Towards genomic analysis in canola multi-environment trials

W. Cowling[1,2], A. Ganesalingam[1], S. Chen[1,2], M. Nelson[1,2], C. Beeck[2], A. Smith[3] and B. Cullis[3,4]

[1]The University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia
[2]Canola Breeders Western Australia Pty Ltd, 119 Canning Highway, South Perth 6151, Australia.
[3]University of Wollongong, Wollongong, NSW 2522, Australia.
[4]CSIRO Mathematics Informatics and Statistics, Canberra, ACT 2600, Australia.
Email: wallace.cowling@uwa.edu.au

ABSTRACT

The development of genomic selection in plant breeding will follow the methods developed in animal breeding, with significant modifications for the unique features of plant breeding. The goal is to predict more accurately total and additive genetic values by including genomic relationships in the analysis. This should result in faster cycles of selection. Pedigree or genetic selection has improved rates of genetic improvement in many animal breeds over the past 30 years, and additional benefits are likely from genomic selection. Unlike animals, plant varieties are tested in replicated plots within and across environments, and analysis must account for spatial variation and other systematic or random errors within trials, and genotype x environment interactions. To be meaningful, the test population must encompass the entire breeder's population, or at least a significant proportion of it. The population must be tested in relevant environments, and the analysis must account for genotype x environment interaction. Recent developments in mixed models for analysis of canola multi-environment trials showed significant improvements in prediction of total, additive and non-additive genetic values for oil content and grain yield, and are the first step towards genomic selection in plant breeding.

Key words: pedigree selection, genomic selection, additive and non-additive genetic value

INTRODUCTION

Animal breeders have more than 30 years of experience with genetic selection based on pedigrees, for example, breeding for market weight in chickens and breeding for milk production in Holstein cattle. [Remarkable] gains in modern breeding programs have been achieved by estimating the genetic merit of selection candidates based on phenotypic and pedigree information…" (Hayes and Goddard, 2010).

The goal of genomic selection is to associate phenotypic data with genetic relationships derived from genome-wide markers to predict genomic breeding values, for example, genomic selection of dairy bulls at 12 months of age "…the success of genomic selection depends on the accuracy with which breeding value can be predicted in the selection candidates" (Goddard and Hayes, 2009). In genetic (pedigree) selection, the correlation of estimated breeding value (EBV) to true genetic value is 0.5. In genomic selection, the correlation of genomic breeding value (GBV) to true genetic value is 0.7 to 0.85, limited by population size, linkage disequilibrium from spurious causes and number of markers (Goddard and Hayes, 2009).

Various biometric models are under investigation for estimation of breeding value in self-pollinating crops (Bauer *et al.*, 2008, 2009; Oakey *et al.* 2006). Traditionally, selection has focussed on the best variety for release, not for breeding value. Recently, genetic (pedigree) selection in multi-location canola trials greatly improved the efficiency of selection for total genetic value and estimated breeding value in the presence of genotype x environment (GxE) interaction (Beeck *et al.*, 2010; Cullis *et al.*, 2010). Methods were developed for fitting factor analytic mixed models with pedigree information. These methods will be discussed as the basis of future developments in genomic selection.

MATERIALS AND METHODS

The analysis was conducted on a multi-environment trial (MET) data-set of 332 canola genotypes from 19 canola trial sites in Australia in 2007 and 2008 (Beeck *et al.* 2010), with pedigree data from 1970 on most lines (Cowling 2007). The traits analysed were grain yield and oil content. Trials were highly unbalanced designs, and concurrence of varieties across trials was high.

The variance structure of the genetic effects (GxE) was modelled by using a factor analytic model, and the total genetic effects were partitioned into additive (estimated breeding value, or EBV) and non-additive effects (Beeck *et al.* 2010). Genetic effects were modelled using a MET factor analytic (FA) analysis. Indices were formed for grain yield and oil content based on GxE interaction, and statistical tools were developed to explore GxE, including heat-maps of genetic *correlation matrices, e-scaled uni-plots, and agglomerative (nested) hierarchical clustering (Cullis et al.* 2010).

These developments are discussed in light of potential developments for genomic selection.

RESULTS

MET/FA analysis of canola breeding trials across southern Australia revealed significant spatial trends (both fixed and random) for seed oil and grain yield within trials (Beeck *et al.* 2010). Genetic variance across sites was modelled with a series of FA models which were fitted to the data for both seed oil and grain yield. For both traits, the addition of pedigree information substantially improved fit compared with the corresponding model excluding pedigree information (Beeck *et al.* 2010). The average proportion of additive variance at sites for grain yield was 0.561, compared to an average of 0.904 for seed oil (Beeck *et al.* 2010). This reflects the higher heritability of seed oil compared with grain yield.

