Plant Breeding, **0**, 1–9 (2012)

doi:10.1111/pbr.12021

Plant Breeding

Efficient marker-assisted backcross conversion of seed-parent lines to cytoplasmic male sterility

Eva Herzog[1] and Matthias Frisch[1,2]

[1]Institute of Agronomy and Plant Breeding II, Justus Liebig University, D-35392, Giessen, Germany; [2]Corresponding author, E-mail: matthias.frisch@uni-giessen.de

With 2 figures and 5 tables

Received July 2, 2012 / Accepted September 29, 2012
Communicated by H.-P. Piepho

Abstract

For many crops, cytoplasmic male sterility (CMS) is a cornerstone of hybrid production. Efficient conversion of elite lines to CMS by marker-assisted backcrossing is therefore desirable. In contrast to gene introgression, for which donor segments around target genes have to be considered, background selection for CMS conversion focuses solely on recovery of the recurrent parent genome. The optimal selection strategies for CMS conversion will consequently differ from those for gene introgression and have not yet been investigated. The objectives of our study were to evaluate and optimize the resource requirements of CMS conversion programmes and to determine the most cost-effective use of single-marker (SM) and high-throughput (HT) assays for this purpose. We conducted computer simulations for CMS conversion of genetic models of sugar beet, rye, sunflower and rapeseed. CMS conversion required fewer resources than gene introgression with respect to population size, marker data points and number of backcross generations. Combining HT assays in early backcross generations with SM assays in advanced backcross generations further increased the cost-efficiency of CMS conversion for a broad range of cost ratios.

Key words: cytoplasmic male sterility — simulation study — high-throughput markers — hybrid breeding — marker-assisted backcrossing

Cytoplasmic male sterility (CMS) in plants is a maternally inherited condition, which inhibits the production of functional pollen. It is mediated by plant mitochondrial genomes and the interaction of mitochondrial and nuclear genes (Chase 2007). In seed crops such as rye, sunflower, rice and rapeseed, CMS plus nuclear restoration of male-fertility in F_1 progeny is essential for large-scale production of hybrid seeds. CMS is a mainstay for hybrid breeding and seed production in sugar beet and rye (Hagihara et al. 2005, Tomerius et al. 2008). For some crops such as *Brassica oleracea*, where the use of CMS in hybrid breeding is a comparatively new system, conversion of existing elite lines to CMS is required. For rapeseed (*Brassica napus*), in which the genetic basis of adapted germplasm is relatively narrow (Gehringer et al. 2007), CMS conversion of newly developed lines is used after the introduction of new genetic variation into the breeding pool. Moreover, it has been recognized in maize and rice that cytoplasmic uniformity can lead to vulnerability to pathogens (Pring and Lonsdale 1989, Dalmacio et al. 1995). For such crops, it may be important to convert existing lines to newly identified CMS systems to reduce maternally inherited disease susceptibility.

New CMS donors used in early cycles of hybrid breeding programmes are often poorly adapted or wild relatives of cultivated crops (Hanson and Bentolila 2004). Complete recovery of the converted elite genotypes is therefore desirable. Typically, elite lines are selected as fertile maintainers and converted to CMS by backcrossing. As thousands of lines often are to be converted, breeders will seek to devote as little resources as possible to the conversion of a single line.

In commercial breeding programmes, dense marker maps are available for major crops. In combination with high-throughput (HT) marker systems based on single nucleotide polymorphisms (SNPs), they can speed up the backcross process by marker-assisted background selection (Gupta et al. 2010).

In the field of single-marker (SM) assays, the Competitive Allele Specific PCR (KASPar) assay has quite recently emerged. KASPar is a SNP detection system, which is cost-effective for genotyping small subsets of SNP markers. It enables the combined use of HT and SM assays for SNP genotyping at different stages in marker-assisted breeding programmes, given that a SNP set exists which is inter-convertible between KASPar and HT marker platforms (Chen et al. 2010, Mammadov et al. 2012). An advantage of HT assays is fast and cost-effective screening of large populations with a high number of marker data points. However, while with HT assays such as SNP chips, all markers need to be analysed in every backcross generation, SM assays allow for analysing only those marker loci which are not yet fixed for the desired alleles in advanced backcross generations. A combination of HT assays in early backcross generations with SM assays in advanced backcross generations has the potential to increase the cost-effectiveness of background selection for gene introgression (Herzog and Frisch 2011).

For gene introgression, background selection focuses on both reduction of donor segments around target genes and recurrent parent genome recovery. In contrast, in CMS conversion programmes, background selection solely focuses on fast and complete recurrent parent genome recovery. Moreover, as no preselection for target genes is conducted, all individuals from a backcross are subjected to background selection. This results in higher selection intensity and hence a greater selection response per backcross generation. However, it will also substantially increase the number of required marker data points. The optimal strategies for using molecular markers for CMS conversion will consequently differ from those for gene introgression and have not yet been investigated for major CMS crops. Depending on the genome size of a crop species, population size, marker density and use of HT and/or SM marker systems need to be optimized.

The goal of our study was to investigate, with computer simulations, CMS conversion in sugar beet, rye, sunflower and rape-

seed with two to four backcross generations. In particular, our objectives were (i) to assess recurrent parent genome recovery with different marker densities and to investigate the effect of increasing population size per backcross generation, (ii) to evaluate the resource requirements for recovering varying target levels of recurrent parent genome while minimizing the number of marker data points, and (iii) to determine the most efficient use of SM and HT assays for different cost ratios of HT/SM.

Material and Methods

Simulations were conducted assuming no interference in crossover formation. Each simulation was replicated 10 000 times to reduce sampling effects and to obtain results with high numerical accuracy and a small standard error. The 10% quantile (Q10), the arithmetic mean and the standard deviation of the probability distribution of the proportion of recipient genome in the entire genome of selected individuals (in percentage) were determined in every backcross generation to measure recurrent parent genome recovery.

Q10 values were included as they allow inferences about the probability to reach a certain level of recurrent parent genome. For example, a Q10 value of 96% can be interpreted as 'with a probability of 0.9 a recurrent parent genome proportion >96% can be achieved'. The arithmetic mean does not allow such probability inferences in advanced backcross generations, when the distribution of recurrent parent genome is getting more skewed.

