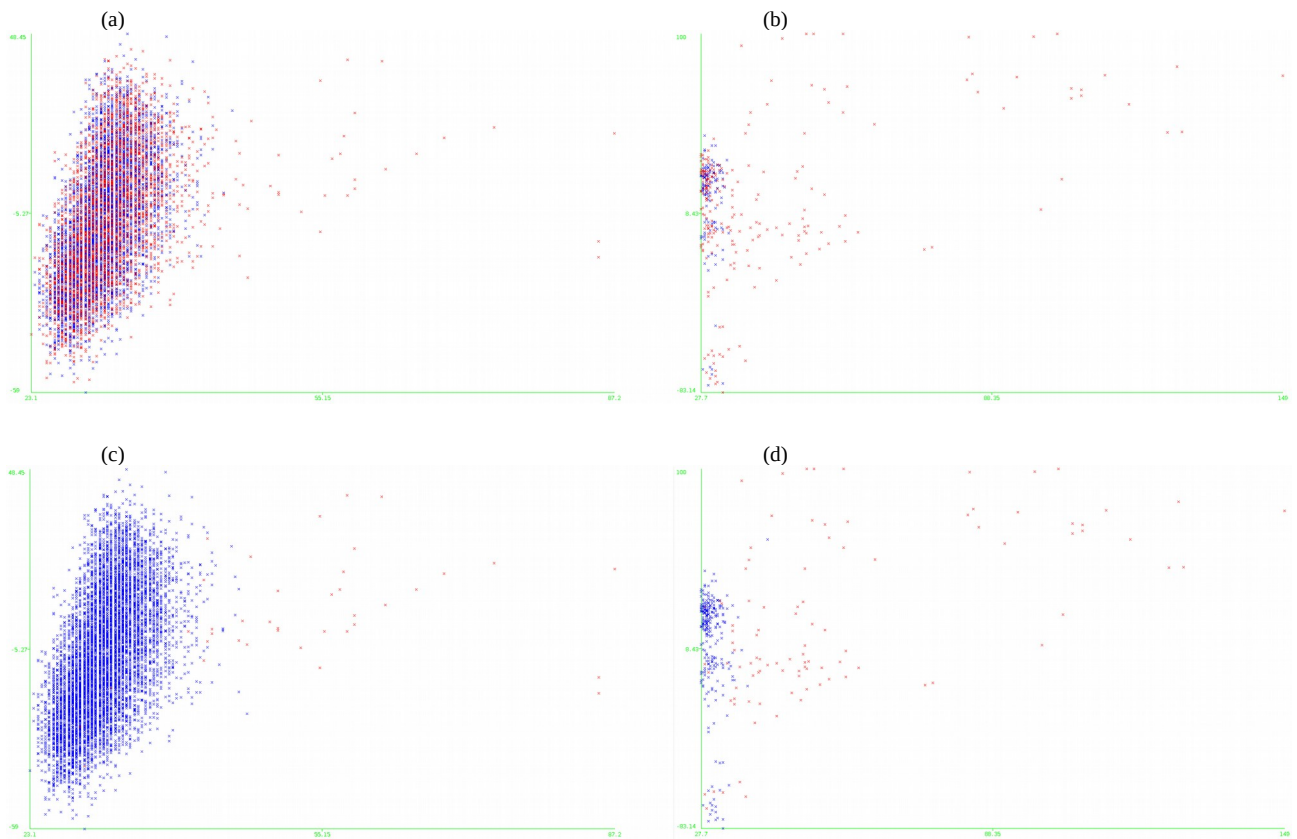


## Plots of alpha score vs. bit score to the other seven datasets

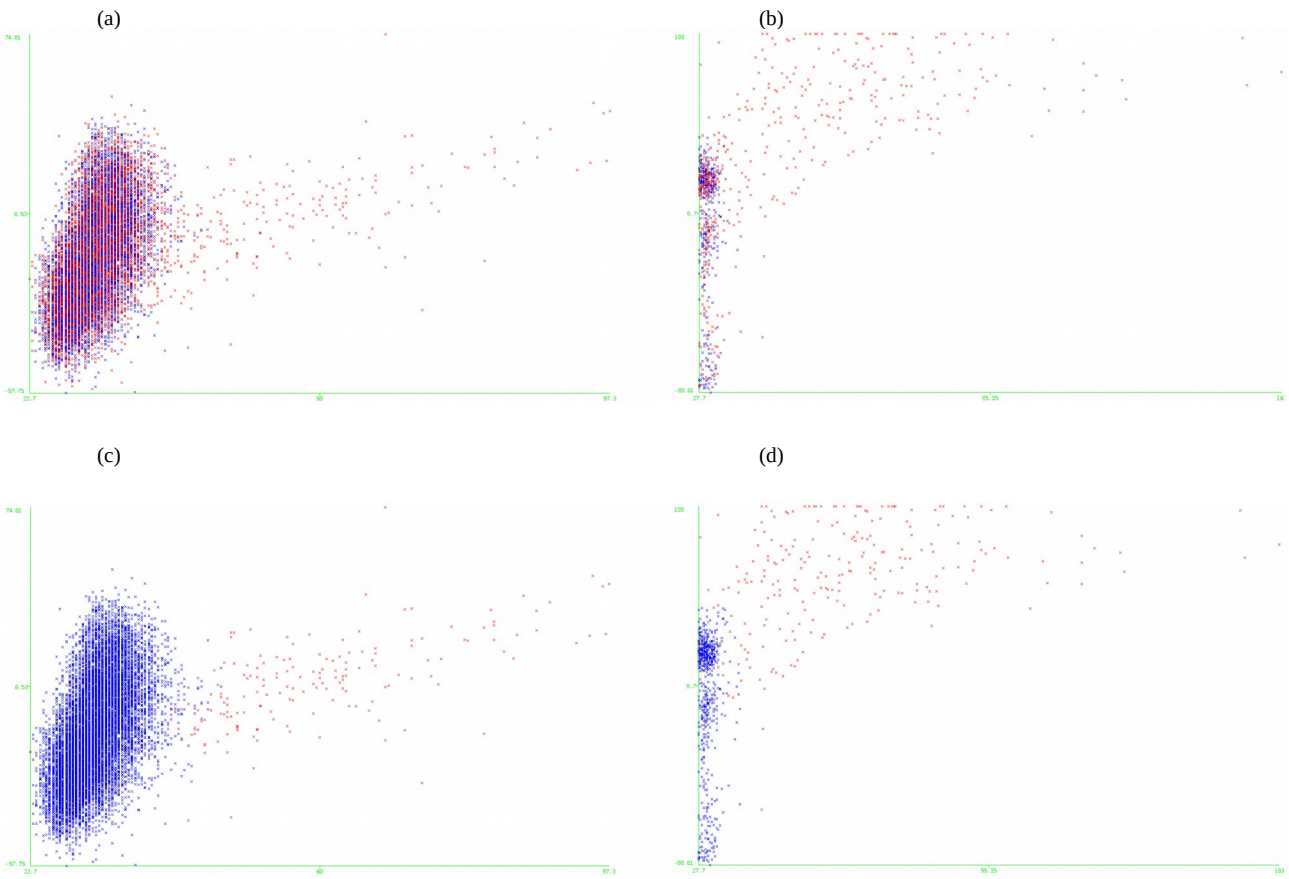
The data points represent the query alignments found by BLAST. For each bacterium, (a) and (b) show the alignments of vsORFs and sORFs, respectively, before applying the ANN+TSA approach, while in (c) and (d) the same data is shown after relabeling the target hits that were considered random by ANN+TSA. In (a) and (b), red data points are target hits, whereas blue data points are decoy hits. In (c) and (d), red data points are those target hits considered as significant by OCCAM (label was kept), while blue data points are decoy hits plus target hits relabeled, i.e., considered random.

In all plots below, axis y is alpha score while axis x is bit score.

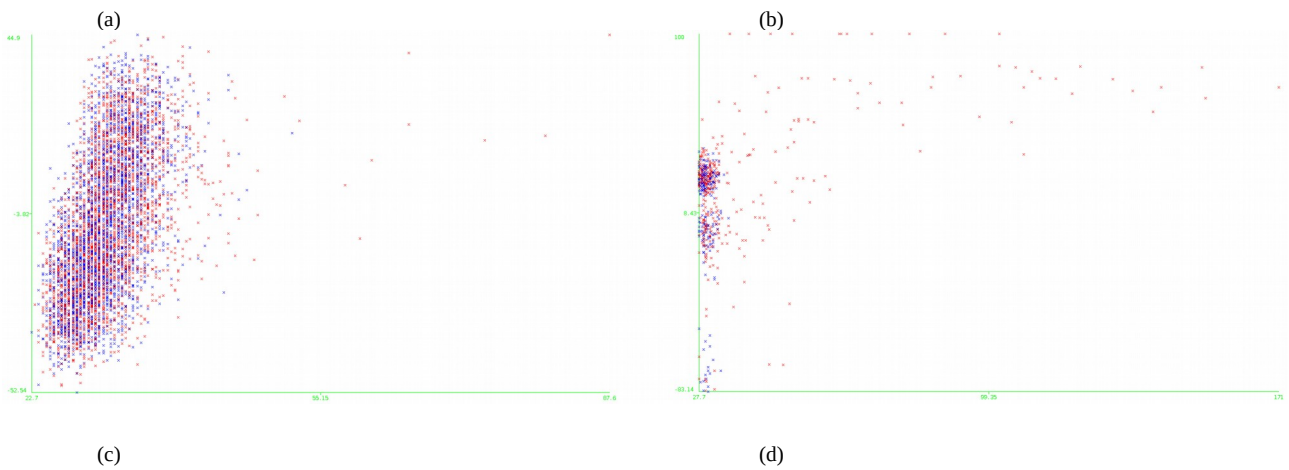
*Actinobacillus pleuropneumoniae* serovar 5b L20 uid58789

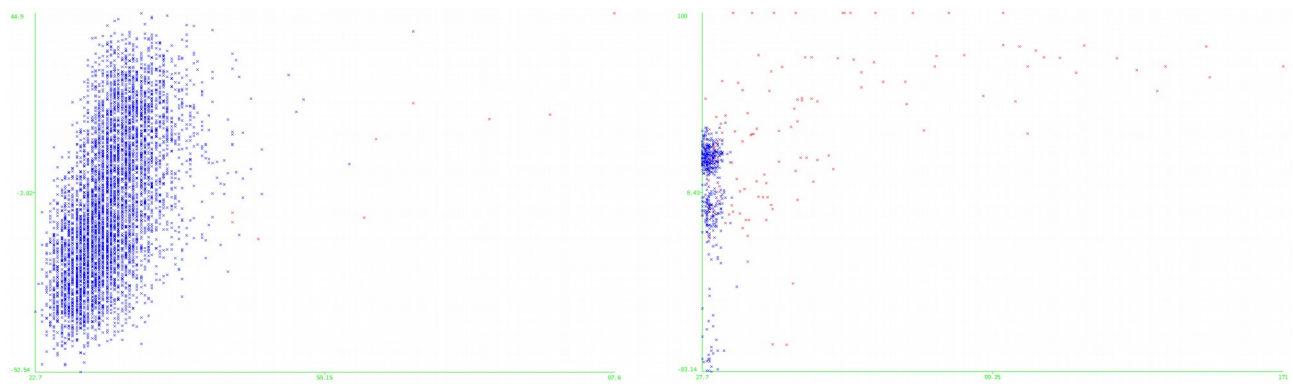


*Bacillus subtilis* 168

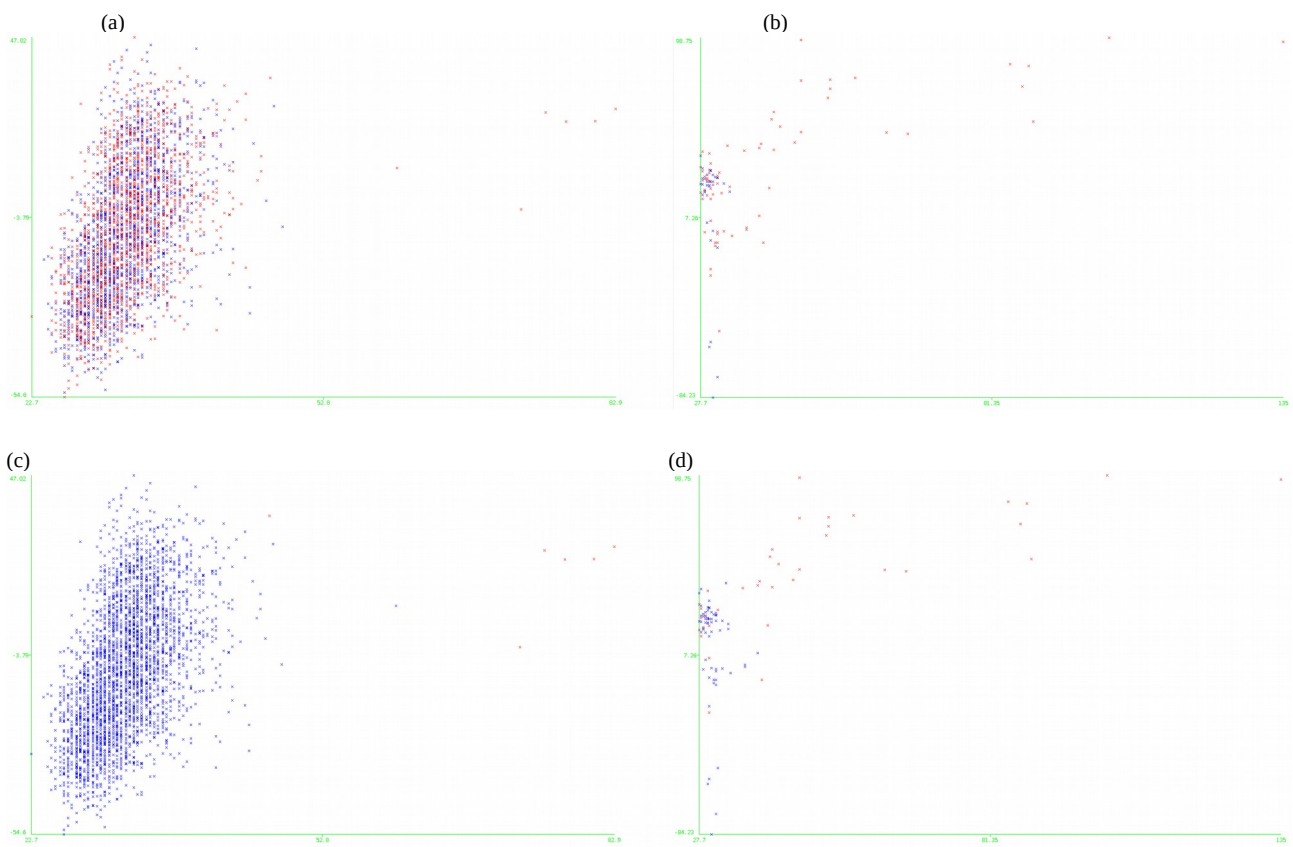


*Caulobacter crescentus* CB15 uid57891

