Vitis **52** (4), 193–196 (2013)

Selection strategies for marker-assisted background selection with chromosome-wise SSR multiplexes in pseudo-backcross programs for grapevine breeding

E. Herzog[1)], R.Töpfer[2)], L. Hausmann[2)], R. Eibach[2)] and M. Frisch[1)]

[1)] Institute of Agronomy and Plant Breeding II, Justus Liebig University, Giessen, Germany

[2)] Julius Kühn Institut (JKI), Federal Research Centre for Cultivated Plants, Institute for Grapevine Breeding Geilweilerhof, Siebeldingen, Germany

Summary

Organizing SSR markers located on one chromosome into PCR multiplexes has the potential to reduce the costs of marker analysis. The optimal selection strategies for such chromosome-wise multiplexes have not yet been investigated. We investigated with computer simulations three different selection strategies for gene introgression with a pseudo-backcross scheme and a marker density of one marker every 10 cM. Selecting individuals with the highest number of chromosomes carrying *V. vinifera* alleles at all background marker loci reduced the number of required multiplexes by 7.24-7.87 % in generations pBC_4-pBC_6 for population sizes n_t = 150-300 individuals per pseudo-backcross generation.

K e y w o r d s : gene introgression, simulation study, multiplex PCR, microsatellite, marker-assisted selection.

Introduction

American and Asian *Vitis* species carrying resistance genes against mildew disease have been employed in interspecific breeding programs (Di Gaspero and Cattonaro 2010, Töpfer *et al.* 2011). Along this line only one example has been described for a systematic development of introgression lines as described by the pioneering work of Alain Bouquet (Pauquet *et al.* 2001). This work turns out to be very time consuming as well as space and labor demanding. The development of molecular markers for early selection of seedlings with traits of agronomic interest is therefore of particular value (Eibach *et al.* 2007). Simple sequence repeats (SSRs) are useful genetic markers, as they are abundant in the genome, highly polymorphic and transferable between *V. vinifera* and related species (Salmaso *et al.* 2008, Vezzulli *et al.* 2008, Blanc *et al.* 2012). However, they have the disadvantage of low throughput compared to single nucleotide polymorphisms (SNPs), which limits their use in large-scale breeding programs. Organizing SSRs into PCR multiplexes considerably reduces the costs of marker analysis (Merdinoglu *et al.* 2005). Patocchi *et al.* (2009) suggested that organizing SSR markers located on one linkage group in one multiplex is applicable and advantageous. Efficient selection strategies for *V. vinifera* genome recovery in pseudo-backcross programs for gene introgression with such chromosome-wise multiplexes have not yet been investigated. The objective of our study was to compare with computer simulations different selection strategies for gene introgression with chromosome-wise multiplexes.

Material and Methods

Computer simulations were carried out assuming no interference in crossover formation. Each simulation was run 10,000 times in order to reduce sampling effects and to obtain stable results with small standard error.

The genetic model for grapevine consisted of 19 chromosomes (2 x 40 cM, 7 x 60 cM, 5 x 80 cM, 5 x 100 cM). This corresponded to a total genome length of 1400 cM. The marker for the dominant target gene was assumed to be a gene-based marker and was located on a 100 cM chromosome at 61 cM from the telomere. Background markers were equidistantly spaced with one marker every 10 cM, the first and last marker of each chromosome being placed on the telomeres.

A pseudo-backcross scheme with changing *V. vinifera* parents in every generation was investigated up to generation pBC_6. The goal was to recover as much *V. vinifera* genome as possible, irrespective from which of the parents. For chromosome-wise multiplexes it was assumed that one multiplex included genotyping all background marker loci located on one chromosome. This resulted in multiplexes comprising 5 to 11 SSRs. In advanced pseudo-backcross generations only those chromosomes were analyzed which did not yet carry *V. vinifera* alleles at all background marker loci.

The donor was heterozygous for the desired allele at the target locus. The *V. vinifera* parents could be distinguished from the donor at all marker loci. Initially, the donor and the first *V. vinifera* parent were crossed to produce n_{F1} F_1 individuals. From this F_1 population, one individual that carried the donor allele at the target locus was selected as parent for generation pBC_1. This individual was crossed to the second *V. vinifera* parent to produce n_1 pBC_1 individuals. From this pBC_1 population, one best individual was selected with the selection strategies described below (see

Correspondence to: Dr. M. Frisch, Institute of Agronomy and Plant Breeding II, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany. E-mail: matthias.frisch@uni-giessen.de

also Fig. 1), and crossed to the next *V. vinifera* parent. This procedure was repeated for $t = 6$ pseudo-backcross generations with constant population sizes $n_{F1} = n_t = 50, 100, 150, 200, 250, 300$.

