A practical application of SNP marker assisted selection in canola (*Brassica napus*) cultivar development programs

J. Brown[1†], Z. Zhang[2], H. Dong[2], J. Kuhl[1], J.B. Davis[1], and A. Job[1]

[1]University of Idaho, USA; [2] Washington State University, USA.

† Presenting author: jbrown@uidaho.edu

INTRODUCTION

Despite recent advances in molecular marker technologies, modern canola breeding still relies on recurrent phenotypic selection. Greatest genetic gain should be in the early selection phases when phenotypic variation is greatest. However early generation selection for quantitative traits is often, at best, a random reduction in breeding lines.

Cross prediction and genomic selection provide alternative methods of reducing breeding accession numbers over traditional phenotypic selection on single plants or small plots. However, cross prediction genetic parameterization require considerable effort and genomic selection requires high quality and reliable phenotypic performance data. In addition, most genomic selection involves training populations considerably wider in genetic diversity than used as breeding parents.

Here cross prediction and a novel SNP based method of estimating parent and cross worth were examined. Accuracy of predictions were gauged against progeny that survived selection in a breeding scheme when larger numbers were evaluated.

MATERIAL & METHODS

Nine spring canola (*Brassica napus*) cultivars were crossed in a full diallel design, and 8 F_3 progeny from each of 72 crosses evaluated for Breeders' Preference in 2015 cross prediction field trials. Data were used to estimate cross prediction probabilities using:

$$\int_T^\infty f(x_i)\mathrm{d}x_i$$

where the function $f(x_i)$ is based on m, the mean of all possible progeny from a cross, and σ_G, the trait genetic standard deviation.

Seed from an additional 81 F_3 progeny from each cross were used to plant small single replicate plots at two locations in northern Idaho. These F_4 plots were evaluated for Breeders' Preference and the best 214 lines selected for testing in replicated yield trials the following year. Based on performance in the 2017 F_5 yield trials, the best 43 breeding lines were identified and tested in 2018 F_6 yield trials, and when results were available the best 7 F_7 lines were selected.

DNA was extracted from the 43 F_6 selected lines and each genotyped by LGC Genomics producing 80,000 Single Nucleotide Pair (SNP) reads on each sample and from these 9,720 polymorphic SNP's were identified.

Genomic parent and cross worth indices were estimated from **selection differentials** ($\Delta S_{i,AT}$ and $\Delta S_{i,CG}$), which are calculated as the difference in percentage of SNP nucleotides within the parent population ($P_{i,AT}$ and $P_{i,CG}$) and the selected population ($S_{i,AT}$ and $S_{i,CG}$). Each parental worth index is then the sum of the selection differential values associated with that parent, over all SNP's (i.e. $\sum\Delta S_{i,Parent}$). Cross worth indices were calculated in a similar way, but the average $\Delta S_{i,(Parent1 + Parent2)}$ values of both parents were used rather than only one parent.

RESULTS

The value of the parent worth estimates based on cross prediction and on SNP $\sum\Delta S_{i,j}$ values were compared to the number of selections made from progeny with the parent in their pedigree (Table 1). The highest ranked parent ('HyClass-930') was a parent in 21% and 33% of the F_5 and F_6, respectively, selected lines. Similarly, 'Star-402' was ranked 2nd and appeared in the pedigree of 20% and 19% of the F_5 and F_6, respectively, selected lines. If the top 4 parents were selected based on parent worth from SNP $\sum\Delta S_{i,j}$ these parent appeared in the pedigree of 76% and 77% of the F_5 and F_6 selected lines.

The association between predictions of cross worth based on cross prediction and SNP $\sum\Delta S_{i,j}$ values also was examined by considering the frequency of selected lines that are observed from progeny of each cross (Table 2). Cross worth from DKL.55-55 x HyCl-930 was ranked 1st on cross prediction and 2nd on SNP $\sum\Delta S_{i,j}$, and 9% and 16% of all F_5 and F_6 selections progeny derived that cross. Similarly, the top 6 crosses based on cross prediction of Breeders' Preference were all within the top 10 crosses based on SNP $\sum\Delta S_{i,j}$ values. These 6 crosses represent 8% of the crosses examined, yet progeny from these cross progeny included over half (52%) of F_5 and 68% of all F_6 selections.

Parent	Cross Prediction Rank	SNP Select Differential Rank	Percentage Selected	
			F_5	F_6
DKL.55-55	6	5	7%	11%
DKL.70-70	5	6	8%	7%
HyClass-930	1	1	21%	33%
IMC.109	3	4	19%	11%
InVigor-5440	7	8	8%	5%
InVigor-L130	4	3	16%	14%
Python	9	7	0%	0%
Star-402	2	2	20%	19%
V12-1	8	9	1%	0%

Table 1. Cross prediction and SNP $\sum\Delta S_{i,j}$ parent worth ranks, and the number of selected progeny at the F_5 and F_6 stages of selection in which each parent appeared in the pedigree.

Parent 1	Parent 2	Breeders' Preference Prediction Rank	SNP Selection Differential Rank	% Selected in the breeding program	
				F_5	F_6
DKL.55-55 x	HyCl-930	1	2	9%	16%
DKL.55-55 x	Star.402	4	1	5%	7%
HyClass-930 x	InVigor.L130	2	10	6%	19%
HyClass-930 x	Star.402	3	4	12%	16%
IMC-109 x	Star.402	6	5	8%	5%
InVi-L130 x	Star.402	5	8	12%	5%

Table 2. Top cross prediction and SNP $\sum\Delta S_{i,j}$ cross worth rankings, and the number of selected progeny at the F_5 and F_6 stages of selection that come from each cross.

CONCLUSIONS

Parental worth from SNP $\sum\Delta S_{i,j}$ values provided predictions as good or better than those from cross prediction. However, the predictions were only available after selections have passed through the breeding scheme, and hence after the fact. There is perhaps good opportunity to utilize the SNP $\Delta S_{i,j}$ values from one data set to determine parent worth from other parents. In a similar manner it may be feasible to use SNP $\Delta S_{i,j}$ array data to rate the worth of parent cross combinations that have not yet been tried. In addition, a database of SNP $\Delta S_{i,j}$ arrays from multiple years could easily be compiled and used to determine indices and worth of future parent. It would take little cost and effort to genotype parents used in crosses each year, and to set up SNP $\Delta S_{i,j}$ arrays between these parents and say the F_5 selection made from these crosses. If parents are not all used in the same number of crosses the SNP genotype of each parents should be weighted according to how many crosses the parent appeared.

In conclusion, parent and cross worth indices using SNP $\sum\Delta S_{i,j}$ arrays would help breeders choose parents and cross combinations that increase the probability of producing superior progeny and cultivars. The technique avoids multiple and complex phenotyping or complex statistical and multivariate transformation that many other genomic selection tools utilize.