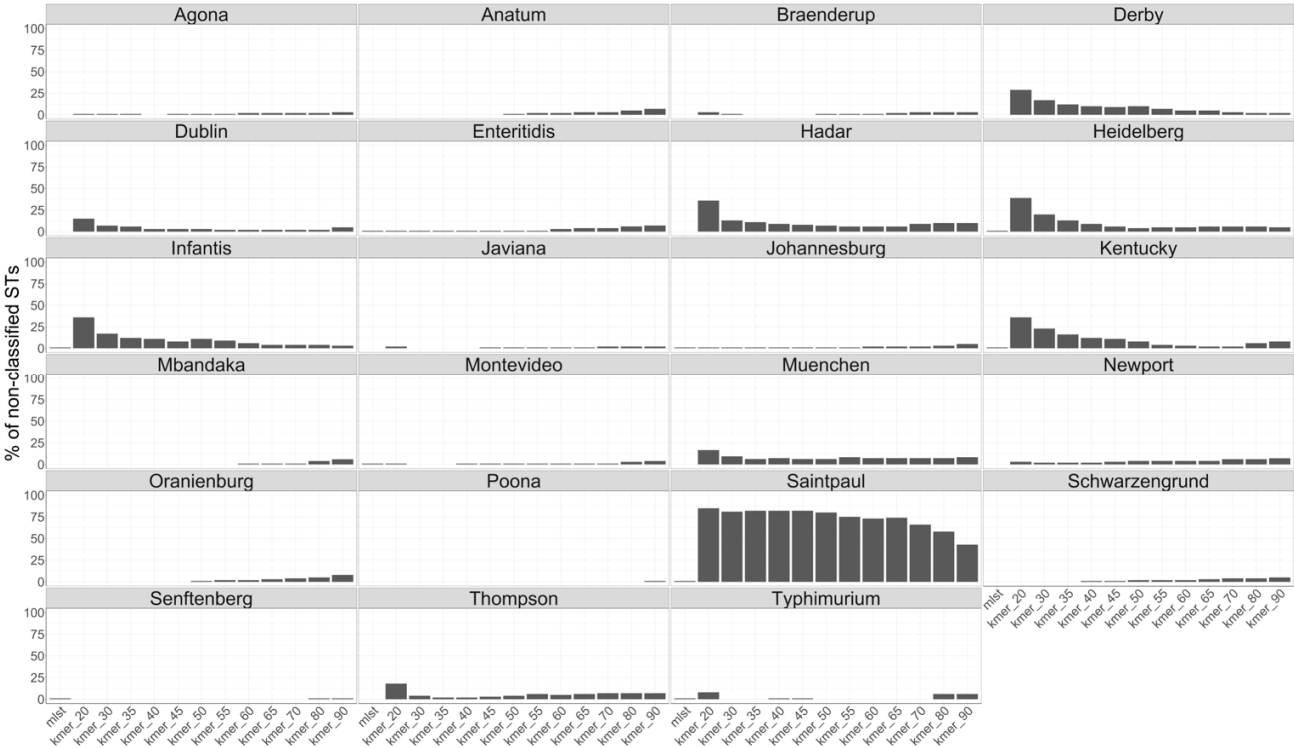