For all selection strategies, carriers of the donor allele at the target locus were pre-selected in the first selection step. These individuals were then subjected to one of three genome-wide background selection strategies. For Strategy 1, a selection index $i = \sum_m x_m$ was created, where summation is over background markers and x_m is the number of *V. vinifera* alleles at the *m*th marker. One individual with the highest value of *i* was selected in the second selection step and crossed to the next *V. vinifera* parent of generation pBC_{t+1}. The individual with the highest proportion of *V. vinifera* alleles at background marker loci was thus selected as parent for the next generation pBC_{t+1}.

For Strategy 2, a selection index $j = \sum_{c\,xc}$ was created, where summation is over chromosomes and $x_c = 1$ if a chromosome carries *V. vinifera* alleles at all background marker loci. All individuals with the highest value of *j* were selected in the second selection step. For these individuals, the value of *i* was determined as described for Strategy 1 in the third selection step. One individual with the highest value of *i* was selected and crossed to the next *V. vinifera* parent of generation pBC_{t+1}. The best individual with the highest number of chromosomes carrying *V. vinifera* alleles at all background marker loci was thus selected as parent for the next generation pBC_{t+1}.

For Strategy 3, a selection index $k = \sum_{c\,xc}$ was created, where summation is over chromosomes and *xc* = length of chromosome c in M if a chromosome carries *V. vinifera* alleles at all background marker loci. All individuals with the highest value of *k* were selected in the second selection step. For these individuals, the value of *i* was determined as described for Strategy 1 in the third selection step. One individual with the highest value of *i* was selected and crossed to the next *V. vinifera* parent of generation pBC_{t+1}. The best individual with the highest cumulative length of chromosomes carrying *V. vinifera* alleles at all background marker loci was thus selected as parent for the next generation pBC_{t+1}.

To quantify the success of the respective pseudo-backcross programs, the 10th percentile (Q_{10}), the arithmetic mean ($\bar{x}$) and the standard deviation (s_x) of the frequency distribution of *V. vinifera* genome in percentage, the average number of chromosomes carrying *V. vinifera* alleles at all background marker loci, and the average number and length of donor fragments were determined in every backcross generation for the selected individuals. In addition, the required number of chromosome-wise multiplexes for the respective pseudo-backcross programs, and the number of individuals n_i selected for evaluation with selection index *i* were determined in every backcross generation.

Results and Discussion

Up to generation pBC_3, Strategy 2 required a higher resource input than Strategy 1 for recovering equivalent levels of *V. vinifera* genome (Fig. 2). Strategy 3 was always inferior to Strategy 2. A $Q_{10} \geq 98$ % required $n_t = 100$ individuals per generation, 1753 multiplexes and three pseudo-backcross generations with Strategy 1 (Fig. 1). With Strategy 2, a $Q_{10} \geq 98\%$ required $n_t = 200$ individuals and 3162 multiplexes. With Strategy 3, a $Q_{10} \geq 98$ % required $n_t = 250$ individuals and 3947 multiplexes, or an additional pseudo-backcross generation.

In generations pBC_1-pBC_3, pre-selection for chromosomes carrying *V. vinifera* alleles at all background marker loci considerably reduced the number of individuals n_i from which the parent for the next generation was selected (Tab. 1). With Strategy 2 only $n_i = 1.8, 2.4, 10.5$ individuals were evaluated for selection index *i*. For Strategy 1, n_i were all individuals carrying the target gene (Tab. 1). These individuals have on average more, but shorter donor fragments than those selected with Strategy 2, and the probability that an individual with a higher overall proportion of *V. vinifera* genome is selected is higher than for Strategy 2.

