Exploring GxE and Genomic Prediction in a Two-Row Barley Pilot Study

Jeffrey L. Neyhart[1] and **Kevin P. Smith**[1]
[1]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Introduction

- Brewers demand locally-grown, two-row malting barley (*Hordeum vulgare* L.)
- Interested regions are often diverse and not serviced by a breeding program
- Genotype-by-environment interactions (GxE) hinder breeding for a wide set of target population of environments[1]
- Various approaches to model GxE in genomic selection show promise, but use highly unbalanced datasets[2, 3, 4]

Objectives

1. **Assess GxE in multi-environment two-row barley genomic selection trials**
2. **Investigate the cross-validation and progeny prediction accuracy between different environments**

Materials & Methods

Population & Phenotyping

- 183-line spring two-row barley training population (S2TP) from five U.S. breeding programs:
  - **USDAA/ARS – Aberdeen (AB)**
  - **Montana State University (MT)**
  - **Washington State University (WA)**
  - **Busch Agricultural Resources (BA)**
  - **North Dakota State University (N2)**
- Prediction set of 50 randomly chosen progeny lines (C1R)
- Field trials in 11 location-year environments (**Table 1**)

Abbreviation	Location
STP	St. Paul, MN
STPFHB	St. Paul, MN (FHB nursery)
CR	Crookston, MN
CRFHB	Crookston, MN (FHB nursery)
CD	Caldwell Farm, Ithaca, NY
KT	Ketola Farm, Ithaca, NY
HF	Helfer Farm, Ithaca, NY
BZ	Bozeman, MT
HT	Huntley, MT

Table 1. Abbreviations and the full information for locations used in trials of the S2TP and C1R.

Genotyping

- All experimental lines were genotyped using genotyping-by-sequencing[5]
- A novel pipeline for read mapping and variant calling was used that employed publically available bioinformatics tools
- Missing data was imputed by the multivariate normal – expectation maximization method implemented in the "rrBLUP" R package [6, 7]

Visualizing and Quantifying GxE

- Additive main effects and multiplicative interaction (AMMI) analysis employed to quantify GxE via the "agricolae" R package [7, 8] and visualized by bi-plots

Measuring Prediction Accuracy

- Marker effects estimated using ridge regression – best linear unbiased prediction (RR-BLUP) via "rrBLUP" R package[6, 7]
- Cross-validation performed between all pairs of environments, as outlined in **Figure 1**, with 500 repetitions
- Measured S2TP – C1R prediction accuracy

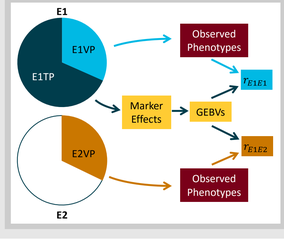


Figure 1. Pairwise environmental cross-validation scheme. In one environment (E1), a random 70% of S2TP lines form a training population (E1TP), which is used to estimate marker effects. The remaining 30% of lines form the validation population (E1VP) and the same lines from a second environment (E2) form another validation population (E2VP). Genotype data of E1VP lines is used to generated genomic estimated breeding values (GEBVs). The GEBVs are correlated with the observed phenotypes of the E1VP and E2VP to obtain the accuracy measures r_{E1E1} and r_{E1E2}, respectively.

Results

Phenotypic data reflected the trait heritability

Figure 2. Density plots of the phenotypic data collected on the S2TP and C1R. Traits included days to heading and height (cm). Plots are split by breeding program/state, and each environment is color-coded. Note that some environments for days to heading included Fusarium headblight (FHB) nurseries. These are designated FHB in the environment abbreviation.

