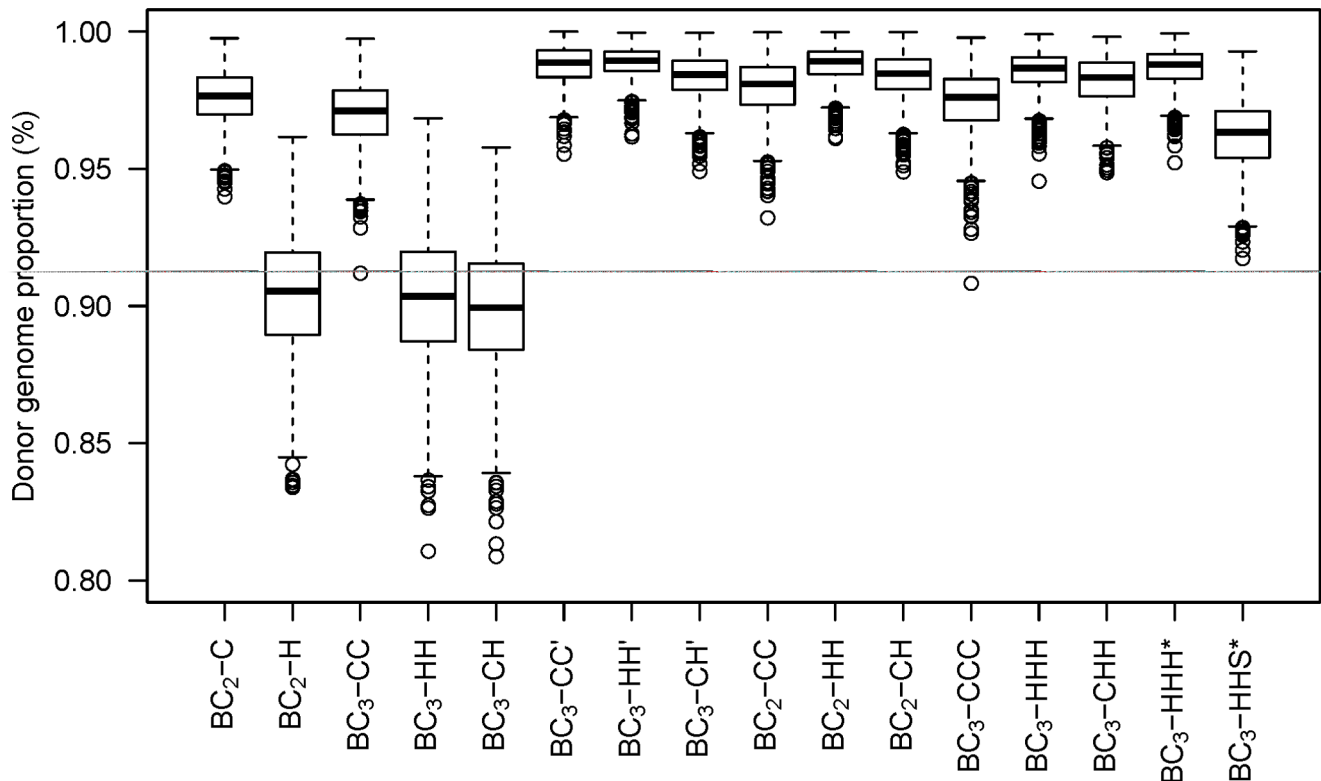


Figure 4. Donor genome proportion of target segments D_s for all investigated S₂ crossing schemes. The boxplots represent the distribution over 1,000 replications of the simulations. The basic crossing schemes BC₂S₂–H, BC₃S₂–HH and BC₃S₂–CH which select for donor chromosome halves are characterized by higher ranges for D_s . doi:10.1371/journal.pone.0092429.g004

factor for the successful application of selection for donor chromosome halves.

A loss of target segments caused by small population sizes was also observed for the basic combined selection strategy CH which selected for complete donor chromosomes in generation BC₁ and for donor chromosome halves in generation BC₂. In addition, the combined strategies CH and CHH resulted in high values for the donor genome proportion on the carrier chromosomes D_c of up to 45% (Table 3). This can be explained by the efficient selection for complete donor chromosomes from the comparatively large BC₁ population of $n_g = n_{\text{tot}} = 360$ individuals (Table 2). The pre-selected complete donor chromosomes are in large part preserved up to line development. The combination of missing target segments with large donor chromosome segments on carrier chromosomes was also reflected in the graphical genotype for scheme BC₃DH–CH, *e.g.*, in IL 47 and 54 (Figure 3A). The selection strategies CH and CHH therefore combine the drawbacks of both selection for complete donor chromosomes and selection for chromosome halves. They are not suitable for crossing schemes with small and constant population sizes, in which the population sizes n_g in the sub-populations are subsequently reduced over the backcross generations. We conclude that for small breeding programs with a constant population size of $n_{\text{tot}} = 360$ and a limited number of HT assays for selection, selection strategies which only select for complete donor chromosomes in the backcross generations should be employed in both DH and S₂ crossing schemes to avoid the loss of target segments.

