

Supplementary Figure 4: Violin plots of correct TNT classifications using different thresholds on relative abundance per ASV for the validation set. The dot represents the mean balanced accuracy, averaged over six different data set splits. n indicates the number of models calculated. The random forest models consisted of 10000 trees with an mtry factor of 5. The mean balanced accuracy ranged from 80.4 – 83.0%. 0.08% was chosen because the distribution became more distinct compared to the slightly better performing 0.06% threshold.