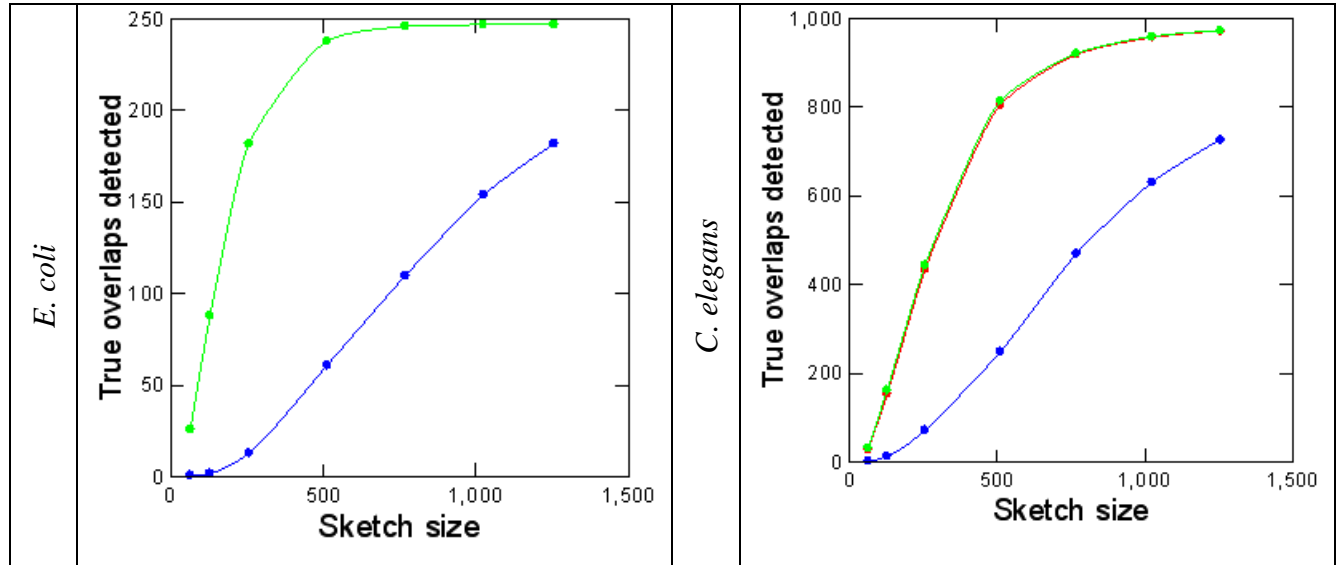


Supplemental Fig. S3



Supplemental Figure S3. Sensitivity of read overlap detection, using the genome-derived list of valid *k*-mers. PacBio reads (pairs of 10 kb sequences with 2kb overlaps) were simulated from the *E. coli* genome (left panel; 250 pairs of reads) or the *C. elegans* genome (right panel; 1000 pairs of reads), and were subjected to standard MHAP (blue), MHAP with masking of low-frequency *k*-mers from Illumina-derived list (red), or MHAP with masking of genome-derived non single-copy *k*-mers (green). Note that the results with Illumina derived (red) and genome-derived (green) list of valid *k*-mers were essentially identical; only the genome-derived (green) is visible in the left graph.