Finding more Carriers of Donor Target Segments for Line Development

To employ selection for donor chromosome halves effectively for reducing the donor genome proportion of the carrier chromosomes D_c without losing the target segments, it is necessary to increase the frequency of carriers of donor target segments for line development. Using larger population sizes is a straightforward solution for this problem, which in addition can improve the overall structure of introgression populations. The crossing schemes with high selection intensity and double population sizes of $n_{\text{tot}} = 720$ individuals resulted in small improvements of the total donor genome proportion D_t of about 1–1.5% compared to the basic crossing schemes (Table 3). The desired increase in the donor proportion of the target segments D_s was achieved. For selection for donor chromosome halves, D_s was increased by 5.2–8.6%. Selection for donor chromosome halves was then even superior to selection for complete donor chromosomes. Moreover, the donor genome proportion on the carrier chromosomes D_c was reduced by up to 7.5%, indicating an improved separation of target segments. The observed improvements were greater for the DH than for the S₂ crossing schemes. Nevertheless, the comparatively small improvements of the introgression populations required 1080 additional HT assays. We therefore conclude that such large population sizes are only suitable for breeding programs with access to DH technology, less stringent resource restrictions and high requirements with respect to the genetic background. If the requirements concerning the structure of the introgression population are not that high, it might be more economic to increase population size only in the final backcross generation and/or to invest in additional HT assays only in this generation.

For crossing schemes with selection in the final backcross generation, the total donor genome proportion D_t was similar to the values of the basic crossing schemes, and about 1% higher than for the crossing schemes with higher selection intensity

(Table 3). However, the average values for the donor genome proportion of the target segments D_s were similar to the crossing schemes with higher selection intensity (Table 3) and the ranges were effectively reduced (Figure 4). Moreover, the number of required HT assays was reduced by 360 for the DH crossing schemes and by 720 for S₂ crossing schemes compared to the crossing schemes with higher selection intensity (Table 3). The decision for doubling population sizes requires the same resources as would be required for generating an additional introgression population. This large effort seems not to be justified by the relatively small improvements compared to the basic crossing schemes. We therefore conclude that selection in the final backcross generation is the more efficient solution for both DH and S₂ crossing schemes.

Selection for donor chromosome halves was the best strategy with selection in the final backcross generation for both DH and S₂ crossing schemes (Table 3). However, for the DH schemes, only small improvements for schemes BC₂DH–HH and BC₃DH–HHH were observed compared to the most efficient basic crossing schemes BC₂DH–C and BC₃DH–CC (Table 3). For these small improvements, 720 additional individuals and HT assays had to be invested. For the S₂ schemes, considerable reductions in S of 174 and 236 segments were observed for schemes BC₂S₂–HH and BC₃S₂–HHH with selection in the final backcross generation compared to the basic crossing schemes BC₂S₂–C and BC₂S₂–CC. D_t was only slightly reduced. However, the donor genome proportion on carrier chromosomes D_c was in general very high for the crossing schemes with selection in the final backcross generation with 38–48%. This indicates a fixation of the selection regions of the final backcross generation (Figure 3B and D). In schemes BC₂S₂–CC and BC₃S₂–HH, complete donor chromosomes and donor chromosome halves still appear as blocks around the target segments. These blocks lead to an overlap of donor segments between ILS that reduces the effective resolution of the introgression population for QTL detection. The overlap also hampers the further use of the ILS in the breeding process, as further steps of separation of the target segments by backcrossing are required. We therefore conclude that the crossing schemes with selection in the final backcross generation have the potential to improve the resulting introgression populations at moderate cost. However, for the DH crossing schemes, the number of required HT assays and individuals has to be reduced. For the S₂ crossing schemes, the fixation of large donor chromosome segments has to be avoided. Optimizations of the respective crossing schemes are presented in the following.

