

Table S4 – Difference in accuracy from Ames×PHZ51 to Ames×B47 for prediction in NAM-H

Sampled hybrids in training set (<i>n</i> =235)	Trait	Difference in prediction accuracy (p-value)		
		All origins	Non-tropical origin	Tropical origin
Same female lines	DTS	-0.007 (0.63)	+0.005 (0.84)	-0.018 (0.37)
	PH	-0.030 (0.27)	-0.032 (0.51)	-0.028 (0.39)
	GY	+0.035 (0.28)	-0.041 (0.24)	<u>+0.100 (0.048)</u>
Random female lines	DTS	-0.041 (0.076)	-0.024 (0.46)	-0.056 (0.10)
	PH	<u>-0.096 (7.7×10⁻⁵)</u>	-0.076 (0.068)	<u>-0.113 (1.0×10⁻⁴)</u>
	GY	<u>+0.082 (9.1×10⁻³)</u>	+0.032 (0.40)	<u>+0.124 (0.010)</u>

Differences in prediction accuracy: mean difference in accuracy from Ames×PHZ51 (baseline) to Ames×B47; All origins: accuracy over all NAM-H populations; Non-tropical origin: accuracy over the 11 NAM-H populations of non-tropical origin (e.g., (B97×B73)×PHZ51 families); Tropical origin: accuracy over the 13 NAM-H populations of tropical origin (e.g., (NC350×B73)×PHZ51 families). Origin (non-tropical or tropical) follows McMullen et al. (2009). In each Ames set (Ames×PHZ51 or Ames×B47), 235 hybrids were selected for training GBLUP models and predicting phenotypes in NAM populations; Shared lines: only the 235 hybrids with female parents common to both Ames sets were used for training in each Ames set (1 replicate); Random subsets: random subsets of 235 hybrids were used for training in each Ames set (50 replicates). For each trait, sampling method, and subsets of NAM populations, significance of estimated differences in prediction accuracy (non-zero difference) was assessed by two-factor ANOVA, using the aov function in R: $accuracy_{ijk} = set_i + population_j + \varepsilon_{ij} + rep_{ik} + \delta_{ijk}$, with training set (set_i) as treatment factor, NAM populations ($population_j$) as block factor, random training subset (rep_{ik}) as subsampling factor, ε_{ij} and δ_{ijk} as sample and subsample errors, respectively.