# Optimizing marker-assisted background selection for rapid introgression of desirable genes

**Randhawa HS[*,1], Mutti JS[*], and Gill KS**
*Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420*
**Authors with equal contribution thus should be considered joint first authors*
*[1]Current address: AAFC 5403-Ist Av. S Lethbridge, AB, T1J4B, Canada*

## INTRODUCTION

Rapid introgression of single genes in a targeted and identity-preserved manner is essential to protect and increase yield and other value-added traits. Life of a cultivar is shortened by sudden population shift and/or mutations in ever-evolving, dynamic pest populations causing serious yield losses. Rapid introgression of single genes is valuable not only to alleviate these constraints on the success of a variety but also to timely benefit from newly available value-added genes. Lack of an efficient and rapid method to introgress genes in an identity-preserved manner is a major bottleneck for a timely deployment of useful genes.

Currently during single gene introgression, recurrent parent genome (RPG) is recovered by repeated (up to six) backcrosses (BC). Theoretical expectations suggest a possible recovery of up to 99% of the RPG after six backcrosses (Allard 1999). Using few randomly selected plants during each backcrossing cycle, the simulated RPG recovery is not expected to be realized especially for the carrier chromosome because of a low probability of double recombinants around a target gene. For example, linkage drag around the *Tm-2* gene of tomato was estimated to be ranged from 4 cM to 51 cM even after >20 BC generations (Young and Tanksley, 1989). The expectation for the non-carrier chromosomes is also the same as the probability of plants carrying all recurrent parent type chromosomes is equally low.

Marker-assisted background selection (MABS) serves as a powerful strategy to select for the target gene (foreground selection) and to remove donor parent segments linked to the target gene along with accelerating the recovery of unlinked segments (background selection) (Tanksley et al. 1989). A computer simulation in tomato predicted that a MABS scheme can reconstruct RPG in just three BC cycles. These simulations also suggested that ~100 BC cycles would be required to reduce the linkage drag with few randomly selected plants during each cycle where as marker-assisted selection (MAS) using a large number of plants will have a probability of 0.95 of finding a recombinant near the target loci in two generations (Tanksley et al. 1989).

The estimated population size and marker data points (MDP's) for a high RPG recovery in two backcrosses, appear unrealistic based on the available numerical calculations and computer simulations probably because of the assumptions made while designing the mathematical models. The assumption of uniform distribution of genes and recombination rate on chromosomes is not accurate. Physical size of the linkage drag will depend upon the recombination rate around the target gene, as the distribution of recombination is highly uneven on chromosomes of higher eukaryotes (Sidhu and Gill 2004, for review). To ensure accuracy, information on the distribution of genes and recombination needs to be incorporated into the simulation models. Simulations incorporating genome structure information are not yet available for a backcrossing program.

The objective of this study was to develop and optimize an efficient, accurate and quick method to transfer single genes into popular genetic backgrounds. First, computer simulations were performed for various approaches of marker-assisted backcross program by incorporating the available information on the genome structure and the distribution of genes and recombination rates on the wheat chromosomes. The simulation results were used to develop a backcrossing scheme, efficiency of which was tested by transferring a stripe rust resistance gene *Yr15* into a susceptible but otherwise very good soft white spring wheat cultivar 'Zak'.

## MATERIALS AND METHODS

Soft white spring wheat cultivar 'Zak' (Kidwell et al. 2002) which is susceptible to stripe rust was used as female parent to introgress *Yr15* gene from Avocet S*6/*Yr15*' (PI 640428). Disease screening was performed using *PST-78,* a highly virulent race prevalent in the Pacific Northwest (PNW). Marker analysis was carried out using wheat SSRs and PCR products were analyzed using IR2 DNA Analyzer (Li-Cor Biosciences, USA). Percentage of the RPG was determined by dividing the number of loci homozygous for the recurrent parent by the total number of loci analyzed.

Computer simulations were performed using Plabsim (Frish et al. 2000) software. The Plabsim software was used to estimate the required population size and the number of MDP's for various backcross cycles. Genetic linkage map of the 251 markers that detected polymorphism between 'Zak' and the *Yr15* donor line, using the cM data from the wheat SSR genetic linkage map was used for computer simulations (Somers et al. 2004). Each simulation was repeated 2,500 times.

## RESULTS AND DISCUSSION

### Simulations to Compare Selection Strategies

The control simulation runs involving selection for only the target gene showed an RPG recovery at a Q10 value of 72.4% (means there is 0.90 probability that an RPG recovery of 72.4% will be attained), with a mean value of 74.9%. This number is very close to the 75% ratio that is expected if no background selection is applied. The same value after six backcross generations reached 98.8%, which is also very similar to the expected value of 99.2%. Simulations were performed for various selection strategies as described by Frisch et al. (1999). In case of the two-stage selection (first select plants carrying the target allele and then which are homozygous for RPG across the genome) a Q10 value of 94% was reached in the $BC_3$ generation provided the background selection was carried out using 11,228 MDP's. Three-stage selection (a second selection step of using flanking markers in addition to two stage selection steps) required 5225 MDP's, whereas four-stage selection (use of markers on carrier chromosome in addition to stage three selection steps) required only 921 MDP's to reach a set goal of ~96% RPG in $BC_3$ (Figure 1a). These simulations clearly showed that the four-stage selection required the least number of MDP's to recover ~96% RPG. Simulation to check the effect of unequal population sizes on the number of MDP's required to achieve 96% RPG in three backcrosses using four stage selection, are given in Figure 1c. A population size of 100, 150 and 200 plants per subsequent backcrosses required ~1000 MDP's compared to 921 if a constant population of 100 plants per backcross was used.

