Genomic prediction using multi-year data: a case study in a hybrid maize breeding program

Kaio Olímpio das Graças Dias[12]

Single-cross hybrids have been used to explore heterosis in selfing and outcrossing species. Because it is infeasible to obtain and evaluate all possible pairwise inbred combinations, predicting the performance of untested single-cross hybrids is essential to increase genetic gains in breeding programs. Highly unbalanced historical data, with limited connectivity of hybrids evaluated across multiple years, is a key component of many commercial breeding programs that should be considered for efficient implementation of genomic prediction (GP) in crop improvement. In addition, if genotype-by-year effects are not properly modeled, GP can divert part of the molecular marker information to predict these effects rather than to estimate breeding values. In this context, appropriate statistical models that account for genetic and residual correlations across environments, years and deals with unbalanced data is required. Moreover, in species with high level of heterosis, such as maize, it is important that the GP models consider not only additive genetic effects, but also the non-additive genetic effects to the prediction the performance of untested hybrids. Data of hybrids from the ongoing Embrapa's maize breeding program evaluated from 2006 to 2013 was used to compare the predictive ability of hybrid performance under different scenarios and statistical models. For this, a statistical-genetics model that accounts for the effects of locations, seasons, regions and years plus additive and dominance effects was proposed via stage-wise approach. Our results discussed novel strategies to improve predictive ability for GP of untested single-cross maize hybrids using unbalanced historical data. The stage-wise analysis performed in this study may be applied for any crop when historical unbalanced data are available. Likewise, this approach has potential to reduce costs and accelerate the release of new hybrids.

---

[1] *Departamento de Genética, Escola Superior de Agricultura "Luiz de Queiroz"-USP*
[2] *joint work with Hans-Peter Piepho, L.J.M Guimarães, P.E.O. Guimarães, S.N. Parentoni, C.T. Guimarães, A.A.F. Garcia & M.M. Pastina*

A complex trait with unstable QTLs can follow from component traits with stable QTLs: an illustration by a simulation study in pepper

Paulo Canas Rodrigues[12]

Complex traits are traits whose phenotypic variation is driven by a set quantitative trait loci (QTLs) that are typically environment dependent. The environment dependence of complex traits can be observed at the phenotypic level as genotype-by-environment interaction (GEI) and at the genetic level as QTL-by-environment interaction (QEI). Genetic improvement of complex traits requires strategies for dealing with GEI and QEI. We illustrate a strategy for modelling of GEI and QEI in complex traits that departs from dissection of a target complex trait in a number of component traits, where each of the component traits is purely genotype dependent. An eco-physiological genotype-to-phenotype model converts the set of genotype specific component traits into the complex target virtual phenotype by integrating the components with environmental inputs over the duration of the growing season. For component traits with a simple genetic basis, consisting of a few additive QTLs, an attractive scenario for marker assisted selection of the corresponding complex trait appears. First, identify the QTLs for the components. For new genotypes, then use molecular markers linked to the QTLs to predict the phenotypes for the components. Subsequently, use an appropriate genotype-to-phenotype model to integrate the components with environmental inputs to produce predictions for the complex target virtual phenotype. In this paper, we demonstrate the viability of our modeling approach for complex traits by a case study in sweet pepper (Capsicum annuum L.). We developed a seven component eco-physiological model for yield in pepper and simulated for a back cross population yield and yield components, where the yield components were given a simple QTL basis. We show how credible patterns of GEI and QEI for yield can be simulated from genotype specific yield components with a simple QTL basis. Our results can be instrumental in breeding strategies for the improvement of complex traits.

---

[1] *Departamento de Estatística, Universidade Federal da Bahia*
[2] *joint work with E. Heuvelink, L.F.M. Marcelis, S. Chapman and & F.A. van Eeuwijk*

Robust estimation in plant breeding: evaluation using simulation and empirical data

Vanda Marisa da Rosa Milheiro Lourenço[12]

Background. Genomic prediction (GP) is used in animal and plant breeding to help identify the best genotypes for selection. One of the most important measures of the e ectiveness and reliability of GP in plant breeding is predictive accuracy. An accurate estimate of this measure is thus central to GP. Moreover, regression models are the models of choice for analyzing eld trial data in plant breeding. However, models that use the classical likelihood typically perform poorly, often resulting in biased parameter estimates, when their underlying assumptions are violated. This typically happens when data are contaminated with outliers. These biases often translate into inaccurate estimates of heritability and predictive accuracy, compromising the performance of GP. Since phenotypic data are susceptible to contamination, improving the methods for estimating heritability and predictive accuracy can enhance the performance of GP. Robust statistical methods provide an intuitively appealing and a theoretically well justied framework for overcoming some of the drawbacks of classical regression, most notably the departure from the normality assumption. We compare the performance of robust and classical approaches to two recently published methods for estimating heritability and predictive accuracy of GP using simulation of several plausible scenarios of random and block data contamination with outliers and commercial maize and rye breeding datasets.

---

[1] *Departamento de Matemática, Universidade Nova Lisboa*
[2] *joint work with E. Heuvelink, L.F.M. Marcelis, S. Chapman and & F.A. van Eeuwijk*