Supplementary Figure 2: Assessment of standard bioinformatics toxin classification methods

Supplementary Figure 2 shows the accuracy of BLAST and HMMER based toxin annotation methods (described in detail in Supplementary Note 1B). Each of the graphs shows accuracy (y-axis) in relation to BLAST/HMMER e-value cut-off values, below which the query sequence was annotated as a 'toxin'. The accuracy was calculated for the Positive, Easy, Moderate and Hard datasets. Balanced results were calculated as the weighted average of all datasets: Balanced Data Accuracy = (Positive Accuracy x 4 + Easy Accuracy + Moderate Accuracy + Hard Accuracy x 2)/8.
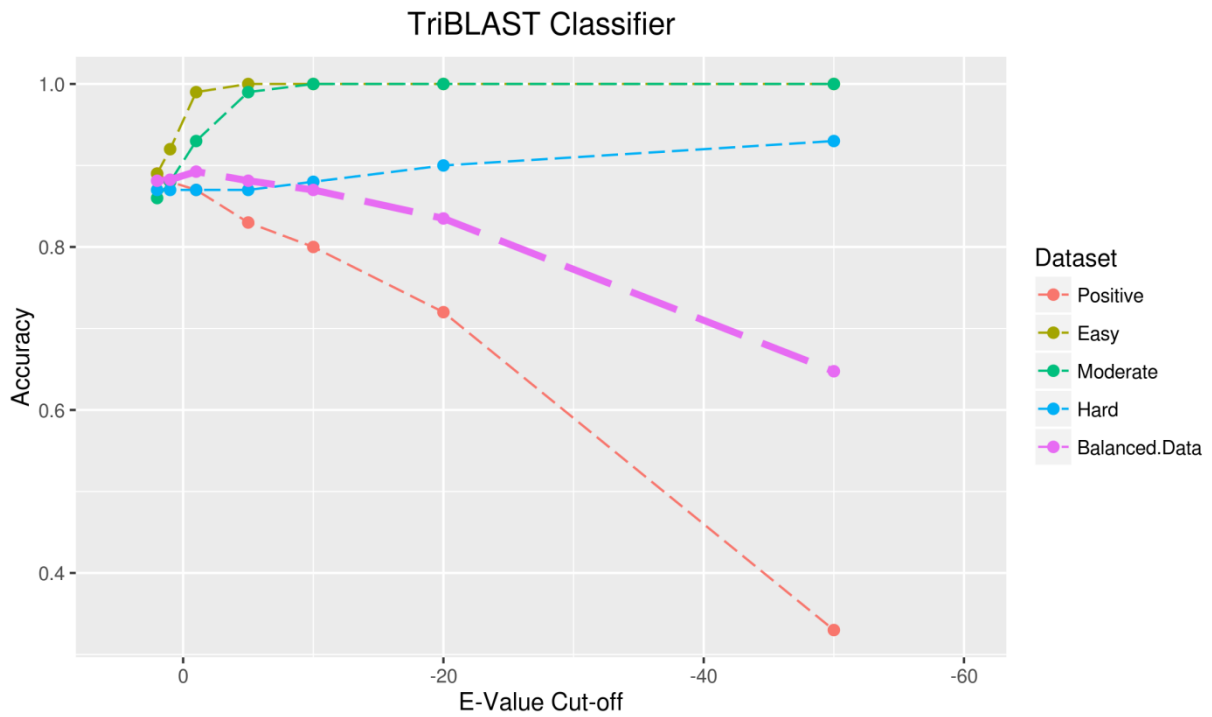
A: NaiveBLAST Classifier, accuracy vs BLAST e-value cutoff

