Association study and multivariate genomic prediction for simultaneous improvement of grain yield and grain protein content in durum wheat

Vitale P[1], Esposito S[1], Taranto F[2], Saia S[3], Pecorella I[1], Rodriguez M[4], Natoli V[5], Pecchioni N[1], De Vita P[1*]

[1]Research Centre for Cereal and Industrial Crops (CREA-CI), CREA—Council for Agricultural Research and Economics, 71122 Foggia, Italy. [2]Institute of Biosciences and Bioresources (CNR-IBBR), 80055 Portici, Italy. [3]Department of Veterinary Sciences, University of Pisa, 56129 Pisa, Italy. [4]Department of Agriculture, University of Sassari, Viale Italia, 39, 07100, Sassari, Italy [5]Genetic Services SRL Contrada Catenaccio, snc 71026 Deliceto (FG) Italy

Introduction

Obtaining genotypes with high grain yield (GY) and protein content (GPC) represents the complete goal for durum wheat breeding. However, the negative correlation between these two traits represents the major obstacle. Thus, we investigated the genetic basis of GY and GPC along with derivative indices (grain yield deviation: GYD and grain protein deviation: GPD) to identify independent associations (QTNs). A genomic prediction (GP) framework was also implemented using both uni- and multivariate approaches (UV and MV) to assess their accuracy.

GWAS

Six multi-locus GWAS models were used to identify SNPs associated with the traits under study (Figure 1).

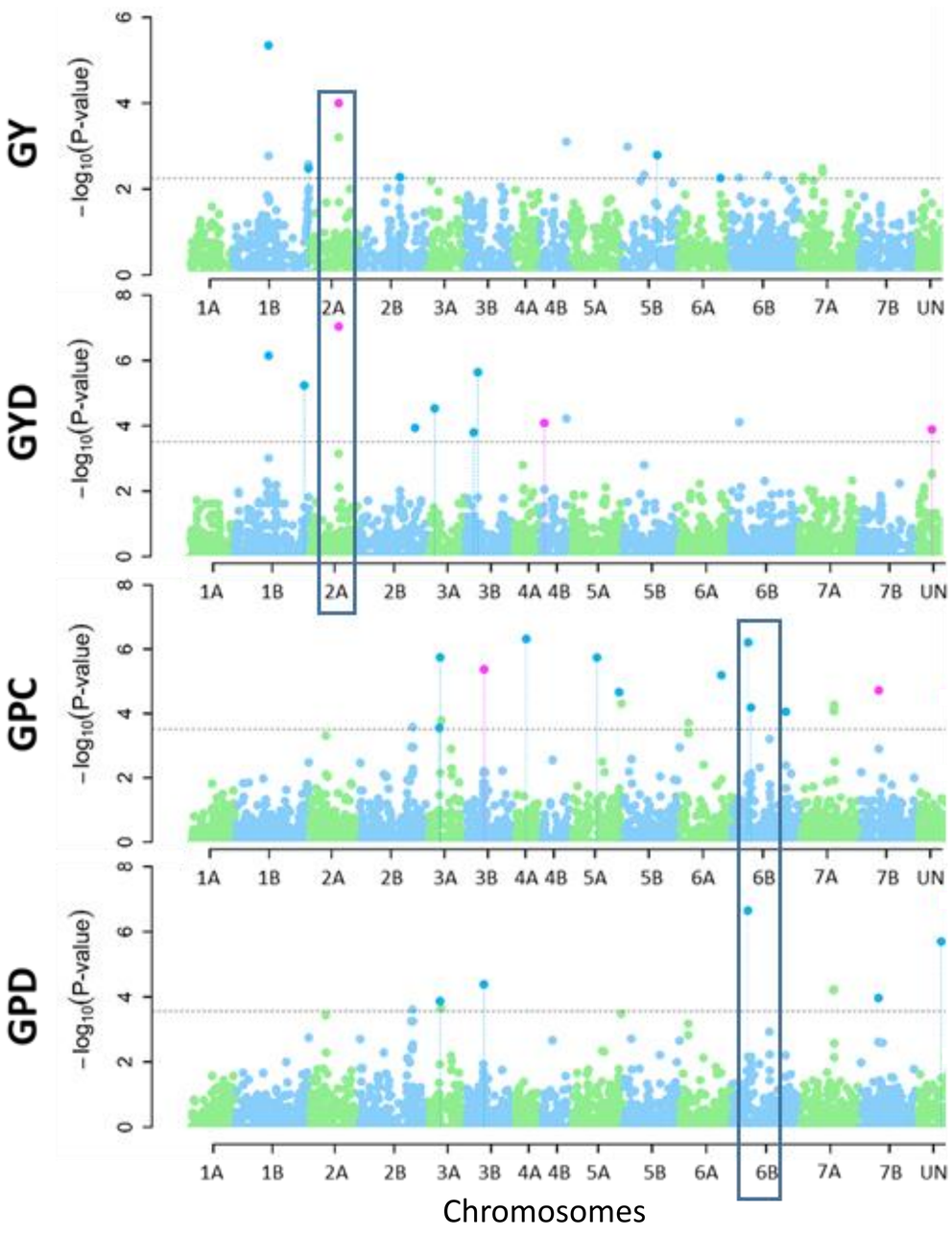

Figure 1. Manhattan plot showing significant QTNs (LOD > 3). Reliable QTNs (more than two models) are shown with pink dots.

