

Supplemental material from “Detecting adaptive differentiation in structured populations with genomic data and common gardens”

Emily B. Josephs, Jeremy J. Berg, Jeffrey Ross-Ibarra, and Graham Coop

January 18, 2019

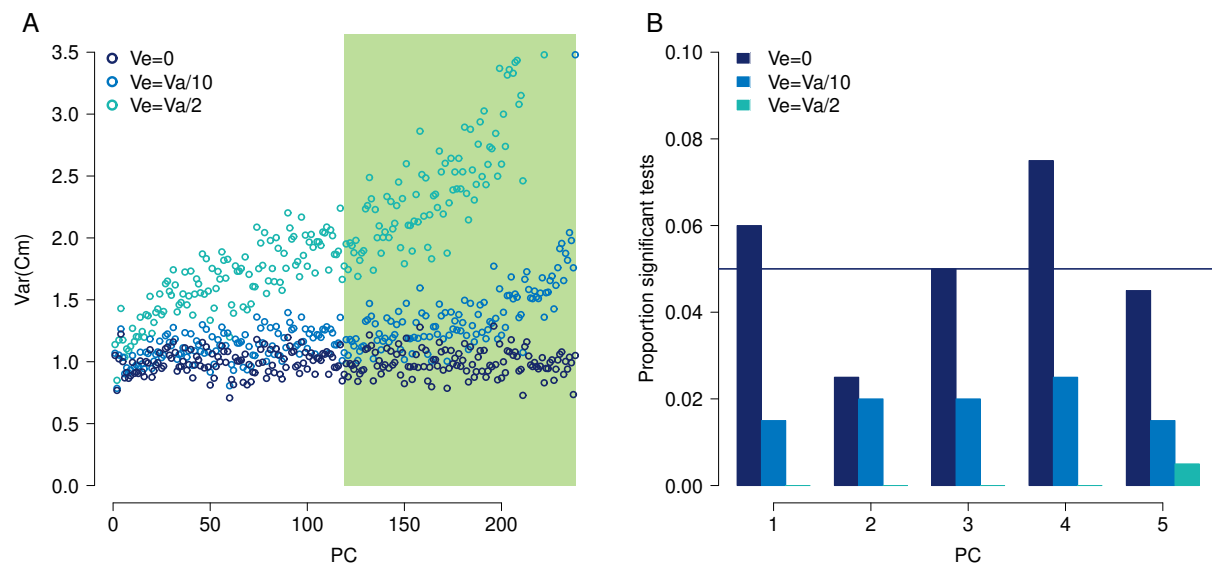


Figure S1: Q_{PC} on simulated neutral traits with varying amounts of V_E . A) $\text{var}(C_m)$ across 200 neutral simulations for varying levels of V_E . The PCs used to estimate V_A within populations (the denominator of Q_{PC}) are shaded green. B) The proportion of 200 neutral simulations that showed evidence of diversifying selection at $p < 0.05$. We expect that, under neutrality, 0.05 of all simulations should appear significant, but we see that as simulated V_E increases, fewer simulations are significant