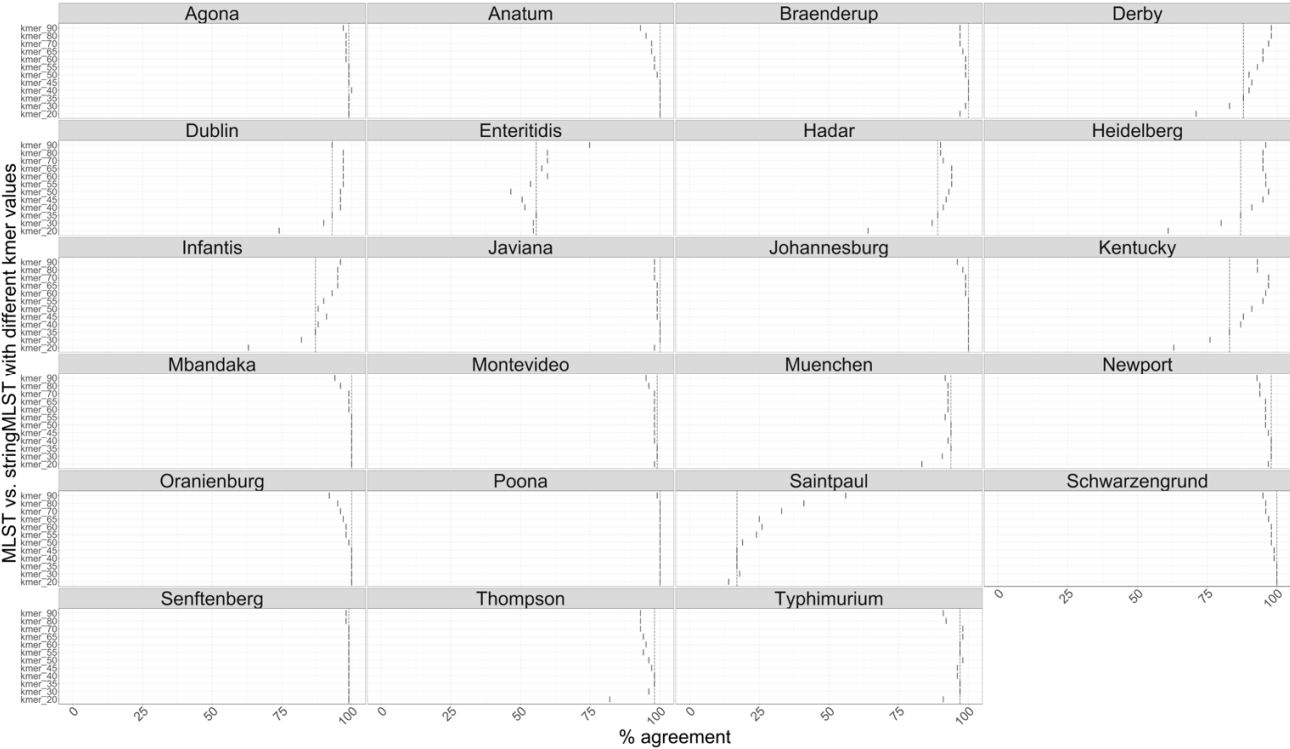


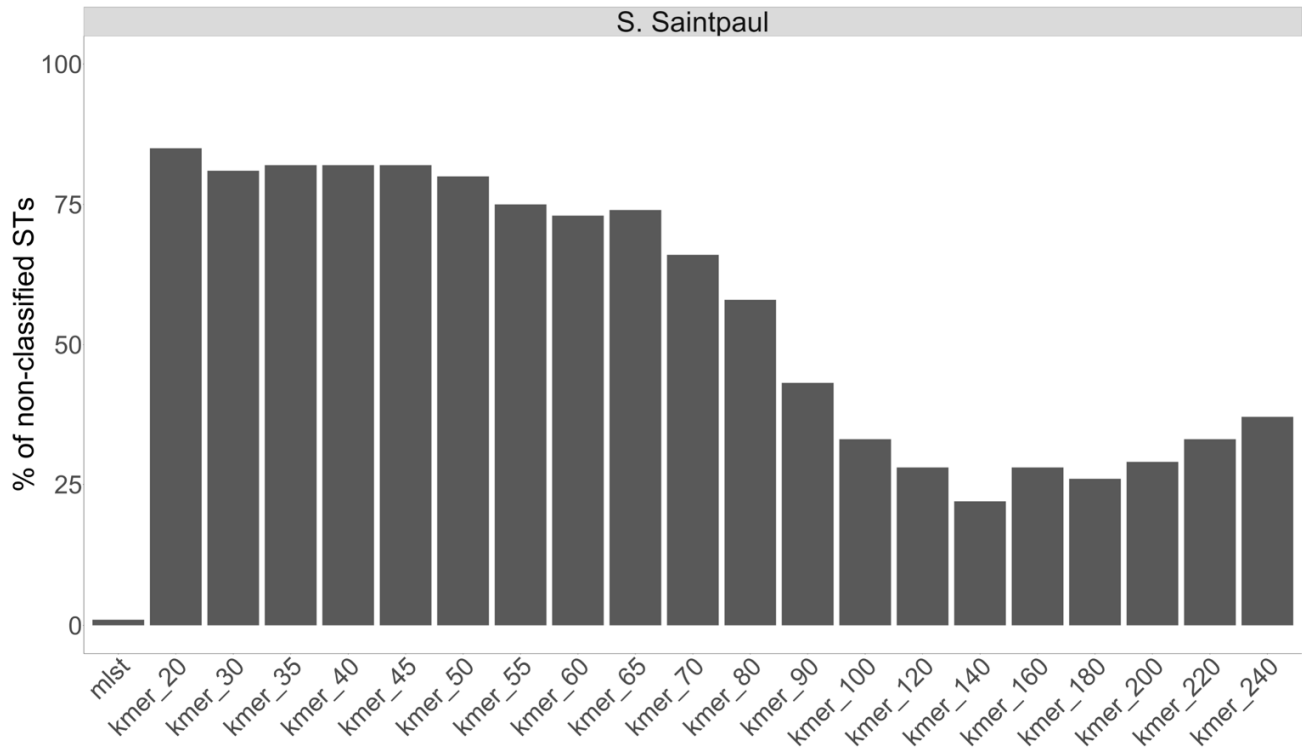