Increasing Population Sizes Over Backcross Generations

With constant population sizes of $n_{\text{tot}} = 360$ individuals, the population size in generation BC₁ was large in relation to the genetic gains that could be achieved by selecting a comparatively small fraction of 10 or 20 individuals (Table 2). Starting with smaller population sizes in generation BC₁ and gradually increasing population sizes in the following backcross generations was therefore an efficient option to reduce the overall number of required individuals and HT assays for selection in the final backcross generation. Larger population sizes in generation BC₃ also enabled selection for target segments, which was investigated as an option to avoid the fixation of large donor chromosome segments especially for the S₂ crossing schemes.

The schemes BC₃DH–HHH* and BC₃DH–HHS* resulted in similar values for all measures (Table 3). However, D_c and D_s were slightly lower for scheme BC₃DH–HHS*. We therefore conclude that selection for target segments already in the final backcross generation is not efficient for DH crossing schemes. In

comparison to the best but also very expensive scheme $BC_3DH-HHH$ with selection in the final backcross generation, scheme $BC_3DH-HHH^*$ can be considered equivalent, but required 360 individuals and HT assays less. In comparison to the more economic basic crossing scheme BC_3DH-CC , scheme $BC_3DH-HHH^*$ improved D_c and D_s and thus the separation of target segments. This is also visible in the graphical genotype (Figure 3). The investment in the additional 360 HT assays seems therefore worthwhile (Table 3).

The scheme $BC_3S_2-HHS^*$ resulted in better values than the schemes BC_3S_2-HHH and $BC_3S_2-HHH^*$. Most notably, it resulted in a much lower D_c of 30% compared to 38%. Scheme $BC_3S_2-HHS^*$ resulted in a D_s that was 2.4% lower compared to schemes BC_3S_2-HHH and $BC_3S_2-HHH^*$ and the ranges for D_s were higher (Figure 4). Nevertheless, it resulted in the lowest values of D_t and D_c and the best separation of target genes of all investigated DH and S_2 crossing schemes with comparable population sizes. The comparatively high value of S of 596 segments in combination with reduced values for D_t can in this case be explained by a greatly improved separation of target segments compared to the other S_2 schemes with selection in the final backcross generation. The improved separation of target segments is also visible in the graphical genotype (Figure 3F). This was achieved with 40 HT assays less (Table 3). We therefore expect that scheme $BC_3S_2-HHS^*$ will result in an improved power of QTL detection, and recommend selection for target segments in the final backcross generation for S_2 crossing schemes.

Compared to the best but expensive comparable DH crossing scheme $BC_3DH-HHH$, the S_2 crossing scheme $BC_3S_2-HHS^*$ resulted in similar values and required 400 HT assays less. Overall, we conclude that increasing population sizes over backcross are advantageous and economic for both DH and S_2 crossing schemes. Moreover, crossing scheme $BC_3S_2-HHS^*$ can provide a cheap alternative to comparable DH crossing schemes.

Conclusions

Our study has shown that introgression populations with complete coverage of the donor genome and reasonably clean recipient background can be developed with a limited number of backcross individuals and HT assays. It has provided further insight on how different crossing and selection schemes influence the structure of the resulting introgression populations. The

guidelines which have been derived for maize are transferable to other crop species with similar number and length of chromosomes. For crops with different genome size, some considerations are discussed in the following.

Rapeseed is a crop with a large genome of 19 chromosomes, for which efficient protocols of microspore culture are available for DH production. For the large genome of rapeseed, it can be expected that the values for the total donor genome proportion D_t will be lower than those observed for the smaller genome of maize. With the investigated selection index i , the selection pressure on the carrier chromosomes will be reduced with increasing genome size and number of chromosomes. It might therefore be an interesting option for rapeseed to put more weight on the background markers on the carrier chromosomes to achieve an efficient reduction of D_c . As with microspore culture many DH lines can usually be derived from one backcross individual, the advantages of DH production should be more pronounced than for maize. However, the optimal selection strategies for DH crossing schemes in rapeseed should then be similar to those for selfing in maize.

Sugar beet is a crop with a small genome of 9 chromosomes, for which the guidelines for selfing should be most relevant. In smaller genomes, equivalent values of D_t can usually be reached with smaller population sizes and with fewer backcrosses. However, the average length of the chromosomes in cM is also much shorter than in maize. This implies that fewer crossovers occur per meiosis, and that it might require more individuals and backcross generations to effectively separate the target segments. The combined effects of genome size and chromosome length will also depend on the desired number and length of the target segments.