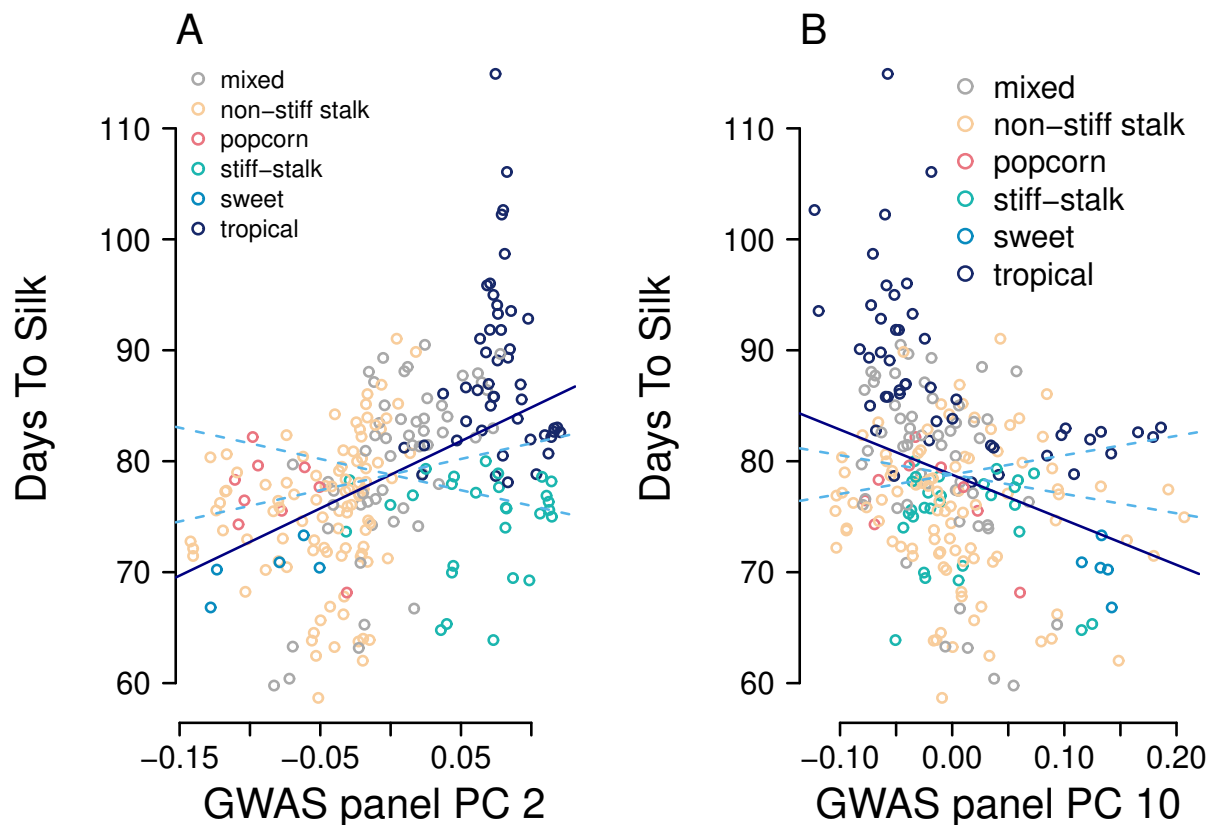


Figure S2: **Selection on days to silk along PC10 in the GWAS panel** Each point represents a line in the GWAS panel, colored by its membership in a subpopulation (same colors as Fig. ??A). A) The solid line shows the linear regression of the trait on PC 2 and the dashed lines show the 95% confidence interval of linear regressions expected under neutrality. Note that the linear regression is not the same as the F test done in Q_{PC} , and that we plot these lines for visualization purposes only. B) The same as (A) but for PC 10.