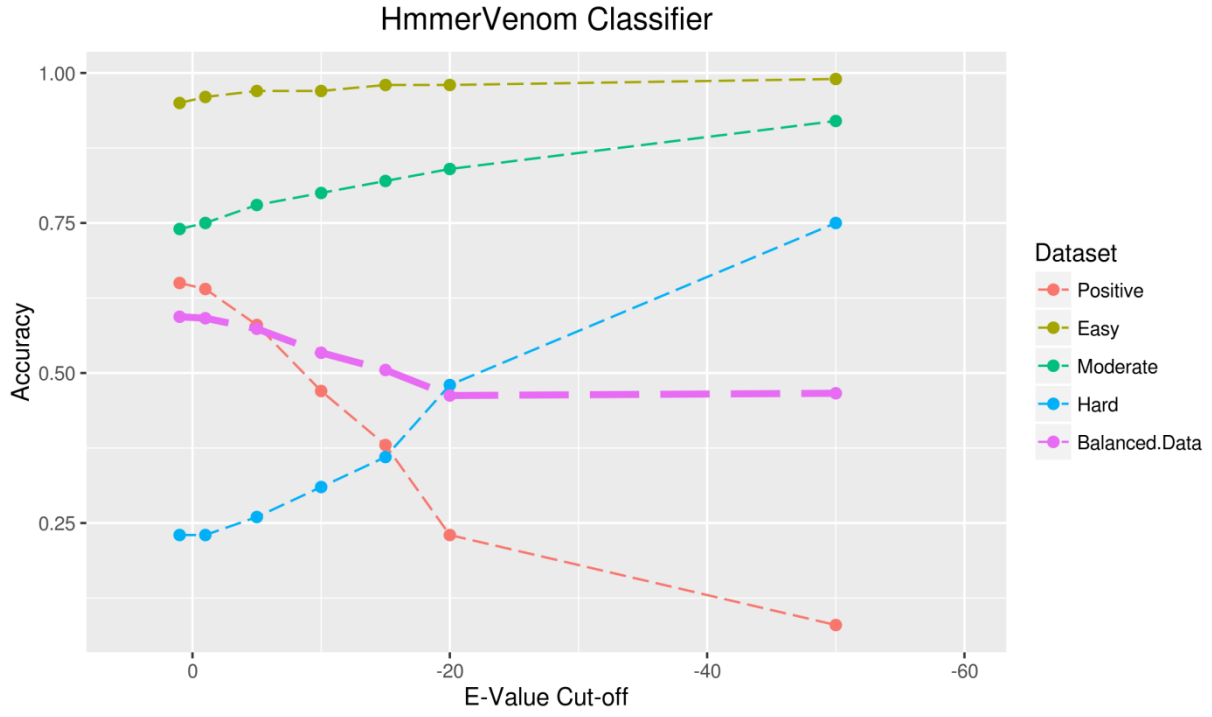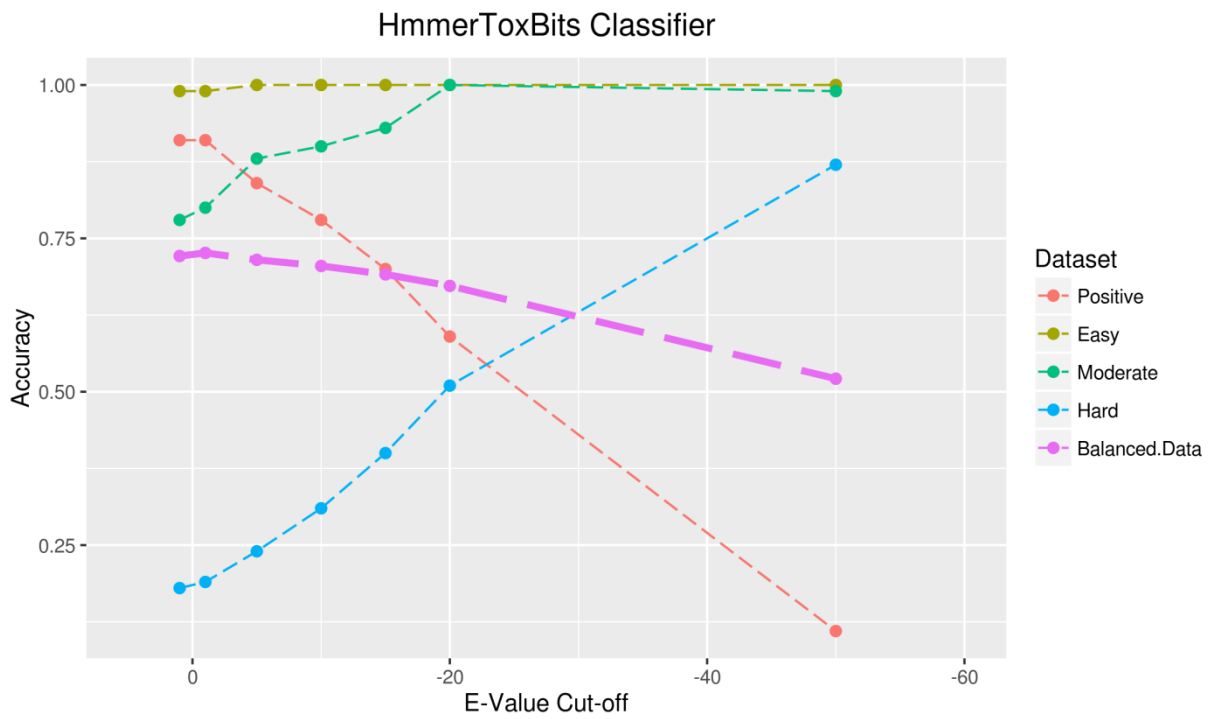B: OneBLAST Classifier, accuracy vs BLAST e-value cutoff
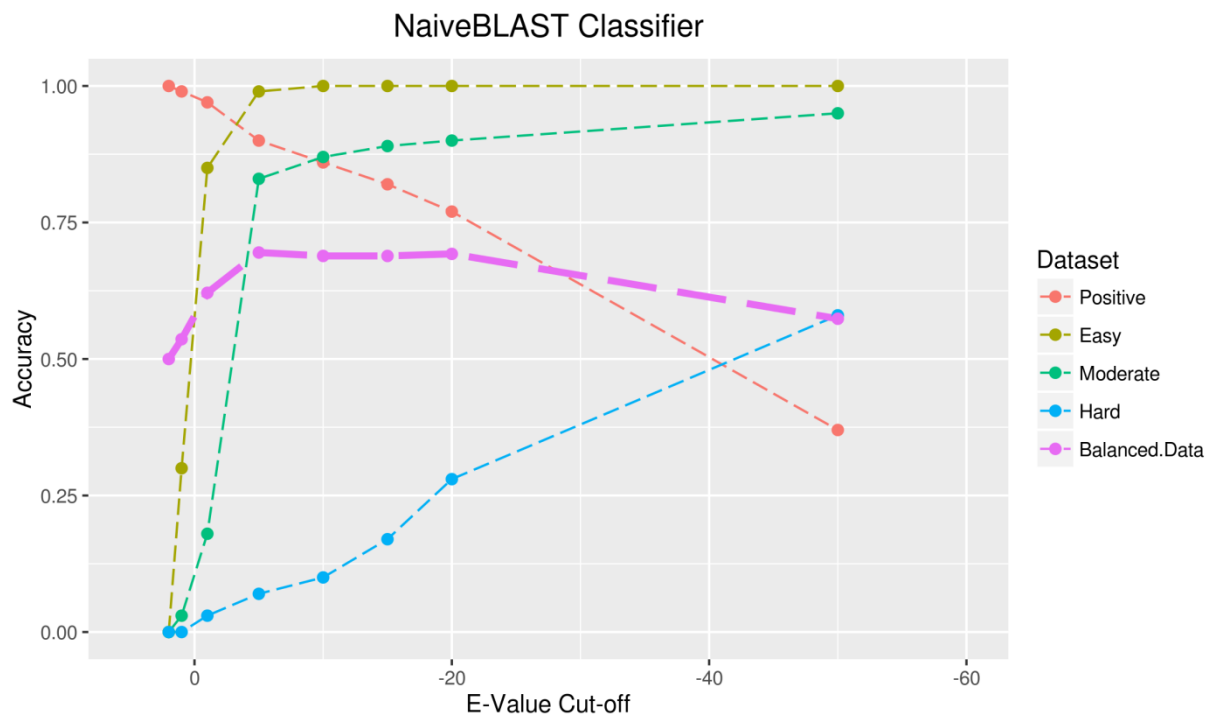
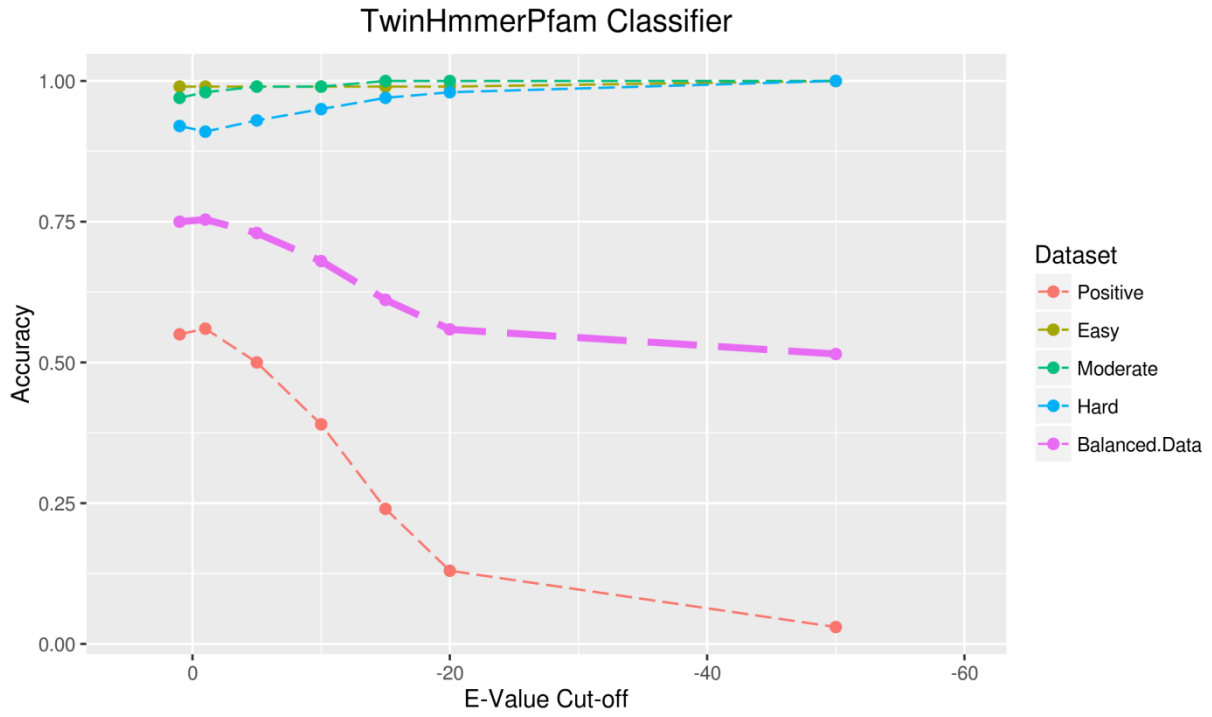OneBLAST Classifier
