

Towards the application of association genetics to canola breeding

W.A. Cowling^{1,2}, C.P. Beeck^{1,2}, A.B. Smith³, M.N. Nelson^{1,2}, B.R. Cullis³

¹School of Plant Biology M084, The University of Western Australia, 35 Stirling Hwy., Crawley, WA 6009, Australia. wcowling@cyllene.uwa.edu.au

²Canola Breeders Western Australia Pty Ltd., 15/219 Canning Hwy., South Perth, WA 6151, Australia. cbeeck@cyllene.uwa.edu.au

³NSW DPI Biometrics, Agricultural Research Institute, Wagga Wagga, NSW 2650, Australia. brian.cullis@dpi.nsw.gov.au

ABSTRACT

Multi-environment trial (MET) analysis, as used in plant breeding, provides a framework on which to build new applications of association genetics. Spatial analysis within sites is used to develop a diagonal model, which is normally inferior to MET analysis with a factor analytic (FA) mixed model framework across sites. Pedigree relationships can be incorporated into MET analysis (MET-ped) to improve estimation of additive and non-additive genetic variance. We used MET analysis to rank canola genotypes for oil content in 7 field trials across southern Australia in 2008. Spatial variation was detected within sites for oil and other quality traits based on near-infrared radiation analysis of seed samples from each plot. The FA(2) model gave a significant improvement in log-likelihood score above FA(1). There was high genotypic correlation in oil content between sites (range 0.69 – 0.93) with FA(2). The ranking of genotypes based on predicted values from the diagonal model was different from the ranking based on the FA(2) model. There were 5 changes to genotypes in the top 26 (top 10%) of lines in the FA(2) model compared with the diagonal model. MET-ped improved the log-likelihood score over FA(2), and there were 3 changes to genotypes in the top 26 lines compared with the FA(2) model. These changes in ranking with MET-ped analysis are important to avoid sub-optimal selection for promotion and crossing. Similar improvements are expected with the addition of medium to high-density genome-wide molecular marker data to MET-ped analysis.

Key words: association mapping - quantitative trait loci

INTRODUCTION

Plant breeders normally test varieties across a range of sites in the target environments, with genotypes replicated within and across sites. In order to maximise information on genotypes in such trials, a multi-environment trial (MET) analysis with a factor analytic mixed model framework, incorporating spatial relationships within sites, has been developed (Smith et al. 2001, 2005; Kelly et al. 2007). MET analysis, combined with information from pedigree relationships, improved estimates of genetic value of yield of sugarcane genotypes, as a result of partitioning of the genetic effect of a line into additive, dominance and residual non-additive effects (Oakey et al. 2007). We applied this approach to canola breeding, based on seed oil content in 7 field trials across southern Australia in 2008. Ultimately, we propose to use whole-genome molecular marker data in MET analysis to improve estimates of genetic value of lines.

MATERIALS AND METHODS

Example data set

The data in this paper were seed oil measurements on harvest samples from 7 canola breeding trials in 2008 conducted by Canola Breeders Western Australia Pty Ltd. The 7 trials contained advanced germplasm in a p-rep trial design (Cullis et al. 2006). Every plot was harvested separately, and subsamples of harvested seed were analysed for seed oil content by near-infrared spectroscopy (NIRS).

Analysis

The dataset for seed oil content was analysed by mixed model techniques for multi-environment trials (METs) (Smith et al. 2001) using adjustments for spatial trends in the diagonal model, before proceeding to MET analysis with factor analytic (FA) models. Analysis was conducted using the software ASReml in the R environment (Butler et al. 2007). The analysis also provided genetic correlations between pairs of trials. Additive genetic effects were estimated by adding pedigree information to the model in the form of an additive relationship matrix (Oakey et al. 2007).

RESULTS

The FA(2) model gave a significant improvement in log-likelihood score above FA(1). There was high genotypic correlation in oil content between sites (range 0.69 – 0.93) with FA(2). The ranking of genotypes based on predicted values from the diagonal model was different from the ranking based on the FA(2) model. There were 5 changes to genotypes in the top 26 (top 10%) of lines in the FA(2) model compared with the diagonal model (Fig. 1). MET-ped improved the log-likelihood score over FA(2), and there were 3 changes to genotypes in the top 26 lines compared with the FA(2) model (Fig. 2).