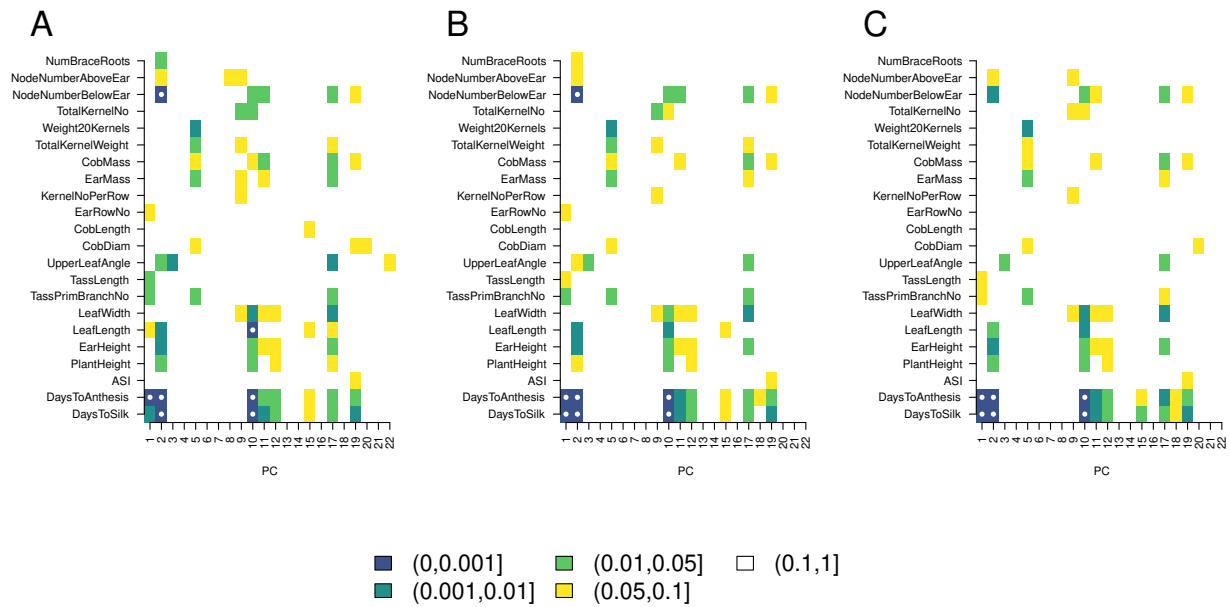


Figure S3: **How the PCs used to estimate V_A affect results of Q_{PC} .** We altered the value of R from Eq. ?? which changes the number of PCs essentially used to estimate V_A and tested for selection with Q_{PC} . A) Q_{PC} results for $R = 23$, so essentially all the PCs past 22 are used in the denominator to estimate V_A . B) The same as A but with $R = 119$, so the later half of PCs are used to test for selection. These are the same as the results presented in Fig. ?. C) The same as A and B, but with $R = 189$, so only the latest 50 PCs are used to estimate V_A .