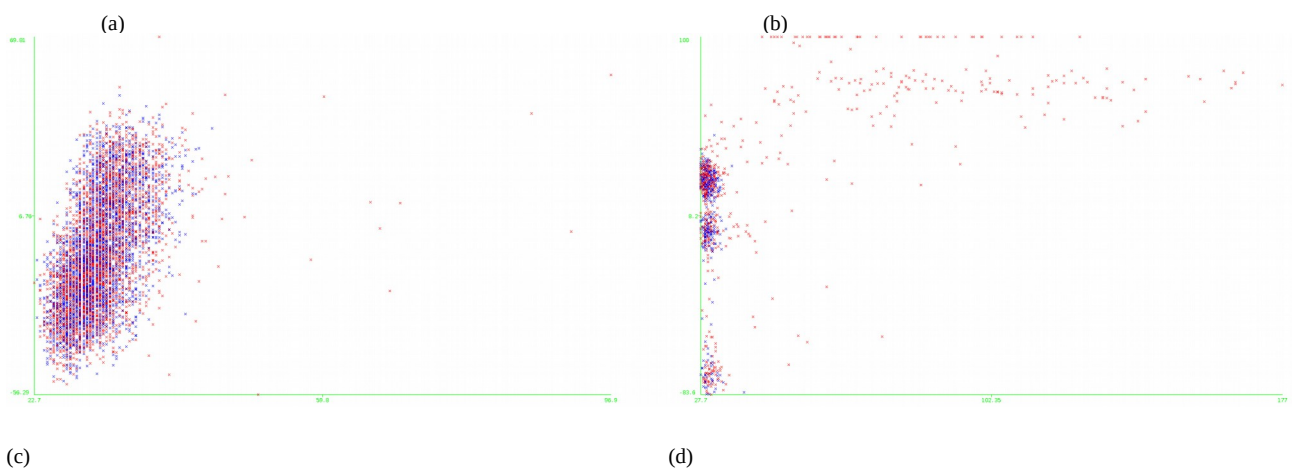


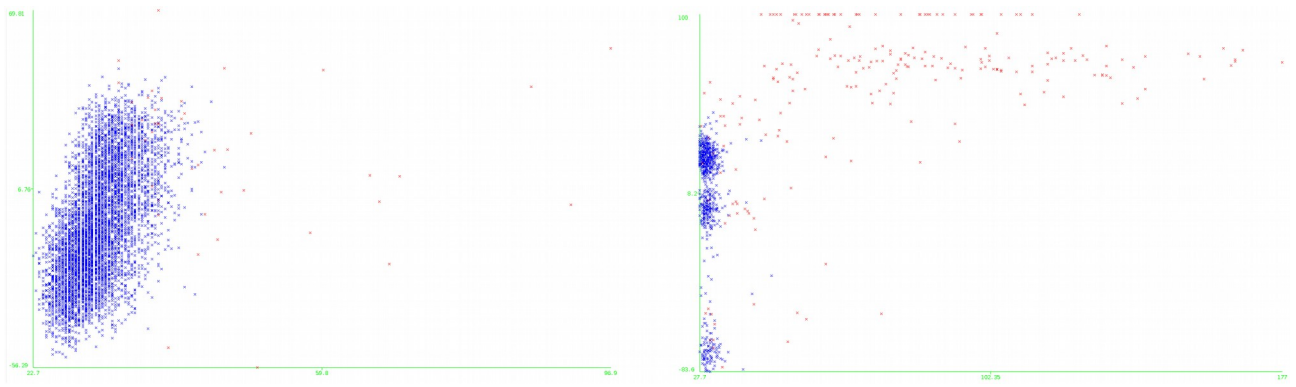


*Mycoplasma hyopneumoniae* 7448 uid58039

