

Accuracy of Eight Genomic Selection Models in the Prediction of Salt Tolerance in Alfalfa

Charles Hawkins, USDA-ARS

Long-xi Yu, USDA-ARS

Soil salinity is a growing problem in world agriculture. In 2014, saline soil cost an estimated \$27 B in lost crop yields worldwide. Continued improvement in crop salt tolerance will require the implementation of new breeding technologies such as genomic selection (GS). This technology utilizes machine learning to predict breeding values for candidates under selection, making selections based on these predictions. GS offers high accuracy and the potential for gains that are more rapid than those of phenotypic selection while being more sustainable than those of other methods. In this work, we report the results of cross-validation of eight GS models on a population of alfalfa with three different phenotypic datasets related to salt tolerance: Yield under salt stress in a field, general health under salt stress in a field, and a set of health and productivity metrics collected from salt-exposed plants grown in a greenhouse. The highest-performing model, SVR with nu-regression, achieved an accuracy of 0.43 for the field health (Fig 1). This level of accuracy is sufficient to make gains in a real-world breeding program. Grid searches were performed to optimize the parameters of the models, but this did not result in improvements over the default values (Fig 2). Multiple levels of stringency for required reads to call a locus were also tested. Less stringency resulted in higher accuracy down to 5 reads per sample. Lowering stringency further than this increased compute time but did not affect accuracy (Fig 3, 4). Association mapping was also performed on the health and yield datasets. The most-significant markers among these data had a $-\log_{10}$ p-value of 4.5 (Fig 5).

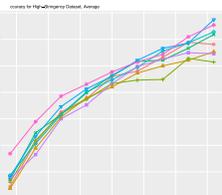


Fig 1:

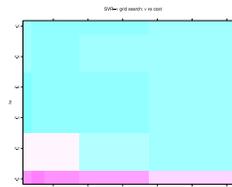


Fig 2:

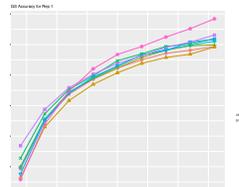


Fig 3:

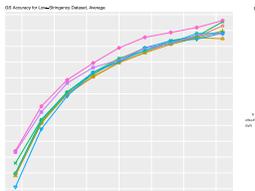


Fig 4:

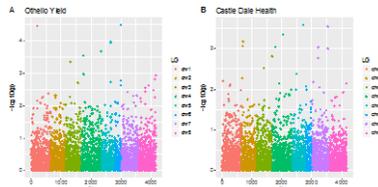


Fig 5: