Prediction of wheat leaf appearance through integration of single nucleotide polymorphisms (SNPs) with a crop model

Bangyou Zheng[1], Hamish Brown[2], Zhigan Zhao[3], Enli Wang[3], Neil Huth[4], Shannon Dillon[3], Jessica Hyles[3], Tina Rathjen[3], Maxwell Bloomfield[5], Corinne Celestina[5], James Hunt[5], Ben Trevaskis[3]

[1] CSIRO Agriculture and Food, Queensland Biosciences Precinct 306 Carmody Road, St Lucia, QLD, 4067 bangyou.zheng@csiro.au
[2] The New Zealand Institute for Plant & Food Research, Private Bag 4604, Christchurch, New Zealand
[3] CSIRO Agriculture and Food, GPO Box 1700, Canberra, ACT 2601, Australia
[4] CSIRO Agriculture and Food, PO Box 102, 4350 Toowoomba, Australia
[5] DAPSS, La Trobe University, 5 Ring Rd, Bundoora, VIC 3086

Abstract

Wheat phenology is a major determinant of the adaptation of wheat to different environments. Although many crop models have been parameterised to predict wheat phenology, wheat breeding and agronomy research would benefit from robust models that quantitatively predict the interaction of genotype and environment based on known genes or single nucleotide polymorphisms (SNPs). In this study, the 90K SNP data of 55 wheat cultivars in the Australian Phenology Panel are integrating into wheat model in APSIM Next Generation (APSIM NG) to predict traits related with wheat phenology. Here we report the results for leaf appearance rate. The datasets collected from controlled environments and field experiments are used to calibrate and validate the wheat model in APSIM NG (APSIM-Release) and the new integrated genomic model (APSIM-GS) with different strategies. The model parameter values are estimated using global optimizations. Both APSIM-Release and APSIM-NG could predict the leaf appearance rate for 55 cultivars with an acceptable level of accuracy?

Keywords

genomic prediction, crop model, genomics, phenomics

Introduction

Phyllochron, defined as the thermal time duration for a leaf to fully emerge, is one of the key determinants for leaf area development and early vigor (Botwright *et al.*, 2002). It is mainly driven by temperature and photoperiod (day length) with genotypic variations (McMaster *et al.*, 2003). The APSIM model uses phyllochron to simulate the dynamic of leaf appearance, leaf area and developmental progress towards flag leaf emergence (or flowering) of wheat (Holzworth *et al.*, 2018). While genetic information has been used to predict wheat phenology parameters in crop models (White *et al.*, 2008; Zheng *et al.*, 2013), no efforts have been made to derive phyllochron and its photoperiod sensitivity from genomic data (e.g. single nucleotide polymorphisms (SNPs)). In this study, we report on using 90K SNPs of wheat to derive parameters for APSIM to predict leaf appearance rate. The new SNP-based model is calibrated and validated using data from controlled environments and field experiments.

Methods

The released version of Wheat model in APSIM NG is developed using the plant modelling framework (PMF) (Brown *et al.*, 2014) The PMF is a collection of generic classes that can be assembled and parameterised in different ways to create different plant models and has been described in detail by Brown et al. (2014) and Brown *et al.*, (2019). In this paper, we only focus on leaf appearance in the early stage (Huan Index < 7). In APSIM NG, phyllochron of wheat is simulated using a base phyllochron (^{o}Cd) at 12 hour photoperiod and sensitivity to photoperiod (i.e. day length). The photoperiod sensitivity is defined as a modifier to base phyllochron, it gradually decreases from 1.6 to 1 when day length changes from 8 to 12 hours.

The observational data were obtained from the wheat validation datasets of APSIM NG, including data from both controlled environment and field experiments conducted at 12 sites in Australia and 2

sites in New Zealand (Palmerston North and Lincoln) (Figure 1). There were 2973 observations of 55 cultivars for leaf appearance including Haun Index (2785), appeared leaf tips (36) and appeared cohort numbers (152), depending on the experiments. Genomic? data were extracted using the 90K Illumina SNP-chip for the 55 cultivars. Missing SNPs were discarded. Finally, there were 8863 SNPs used in the analysis.
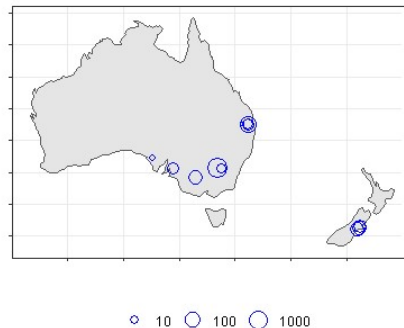

Figure 1. The spatial distribution of observations in the field and controlled environments

The whole dataset was split into two groups for calibration and validation with 13 strategies (i.e. All for all experiments, CE for controlled environments, Other for non-controlled environments, Sample 1-10 for controlled environments plus 50% others in the calibration group). One group of data were used to optimize the parameter values (i.e. base phyllochron and photoperiod sensitivity) for 55 cultivars, and the other group for model validation. A wide range of values for the two parameters were generated with each strategy. The parameter sets with minimum MSV (mean squared variation) (Kobayashi and Salam, 2000) between observed and simulated leaf appearance rates were selected. The use of MSV is to avoid the systematic errors caused by simulations in emergence date. The parameter values of 55 cultivars are used together with SNP data to: 1) develop a random forest model to predict APSIM parameters from SNP, and 2) further using the SNP-derived parameters to run APSIM to simulate leaf appearance rate (APSIM-GS). The two steps are re-optimized for 55 cultivars in one go (i.e. Global Optimization). Consequently, a ML-RF model was developed to generate the base phyllochron and Pp sensitivity of 55 cultivars from SNP data.

Results
There was a good agreement between the observations and predictions in the calibration group (e.g. calibration strategy "All", Figure 2) for APSIM-Release. The MSVs ranged from 0 to 0.7 depending on the cultivars and calibration strategies (Figure 3). A few cultivars in the calibration group had higher MSV?s if controlled environments were used for calibration. In the validation group, stable MSVs were observed across calibration strategies except MSVs when non-CE datasets were used for calibration (i.e. "Other", Figure 3).
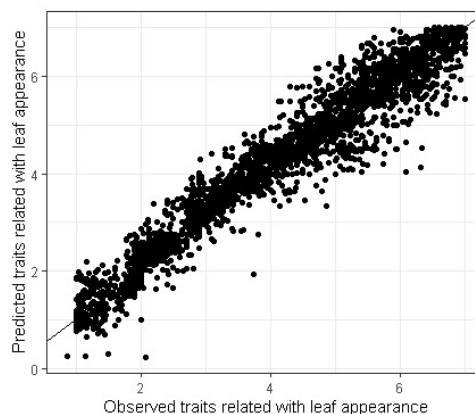
Figure 2. Comparison of observed and predicted traits related with leaf appearance for all cultivars when all datasets are used in the calibration with release model of APSIM Next Generation.
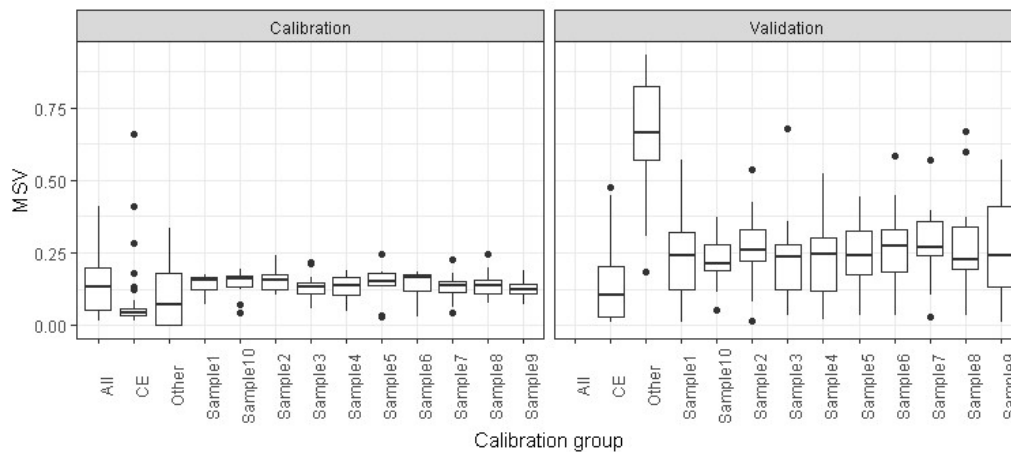

Figure 3. The mean squared variation for traits related with leaf appearances in the calibration and validation groups in 13 calibration strategies when release model of APSIM Next Generation is used.

The performance was not significantly changed when SNP data integrated with processed based model APSIM NG (APSIMNG-GS) for all calibration strategies (Figure 4 and Figure 5). A few outliers were removed in CE datasets in the calibration group. The MSVs ranged from 0 to 0.4 depending on the cultivars and calibration strategies in the calibration group (Figure 5). Similar with APSIMNG-Release, in the validation group, stable MSVs were observed across calibration strategies except MSVs when non-CE datasets were used for calibration (i.e. "Other", Figure 5).

Comparing with APSIMNG-Release which had independently parameter sets for each cultivar, APSIMNG-GS only had one single model and parameter set for all 55 cultivars. The parameter values for APSIMNG-Release (i.e. base phyllochron and photoperiod sensitivity) was estimated from genomic selection models.
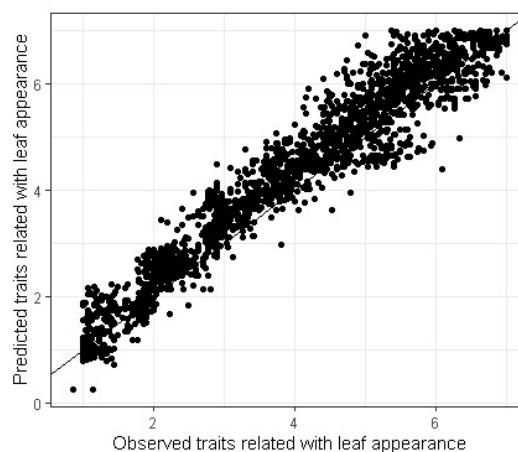

Figure 4. Comparison of observed and predicted traits related with leaf appearance for all cultivars when all datasets are used in the calibration with genomic selection integrated into APSIM Next Generation (APSIMNG-GS).
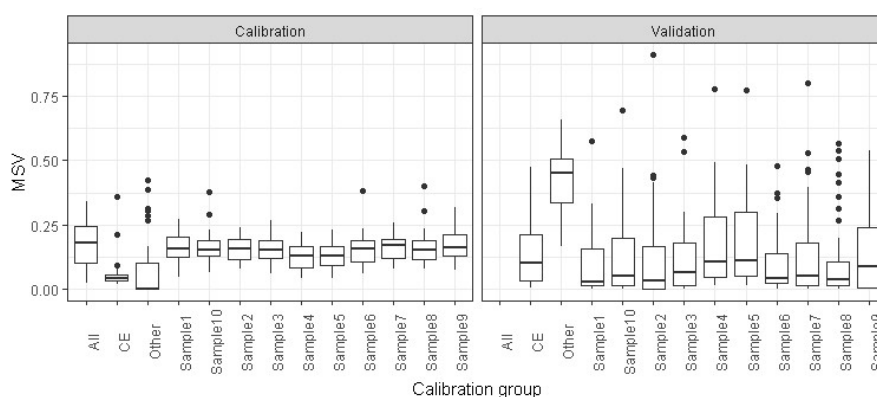
Figure 5. The mean squared variation for traits related with leaf appearances in the calibration and validation groups in 13 calibration strategies when genomic selection is integrated into APSIM Next Generation (APSIMNG-GS).

Conclusion

The traits related with leaf appearance were predicted with release version of APSIM NG and a new SNP based model (genomic prediction + process-based model). Both models can predict leaf appearance of wheat in the diverse environments in different calibration strategies. The whole genomic selection with SNP data is sufficiently robust to predict phenotypic traits in diverse environments.

Acknowledgements

References

Botwright TL, Condon AG, Rebetzke GJ, Richards RA. **2002**. Field evaluation of early vigour for genetic improvement of grain yield in wheat. *Australian Journal of Agricultural Research* **53**, 1137–1145.

Brown TB, Cheng R, Sirault XR, et al. 2014. TraitCapture: genomic and environment modelling of plant phenomic data. *Current Opinion in Plant Biology* **18**, 73–79.

Brown HE, Huth NI, Holzworth DP, Teixeira EI, Wang E, Zyskowski RF, Zheng B. **2019**. A generic approach to modelling, allocation and redistribution of biomass to and from plant organs. *in silico Plants* **1**.

Holzworth DP, Huth N, deVoil PG, et al. 2014. APSIM – evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software* **62**, 327–350.

Holzworth D, Huth NI, Fainges J, et al. 2018. APSIM Next Generation: overcoming challenges in modernising a farming systems model. *Environmental Modelling & Software* **103**, 43–51.

Kobayashi K, Salam MU. **2000**. Comparing Simulated and Measured Values Using Mean Squared Deviation and its Components. *Agronomy Journal* **92**, 345.

McMaster GS, Wilhelm WW, Palic DB, Porter JR, Jamieson PD. **2003**. Spring Wheat Leaf Appearance and Temperature: Extending the Paradigm? *Annals of Botany* **91**, 697–705.

White JW, Herndl M, Hunt LA, Payne TS, Hoogenboom G. **2008**. Simulation-based analysis of effects of *Vrn* and *Ppd* loci on flowering in wheat. *Crop Science* **48**, 678–687.

Zheng BY, Biddulph B, Li D, Kuchel H, Chapman SC. **2013**. Quantification of the effects of *VRN1* and *Ppd-D1* to predict spring wheat (*Triticum aestivum*) heading time across diverse environments. *Journal of Experimental Botany* **64**, 3747–3761.

Zheng BY, Chenu K, Dreccer MF, Chapman SC. **2012**. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (*Triticum aestivum*) varieties? *Global Change Biology* **18**, 2899–2914.