

Genomic Prediction and Genotype x Environment characterization in unbalanced historical cereal variety trials



Madhav Bhalta, Lucia Gutierrez
University of Wisconsin-Madison, Madison, Wisconsin, USA

Introduction

Utilization of historical multi-year, multi-location (environment) variety trials could help in developing robust genomic prediction models enhancing genetic selection in breeding programs. However, historical variety trial data are generally highly unbalanced as the same sets of genotypes may not be tested across several years and locations. We examined multiple strategies to utilize unbalanced historical datasets from oat and wheat breeding programs in the Midwestern region of the United States to characterize genotype x environment interactions, identify mega-environments with similar ranking of genotypes across multiple environments, and perform genomic selection integrating GxE interactions in each environment, mega-environment, and all environments combined.

Goals and Objectives

- This project was designed to build a variety decision tool for small grains such as oat and wheat for each geographical region in the Mid-Western region of the United States.
- Deploy the prediction model in a web-based tool for selecting the top-ranking lines in a specific location using a Zip code system.



Materials and Methods: Oat example



Fig. 1. 70 locations and 11 states (IA, IL, IN, MI, MN, ND, NE, WI, OH, SD, WY) from 16 states in 11 states were used in this study.

Data collected

- Oat lines: 2,903
- Year: 1996-2018 (23 years)
- phenotyped for yield 1,021 lines
- Genotype with 10,077 GBS-derived SNP markers
- Historical variety testing data were from Triticeae Toolbox (T3/Oat; <https://triticeae.toolbox.org/>)

Analysis: Filling-up the missing information

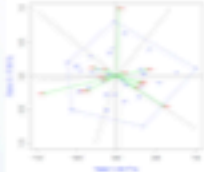


Fig. 2. Data available from T3/Oat.



Fig. 3. Historical oat lines.

- Mixed model with additive relationship matrix and environmental covariates were fitted to predict performance for all genotype (Lado et al., 2016)
- GGE biplot was used to identify mega-environments based on outstanding genotypes (Gade et al., 2000, 2002)



Funding: USDA-SARE-NCR



Results

Fig. 4. Correlation heatmap among oat growing environments across multiple years using long-term historical data.

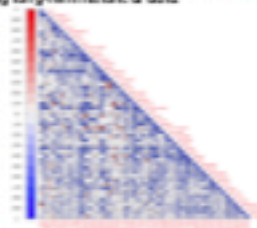


Fig. 5. Map showing each location nested and grouped into one of the six Mega-environments identified using GGE biplot analysis.



Fig. 6. Six Mega-environments with similar ranking within each geographical region were identified from oat yield from using GGE biplot.



Table 1. Example showing predicted top performing oat cultivars for a site in Iowa, U.S.

Genotype	Location	Zip code	Origin	Released	Ranking
MN-Pearl	Nashua, IA	50658	UMN	2019	1
DEON	Nashua, IA	50658	UMN	2014	2
SADDLE	Nashua, IA	50658	SDS	2019	3

Fig. 7. Use Zip-code to identify top cultivar for the area.



Conclusions

- The variety selector tool will help growers and scientists in identifying best ranking genotypes.
- This study will help in developing breeding strategies in the Mid-west region.
- Provide further insights on utilizing large historical datasets.

Future Directions

- Divide the GxE models with environmental covariates to predict "new" environments
- Increase data area coverage by having growers/scientist evaluate 3-3 cultivars including the best predicted and the usually planted for each farm.
- Use farmers' results to improve the GxE models by increasing location coverage even at lower depth.