We investigated four different genetic models that represent different crop species for which CMS is used in hybrid seed production. Model 1 represented sugar beet (*Beta vulgaris*) and cabbage (*B. oleracea*) and had n = 9 chromosomes of 100 cM length (cf. Weber et al. 1999, cf. Iniguez-Luy et al. 2009). Model 2 represented rye (*Secale cereale*) and had n = 7 chromosomes of 100 cM length (cf. Gustafson et al. 2009). Model 3 represented sunflower (*Helianthus annuus*) and had n = 17 chromosomes of 80 cM length (cf. Tang et al. 2002). Model 4 represented rapeseed (*B. napus*) and had n = 19 chromosomes of 140 cM length (cf. Piquemal et al. 2005). These models are hereafter referred to as sugar beet, rye, sunflower and rapeseed, respectively.

Markers for genome-wide background selection were assumed to be equally spaced. We considered different marker densities: two markers per chromosome (2M/chr), three markers per chromosome (3M/chr), as well as marker distances between two adjacent loci of 20, 10, 5 and 2 cM. For 2M/chr and 3M/chr, markers divided the chromosomes in 3 or 4 equal parts, respectively. For marker densities of 20, 10, 5 and 2 cM, the first marker and last marker of each chromosome were placed on the telomeres.

Each backcross scheme started by crossing two homozygous parents (CMS donor and recipient), which were polymorphic at all loci. The CMS recipient carried the desirable alleles at all loci of the genome, while the donor carried no desirable alleles. The CMS recipient was assumed to be a fertile maintainer. An F_1 individual was created by crossing CMS donor and recipient. This F_1 individual was backcrossed to the recipient to create n_1 BC_1 individuals. The n_1 BC_1 individuals were subjected to genome-wide background selection. A selection index $i = \Sigma_m x_m$ was constructed, where summation is over markers and x_m is the number of recurrent parent alleles at the mth marker. The plant with the highest value of i was selected and backcrossed to the recipient. For each of the four genetic models, we investigated two to four backcross generations t and constant population sizes of n_t ranging from 10 to 200 individuals.

For calculating the relative costs of different marker strategies, the resource requirements for target Q10 values of 96% in generation BC_2 and 99% in generation BC_3 with a marker density of 10 cM were determined. One HT assay included genotyping one individual for all markers on the linkage map. One SM assay corresponded to one locus and thus one marker data point. For estimating the total number of required marker data points for SM assays, only marker loci not yet fixed for the recipient allele were analysed in advanced backcross generations. We took into account cost ratios of HT : SM of 200 : 1, 100 : 1, 50 : 1, 20 : 1 and 10 : 1. To give an example in absolute costs, a cost ratio of HT/SM of 100 : 1 corresponded to costs of € 50 for analysing all SNP background marker loci with a SNP chip, and € 0.5 for analysing one SNP marker locus with a KASPar assay. We compared the costs of using only HT assays in all generations of the backcross conversion programme (strategy HT) to the costs of using only SM assays in all generations of the backcross conversion programme (strategy SM). In this case, the costs for strategy SM were set to 1.

In addition, for two-generation programmes, we compared the costs of a combined strategy that relied on HT assays in generation BC_1 and SM assays in generation BC_2 (strategy Combined A) to the costs of strategy SM and strategy HT. In this case, the costs of strategy SM and strategy HT were set to 1, respectively. For three-generation programmes, we compared the costs of a strategy using HT assays in generation BC_1 and SM assays in generations BC_2 and BC_3 (strategy Combined B) to the costs of strategy HT. We also compared the costs of a strategy using HT assays in generations BC_1 and BC_2, and SM assays in generation BC_3 (strategy Combined C) to the costs of strategy HT. In both cases, the costs for strategy HT were set to 1.

Results

For a marker density of 20 cM and constant population sizes of n_t = 40, 80, 120, 160, 200 individuals per backcross generation, the Q10 values recovered in generations BC_1 and BC_2 were higher for genetic models with shorter genomes (Tables 1–4). Q10 values for rye were 2.9–3.8% higher than for rapeseed, while for sugar beet and sunflower, intermediate Q10 values were recovered. The differences in Q10 values between the genetic models that were observed in generation BC_2 diminish in advanced backcross generations.

Genetic models with shorter genomes had fewer and shorter fragments of donor genome in generations BC_1 and BC_2 (Tables 1–4). In generations BC_1 and BC_2, the length of donor fragments is decreasing to a greater extent with increasing n_t in genetic models with shorter genomes. The average length of donor fragments is decreasing by about 39% in sugar beet, 30% in rye, 28% in sunflower and 20% in rapeseed if n_t is increased from 40 to 200 individuals in generation BC_2. The average length of donor fragments ranged between 32% and 46% of marker distance for rye and 88–110% for rapeseed in generation BC_2. In advanced backcross generations, the differences in the length of donor fragments between population sizes and genetic models diminish considerably.

Genetic models with shorter genomes required fewer marker data points (Tables 1–4). For a population size of n_t = 200, rapeseed required about four times as many marker data points as rye, about three times as many marker data points as sugar beet, and about twice as many marker data points as sunflower. For all genetic models, the major proportion of marker data points had to be analysed in generations BC_1 and BC_2. For example, for sugar beet, 98.2–99.8% of marker analyses had to be conducted in generations BC_1 and BC_2. From generation BC_3 –BC_4, marker data points were no longer or only marginally increasing, indicating complete fixation. This also held true for higher marker densities of 10, 5 and 2 cM (data not shown).

For all four genetic models, Q10 values of >90% could be recovered in generation BC_2 with low marker densities of 2M/chr or 3M/chr and n_t = 10–20 individuals per backcross generation (Fig. 1). Q10 values increased considerably for all four investigated genetic models when population size was increased from n_t = 10 to n_t = 40–50 individuals per backcross generation, irrespective of marker density.