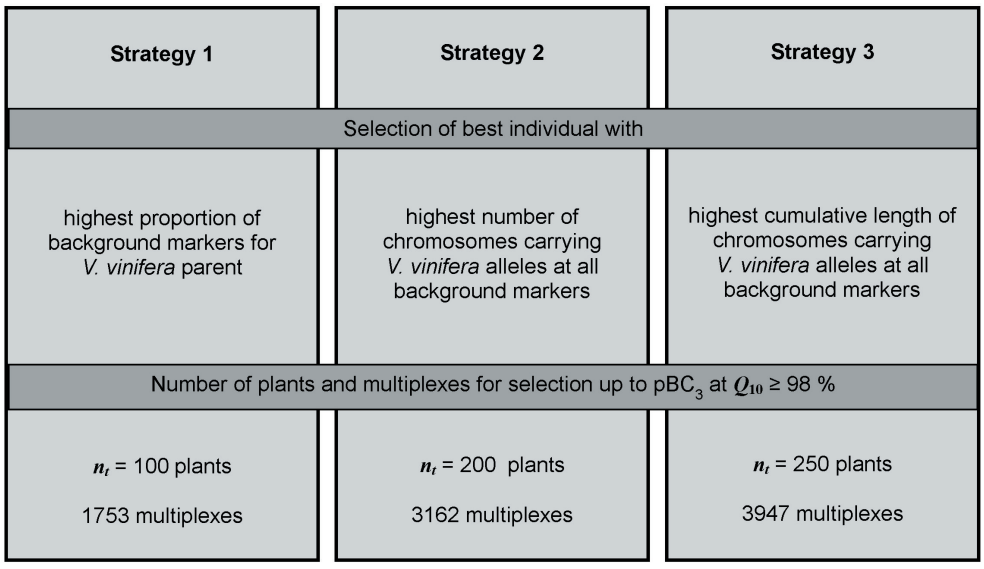

Fig. 1: Strategies and effort required to get $Q_{10} \geq 98$ % of *V. vinifera* genome in generation pBC_3. Number of plants per generation and number of chromosome-wise multiplexes are indicated for each strategy.

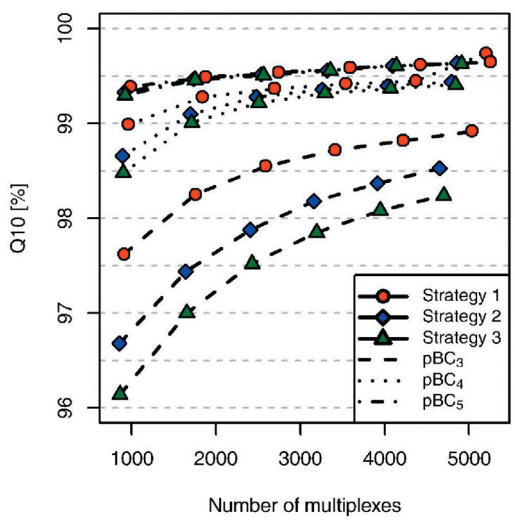

Fig. 2: Tenth percentile Q_{10} values of *V. vinifera* genome recovered in generations pBC_3-pBC_5 for Strategies 1, 2, 3 with constant population sizes n_t = 50, 100, 150, 200, 250, 300 individuals per pseudo-backcross generation plotted against the required number of chromosome-wise multiplexes.

In generations pBC_4-pBC_6, the differences in *V. vinifera* genome between the selection strategies disappeared (Fig. 2 and Tab. 1). For population sizes of $n_t \geq 150$ individuals per pseudo-backcross generation, the differences in Q_{10} values were < 0.1% between Strategy 1 and 2 (Fig. 2). Moreover, Strategy 1 then required more chromosome-wise multiplexes for equivalent levels of *V. vinifera* genome than Strategy 2 (Tab. 2). For n_t = 150-300 individuals per pseudo-backcross generation, 7.24-7.87 % of multiplexes were saved in generations pBC_4-pBC_6 with Strategy 2 compared to Strategy 1 (Tab. 2). In generation pBC_4, all non-carrier chromosomes carried *V. vinifera* alleles at all background marker loci on average, and selection focused on individuals with the shortest donor segment around the target gene for all three strategies (Tab. 1). The number of required multiplexes was then increasing at approximately the same rate for Strategy 1 and Strategy 2. However, Strategy 2 was more efficient than Strategy 1 in selecting for chromosomes which carried *V. vinifera* alleles at all background marker loci in generations pBC_1-pBC_3. While for Strategy 2, 9.3, 15.6, and 18.0 chromosomes on average carried *V. vinifera* alleles at all background marker loci, only 8.1, 14.5, 17.6 chromosomes carried *V. vinifera* alleles at all background marker loci with Strategy 1. These early savings resulted in an overall saving of chromosome-wise multiplexes in advanced pseudo-backcross generations.