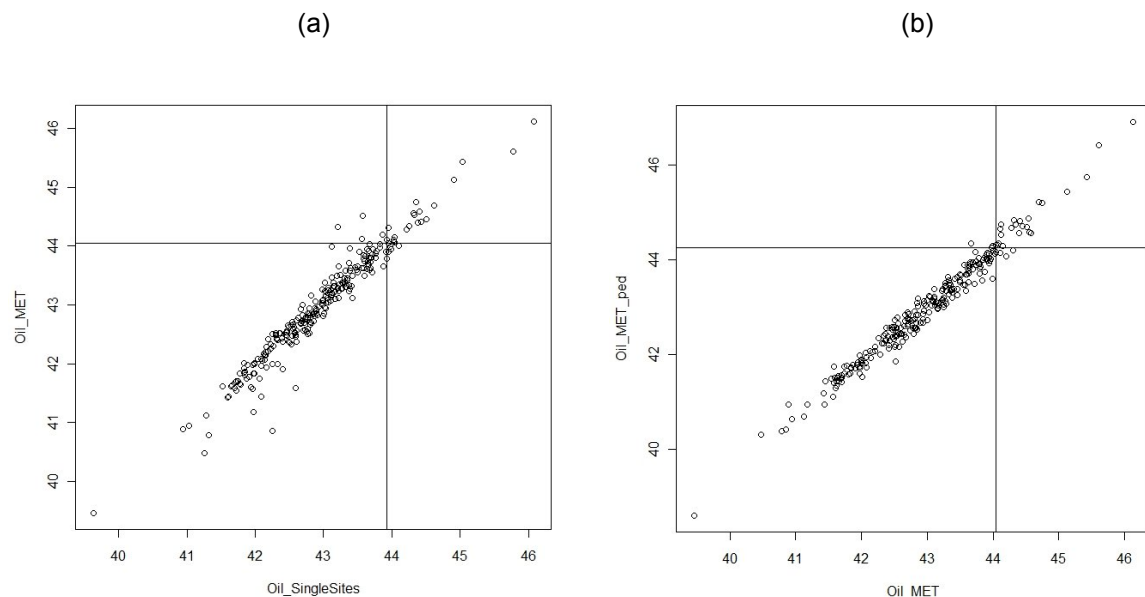


Fig.1. Ranking of canola genotypes for seed oil content (%) in comparisons of (a) the diagonal model (average across sites) vs the MET FA(2) model across 7 sites, where 5/26 of the top 10% genotypes were selected by the MET FA(2) model, but not by the diagonal model; and (b) the MET FA(2) vs the MET-ped model, where 3/26 of the top 10% genotypes were selected by the MET-ped model, but not by the MET FA(2) model.

DISCUSSION

The principles of association genetics can be applied to canola breeding, by using systems of trial design and analysis that incorporate spatial analysis within sites and MET FA models across sites. We found a significant improvement in estimation of genetic value of lines through the use of pedigree information in the MET analysis, similar to that found in sugar cane by Oakey et al. (2007). This is most likely the result of a better estimation of non-additive genetic effects in the MET-ped model, which in the case of inbred crops are mostly due to epistasis (Oakey et al. 2007). Several lines were selected for further testing or crossing by the MET-ped model that were not selected by the diagonal model. We expect similar benefits to arise from the addition of medium to high-density genome-wide molecular marker data to MET-ped analysis. Such developments will aid the selection of genetically superior lines.

ACKNOWLEDGEMENTS

Part salaries of WAC and MNN were supported by grants from the Grains Research and Development Corporation.

REFERENCES

- Cullis, B.R., A.B. Smith, and N.E. Coombes, 2006: On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics* 11, 381.
- Butler, D.G., B.R. Cullis, A.R. Gilmour, and B.J. Gogel, 2007: Analysis of mixed models for S language environments, ASReml-R reference manual Release 2. QLD Department of Primary Industries and Fisheries, Brisbane, QLD. Training and Development Series, No QE02001.
- Kelly, A.M., A.B. Smith, J.A. Eccleston, and B.R. Cullis, 2007: The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Science* 47, 1063-1070.
- Oakey, H., A.P. Verbyla, B.R. Cullis, X. Wei, and W.S. Pitchford, 2007: Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theor. Appl. Genet.* 114, 1319–1332.
- Smith, A.B., B.R. Cullis, and R. Thompson, 2001: Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57, 1138-1147.
- Smith, A.B., B.R. Cullis, and R. Thompson, 2005: The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *Journal of Agricultural Science* 143, 1-14.