Table 1: Sugar beet: recovered proportion of recurrent parent genome (Q10, $\bar{x}$, s_x), required number of marker data points (MDP) for single-marker assays, number of donor fragments ($\bar{x}$, s_x) and length of donor fragments in cM ($\bar{x}$, s_x) in generations BC_1–BC_4 with genome-wide background selection with constant population sizes n_t = 40, 80, 120, 160, 200 and equally spaced markers (marker density 20 cM) (Note that the number of required high-throughput assays can be easily obtained by multiplying n_t by the number of backcross generations)

		Recurrent parent genome (%)				No. of donor fragments		Length of donor fragments (cM)	
n_t	BC_t	Q_{10}	$\bar{x}$	s_x	MDP	$\bar{x}$	s_x	$\bar{x}$	s_x
40	BC_1	84.56	88.08	2.81	2160	6.48	1.71	33.10	28.16
	BC_2	97.23	98.71	1.07	2682	2.00	1.31	11.60	10.20
	BC_3	99.32	99.80	0.32	2732	0.59	0.78	6.01	4.49
	BC_4	99.57	99.90	0.23	2732	0.32	0.58	5.70	4.36
80	BC_1	86.30	89.60	2.61	4320	6.06	1.68	30.90	26.79
	BC_2	98.19	99.27	0.75	5233	1.44	1.16	9.19	7.75
	BC_3	99.34	99.82	0.31	5276	0.54	0.74	6.08	4.56
	BC_4	99.62	99.91	0.22	5276	0.28	0.55	5.79	4.38
120	BC_1	87.32	90.34	2.45	6480	5.86	1.65	29.65	25.95
	BC_2	98.58	99.46	0.62	7754	1.20	1.09	8.06	6.63
	BC_3	99.38	99.83	0.30	7790	0.50	0.72	6.02	4.55
	BC_4	99.65	99.92	0.21	7790	0.27	0.53	5.64	4.38
160	BC_1	87.98	90.89	2.36	8640	5.71	1.64	28.71	25.25
	BC_2	98.81	99.55	0.55	10245	1.10	1.05	7.34	5.91
	BC_3	99.40	99.83	0.30	10273	0.51	0.73	5.89	4.42
	BC_4	99.65	99.91	0.21	10273	0.27	0.54	5.59	4.36
200	BC_1	88.46	91.30	2.29	10800	5.60	1.63	27.96	24.76
	BC_2	98.92	99.60	0.51	12723	1.01	1.01	7.07	5.58
	BC_3	99.40	99.84	0.29	12745	0.49	0.71	5.93	4.47
	BC_4	99.65	99.92	0.21	12745	0.27	0.53	5.65	4.35

Table 2: Rye: recovered proportion of recurrent parent genome (Q10, $\bar{x}$, s_x), required number of marker data points (MDP), number of donor fragments ($\bar{x}$, s_x) and length of donor fragments ($\bar{x}$, s_x) in generations BC_1–BC_4 with genome-wide background selection with constant population sizes n_t = 40, 80, 120, 160, 200 and equally spaced markers (marker density 20 cM) (Note that the number of required high-throughput assays can be easily obtained by multiplying n_t by the number of backcross generations)

		Recurrent parent genome (%)				No. of donor fragments		Length of donor fragments (cM)	
n_t	BC_t	Q_{10}	$\bar{x}$	s_x	MDP	$\bar{x}$	s_x	$\bar{x}$	s_x
40	BC_1	85.75	89.73	3.15	1680	4.71	1.51	30.50	26.45
	BC_2	98.05	99.28	0.87	2031	1.11	1.04	9.11	7.80
	BC_3	99.31	99.82	0.35	2047	0.42	0.66	5.93	4.47
	BC_4	99.66	99.91	0.24	2047	0.22	0.48	5.46	4.31
80	BC_1	87.79	91.35	2.85	3360	4.32	1.45	28.01	24.64
	BC_2	98.79	99.60	0.58	3953	0.78	0.89	7.23	5.79
	BC_3	99.35	99.84	0.33	3961	0.38	0.63	5.88	4.43
	BC_4	99.69	99.92	0.23	3961	0.20	0.46	5.57	4.34
120	BC_1	88.82	92.21	2.67	5040	4.14	1.41	26.34	23.47
	BC_2	98.93	99.66	0.51	5844	0.72	0.87	6.59	4.99
	BC_3	99.38	99.85	0.33	5847	0.36	0.62	5.89	4.50
	BC_4	99.71	99.92	0.23	5847	0.20	0.46	5.53	4.37
160	BC_1	89.55	92.78	2.56	6720	3.98	1.42	25.41	22.59
	BC_2	99.01	99.69	0.47	7715	0.68	0.83	6.40	4.72
	BC_3	99.39	99.85	0.32	7716	0.36	0.61	5.87	4.49
	BC_4	99.73	99.93	0.22	7716	0.18	0.45	5.60	4.31
200	BC_1	90.14	93.18	2.44	8400	3.88	1.38	24.58	21.87
	BC_2	99.02	99.70	0.46	9576	0.65	0.82	6.39	4.67
	BC_3	99.37	99.85	0.32	9577	0.35	0.60	6.09	4.54
	BC_4	99.72	99.93	0.22	9577	0.18	0.43	5.76	4.38

For all genetic models, Q10 values of $\geq$96% could be reached in generation BC_2. Minimum required marker densities for a Q10 value of 96% were 3M/chr for sugar beet and rye, 2M/chr for sunflower and 20 cM for rapeseed. For sugar beet and rye, there was a limit of recurrent parent genome that could be recovered, indicated by a plateau in the Q10 curves for marker densities of 2M/chr, 3M/chr and 20 cM. The population sizes per backcross generation for which the limit was reached depended on marker density and lay between n_t = 70–200 for sugar beet and between n_t = 50–150 for rye. For sunflower and rapeseed, the plateau was not reached with the highest investigated population size of n_t = 200.

The differences in Q10 values between marker densities were bigger in sugar beet and rye than in sunflower and rapeseed (Fig. 1). For example, for n_t = 100 individuals per backcross generation, the differences in Q10 values between a marker density of 3M/chr and 20 cM were 1.7% for sugar beet, 2.3% for rye, 0.4% for sunflower and 1.0% for rapeseed. The maximum

Table 3: Sunflower: recovered proportion of recurrent parent genome (Q10, $\bar{x}$, s_x), required number of marker data points (MDP), number of donor fragments ($\bar{x}$, s_x) and length of donor fragments ($\bar{x}$, s_x) in generations BC_1–BC_4 with genome-wide background selection with constant population sizes n_t = 40, 80, 120, 160, 200 and equally spaced markers (marker density 20 cM) (Note that the number of required high-throughput assays can be easily obtained by multiplying n_t by the number of backcross generations)