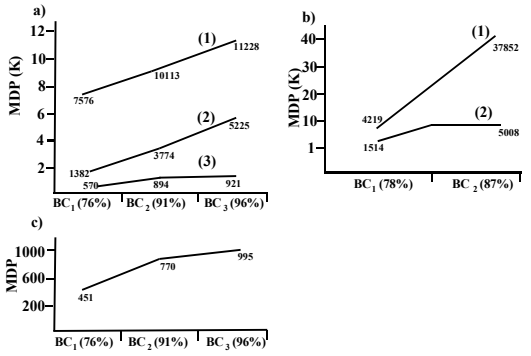



Figure 1. Computer simulations to estimate the number of marker data points (MDP's) and population size required for marker-assisted background selection program in wheat. In (a-c) the X-axis represents the number of backcross generations along with the percent RPG recovered and the y-axis represents the number of MDP's required in each generation. (a) Plabsim results for number of MDP's required using two-stage (1), three-stage (2) and four-stage (3) selection strategy in three backcross generations. (b) Number of MDP's required in three-stage (1) and four-stage (2) selection strategies for two backcross generations. (c) Four-stage selection for three backcross generations with population size of 100, 150 and 200, respectively during the three generations.

After establishing the four-stage selections to be the most efficient based on the number of MDP's required, simulations were performed to obtain the number of plants and MDP's needed to recover the same proportion of RPG in two backcrosses. Simulations run with a population size of 200 plants per BC required ~800 MDP's but the Q10 value of only 86% was recovered. To achieve an RPG recovery of 96%, the population size was increased from 200 to 5,000 plants per BC. A 25-fold increase in population size per BC resulted in only ~1.5% increase in RPG whereas number of MDP's increased 16-fold using four-stage selection. Simulation results with 1000 and 5000 plants per backcross are summarized in Figure 1b. A two-stage selection resulted in ~0.5 to 1% increase in RPG recovery compared to three and four-stage selection strategies with 5,000 plants per BC but number of required MDP's increased anywhere from 5-fold compared to three-stage to 48-fold compared to four-stage selection (data not shown). According to computer simulations, an RPG recovery at Q10 value of more than 92% is not achievable in two BC's regardless of the selection strategy.

### Proposed Marker Assisted Background Selection Scheme

Based on the computer simulation results and by taking into account the structural and functional organization of the wheat genome, we developed a background selection scheme (Figure 2). There are 48 gene rich regions (GRRs) in wheat that account for only about 29% of the genome but contain more than 85% of the genes (Erayman et al. 2004). Two markers flanking each of the GRR were selected except for the GRRs like 1S0.8 (46.1 cM) and 1L0.9 (50.0 cM) that has a very high rate of recombination. In those cases, another marker mapping in the middle of the GRR was also selected.

In order to increase the probability of obtaining a double recombinant in two BC generations and a high proportion of RPG, we propose to use 1000-3000 plants for each of the BC generations. Since probability of finding a double recombinant in a single generation is low, we propose to identify recombinants between the target gene and one of the flanking markers in $BC_1$ after phenotypic screening for the trait and for the second flanking marker in $BC_2$. The selected $BC_1$ plants should then be screened using the available markers for the carrier chromosome in order to select plants with the maximum recurrent parent alleles. After selecting recombinants for the second flanking marker, the selected $BC_2$ plants should be screened with the selected markers for the remaining chromosomes, except for the markers that were homozygous in $BC_1$. Two to four $BC_2$ plants should be selected and then made homozygous either using double haploid methods or by selfing. Progeny of the selected plants should then be subjected to field evaluation in order to select the most desirable genotype.

### Model Testing: Introgression of *Yr15* gene into cultivar 'Zak'

The polymorphism survey of the two parents ('Zak' and the donor parent Avocet S*6/*Yr15*) was performed with 639 SSR markers covering the entire wheat genome. About 50% (314/639) of these markers detected polymorphism that included 35 markers for chromosome *1B* that carries *Yr15* gene. Using the marker selection strategy outlined earlier, 251 of these markers were selected for the background selection as well as for

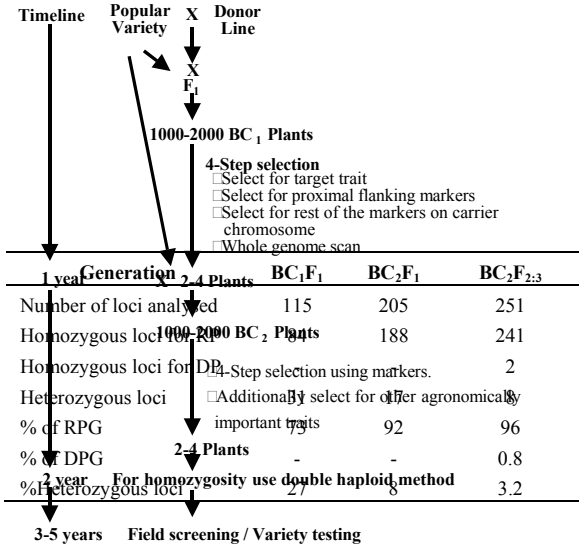
constructing the genetic linkage map used for the simulation studies.