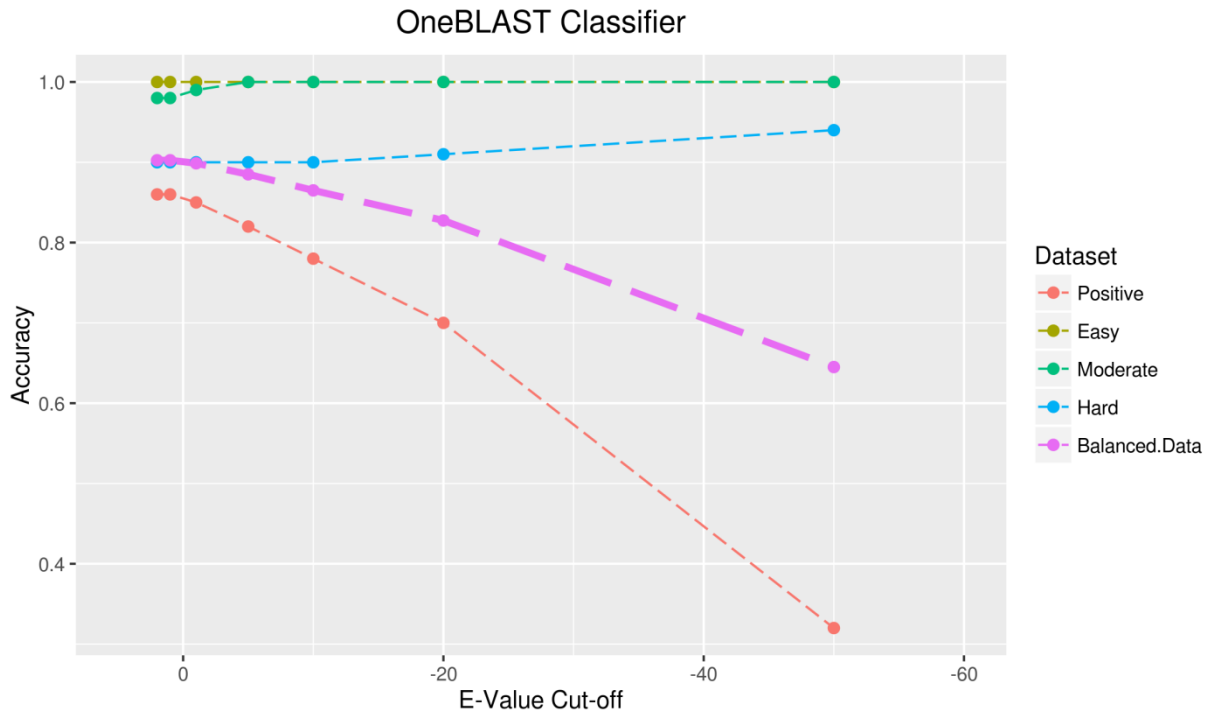
C: TriBLAST Classifier, accuracy vs BLAST e-value cutoff

TriBLAST Classifier

D: hmmerVenom Classifier, accuracy vs Hmmsearch e-value cutoff

HmmerVenom Classifier

E: hmmerToxBits Classifier, accuracy vs Hmmsearch e-value cutoff

HmmerToxBits Classifier

F: twinHmmerPfam, accuracy vs Hmmsearch e-value cutoff

TwinHmmerPfam Classifier

G: twinHmmerToxBits, accuracy vs Hmmsearch e-value cutoff

TwinHmmerToxBits Classifier