A.



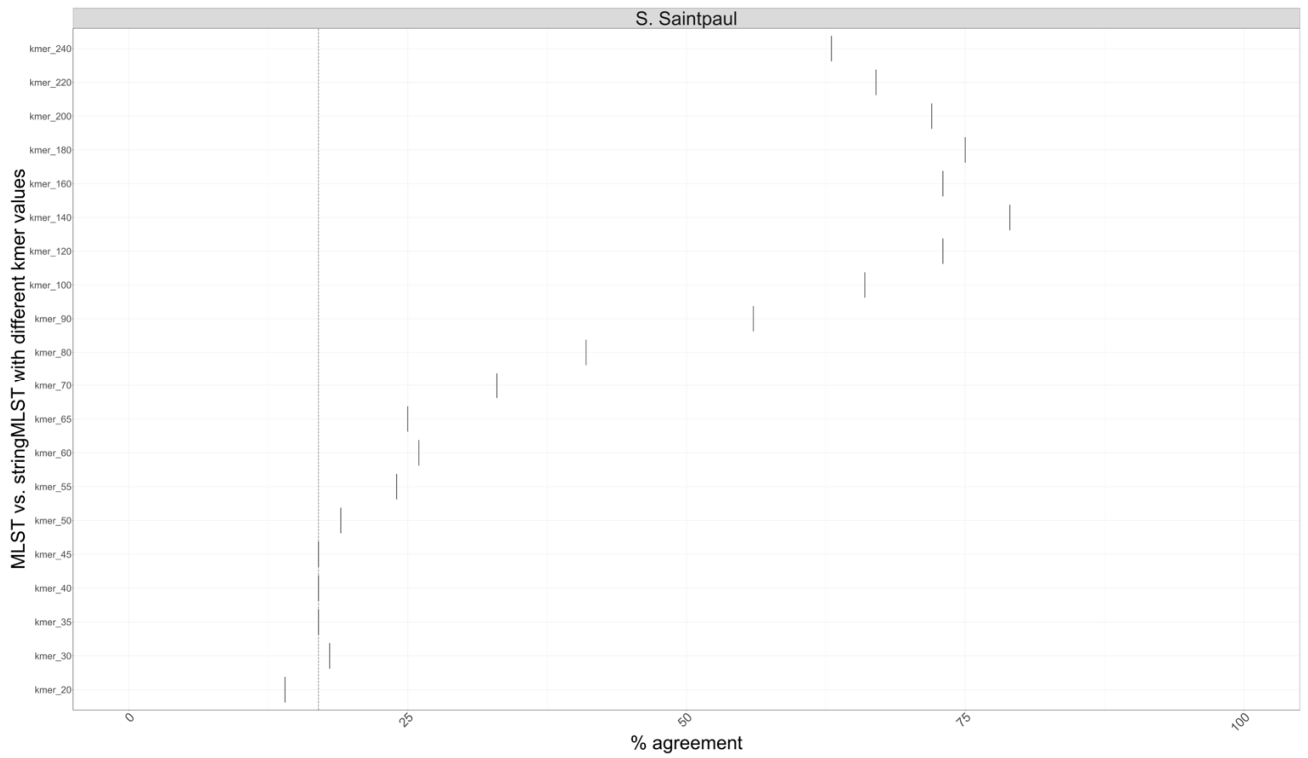
B.