The addition of pedigree information to the analysis permitted independent estimates of additive and non-additive genetic variance. Non-additive genetic variance represents the failure of lines to perform at sites according to predictions from ancestry (Cullis *et al.* 2010). Plant breeding experiments have a vital advantage over animal breeding - it is possible to assess the performance of both parents and progeny in the same experiments, and “borrow” information from relatives, thereby improving the estimates of additive (EBV) and non-additive genetic effects. However, plant breeding experiments have a complication due to replication of varieties within and across environments, which enhances confidence of estimates of genetic variance, but new models of analysis must be developed to allow for GxE and replication.

There were two major groups of environments for additive effects for canola yield in southern Australia, and negative correlations between some sites, indicating a GxE cross-over in additive effects for yield. Examples of varieties with specific adaptation to low rainfall and drought stressed sites, or to high rainfall sites where blackleg disease was prevalent, are presented in Cullis *et al.* (2010). The EBV of variety CB Telfer was low compared with variety N04D-3521, even though the two varieties had similar high total genetic values in the low rainfall environment group (Cullis *et al.* 2010).

Pedigree selection in canola breeding revealed distinct patterns of behaviour across environments for additive, non-additive, and total genetic effects (Cullis et al. 2010). Pedigree selection, and in future genomic selection, should help to accelerate genetic progress in canola breeding by reducing cycle time and providing more effective selection of parents for target environments.

DISCUSSION

The value of genetic (pedigree) selection in canola breeding, which was clear from MET/FA modelling (Cullis *et al.* 2010), resulted from the reduction in uncertainty by “borrowing” information from relatives to improve estimates of genetic values across environments. In theory, additional benefits should arise from genomic selection, which will improve the accuracy of estimates of breeding value and total genetic value.

The economic value of genomic selection to plant breeding has yet to be evaluated – already there is significant value derived from genetic (pedigree) analysis in animals (Hayes and Goddard, 2010) and canola (Cullis *et al.* 2010). These studies demonstrate that it is important to

invest in high quality phenotyping (field trials in target environments) and biometrical systems for genetic (pedigree) analysis.

Technical issues need to be addressed in genomic selection in plant breeding, such as research into the role of population size, number of markers, and the extent to which whole genome markers (such as SNPs) explain genetic variance in the data, as in animal breeding (Hayes and Goddard 2010). Appropriate biometrical models must be developed for incorporating molecular genetic data into the MET/FA analysis. Once this research has taken place, a cost/benefit analysis will demonstrate if the cost of molecular genotyping improves the value of MET/FA genomic analysis sufficiently to warrant its use over pedigree analysis.

ACKNOWLEDGEMENTS

GRDC provided significant support through funding of the project Statistics in the Australian Grains Industry (SAGI).

REFERENCES

Bauer, AM, Hoti F., Reetz TC, Schuh W-D., Leon J, Silanpaa MJ (2009) Bayesian prediction of breeding values by accounting for genotype-by-environment interaction in self-pollinating crops. Genetics Research 91: 193-207.

Bauer, AM, Reetz TC, Leon J (2008) Predicting breeding values of spring barley accessions by using the singular value decomposition of genetic similarities. Plant Breeding 127: 274-278.

Beeck CP, Cowling WA, Smith AB, Cullis BR (2010) Analysis of yield and oil from a series of canola breeding trials, part I: Fitting factor analytic mixed models with pedigree information. Genome 53: 992–1001.

Cowling WA (2007) Genetic diversity in Australian canola and implications for crop breeding for changing future environments. Field Crops Res 104: 103–111.

Cullis BR, Smith AB, Beeck CP, Cowling WA (2010) Analysis of yield and oil from a series of canola breeding trials, part II: Exploring variety by environment interaction using factor analysis. Genome 53: 1002–1016.

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10: 381–391.

Hayes BJ, Goddard ME (2010) Genome-wide association and genomic selection in animal breeding. Genome 53: 876–883.

Oakey H, Verbyla A, Pitchford W, Cullis B, and Kuchel H (2006) Joint modelling of additive and non-additive genetic line effects in single field trials. Theor. Appl. Genet. 113: 809-819.