GWAS-Assisted Genomic Prediction for Forage Yield Under Deficit Irrigation Management in Elite Alfalfa (*Medicago sativa* L.) Germplasm

Harpreet Kaur, New Mexico State University
Cesar Medina Culma, USDA-ARS
Christopher Pierce, New Mexico State University
Arshdeep Singh Gill, New Mexico State University
Jon Reich, Alforex Seeds
Steve Damon, Corteva Agriscience
Long-Xi Yu, USDA-ARS
Ian Ray, New Mexico State University

Arid and semiarid regions of the U.S. possess limited water resources for irrigation. A major objective for alfalfa (*Medicago sativa* L.) breeding programs in these environments is to develop cultivars that can remain productive under highly variable soil moisture conditions. The goals of this study were to identify genomic regions, and to evaluate genomic prediction (GP) models, associated with forage yield in an elite New Mexico Genome Selection (NMGS) alfalfa population grown under deficit irrigation management. Half-sib families derived from 200 to 215 semi- and non-dormant NMGS plants were evaluated for forage yield over multiple harvests in 2018, 2019 and 2020 near Las Cruces, NM and Sloughhouse, CA under summer irrigation termination and continuous deficit drip irrigation management, respectively. Among the maternal parents of the NMGS half-sib families evaluated, 12,884 genome-wide single nucleotide polymorphism (SNP) markers were identified using a *Medicago truncatula* reference genome assembly. These SNP markers, and best linear unbiased estimates for forage yield, were used to conduct a genome-wide association study (GWAS) based on six polyploid gene action models as provided by the R-package, GWASpoly. Few markers (4) were significantly associated with forage biomass in NM and CA. All SNP markers were also used for GP analyses for individual harvests and seasonal totals according to ridge regression BLUP (rrBLUP), genomic BLUP (GBLUP), support vector machine (SVM)-linear, SVM-gaussian, random forest, Bayes A, Bayes B, Bayes C, and Bayesian Lasso methods. For all GP methods, the prediction accuracies (i.e., coefficient of correlation, r) between predicted and observed yield were low to moderate. For instance, average r ranged between -0.11 to 0.32 for the rrBLUP model at the two locations. GWAS-assisted GP using a subset of SNP markers possessing GWASpoly scores>1 greatly enhanced prediction accuracies for all methods (e.g., r for rrBLUP ranged from 0.74 to 0.88). Additionally, moderate to high prediction accuracies ranging from 0.52 to 0.82 over all harvests in two locations were obtained using a weighted-GBLUP method with a complete set of markers weighted according to their GWASpoly scores. The data were also analyzed using a recently developed *Medicago sativa* reference genome assembly and similar results were obtained.