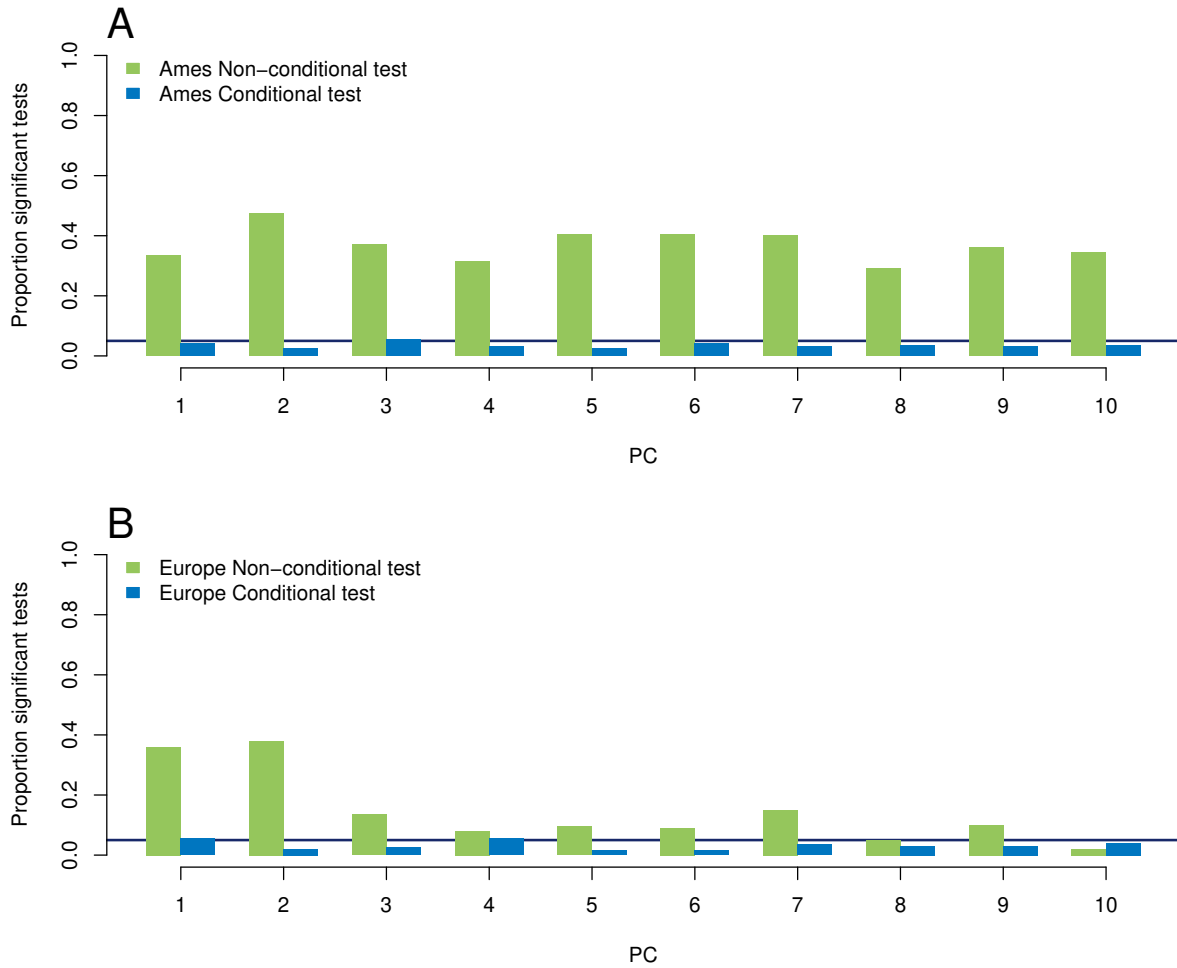


Figure S4: **Simulations of Q_{PC} on polygenic scores where 50 SNPs determine phenotype.** A) The proportion of 200 neutral simulations that were significant at the $p < 0.05$ level for the non-conditional Q_{PC} test and the conditional Q_{PC} test. A horizontal line is plotted at 0.05, to show the proportion of significant tests expected under the null hypothesis. B) The same information, this time for the European landraces.

