**Supplemental Material**

Environment-Specific Genomic Prediction Ability in Maize Using Environmental Covariates Depends on Environmental Similarity To Training Data

Anna R. Rogers and James B. Holland

Histogram

Description automatically generated with low confidence

**Figure S1.** Increase in prediction ability within environments by adding G×E effects using the PCA(Markers)\* Env GxE term to a model with 10,154 marker dominance coefficients and 5-day window weather variable summaries. Prediction abilities are compared within training sets including 40% to 90% of the full data set, sampled at random.

Chart, box and whisker chart

Description automatically generated

**Figure S2.** Distributions of slopes of regression of observed values on predicted values within each environment and test set for models including 10,153 marker dominance coefficients and 377 environmental covariates, along with no G×E effects or G×E effects computed using principal components of the environmental data (PCA(Env)\*Markers) or using principal components of the marker data (PCA(Markers)\*Env). Each model was evaluated in test sets selected by random sampling (CV1), partial replication across environments (CV2), leaving out a single year of data (LO1Y), stratification by environment clusters (LORE), by hybrid clusters (LORH), leaving out single environments (LO1E), and bidirectional censoring schemes leaving out both a year and 10% of hybrids (CV1 + LO1Y), a year and related hybrids (LORH + LO1Y), environment clusters with 10% of hybrids (CV1 + LORE), and environment and hybrid clusters (LORH + LORE).

**Table S1.** Descriptions of soil parameters taken from SSURGO metadata column descriptions

|  |  |
| --- | --- |
| **Parameter** | **Description** |
| slope\_r | Slope gradient of field |
| tfact | Soil loss tolerance factor, denoting the maximum amount of erosion where the soil can be considered a medium for plant growth |
| sieveno40\_r | Fraction of soil that passes through a number 40 sieve (0.42 mm square opening) |
| sieveno200\_r | Fraction of soil that passes through a number 200 sieve (0.074 mm square opening) |
| sandvc\_r | Percentage of very coarse sand (1.0-2.0 mm diameter) in sample |
| sandco\_r | Percentage of coarse sand (0.5-1.0 mm diameter) in sample |
| sandmed\_r | Percentage of medium sand (0.25-0.5 mm diameter) in sample |
| sandfine\_r | Percentage of fine sand (0.10-0.25 mm diameter) in sample |
| sandvf\_r | Percentage of very fine sane (0.05-0.10 mm diameter) in sample |
| siltco\_r | Percentage of coarse silt (0.02-0.05 mm diameter) in sample |
| siltfine\_r | Percentage of fine silt (0.002-0.02 mm diameter) in sample |
| claytotal\_r | Percentage of clay (particle size < 0.002 mm diameter) in sample |
| ksat\_r | Amount of water that moves through a unit area of saturated soil in unit time |
| awc\_r | Amount of water available to plants, adjusted |
| wthirdbar\_r | Volumetric content of soil water retained at a 1/3 bar tension |
| wfifteenbar\_r | Volumetric content of soil water retained at a 15 bar tension |
| kwfact | Quantification of susceptibility of soil particles to detachment and movement by water to measure erosion factors |
| caco3\_r | Quantity of Carbonate () expressed as |
| gypsum\_r | Percent by weight of hydrated calcium sulfate in sample |
| sar\_r | Relative amount of Sodium to Calcium and Magnesium extracted from water saturated soil sample |
| cec7\_r | Amount of exchangeable cations at pH 7.0 |
| sumbases\_r | The sum of extractable bases at pH 7.0 |
| extracid\_r | Measure of soil exchangeable hydrogen ions |
| resdepb\_r | Distance from soil surface to lower boundary of restrictive layer |

**Table S2**. Training and test set sample sizes, minima, maxima, and quantiles of the distribution of genetic similarities and environmental covariable similarities between training and test sets for each fold of each cross validation scheme. See “Table\_S2\_Train\_Test\_Relationships.xlsx”

**Table S3.** Across-environment prediction accuracy computed as correlation between BLUEs and predicted values for models including 10,153 marker dominance coefficients and 377 environmental covariates, along with no G×E effects or G×E effects computed using principal components of the environmental data (PCA(Env)\* Markers) or using principal components of the marker data (PCA(Markers) \* Env) for each sampling method. Bidirectional schemes are shaded grey.

|  |  |  |  |
| --- | --- | --- | --- |
| Method | G + E | PC(Env) \* Markers | PC(Markers) \* Env |
| CV1 | 0.777 | 0.794 | 0.808 |
| CV2 | 0.786 | 0.803 | 0.819 |
| LO1Y | 0.286 | 0.267 | 0.262 |
| LORE | 0.216 | 0.207 | 0.203 |
| LORH | 0.734 | 0.743 | 0.750 |
| LO1E | 0.476 | 0.480 | 0.470 |
| CV1 + LO1Y | 0.246 | 0.275 | 0.236 |
| LORH + LO1Y | 0.252 | 0.205 | 0.237 |
| CV1 + LORE | 0.161 | 0.171 | 0.173 |
| LORH + LORE | 0.167 | 0.158 | 0.175 |

**Table S4.** Mean within-environment prediction accuracies for models including 10,153 marker dominance coefficients and 377 environmental covariates, along with no G×E effects or G×E effects computed using principal components of the environmental data (PCA(Env)\* Markers) or using principal components of the marker data (PCA(Markers) \* Env) for each sampling method. Bidirectional schemes are shaded grey.

|  |  |  |  |
| --- | --- | --- | --- |
| Method | G + E | PC(Env) \* Markers | PC(Markers) \* Env |
| CV1 | 0.436 | 0.482 | 0.517 |
| CV2 | 0.464 | 0.510 | 0.547 |
| LO1Y | 0.363 | 0.365 | 0.353 |
| LORE | 0.452 | 0.430 | 0.439 |
| LORH | 0.296 | 0.311 | 0.307 |
| LO1E | 0.476 | 0.480 | 0.470 |
| CV1 + LO1Y | 0.347 | 0.352 | 0.340 |
| LORH + LO1Y | 0.252 | 0.258 | 0.234 |
| CV1 + LORE | 0.401 | 0.380 | 0.386 |
| LORH + LORE | 0.285 | 0.278 | 0.269 |