Supplemental Figure 1. A. Bayesian clustering of 368 half-sib (HS) families in the training set using Admixture. Each HS family is represented by a thin vertical line divided into 3-colored segments representing the individual’s estimated membership probability for each of the 3 clusters. Eastern Lowland consists of 89 HS families, Southern Lowland consists of 168 HS families, and Lowland x Upland consists of 111 HS families from Liberty. B. Bayesian clustering of 489 individuals from the northern association panel (Evans et al., 2017) and 178 individuals from the Validation data set #2 (VDS #2). Each individual is represented by a thin vertical line divided into 3-colored segments representing the individual’s estimated membership probability to each of the 3 clusters.

Supplemental Figure 2. Distribution of minor allele frequency (MAF) of polymorphic markers in each of four independent panels used for training the GS prediction procedures. A. All populations, B. Southern Lowland, C. Eastern Lowland, and D. Lowland x Upland. The vertical dashed lines represent the 95% percentile of MAF.

Supplemental Figure 3. Pattern of linkage disequilibrium decay represented by squared correlation (r2) in each of four independent panels used for training the genomic prediction procedures calculated using R package SNPRelate. A. All populations, B. Southern Lowland, C. Eastern Lowland, and D. Lowland x Upland. The blue lines are smoothed running averages.

Supplemental Figure 4. Distribution of reference allele frequency for 178 individuals in Validation data set #2. Markers that segregate at a segregation ratio of 1:2:1 were used for the purpose of discriminating between tetraploid and octoploid. Allele frequencies were calculated by using allele read depth at each position for each individual. A total of 16 known octoploid and 67 known tetraploid individuals were used as the reference populations for each ploidy.