GxE interaction effects were larger for height than heading in the S2TP

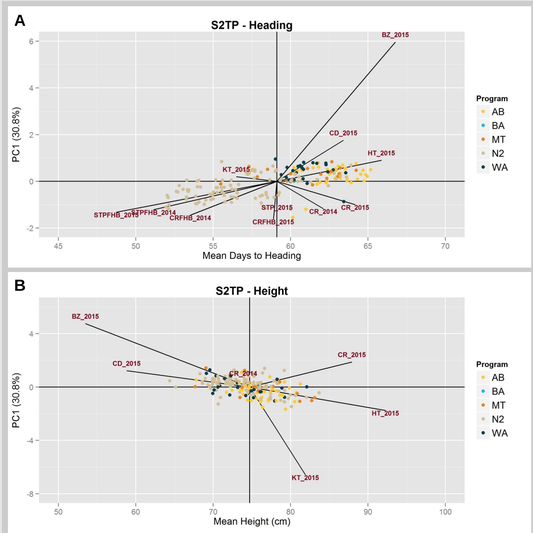
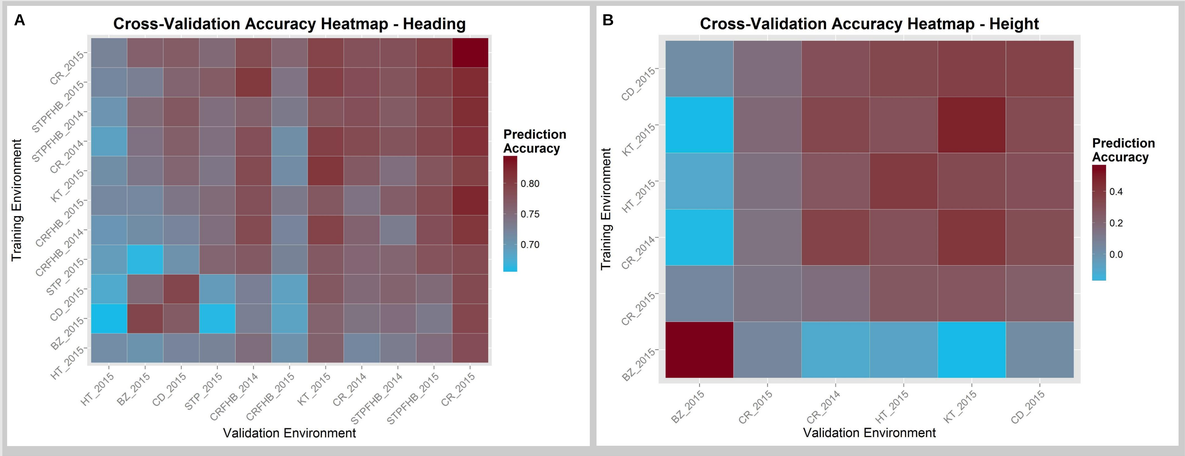


Figure 3. Additive main effects and multiplicative interaction (AMMI) analysis bi-plots for S2TP heading (**A**) and height (**B**) data. The abscissa depicts the mean trait value and the ordinate plots the first principal component (PC1). Environments are displayed as text and genotypes as points, colored according to origin breeding program. Distance from the coordinate of either an environment or a genotype to the line PC1 = 0 measures the interaction effect.

No environment was generally a better predictor for heading, but some were consistent in validation ability. For height, some environments were both inaccurate predictors and validators



Progeny prediction accuracy for heading depended more on the environment being predicted, while for height the opposite was true – training environment was more important

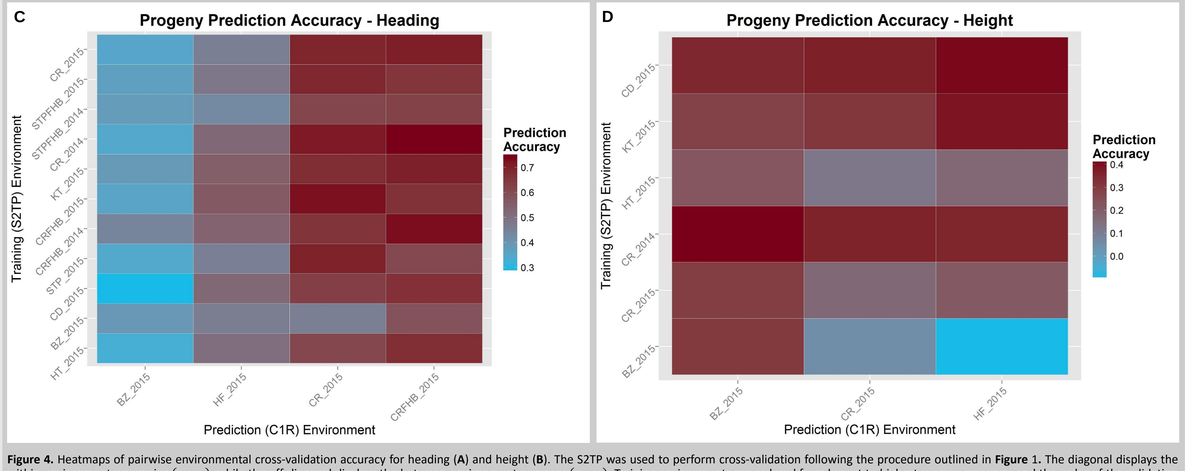


Figure 4. Heatmaps of pairwise environmental cross-validation accuracy for heading (**A**) and height (**B**). The S2TP was used to perform cross-validation following the procedure outlined in **Figure 1.** The diagonal displays the within-environment accuracies (r_{E1E1}), while the off-diagonal displays the between-environment accuracy (r_{E1E2}). Training environments are ordered from lowest to highest average accuracy, and the order of the validation environments mirrors that of the training environments. The progeny prediction accuracies (i.e. the predictive ability of the S2TP on the C1R) are also displayed as heatmaps for heading (**C**) and height (**D**). The order of training environments in **C** and **D** mirrors that in **A** and **B**, respectively. In all heatmaps, higher accuracies are shown in red, while lower accuracies are shown in blue. Environments that were better predictors have consistently high accuracy across a row, while environments that were better predicted have consistently high accuracy down a column.