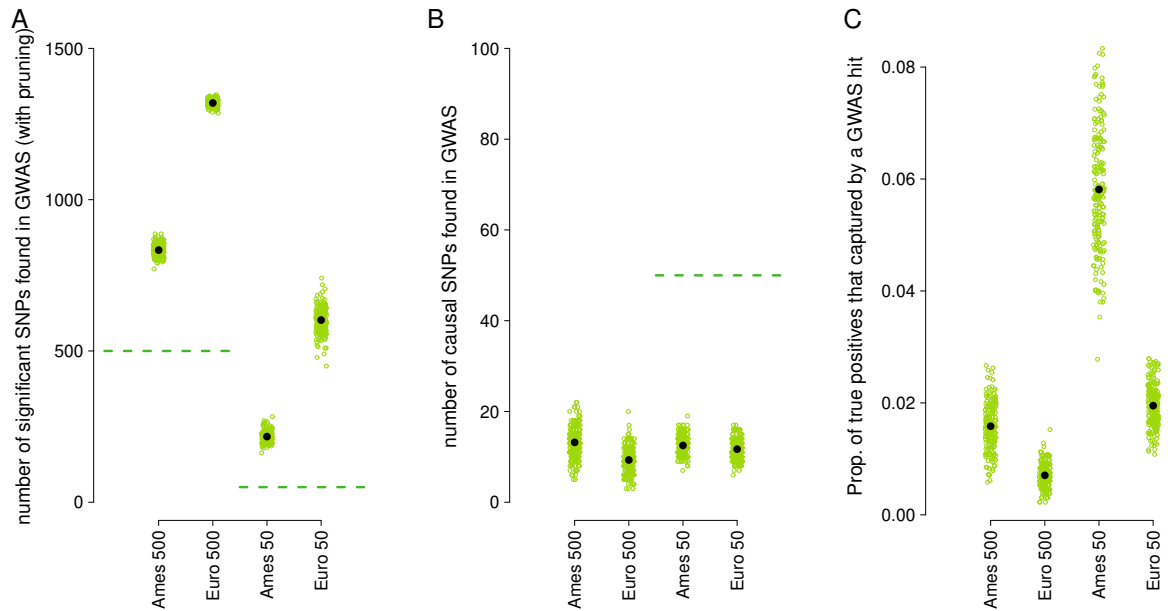


Figure S5: **Accuracy of GWAS on simulated traits.** A) The number of significant loci that were detected in GWAS ($p < 0.005$) in each of 200 simulations done using SNPs shared with the Ames panel and shared with the European panel done with 500 or 50 causal loci. In this and subsequent plots, each green dot represents one simulation, the black dots represent the mean across all simulations, and dotted lines represent the expectation if GWAS perfectly found all causal loci with no false positives. B) The number of causal loci from the simulations that were identified in the GWAS at $p < 0.005$. C) The proportion of simulated causal loci that were in an LD window identified by the GWAS.

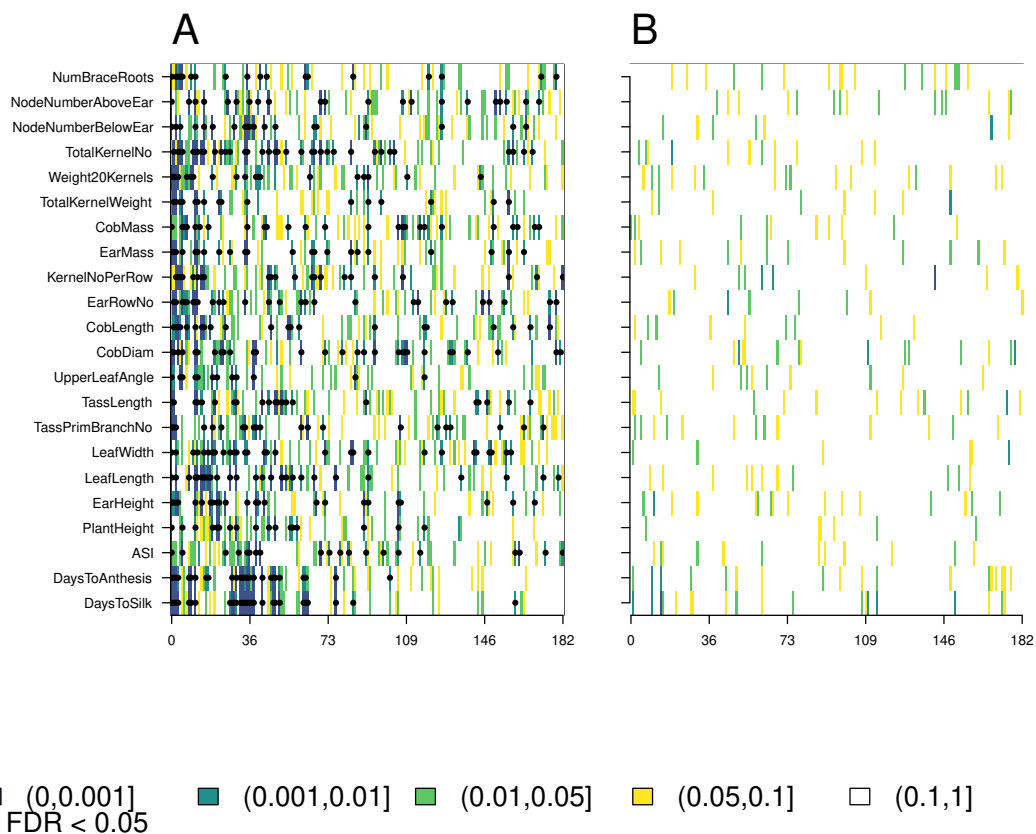


Figure S6: **P** values from applying Q_{PC} to polygenic scores in the Ames panel. A) Results from the non-conditional test B) Results from the conditional test