		Recurrent parent genome (%)				No. of donor fragments		Length of donor fragments (cM)	
n_t	BC_t	Q_{10}	$\bar{x}$	s_x	MDP	$\bar{x}$	s_x	$\bar{x}$	s_x
40	BC_1	82.19	85.06	2.31	3400	11.81	2.17	34.39	26.21
	BC_2	95.46	97.08	1.24	4424	4.94	1.83	16.11	14.47
	BC_3	99.20	99.69	0.36	4620	1.23	1.14	6.84	5.53
	BC_4	99.57	99.87	0.21	4626	0.59	0.80	5.75	4.42
80	BC_1	83.61	86.23	2.12	6800	11.35	2.13	32.99	25.73
	BC_2	96.45	97.84	1.04	8692	4.17	1.69	14.12	12.53
	BC_3	99.37	99.77	0.29	8976	1.01	1.03	6.24	4.63
	BC_4	99.59	99.89	0.20	8976	0.53	0.75	5.82	4.41
120	BC_1	84.37	86.88	2.03	10200	11.05	2.13	32.30	25.48
	BC_2	96.94	98.20	0.94	12904	3.76	1.64	13.01	11.44
	BC_3	99.40	99.78	0.28	13249	0.98	1.00	6.10	4.49
	BC_4	99.60	99.89	0.19	13249	0.52	0.75	5.63	4.35
160	BC_1	84.93	87.36	1.98	13600	10.84	2.11	31.72	25.24
	BC_2	97.29	98.45	0.86	17076	3.47	1.60	12.15	10.65
	BC_3	99.41	99.79	0.27	17464	0.94	1.00	6.06	4.59
	BC_4	99.62	99.90	0.19	17464	0.49	0.71	5.75	4.45
200	BC_1	85.30	87.68	1.95	17000	10.72	2.07	31.26	25.02
	BC_2	97.52	98.61	0.81	21244	3.24	1.55	11.64	10.04
	BC_3	99.41	99.79	0.27	21670	0.93	0.98	6.04	4.51
	BC_4	99.61	99.89	0.19	21670	0.49	0.72	5.81	4.45

Table 4: Rapeseed: recovered proportion of recurrent parent genome (Q10, $\bar{x}$, s_x), required number of marker data points (MDP), number of donor fragments ($\bar{x}$, s_x) and length of donor fragments ($\bar{x}$, s_x) in generations BC_1–BC_4 with genome-wide background selection with constant population sizes n_t = 40, 80, 120, 160, 200 and equally spaced markers (marker density 20 cM) (Note that the number of required high-throughput assays can be easily obtained by multiplying n_t by the number of backcross generations)

		Recurrent parent genome (%)				No. of donor fragments		Length of donor fragments (cM)	
n_t	BC_t	Q_{10}	$\bar{x}$	s_x	MDP	$\bar{x}$	s_x	$\bar{x}$	s_x
40	BC_1	81.07	83.37	1.89	6080	19.60	2.87	45.13	38.58
	BC_2	94.19	95.65	1.11	8116	10.57	2.60	21.92	20.56
	BC_3	98.68	99.29	0.44	8644	3.69	1.86	10.27	9.21
	BC_4	99.64	99.86	0.16	8712	1.28	1.17	5.75	4.46
80	BC_1	82.19	84.36	1.78	12160	19.16	2.86	43.45	37.65
	BC_2	95.12	96.39	0.99	15992	9.59	2.53	20.00	18.78
	BC_3	99.07	99.54	0.34	16861	2.88	1.71	8.42	7.34
	BC_4	99.66	99.87	0.15	16926	1.20	1.13	5.74	4.39
120	BC_1	82.86	84.89	1.69	18240	18.87	2.87	42.59	37.14
	BC_2	95.62	96.79	0.91	23797	9.09	2.46	18.77	17.61
	BC_3	99.24	99.64	0.30	24955	2.53	1.64	7.47	6.38
	BC_4	99.67	99.88	0.15	25011	1.15	1.11	5.70	4.36
160	BC_1	83.26	85.25	1.64	24320	18.71	2.84	41.93	36.68
	BC_2	95.91	97.04	0.86	31551	8.64	2.38	18.23	17.00
	BC_3	99.34	99.70	0.26	32970	2.29	1.55	6.96	5.68
	BC_4	99.68	99.88	0.15	33014	1.09	1.09	5.77	4.40
200	BC_1	83.58	85.51	1.58	30400	18.58	2.84	41.50	36.50
	BC_2	96.16	97.22	0.82	39284	8.39	2.37	17.64	16.35
	BC_3	99.40	99.73	0.24	40941	2.18	1.53	6.63	5.33
	BC_4	99.68	99.88	0.14	40974	1.09	1.10	5.69	4.36

Q10 values recovered in generation BC_2 depended on genome length and were 99.60% for sugar beet, 99.99% for rye, 98.09% for sunflower and 96.50% for rapeseed. Increasing marker density from 10 to 5 or 2 cM did not substantially increase Q10 values. This held true for all investigated genetic models. Moreover, marker densities of 5 and 2 cM incurred very high numbers of marker data points (data not shown).

The optimum designs that minimized the required number of marker data points for target Q10 values of 96–99% in generation BC_2 employed marker densities of 2M/chr–10 cM for sugar beet and rye (Table 5). For the sunflower model, a Q10 value of 98% could only be reached with a marker density of 5 cM and 68 000 marker data points. For rapeseed, a Q10 value of 96% in generation BC_2 could only be reached with a marker density of 20 cM and about 33 000 marker data points. Higher target Q10 values could not be reached in generation BC_2 for this model. For all four genetic models, two-generation programmes incurred substantially more marker data points than