Discussion

Phenotypic Data

- Heading showed bimodal distributions in most S2TP trials (**Figure 2A**) and reflected the structure of the population
- C1R displayed a similar heading distribution (**Figure 2C**), likely due to structure in the population
- Height was normally distributed for the S2TP (**Figure 2B**) and C1R (**Figure 2D**), consistent with its quantitative nature

GxE Interactions

- Nominal environment interaction effects were observed for heading, with the exception of BZ_2015 (**Figure 3A**)
- Most environments displayed interaction effects for height (**Figure 3B**), but without a dominating environment

Cross-Validation (CV) Between Environments

- Heading was more accurately predicted ($\bar{x}_{r_{E1E1}} = 0.780$) by CV within environments than height ($\bar{x}_{r_{E1E1}} = 0.383$)
- CV accuracy between environments for heading (**Figure 4A**) depended on the validation environment
- Pairs of environments generally displayed reciprocal accuracy for height (**Figure 4B**)
- AMMI distance between environments was a significant predictor of CV accuracy for height (r = -0.499, p < 0.01)

Practical Progeny Prediction

- S2TP – C2R prediction accuracy for heading (**Figure 4C**) was largely dependent on the progeny environment
- S2TP – C2R prediction accuracy for height (**Figure 4D**) depended more on the training environment

Conclusions

1. **GxE interactions were more apparent in the lower-heritability trait**
2. **GxE models may help inform the prediction accuracy between environments**
3. **Progeny prediction is less accurate than cross-validation, but follows trait patterns**

Future Direction

- Trait and marker data will be publically available (T3, triticeaetoolbox.org/barley/) to support the breeding of locally-adapted two-row barley varieties
- S2TP and the C1R will be grown in a total of 54 location-year environments from 2014 to 2017 (**Figure 5**)
- Other traits such as yield, maturity, and lodging will be measured
- Environmental clustering and prediction accuracy, as well as training population optimization, will be explored

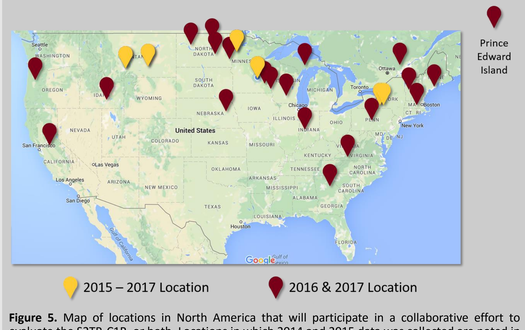


Figure 5. Map of locations in North America that will participate in a collaborative effort to evaluate the S2TP, C1R, or both. Locations in which 2014 and 2015 data was collected are noted in gold. These will also be locations in 2016 and 2017. All other locations are noted in maroon.

Acknowledgements

Special thanks to **Ed Schiefelbein, Guillermo Velasquez, Karen Beaubien, Tyler Tiede, Celeste Falcon, Alex Ollhoff, Lu Yin**, and all other lab members. Thanks also to **Tom Kono, Michael Kantar, and Paul Hoffman**. A debt of gratitude to **Kevin Smith** for encouragement, patience, and advising during my current and ongoing graduate work.

Collaborators:

I am indebted to the support of our collaborators, including **Shiaoman Chao** at USDA/ARS – Fargo, **Mark Sorrells, David Benscher, and Daniel Sweeney** at Cornell University, and **Jamie Sherman, Ken Kephart, and Liz Elmore** at Montana State University.

Funding:

This research is supported by the Minnesota Department of Agriculture and Rahr Malting Co.

References

[1]Bernardo, R. 2010. Breeding for Quantitative Traits in Plants. 2nd ed. Stemma Press, Woodbury, Minnesota.
[2]Lado, B., P.G. Barrios, M. Quincke, P. Silva, and L. Gutiérrez. 2015. Modeling genotype by environment interaction for genomic selection with unbalanced data from a wheat (Triticum aestivum L.) breeding program. Crop Sci.
[3]Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.-L. Jannink, R.P. Singh, E. Autrique, and G. de los Campos. 2015. Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker x Environment Interaction Genomic Selection Model. G3; Genes|Genomes|Genetics 5(4): 569–82.
[4]Heslot, N., D. Akdemir, M.E. Sorrells, and J.L. Jannink. 2013. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor. Appl. Genet.: 1–18.
[5]Elshire, R.J., J.C. Glaubitz, Q. Sun, J. a. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6(5): e19379.
[6]Endelman, J.B. 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome J. 4(3): 250.
[7]R Development Core Team 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
[8]Felipe de Mendiburu (2015). agricolae: Statistical Procedures for Agricultural Research. R package version 1.2-3. http://CRAN.R-project.org/package=agricolae

Scan for PDF and additional information