Supplementary Materials to Harnessing Machine Learning to Identify Antimicrobial Peptides in *Drosophila melanogaster*

Nilanjan Roy^{a,*}

^aDepartment of Molecular Biosciences, University of Kansas *Corresponding author: nilanjan.roy@ku.edu



Figure S1: Error rate in the Random forest model

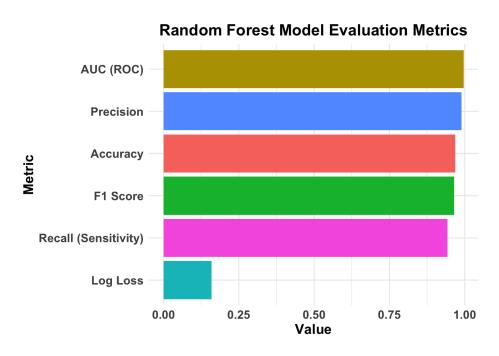


Figure S2: Different evaluation matrices of the Random forest model

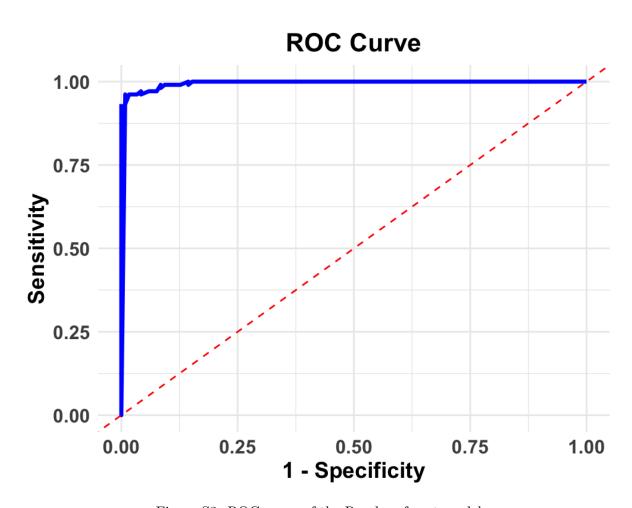
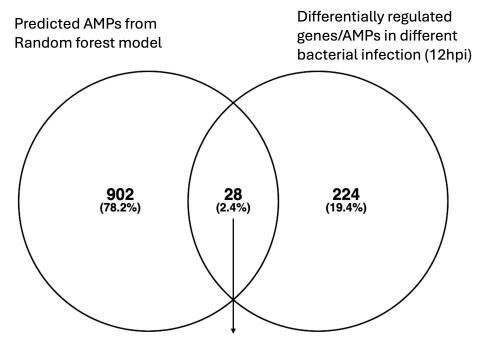


Figure S3: ROC curve of the Random forest model



Among these, 12 genes with previously unknown function got predicted as AMPs with Random forest model and are differentially regulated upon bacterial infection

Figure S4: Screeing the predicted AMPs with expression data in different bacterial infections

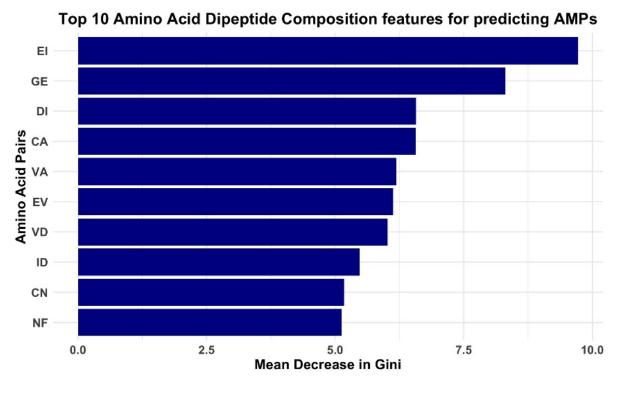


Figure S5: Top 10 amino acid dipeptide composition features for predicting AMPs