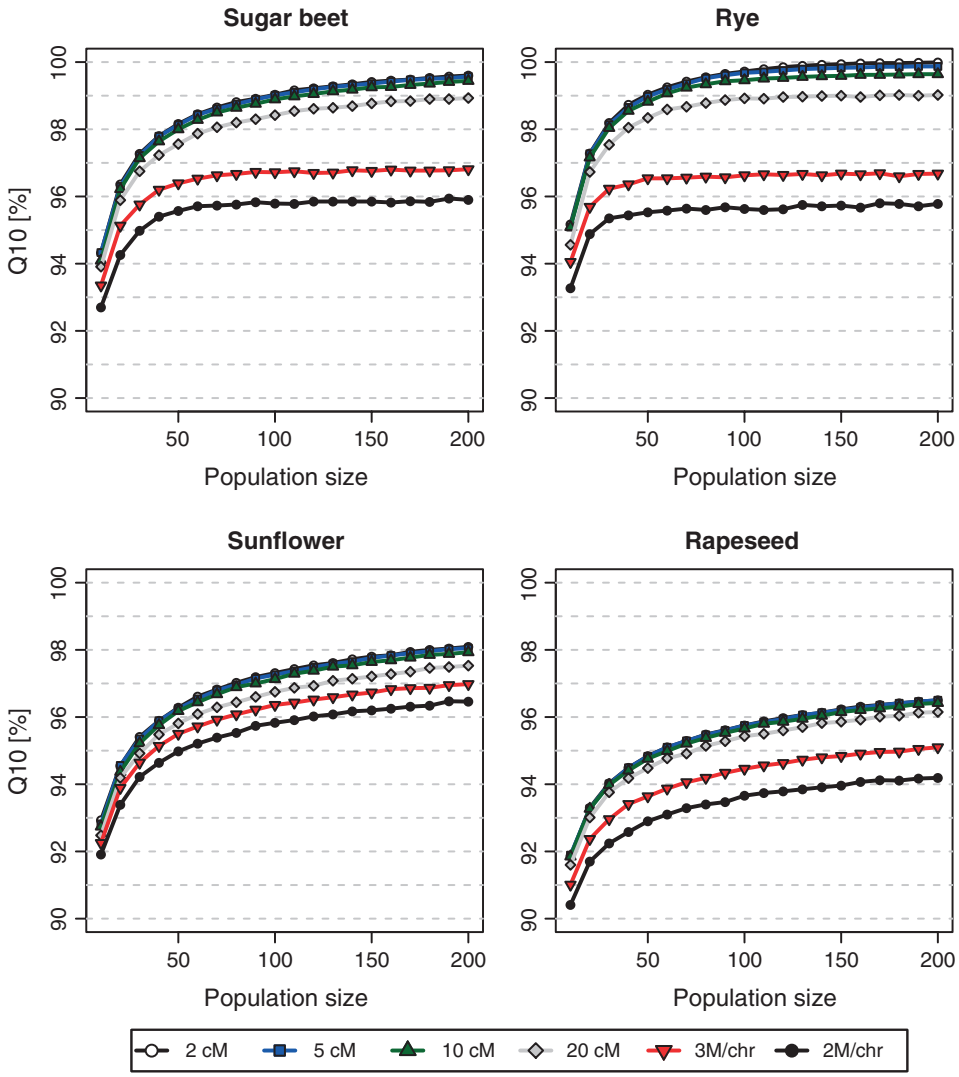

Fig. 1: Q10 values recovered in generation BC_2 with genome-wide background selection (marker densities 2M/chr, 3M/chr and 2, 5, 10, 20 cM) with constant population sizes n_t = 20–300 and equally spaced markers for four genetic models

three-generation programmes. The shorter genomes of sugar beet and rye required 3.6–11.2 times as many marker data points for two-generation programmes as for three-generation programmes. For sunflower and rapeseed, two-generation programmes required 28.8–60.2 times as many marker data points as three-generation programmes.

For two-generation programmes, strategy HT was 0.05–4.10 times as expensive as strategy SM for recovering a target Q10 value of 96%, depending on genetic model and cost ratio of HT : SM (Fig. 2a). Strategy Combined A was more cost-effective than strategy HT, indicated by the smaller range of relative costs and the smaller slopes of the cost curves (Fig. 2b). Which marker strategy was cheapest depended on the cost ratio of HT : SM and the genetic model. For sugar beet, strategy HT was the cheapest strategy for cost ratios of HT : SM of 10 : 1–35 : 1 (Fig. 2c). For cost ratios ranging between 35 : 1 and 100 : 1, strategy Combined A was cheapest (Fig. 2b). For cost ratios of HT : SM >100 : 1, strategy SM was cheapest. If the choice was between either strategy HT or strategy SM, strategy HT should be used for cost ratios of HT : SM of 10 : 1–60 : 1 (Fig. 2a). For longer genomes, using HT assays became relatively cheaper compared with SM assays. For rapeseed, strategy HT was the cheapest strategy for recovering a target Q10 value of 96% for cost ratios of HT : SM of up to 190 : 1 (Fig. 2a). For cost ratios >190 : 1, strategy Combined A was cheapest (Fig. 2b). Strategy SM was never cheaper than strategy Combined A.

For three-generation programmes with a target Q10 value of 99%, the use of HT assays became less efficient compared with SM assays (Fig. 2d), indicated by steeper cost curves. Strategy Combined C was equivalent to or cheaper than strategy HT for nearly all investigated scenarios (Fig. 2f).

Discussion

Genetic models

Computer simulations and model calculations are considered robust and useful tools for the optimization of breeding programmes (Prigge et al. 2008, Tomerius et al. 2008). However, the validity of simulated results for real breeding applications is influenced by the theoretical assumptions for the underlying genetic model.

We used a Poisson procedure for modelling crossover formation during meiosis, assuming no interference in crossover formation as

Table 5: Optimum designs for recovering Q10 values of 96–99% [marker density, population size, no. of backcross generations, no. of marker data points (MDP)] in two vs. three backcross generations with genome-wide background selection if the number of MDP is minimized (Note that the number of required high-throughput assays can be easily obtained by multiplying n_t by the number of backcross generations)

Genetic model	Q10 (%)	n_t	Marker density	No. of MDP
No. of BC generations = 2				
Sugar beet	96	40	3M/chr	1305
9 × 100 cM	97	40	20 cM	2684
	98	70	20 cM	4600
	99	120	10 cM	14 172
Rye	96	30	3M/chr	749
7 × 100 cM	97	30	20 cM	1541
	98	40	20 cM	2030
	99	60	10 cM	5475
Sunflower	96	120	2M/chr	5059
17 × 80 cM	97	130	20 cM	13 948
	98	190	5 cM	68 411
	99	–	–	–
Rapeseed	96	170	20 cM	33 476
19 × 140 cM	97	–	–	–
	98	–	–	–
	99	–	–	–
No. of BC generations = 3				
Sugar beet	96	–	–	–
9 × 100 cM	97	10	2M/chr	240
	98	10	20 cM	746
	99	20	20 cM	1421
Rye	96	–	–	–
7 × 100 cM	97	10	2M/chr	181
	98	10	20 cM	563
	99	10	10 cM	1028
Sunflower	96	–	–	–
17 × 80 cM	97	10	2M/chr	485
	98	20	3M/chr	1394
	99	30	20 cM	3513
Rapeseed	96	10	2M/chr	556
19 × 140 cM	97	20	3M/chr	1618
	98	20	20 cM	4448
	99	80	20 cM	16 863

proposed by Haldane (1919). This approach has the advantage of applicability for a broad range of scenarios, as has been discussed in detail in the study by Frisch and Melchinger (2001). Further necessary simplifications for the sake of generality include the assumptions of perfect fertility, no natural selection at gamete or zygote level, unchanged recombination frequencies and Mendelian segregation in any cross. This will not hold true in all cases, especially if CMS donors are unadapted wild relatives. For such wide crosses, the simulations might underestimate the actual resource requirements and/or overestimate recovered Q10 values. On the other hand, in advanced cycles of hybrid breeding programmes, adapted lines often are available as CMS donors, which might be similar to the recipient lines. In these cases, complete recovery of an elite genotype might be achieved with less resources or in shorter time.

