1 Supplement

The results we reported are based on the most informative simulated optical maps. These were determined by trying all (139 choose 3) enzyme combinations for each assembly of each genome and then using the best performing set of three enzymes for each assembly/genome combination. This approach was only possible because reference genomes were available, both for simulating all the optical maps and for evaluating the assembly quality relative to the reference genome. This raises the question of how stable these results are relative to enzyme selection, and (until the aformentioned future work on algorithmic enzyme selection) how well might this approach work if enzymes are choosen at random rather than *a posteriori*. Figures 1 and 2 are the ROC plots for all (139 choose 3) enzymes for SOAPdenovo and IDBA assembly of *Francisella tularensis*, respectively. The heat maps give an idea of the probability of getting a particular true positive rate and false positive rate with a specific choice of enzymes. These plots show the sensitivity and specificity of misassembly detection using optical mapping data alone. The paired-end sequence data was not used. As can be seen in the plots, if a set of enzymes were chosen at random then optical mapping would still likely be informative and produce a meaningful classifier, however there are combinations with significantly better sensitivity and specificity than the hot spots, suggesting a need for an algorithmic enzyme selection method.

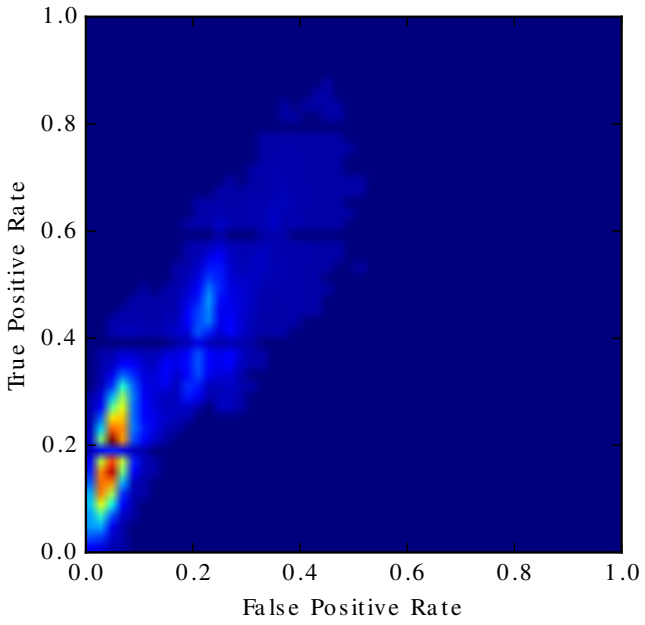

Figure 1: ROC plot illustrating the density of optical map alignment based missassembly detection classification rates for the SOAPdenovo assembly of *Francisella tularensis*. The color intensity at each point indicates the number of three enzyme based classifiers having that classification rate. The plot includes results for optical maps with all three enzyme combinations using a set of 135 enzymes randomly drawn from the REBASE database. The velvet assembly (which is not shown) has a similar pattern. Hot spots represent the likely classification rate for enzymes choosen at random.

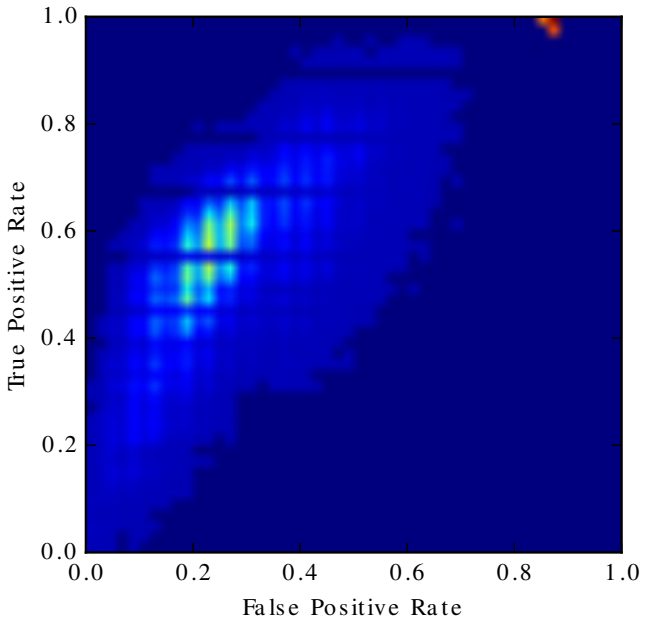

Figure 2: ROC plot illustrating the density of optical map alignment based missassembly detection classification rates for the IDBA assembly of *Francisella tularensis*. The color intensity at each point indicates the number of three enzyme based classifiers having that classification rate. The plot includes results for optical maps with all three enzyme combinations using a set of 135 enzymes randomly drawn from the REBASE database. Both SPAdes assemblies as well as ABySS (which are not shown) have a similar pattern. Hot spots represent the likely classification rate for enzymes choosen at random.