



**Figure S3.** Matthew's Correlation Coefficient calculated using different cut-offs for insertion index in all 13 genomes in this study. True positives are genes that are predicted as essential and whose orthologs are classed as essential in the EcoGene database of essential genes for *E. coli* K-12 and false positives are genes that are predicted as essential, but are classed as non-essential in EcoGene. The cut-offs for DBSCAN and fitting Gamma distributions are shown in each figure. DBSCAN outperforms Gamma fits in separating essential genes from non-essential genes.