

**Figure S7. ORF length bias in false negatively predicted ORFs.**

Plotted in bar-graphs for moderate GC% species *B. subtilis* 168, *E. coli* K12 MG1655, *L. lactis* KF147 and *L. plantarum* WCFS1 is the relative number of ORFs present in the complete reference genome (purple), in the subset of incorrectly predicted ORFs for start codon coordinates (blue), and in the subset of false negatively predicted ORFs (red), over nine ORF length fractions (ORF length measured in nucleotides). The two subsets represent ORFs which were consistently mis-predicted by the combination of four AGES (BASys, ISGA, RAST and xBASE) because they either had incorrectly predicted start codon coordinates, or by being a false negative predicted ORF.