A reliable QTN (r^2>10) on chr. 2A was shared only for GY and GYD and co-mapped with a QTL identified by Mengistu et al. (2016). Similarly, another QTN on chr. 6B was found specific for GPC and GPD but not for GY and GYD.

Genomic Prediction

Univariate GBLUP was performed for all traits. Subsequently, multivariate GBLUP was applied in two cross-validation schemes (CV1 and CV2) (Figure 2).

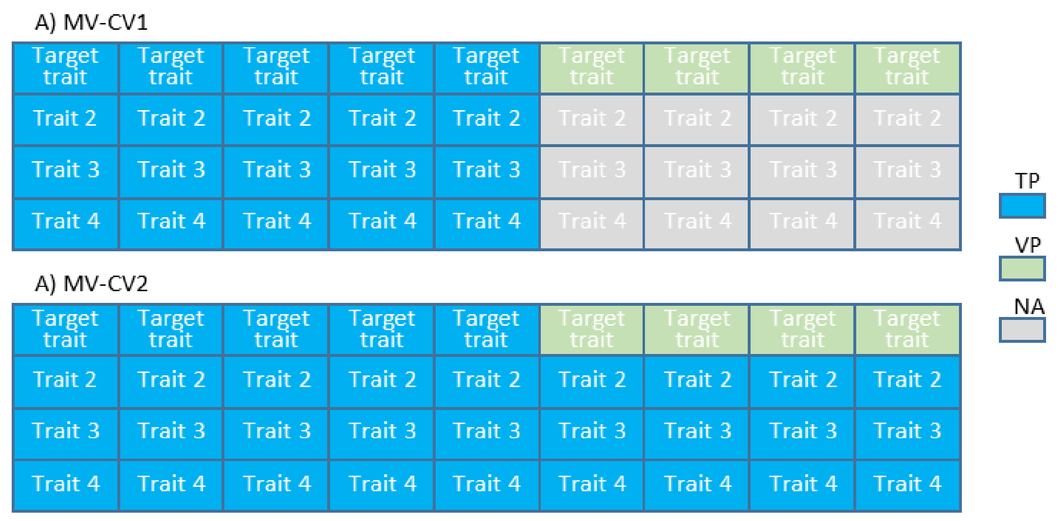

Figure 2. Cross-validation schemes, TP: training population, VP: validation population, NA: not available

For all traits, the multivariate CV1 was ineffective, even reducing prediction accuracy (PA). By contrast, by employing multivariate CV2, PA increased for all traits (up to 30%), except for GYD (Figure 3).

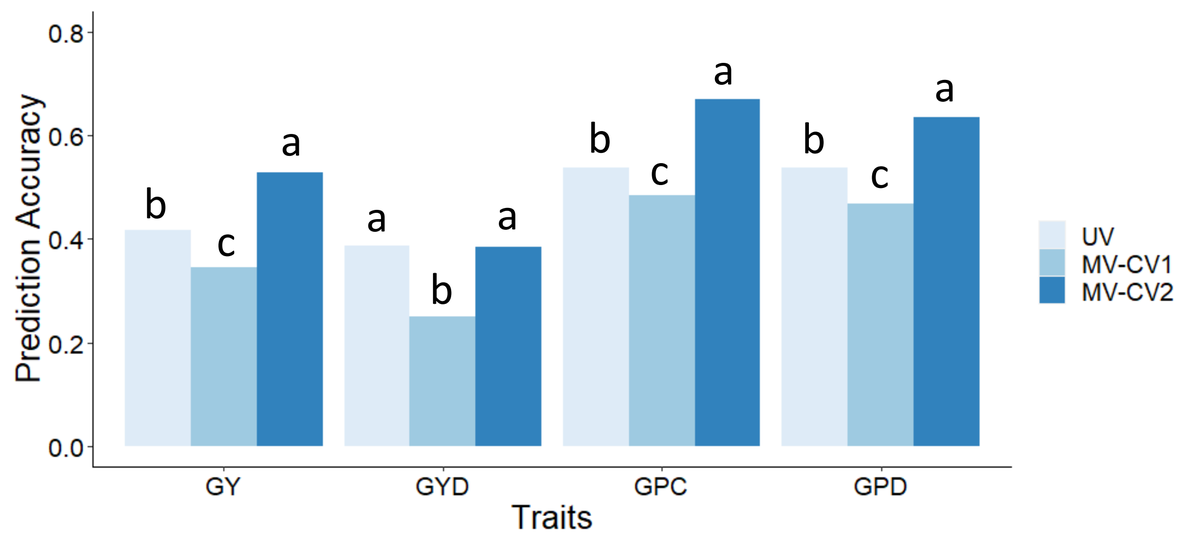

Figure 3. Prediction accuracy for all traits

Conclusions

The ML-GWAS revealed independent QTNs for GY and GPC in the same regions previously described in the literature. We recognized that CV1 was not useful to improve PA, by contrast, CV2 was effective to increase PA for almost all traits by exploiting the additional phenotypical information from other traits.

Email: paolo.vitale@unifg.it

References

D. K. Mengistu et al. Plant Biotechnol. J. 14 18006812 (2016).