The reader should be aware that the presented simulation approach does not cover every detail of the complex biological processes, which might underlie any specific cross. Conclusions drawn from simulated data should therefore be interpreted as guidelines and might require adjustment in specific breeding programmes.

Population size

In a simulation study on the introgression of one dominant target gene, Prigge et al. (2009) employed the same genetic model for sugar beet that was used in the present study. With a marker density of 20 cM and n_t = 40–200 individuals per backcross generation, they recovered Q10 values in generation BC_2 that were approximately 3–4% lower than in the present study (Table 1). The greater selection response in CMS conversion can be explained by the lack of preselection for the target gene and the lack of donor genome attached to the target gene. Consequently, CMS conversion required considerably smaller population sizes than gene introgression.

In generation BC_2, Q10 values increased considerably for all four genetic models when population size was increased from n_t = 10 to n_t = 40–50 individuals (Fig. 1). For sugar beet and rye, a plateau in the Q10 curves was observed. This limit of recurrent parent genome recovery is caused by the limited estimation accuracy of a given marker density. The wider adjacent markers are spaced, the more likely it is that segments of recurrent parent genome between markers go unnoticed. Sugar beet and rye had fewer and shorter donor fragments in generation BC_2, which were still considerably decreased with increasing population size n_t (Tables 1–4). For rye, for which the plateau is reached at n_t = 120 with a marker density of 20 cM, the average length of donor fragments is only 6.59 cM and consequently only about 33% of the distance between two adjacent markers (Table 2). As a consequence, the plateau is reached with smaller population sizes for lower marker densities. Increasing population size beyond the number of individuals for which the plateau is reached (Fig. 1) is not economic.

We conclude that recurrent parent genome recovery is maximized for all four genetic models with population sizes of $n_t \geq$ 40–50 individuals per backcross generation. For rye and sugar beet, population sizes should not exceed n_t = 50–150 and n_t = 70–200 individuals, respectively, depending on marker density. For sunflower and rapeseed, population sizes of $n_t > 200$ still have positive effects.

Marker density

It has been estimated for backcross programmes that a target Q10 value of at least 96% should minimize the risk of undesirable effects from unadapted donor genome (Prigge et al. 2009). For sugar beet and rye, a Q10 level of about 96% could be recovered in generation BC_2 with a marker density of 3M/chr and n_t = 40–60 individuals per backcross generation (Fig. 1). For sunflower, the Q10 value of 96% could be reached with a marker density of 2M/chr, indicating that two markers per chromosome are sufficient for controlling short chromosomes (Fig. 1). We therefore conclude that for CMS conversion, a threshold Q10 value of 96% in generation BC_2 can in most cases be reached with 2–3 markers per chromosome.

The differences in Q10 values between marker densities were bigger in sugar beet and rye than in sunflower and rapeseed (Fig. 1). For example, for a population size per backcross generation of n_t = 100, the differences in Q10 values between a marker density of 3M/chr and 20 cM were 1.7% for sugar beet, 2.3% for rye, 0.4% for sunflower and 1.0% for rapeseed. If marker density was increased from 3M/chr to 20 cM, the increase in the number of markers per chromosome was greater in genetic models with longer chromosomes, which partly accounts for the big gap in Q10 values. Moreover, increasing marker density shifted the frequency distribution of recurrent parent genome to the right and decreased the variance of the distribution in all four genetic models. The extent of these changes depended on chromosome number and length. The differences between marker densities were bigger for genetic models with a lower number of

