



Supplementary Figure S8. The flow diagram here represents the distribution of the maize B73v5 genes. The initial dataset of 39,756 maize B73v5 genes had 874 missing values for the target label core/non-core, therefore the remaining dataset of 38,882 genes were used for subsequent analysis. Among the 38,882 genes, seventy-two percent of the genes marked as core in the maize reference genome version B73v5 were considered as the positive sample, and twenty-eight percent annotated as non-core (near-core, dispensable and private genes) were considered as the negative sample for modeling.