A total of 1131 $BC_1F_1$ plants were screened for stripe rust and 156 resistant plants were first genotyped with flanking markers *gwm11* and *gwm33* in order to identify plants with recombination between *Yr15* and one or both of these markers. Twelve plants that were homozygous for recurrent parent type allele for the distal flanking marker *gwm33* were identified. No double crossover plant was recovered as all 12 plants were heterozygous for the proximal flanking marker *gwm11*. These 12 plants were genotyped with the remaining 33 *1B* specific markers to select four plants with the maximum number of homozygous loci. These four plants were used as male parent to generate 1056 $BC_2F_1$ plants, of which only 204 showed immune response when screened with the race *PST-78*. Genotyping of the resistant plants with the flanking marker *gwm11* identified five plants that were homozygous for the marker. Genotyping of these plants with the remaining 205 selected markers identified a plant that was homozygous for 188 markers and was heterozygous for 17 (Table 1). Selfed progeny of this plant was screened for resistance to stripe rust to identify 150 resistant plants out of the 300 that were screened. Screening of the $F_{2:3}$ progeny of the resistant plants identified 37 homozygous resistant families.

Figure 2. Strategy for marker-assisted background selection for wheat

Table1. Proportion of recurrent and donor parents genome in different generations of backcrosses

Multiple plants for each of these resistant families were screened with the 17 markers for which the selected $BC_2F_1$ plant was heterozygous. Two plants with ~94%. RPG was selected and progeny of these plants were used




| Generation | $BC_1F_1$ | $BC_2F_1$ | $BC_2F_{2:3}$ |
|---|---|---|---|
| Number of loci analyzed | 115 | 205 | 251 |
| Homozygous loci for RP | 84 | 188 | 241 |
| Homozygous loci for DP | - | - | 2 |
| Heterozygous loci | 31 | 17 | 8 |
| % of RPG | 73 | 92 | 96 |
| % of DPG | - | - | 0.8 |
| %Heterozygous loci | 27 | 8 | 3.2 |

for field evaluation.

The phenotypic and quality data from uniform cereal variety testing trails conducted in 2007 averaged over 17 locations, suggested that the selected line, derived using MABS was very similar to the recurrent parent with the mean average yield, test weight and protein content of 51, 60 and 12.7; compared to 51.7, 58.8, and 12.1 for 'Zak' (http://variety.wsu.edu/2007/index.htm#annualreport).

## CONCLUSION

Avoiding linkage drag on the carrier chromosome is the most difficult part of the backcross breeding. Because of a very low frequency of occurrence, double crossovers for the flanking markers and the gene of interest are less likely to be selected even after many backcross cycles involving random selection of plants. As suggested by simulations and shown by our results, it is possible to obtain double recombinants and to recover a high percentage of RPG just in two BC cycles. Additional selfing generations required to achieve a desired level of homozygosity is still a 'bottleneck' to further reduce the time taken by the marker-assisted background selection approach. Although not utilized in this study, we propose to alleviate that constraint by the use of doubled haploid methods to instantly fix homozygosity.

## REFERENCES


Allard, R.W. 1999. Principles of plant breeding. 2nd ed. John Wiley and Sons, Inc. New York, NY.

Erayman, M., D. Sandhu, D. Sidhu, M. Dilbirligi, P.S. Baenziger, and K.S. Gill. 2004. Demarcating the gene-rich regions of the wheat genome. Nucleic Acids Research 32: 3546-3565.

Frisch, M., M. Bohn, and A.E. Melchinger. 1999. Comparison of Selection Strategies for Marker-Assisted Backcrossing of a Gene Crop Science 39: 1295-1301.

Frisch, M., M. Bohn, and A.E. Melchinger. 2000. PLABSIM: software for simulation of marker-assisted backcrossing. Journal of Heredity 91:86-87.

Kidwell, K.K., G.B. Shelton, V.L. Demacon, et al. 2002. Registration of 'Zak' Wheat. Crop Science 42: 661a-662a.

Sidhu, D. and K.S. Gill. 2004. Distribution of genes and recombination in wheat and other eukaryotes. Plant Cell Tissue and Organ culture 79: 257-270.

Somers, D. J., P. Isaac and K. Edwards, 2004 A high-density microsatellite consensus map for bread wheat (Triticum aestivum L.). Theor. Appl. Genet. 109: 1105–1114.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7: 257-264.

Young, N.D. and S.D. Tanksley. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theoretical and Applied Genetics 77: 95-101.