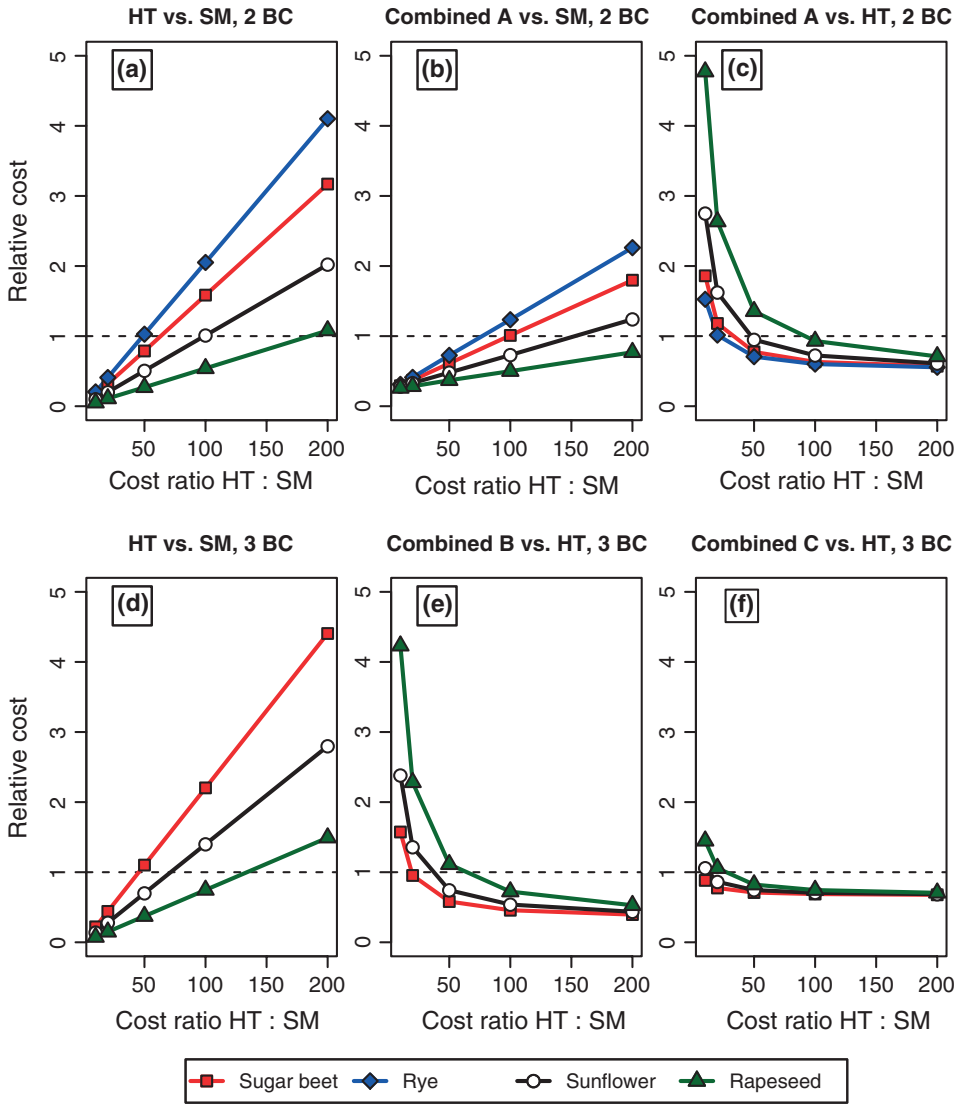

Fig. 2: Relative cost of different strategies of marker analysis plotted against the cost ratio of one high-throughput assays to one single-marker assay (Cost ratio HT : SM) for recovering Q10 values of 96% in generation BC_2, and Q10 values of 99% in generation BC_3 with a marker density of 10 cM HT, a strategy using only high-throughput assays in all backcross generations; SM, a strategy using only single-marker assays in all backcross generations; Combined A, a strategy using HT assays in generation BC_1 and SM assays in generation BC_2; Combined B, a strategy using HT assays in generation BC_1 and SM assays in generations BC_2 and BC_3; Combined C, a strategy using HT assays in generations BC_1 and BC_2 and HT assays in generation BC_3. HT, high-throughput; SM, single-marker

chromosomes. If chromosome number was comparable, the differences were bigger for genetic models with longer chromosomes. For sunflower, 2–3 equally spaced markers per chromosome seemed sufficient to get acceptable genome coverage for recurrent parent genome recovery. For rapeseed, sugar beet and rye, at least 6–8 equally spaced markers per chromosome, corresponding to a marker density of 20 cM, provide an adequate selection response.

Given that differences in Q10 values between marker densities were bigger (Fig. 1) and donor fragments on average shorter (Tables 1–4) in genetic models with shorter genomes, we conclude that it pays off more to invest in higher marker densities for sugar beet and rye than for sunflower and rapeseed.

For all four genetic models, hardly any differences in Q10 values could be observed between marker densities of 2, 5 and 10 cM (Fig. 1). However, marker densities of 5 and 2 cM incurred very high numbers of marker data points (data not shown). This was also observed in a previous simulation study on gene introgression (Herzog and Frisch 2011). The reason is that selection response is not limited by precise estimation of the genetic contribution of the recurrent parent, but by the limited number of recombination events that occur in two- or three-generation backcross programmes. We therefore conclude that it is not efficient to increase effective marker density beyond 10 cM, even though marker maps with higher density are available for SNPs.

Marker fixation

For all four genetic models, the major proportion of marker data points was incurred in generations BC_1 and BC_2. From generation BC_3–BC_4, the number of required marker data points is only marginally increasing (Tables 1–4). Accordingly, the population sizes at which the plateau of recurrent parent genome recovery is reached are diminishing in generations BC_3 and BC_4 due to marker fixation (data not shown). As a consequence, dif-

ferences in Q10 values and average length of donor fragments between the four genetic models disappear in generation BC_4. This indicates that recurrent parent genome recovery was no longer controlled for by markers and resulted in a reduction in selection response.

For gene introgression, Prigge et al. (2009) reported that the optimum backcross designs were characterized by increasing marker densities and population sizes. Due to the faster rate of marker fixation in CMS conversion programmes, we conclude that keeping a constant population size in each backcross generation, or increasing population size in advanced backcross generations, is only efficient for CMS conversion if it is also accompanied by an increase in marker density. Additional markers could be placed between the original markers analysed in previous generations to increase the precision of selection. For sugar beet, rye and sunflower, marker densities of 3M/chr in generations BC_1 and BC_2, and 20 cM in advanced backcross generations could decrease the loss of selection response. For rapeseed, we suggest that CMS conversion programmes could start with 20 cM in generations BC_1 and BC_2, followed by 10 cM in advanced backcross generations.

CMS conversion designs for different genetic models

In the present study, Q10 values of 96–98% could be reached in generation BC_2 for sugar beet and rye with a marker density of 20 cM and n_t = 30–70 individuals (Table 5). We therefore conclude that for these crops, two-generation programmes are suitable for CMS conversion.

If Q10 values >96% were aimed for in generation BC_2, sunflower required n_t = 130–190 individuals per backcross generation and marker densities of 20–5 cM. Moreover, a target Q10 value of 98% in generation BC_2 required about 68 000 marker data points (Table 5). For rapeseed, a Q10 value of 96% in generation BC_2 could only be reached with n_t = 170 individuals per backcross generation, a marker density of 20 cM, and about 33 000 marker data points. We conclude that for target Q10 values of 96–99%, three-generation conversion programmes are required for the longer genomes of sunflower and rapeseed.

With the exception of a Q10 value of 98% for sunflower, all Q10 levels could be reached with marker densities of 2M/chr–10 cM. Increasing marker density beyond 10 cM incurs high numbers of marker data points, but will not help to save additional backcross generations (cf. Fig. 1). This confirms that a marker density of 10 cM is sufficient for almost all backcross designs, as has also been previously observed (Herzog and Frisch 2011).

For all four genetic models, two-generation programmes required considerably more marker data points than the three-generation programmes (Table 5). We therefore conclude that three-generation CMS conversion programmes are also advantageous for shorter genomes if the focus of cost reduction is on the cost of marker analysis.

Relative costs of HT and SM assays

Different strategies of using HT and SM assays for CMS conversion with a marker density of 10 cM were compared by calculating their relative costs for cost ratios of HT/SM ranging from 200 : 1 to 10 : 1 (Fig. 2). For a Q10 value of 96% in generation BC_2, the relative costs of strategy HT compared with strategy SM ranged from 0.10 to 2.02 for sunflower (Fig. 2a). In a gene introgression study on maize with the same parameters, the relative costs ranged from 0.09 to 1.85 (Herzog and Frisch 2011). These genetic models are comparable with respect to genome length (1360 vs. 1600 cM) and number of background markers. For a given population size, the number of SM and HT assays are approximately in the same ratio for gene introgression and CMS conversion in generations BC_1 and BC_2. It can therefore be assumed that the relative costs we determined in the present study are to a certain extent also valid for background selection in gene introgression programmes.