Increasing the number of individuals per pseudo-backcross generation from n_t = 150 to n_t = 300 resulted in an additional *V. vinifera* genome recovery of 1.2-0.7 % for Strategy 1 in generations pBC_1-pBC_3 (data for pBC_1 and pBC_2 not shown, for pBC_3 see Fig. 2). In contrast, increasing the number of individuals per pseudo-backcross generation beyond n_t = 150-200 had little effect on *V. vinifera* genome recovery in advanced backcross generations for both Strategy 1 and Strategy 2 (Tab. 2, see also Fig. 2).

Table 1

Recovered level of *V. vinifera* genome (Q_{10}, $\bar{x}$, s_x) in percentage, required number of chromosome-wise multiplexes (CM), average number of chromosomes carrying *V. vinifera* alleles at all background marker loci (CCV), number of individuals evaluated for selection index *i*, number ($\bar{x}$, s_x) and length ($\bar{x}$, s_x) of donor fragments (cM) in generations pBC_1-pBC_6 for Strategies 1, 2, 3 with constant population size n_t = 150 individuals per pseudo-backcross generation

Strategy	pBC_t	*V. vinifera* genome (%)			CM	CCV	n_t	No. donor fragments		Length donor fragments (cM)	
		Q_{10}	$\bar{x}$	s_x				$\bar{x}$	s_x	$\bar{x}$	s_x
1	pBC_1	82.24	84.59	1.92	1425	8.1	75.0	12.9	2.2	33.55	26.22
	pBC_2	94.90	96.24	1.04	2246	14.5	74.9	5.8	1.8	18.14	17.62
	pBC_3	98.55	99.14	0.43	2585	17.6	75.0	2.2	1.1	10.91	9.71
	pBC_4	99.37	99.64	0.20	2692	18.4	75.0	1.4	0.6	7.05	5.08
	pBC_5	99.54	99.75	0.15	2740	18.7	74.9	1.3	0.5	5.51	3.79
	pBC_6	99.64	99.80	0.13	2761	18.9	74.9	1.2	0.4	4.91	3.26
2	pBC_1	79.10	82.75	2.86	1425	9.3	1.8	11.5	1.7	41.96	29.21
	pBC_2	92.89	95.07	1.66	2152	15.6	2.4	4.7	1.5	29.40	26.45
	pBC_3	97.88	98.82	0.69	2407	18.0	10.5	1.8	0.9	18.69	18.39
	pBC_4	99.28	99.60	0.24	2481	18.3	54.9	1.4	0.6	8.25	6.30
	pBC_5	99.52	99.74	0.16	2534	18.7	47.9	1.2	0.5	5.84	4.02
	pBC_6	99.63	99.79	0.13	2558	18.8	61.9	1.2	0.4	5.06	3.34
3	pBC_1	78.29	82.19	3.06	1425	9.2	1.1	11.6	1.7	43.04	28.95
	pBC_2	92.02	94.47	1.87	2163	15.5	1.2	4.8	1.5	32.23	27.28
	pBC_3	97.52	98.63	0.80	2428	18.0	8.9	1.8	0.9	20.95	20.68
	pBC_4	99.22	99.57	0.27	2506	18.3	55.3	1.4	0.6	8.67	6.88
	pBC_5	99.51	99.74	0.16	2560	18.6	48.2	1.2	0.5	5.90	4.03
	pBC_6	99.62	99.79	0.13	2586	18.8	61.4	1.2	0.4	5.10	3.35

Table 2

Recovered level of *V. vinifera* genome (Q_{10}) and percentage of saved chromosome-wise multiplexes (CM (%)) for Strategy 2 compared to Strategy 1 in generations pBC_4-pBC_6 with constant population sizes n_t = 150, 200, 250, 300