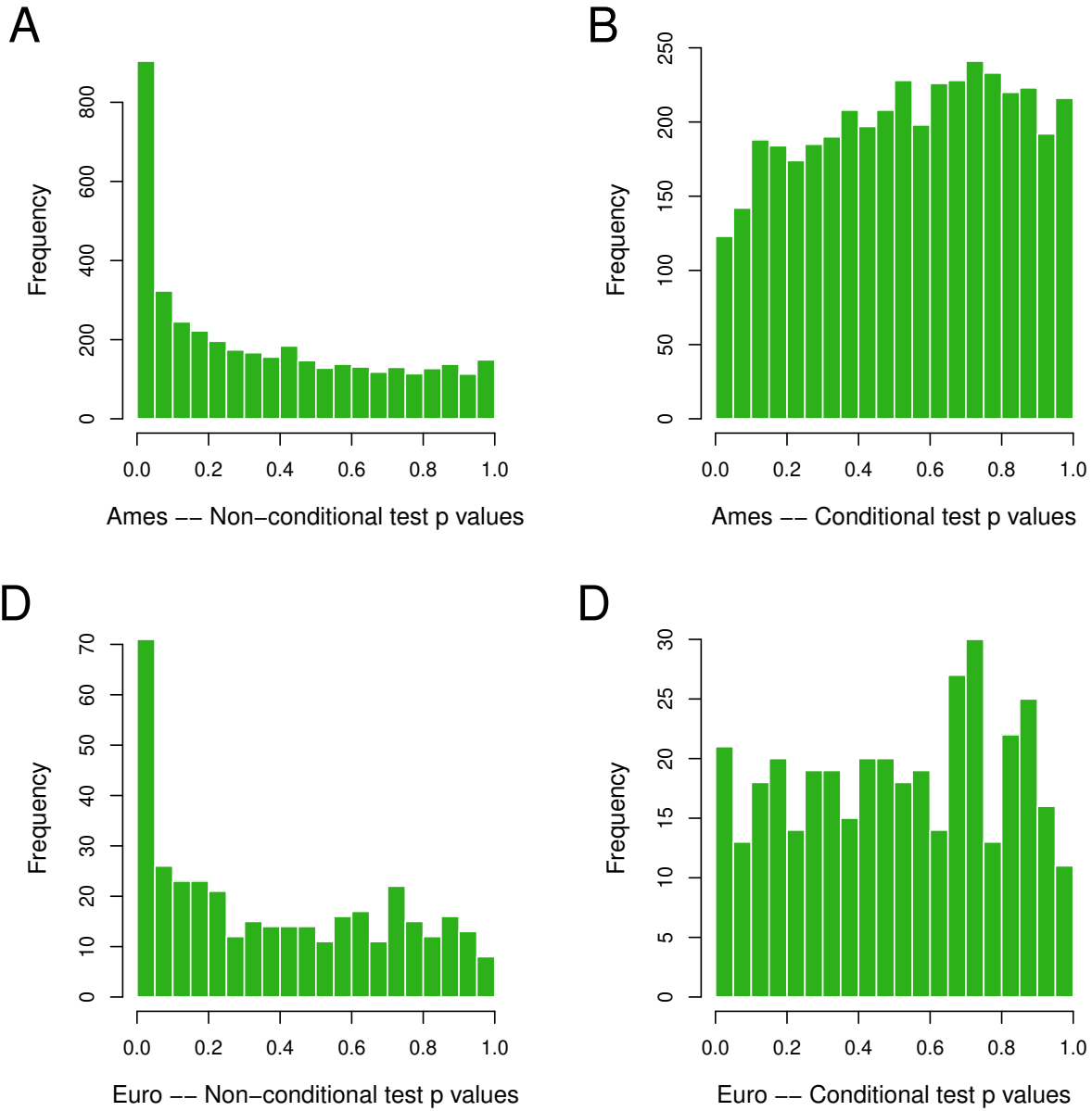


Figure S7: **Histograms of P values from applying Q_{PC} to polygenic scores.** A) P values for the non-conditional test in the Ames panel B) P values for the conditional test in the Ames panel. C) P values for the non-conditional test in the European landrace panel. D) P values for the conditional test in the European landrace panel.