For sugar beet and a target Q10 value of 96% in generation BC_2, strategy HT was cheapest up to a cost ratio of HT/SM of 35 : 1 (Fig. 2c). From a cost ratio of HT/SM of 35 : 1–100 : 1, strategy Combined A was cheapest (Fig. 2b). For higher cost ratios of HT/SM, strategy SM was the cheapest option. If the choice is between either strategy HT or strategy SM, strategy HT should be used up to a cost ratio of HT/SM of 60 : 1 (Fig. 2a). For sunflower and rapeseed, strategies involving HT assays became relatively cheaper. We therefore conclude that the use of HT assays for background selection is cost-efficient for two-generation CMS conversion programmes and crops with long genomes such as sunflower and rapeseed.

For three-generation programmes, strategy HT became less efficient compared with strategy SM, indicated by steeper cost curves (Fig. 2d). For sugar beet and a target Q10 value of 99% in generation BC_3, strategy HT was only cheaper than strategy SM up to a cost ratio of HT/SM of 45 : 1. This can be explained by the fact that for sugar beet, 98–99% of marker data points are incurred in generations BC_1 and BC_2 and most markers are already fixed in generation BC_3 (Tables 1–4).

For three-generation programmes, strategy Combined C was equivalent to or cheaper than strategy HT for nearly all investigated scenarios (Fig. 2f). Combining HT and SM assays in one backcross programme can pose a challenge as HT and SM platforms often require different types of markers. Recently, KASPar assays have become available, which allow for inexpensive analysis of small sets of SNPs (Chen et al. 2010). It has been shown that SNP markers can be inter-converted between KASPar and HT assays (Mammadov et al. 2012). Combinations of HT and SM thus have the potential to make marker-assisted background selection more cost-effective. We conclude that for three-generation CMS conversion programmes, HT assays should be used in generations BC_1 and BC_2, and SM assays in generation BC_3 for all investigated genetic models.

Acknowledgements

We thank the anonymous reviewers and the editor for their helpful suggestions. We greatly appreciated the comments of one reviewer that considerably improved the manuscript. We thank Gregory Mahone for proof-reading the manuscript.

References

Chase, C. D., 2007: Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet. **23**, 81—90.

Chen, W., J. Mingus, J. Mammadov, J. E. Backlund, T. Greene, S. Thompson, and S. Kumpatla, 2010: KASPar: a simple and cost-effective system for SNP genotyping. In: Final program, abstract and exhibit guide of the XVIII international conference on the status of plant and animal genome research, San Diego, CA, 9–13 Jan 2010.

Dalmacio, R., D. S. Brar, T. Ishii, L. A. Sitch, S. S. Virmani, and G. S. Khush, 1995: Identification and transfer of a new cytoplasmic male sterility source from *Oryza perennis* into indica rice (*O. sativa*). Euphytica **82**, 221—225.

Frisch, M., and A. E. Melchinger, 2001: Length of the intact donor chromosome segment around a target gene in marker-assisted backcrossing. Genetics **157**, 1343—1356.

Gehringer, A., R. Snowdon, T. Spiller, P. Basunanda, and W. Friedt, 2007: New oilseed rape (*Brassica napus*) hybrids with high levels of heterosis for seed yield under nutrient-poor conditions. Breed. Sci. **57**, 315—320.

Gupta, P. K., P. Langridge, and R. R. Mir, 2010: Marker-assisted wheat breeding: present status and future possibilities. Mol. Breeding **26**, 145—161.

Gustafson, J. P., X. F. Ma, V. Korzun, and J. W. Snape, 2009: A consensus map of rye integrating mapping data from five mapping populations. Theor. Appl. Genet. **118**, 793—800.

Hagihara, E., N. Itchoda, Y. Habu, S. Iida, T. Mikami, and T. Kubo, 2005: Molecular mapping of a fertility restorer gene for Owen cytoplasmic male sterility in sugar beet. Theor. Appl. Genet. **111**, 250—255.

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

Hanson, M. R., and S. Bentolila, 2004: Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell **16**, S154—S169.

Herzog, E., and M. Frisch, 2011: Selection strategies for marker-assisted backcrossing with high-throughput marker systems. Theor. Appl. Genet. **123**, 251—260.

Iniguez-Luy, F. L., L. Lukens, M. W. Farnham, R. M. Amasino, and T. C. Osborn, 2009: Development of public immortal mapping populations, molecular markers and linkage maps for rapid cycling *Brassica rapa* and *B. oleracea*. Theor. Appl. Genet. **120**, 31—43.

Mammadov, J., W. Chen, J. Mingus, S. Thompson, and S. Kumpatla, 2012: Development of versatile gene-based SNP assays in maize (*Zea mays* L.). Mol. Breeding **29**, 779—790.

Piquemal, J., E. Cinquin, F. Couton, C. Rondeau, E. Seignoret, I. Doucet, D. Perret, M. J. Villeger, P. Vincourt, and P. Blanchard, 2005: Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theor. Appl. Genet. **111**, 1514—1523.

Prigge, V., H. P. Maurer, D. J. Mackill, A. E. Melchinger, and M. Frisch, 2008: Comparison of the observed with the simulated distributions of the parental genome contribution in two marker-assisted backcross programs in rice. Theor. Appl. Genet. **116**, 739—744.

Prigge, V., A. E. Melchinger, B. S. Dhillon, and M. Frisch, 2009: Efficiency gain of marker-assisted backcrossing by sequentially increasing marker densities over generations. Theor. Appl. Genet. **119**, 23—32.

Pring, D. R., and D. M. Lonsdale, 1989: Cytoplasmic male sterility and maternal inheritance of disease susceptibilty in maize. Annu. Rev. Phytopathol. **27**, 483—502.

Tang, S., J. K. Yu, M. B. Slabaugh, D. K. Shintani, and S. J. Knapp, 2002: Simple sequence repeat map of the sunflower genome. Theor. Appl. Genet. **105**, 1124—1136.

Tomerius, A. M., T. Miedaner, and H. H. Geiger, 2008: A model calculation approach towards the optimization of a standard scheme of seed-parent line development in hybrid rye breeding. Plant Breeding **127**, 433—440.

Weber, W. E., D. C. Borchardt, and G. Koch, 1999: Combined linkage maps and QTLs in sugar beet (*Beta vulgaris* L.) from different populations. Plant Breeding **118**, 193—204.