Step One: - VCFtools
An initial filter was completed to remove low depth, too much missing data, and low allele frequency.

Step Two: Custom Perl Script bb.vcf
All “heterozygous” sites that do not have a ratio of 0.3 to 0.7 were set to NA.

Step Three: VCFtools
Sites where all samples were set to NA in previous step are removed in this filter step.

Step Four: Tabix
The next step was imputation. However, this step is faster when completed by splitting the file and completing imputation one chromosome at a time. Tabix was used to split the larger file in to nine individual files representing each chromosome.

Step Five: Java – Beagle
Missing genotypes were imputed with beagle. Each chromosome was completed independently.

Step Six: VCFtools
A final filtering step was completed to remove any sites that become monomorphic following imputation.

Step Seven: Tassel
The genotype format was convert from vcf to hapmap format. Hapmap format is the genotype format used by the GWA analysis software GAPIT.

Step Eight: R - GAPIT
The GWA software - GAPIT converts all genotypic data to numeric format. To save time during GWA analyses, all files were converted to numeric format prior to association analyses.

Step Nine: R – GAPIT
Relatedness and Population structure are calculated with a kinship and PCA analysis, respectively. These analyses are time consuming. As such a random subset of 125,000 markers was used to calculate the kinship and PCA. The resulting files were used in subsequent GWA analyses.