



Figure S1. Exploration of the optimal number of genetic subpopulations (K) within the set of soybean accessions included in this study. A difference in cross-validation error between levels of K was used as a criteria. A plateau in  $\Delta$  cross-validation error at K=9 was used to infer K.