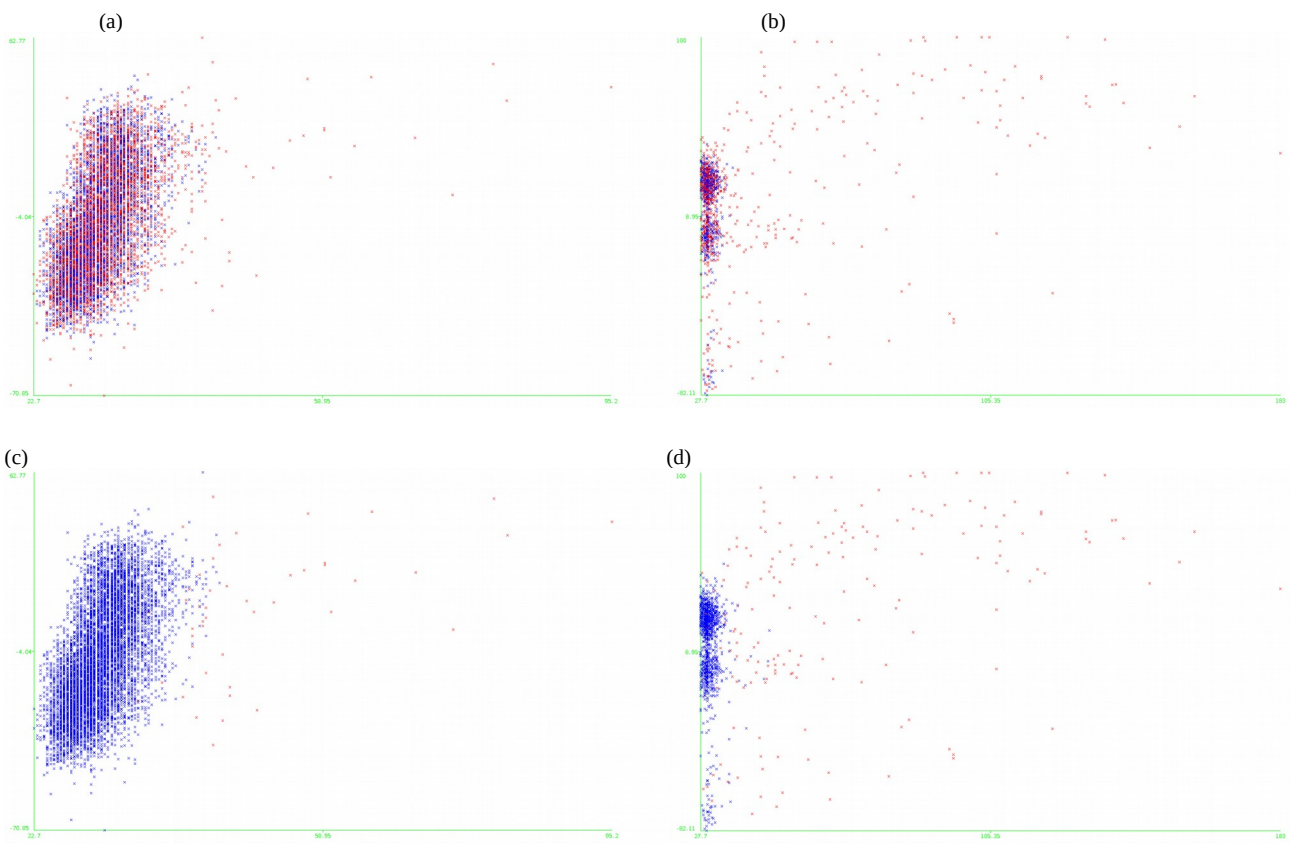


*Mycoplasma hyopneumoniae* 7448 uid58039





*Pseudomonas aeruginosa* PAO1



*Staphylococcus aureus* MRSA