Generation	n_t = 150		n_t = 200		n_t = 250		n_t = 300	
	Q_{10}	CM (%)	Q_{10}	CM (%)	Q_{10}	CM (%)	Q_{10}	CM (%)
pBC_4	99.28	7.84	99.36	7.83	99.40	7.63	99.44	7.87
pBC_5	99.52	7.52	99.56	7.50	99.61	7.33	99.64	7.52
pBC_6	99.63	7.35	99.67	7.37	99.68	7.24	99.68	7.47

We conclude that if SSR markers are analyzed as chromosome-wise multiplexes, selecting for individuals with the highest proportion of *V. vinifera* genome at background marker loci is the most efficient selection strategy for short gene introgression programs of up to three pseudo-backcross generations. For such short gene introgression programs, population sizes of $n_t \geq 300$ individuals per pseudo-backcross generation maximize *V. vinifera* genome recovery. For gene introgression programs of four to six pseudo-backcross generations, pre-selecting individuals with the highest number of chromosomes carrying *V. vinifera* alleles at all background marker loci has the potential to considerably reduce the number of required SSR multiplexes. For these longer gene introgression programs, population sizes of n_t = 150-200 individuals are sufficient.

Acknowledgements

We thank N. Hofheinz for proof-reading the manuscript.

References

Blanc, S.; Wiedemann-Merdinoglu, S.; Dumas, V.; Mestre, P.; Merdinoglu, D.; 2012: A reference genetic map of *Muscadinia rotundifolia* and identification of *Ren5*, a new major locus for resistance to grapevine powdery mildew. Theor. Appl. Genet. **125**, 1663-1675.

Di Gaspero, G.; Cattonaro, F.; 2010: Application of genomics to grapevine improvement. Aust. J. Grape Wine Res. **16**, 122-130.

Eibach, R.; Zyprian, E.; Welter, L.; Töpfer, R.; 2007: The use of molecular markers for pyramiding resistance genes in grapevine breeding. Vitis **46**, 120-124.

Merdinoglu, D.; Butterlin, G.; Bevilacqua, L.; Chiquet, V.; Adam-Blondon, A. F.; Decroocq S.; 2005: Development and characterization of a large set of microsatellite markers in grapevine (*Vitis vinifera* L.) suitable for multiplex PCR. Mol. Breed. **15**, 349-366.

Patocchi, A.; Fernández-Fernández, F.; Evans, K.; Gobbin, D.; Rezzonico, F.; Boudichevskaia, A.; Dunemann, F.; Stankiewicz-Kosyl, M.; Mathis-Jeanneteau, F.; Durel, C. E.; Gianfranceschi, L.; Costa, F.; Toller, C.; Cova, V.; Mott, D.; Komjanc, M.; Barbaro, E.; Kodde, L.; Rikkerink, E.; Gessler, C.; van de Weg, W. E.; 2009: Development and test of 21 multiplex PCRs composed of SSRs spanning most of the apple genome. Tree Genet. Genomes **5**, 211-223.

Pauquet, J.; Bouquet, A.; This, P.; Adam-Blondon, A. F.; 2001: Establishment of a local map of AFLP markers around the powdery mildew resistance gene *Run1* in grapevine and assessment of their usefulness for marker assisted selection. Theor. Appl. Genet. **103**, 1201-1210.

Salmaso, M.; Malacarne, G.; Troggio, M.; Faes, G.; Stefanini, M.; Grando, M. S.; Velasco, R.; 2008: A grapevine (*Vitis vinifera* L.) genetic map integrating the position of 139 expressed genes. Theor. .Appl. Genet. **116**, 1129-1143.

Töpfer, R.; Hausmann, L.; Harst, M.; Maul, E.; Zyprian, E.; Eibach, R.; 2011: New horizons for grapevine breeding. In: H. Flachowsky, M. V. Hanke (Eds): Methods in Temperate Fruit Breeding, 79-100. Global Science Books; Fruit, Vegetable and Cereal Science and Biotechnology **5** (Special Issue 1), 79-100.

Vezzulli, S.; Troggio, M.; Coppola, G.; Jermakow, A.; Cartwright, D.; Zharkikh, A.; Stefanini, M.; Grando, M.S.; Viola, R.; Adam-Blondon, A. F.; Thomas, M.; This, P.; Velasco, R.; 2008: A reference integrated map for cultivated grapevine (*Vitis vinifera* L.) from three crosses, based on 283 SSR and 501 SNP-based markers. Theor. Appl. Genet. **117**, 499-511.

Received February 2, 2013