Simulations can considerably facilitate the planning process for the development of introgression populations in different crop species. The derived guidelines can help breeders and geneticists to enhance the genetic variation of narrow based breeding populations of crops.

Author Contributions

Conceived and designed the experiments: EH KCF TP DS MO MF. Performed the experiments: EH KCF. Analyzed the data: EH. Wrote the paper: EH MF.

References

- Zamir D (2001) Improving plant breeding with exotic genetic libraries. *Nat Rev Genet* 2: 983–989.
- McCouch S (2004) Diversifying selection in plant breeding. *PLoS Biol* 2: e347.
- Eshed Y, Zamir D (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. *Euphytica* 79: 175–179.
- Keurentjes JJ, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, et al. (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* 175: 891–905.
- Törjék O, Meyer RC, Zehndorf M, Teltow M, Strompen G, et al. (2008) Construction and analysis of 2 reciprocal *Arabidopsis* introgression line populations. *J Hered* 99: 396–406.
- Lin S, Sasaki T, Yano M (1998) Mapping quantitative trait loci controlling seed dormancy and heading date in rice, *Oryza sativa* L., using backcross inbred lines. *Theor Appl Genet* 96: 997–1003.
- Cheema KK, Bains NS, Mangat GS, Das A, Vikal Y, et al. (2008) Development of high yielding IR64 × *Oryza rufipogon* (Griff.) introgression lines and identification of introgressed alien chromosome segments using SSR markers. *Euphytica* 160: 401–409.
- Matus I, Corey A, Filichkin T, Hayes P, Vales M, et al. (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.
- Schmalenbach I, Körber N, Pillen K (2008) Selecting a set of wild barley introgression lines and verification of QTL effects for resistance to powdery mildew and leaf rust. *Theor Appl Genet* 117: 1093–1106.
- Liu S, Zhou R, Dong Y, Li P, Jia J (2006) Development, utilization of introgression lines using a synthetic wheat as donor. *Theor Appl Genet* 112: 1360–1373.
- Pumphrey MO, Bernardo R, Anderson JA (2007) Validating the *Fhb1* QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. *Crop Sci* 47: 200–206.
- Ribaut JM, Ragot M (2007) Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *J Exper Bot* 58: 351–360.
- Szalma S, Hostert B, LeDeaux J, Stuber C, Holland J (2007) QTL mapping with near-isogenic lines in maize. *Theor Appl Genet* 114: 1211–1228.
- Falke K, Sušić Z, Hackauf B, Korzun V, Schondelmaier J, et al. (2008) Establishment of introgression libraries in hybrid rye (*Secale cereale* L.) from an Iranian primitive accession as a new tool for rye breeding and genomics. *Theor Appl Genet* 117: 641–652.
- Syed N, Pooni H, Mei M, Chen Z, Kearsley M (2004) Optimising the construction of a substitution library in *Arabidopsis thaliana* using computer simulations. *Mol Breeding* 13: 59–68.
- Falke KC, Miedaner T, Frisch M (2009) Selection strategies for the development of rye introgression libraries. *Theor Appl Genet* 119: 595–603.
- Welcker C, Sadok W, Dignat G, Renault M, Salvi S, et al. (2011) A common genetic determinism for sensitivities to soil water deficit and evaporative

- demand: meta-analysis of quantitative trait loci and introgression lines of maize. *Plant Physiol* 157: 718–729.
18. Belcher AR, Zwonitzer JC, Santa Cruz J, Krakowsky MD, Chung CL, et al. (2012) Analysis of quantitative disease resistance to southern leaf blight and of multiple disease resistance in maize, using near-isogenic lines. *Theor Appl Genet* 124: 433–445.
 19. Mano Y, Omori F (2013) Flooding tolerance in interspecific introgression lines containing chromosome segments from teosinte (*Zea nicaraguensis*) in maize (*Zea mays subsp. mays*). *Ann Bot-London* 112: 1125–1139.
 20. Lübberstedt T, Frei UK (2012) Application of doubled haploids for target gene fixation in backcross programmes of maize. *Plant Breeding* 131: 449–452.
 21. R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
 22. Stam P (1979) Interference in genetic crossing over and chromosome mapping. *Genetics* 92: 573–594.
 23. Mahone GS, Borchart D, Presterl T, Frisch M (2012) A comparison of tests for QTL mapping with introgression libraries containing overlapping and nonoverlapping donor segments. *Crop Sci* 52: 2198–2205.
 24. Falke K, Frisch M (2011) Power and false-positive rate in QTL detection with near-isogenic line libraries. *Heredity* 106: 576–584.