#064

Prediction of canola phenology through integration of genomic prediction and crop growth model

***Bangyou Zheng*[1]**

Jeremy Whish[1]
Julianne Lilley[2]
Alexandre Boyer[2]
Alec Zwart[2]
Chris Helliwell[2]
Shannon Dillon[2]

[1] CSIRO Agriculture & Food, St Lucia, Australia

[2] CSIRO Agriculture & Food, Canberra, Australia

Background:

Canola phenology is a major determinant of the adaptation of canola to different environments and the productivity of canola can be maximised by targeting phenology to the optimal flowering window. Although many crop models have been parameterised to predict canola phenology, these models are reliant on estimation of model parameter values from phenology scores for a given variety obtained in multiple environments.

Objective:

Canola breeding and production would benefit from robust models that predict phenology based on genomic data (single nucleotide polymorphisms (SNPs)), shortening the timeframe for optimising flowering time for new varieties. Here, we present a framework that blends genomic prediction and crop growth model to predict phenology for a given variety across a range of environments based on its genome.

Methods:

Genomic SNP data for (34196 loci) were assayed in a diverse canola panel of 350 Australian and global varieties. For the same panel, key phenology developmental stages (green bud and first flower) were scored in 18 environments (time of sowing by year and location) spanning sowing dates from April to May, corresponding to over 22K observations. SNP and phenology data were subsequently integrated by fusing machine learning (ML) and crop modelling to calibrate (train) and validate (test) a canola model in APSIMNG (CGM), and in parallel a new integrated genomic crop growth model (GP-CGM), which enables prediction of flowering time for any related variety in Australian growing conditions based on its genome. In GP-CGM, the model parameter values are estimated using semi one-step optimisations directly from genotypic information (i.e. SNP data) to phenotypic values (i.e. observed phenology).

Results:

Prediction accuracies of GP-CGM were R = 0.95 for observed environment and genotype, and R = 0.93 for unobserved environment and observed genotypes (decreases of 0.2-0.4 comparing with CGM. The prediction accuracies of new genotypes (UG) were further reduced to R = 0.87 for known environment (OEUG), and R = 0.86 for unknown environments (UEUG).

Conclusions:

We have developed a hybrid method that fuses ML and process-based crop modelling to predict flowering time for any variety, by exploiting genomics and our existing understanding of flowering processes. The GP-CGM model can be generalised to a wide range of Australian environments. It accurately predicts phenology for previously observed or new varieties in previously observed and unobserved environments with prediction accuracies (R) greater than 0.86. The new GP-CGM model will enable growers and breeders to optimise management of new varieties in their region based on phenology years sooner than currently possible. The model can further be used to explore the complexity of interactions among genotypes and environments and extended to other crop models. This genomic based tool is now being integrated into the Canola Flowering Calculator (https://www.canolaflowering.com.au/) where we are aiming to update the calculator annually with genomic based predictions for varieties tested via the NVT network.