*Genotype-by-environment effects on greenup and flowering as functions of environmental cues*

To further explore how GxE in greenup and flowering in these genetic subpopulations covaried with distinct environmental cues, we conducted multivariate adaptive shrinkage (mash) analyses. To do this, we specified additional hypothesis-based covariance matrices derived from correlations in environmental cues in the greenup or flowering date windows for the Midwest, Gulf, or both subpopulations. Just as different genetic subpopulations can have different genetic covariances between phenotypes at different gardens (Figure 1B), distinct SNPs can have different effect patterns, or covariances, on phenotypes at different gardens. Mash allows – indeed, even generates – multiple covariance matrices that particular SNP effects can load on to, and assigns mixture proportions for each SNP onto each covariance matrix using maximum likelihood. Mash then uses Bayes’ theorem to shrink effects for each SNP towards the set of covariance matrices in accordance to their mixture proportions. Thus, a SNP affecting flowering may have a high mixture proportion on a covariance matrix created from a specific environmental cue. In that case, we can infer that the effect of this SNP, or variation linked to this SNP, on flowering is due to a response to that environmental cue. These user-specified and data-driven covariance matrices are an important advantage mash offers for studying patterns of GxE: the user-specified covariance matrices allow hypothesis testing of specific environmental drivers for specific SNPs, while the data-driven covariance matrices allow exploration of additional unexplained patterns of covariation.

We used mash to examine GxE in greenup and flowering across eight common gardens and in three genetic subgroups: the Gulf, the Midwest, and both subpopulations, for a total of six mash runs. The hypothesis-driven matrices were created from environmental correlations during or prior to the phenological event in the Gulf, Midwest, or both subpopulations (Table X). We first looked at the log-likelihoods of each of these six mash runs with and without our hypothesis-driven matrices. For both subpopulations, including the hypothesis covariance matrices significantly improved the model fit (greenup LR = 774 flowering LR = 2942). For the single subpopulations, the hypothesis covariance matrices improved the model fit for the Midwest for greenup and for the Gulf for flowering, but did not improve it for the other phenotype (Midwest greenup LR = 866; flowering LR = 3063; Gulf greenup LR = 318; flowering LR = 1279).

We used distinct sets of user-specified, hypothesis-driven covariance matrices for greenup and flowering, but the same set for all genetic subgroups. Thus, we could directly compare the total weight mash placed on each of these matrices between subgroups. All three subgroups differed in which hypothesized covariance matrices had large weights (Fig 2a). For greenup for the Midwest subgroup, mash placed 28.6% of the weight on the covariance matrix created by correlating the temperature average in the 10 days prior to greenup. In contrast, hypothesized matrices created by correlating average temperature and cumulative GDD over 18 days had nonzero weights in the Gulf and both subgroups. For flowering, the Midwest had the largest hypothesis-driven weights for cumulative GDD matrices, while the Gulf had the largest hypothesis-driven weights for daylength and daylength change matrices. Thus, we found that distinct environmental drivers better captured most SNP effects in these two genetic subpopulations. Mash on both subgroups gave large hypothesis-driven weights for daylength, daylength change, and cumulative GDD matrices, indicating that we could detect both sets of cues in the combined population using this technique. Overall, the hypothesized covariance matrices had larger weights for the flowering phenotypes than for the greenup phenotypes for all three genetic subgroups (Gulf: 8.4% greenup; 31.8% flowering; Midwest: 28.6% greenup, 45.1% flowering; Both: 8.5% greenup, 47.3% flowering), indicating that our hypothesized environmental drivers captured more variation in SNP effects for flowering than they did for greenup (Fig 2b).

**Figure 2.** Total posterior weight placed on each covariance matrix type specified in the six mash models. Hypothesized covariance matrices (green) were created from environment-specific correlations across eight common gardens, and are described in Table X. Data-driven matrices (teal) are specific to each mash model, and canonical matrices (purple) have simple interpretations, such as equal effects across all common gardens, or effects specific to a single common garden. Covariance matrices included in mash that had zero posterior weight in all three genetic subgroups are not shown.