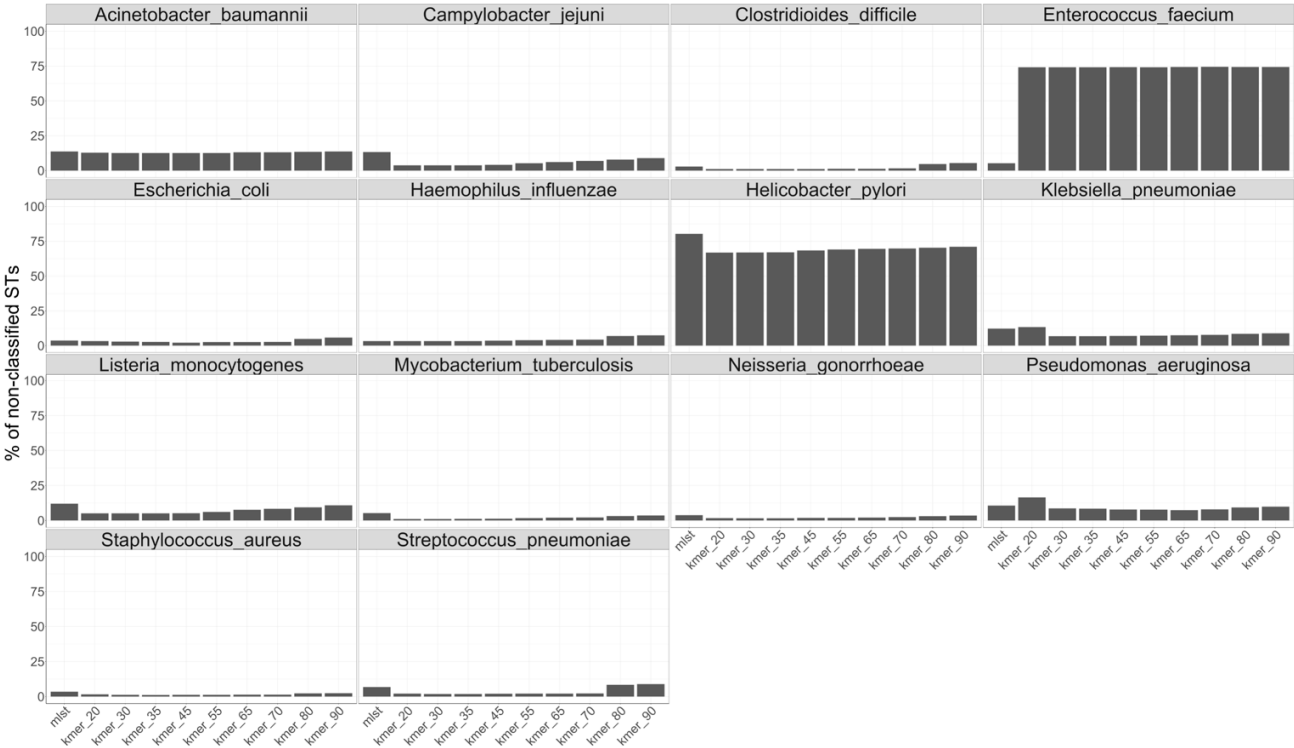
C.



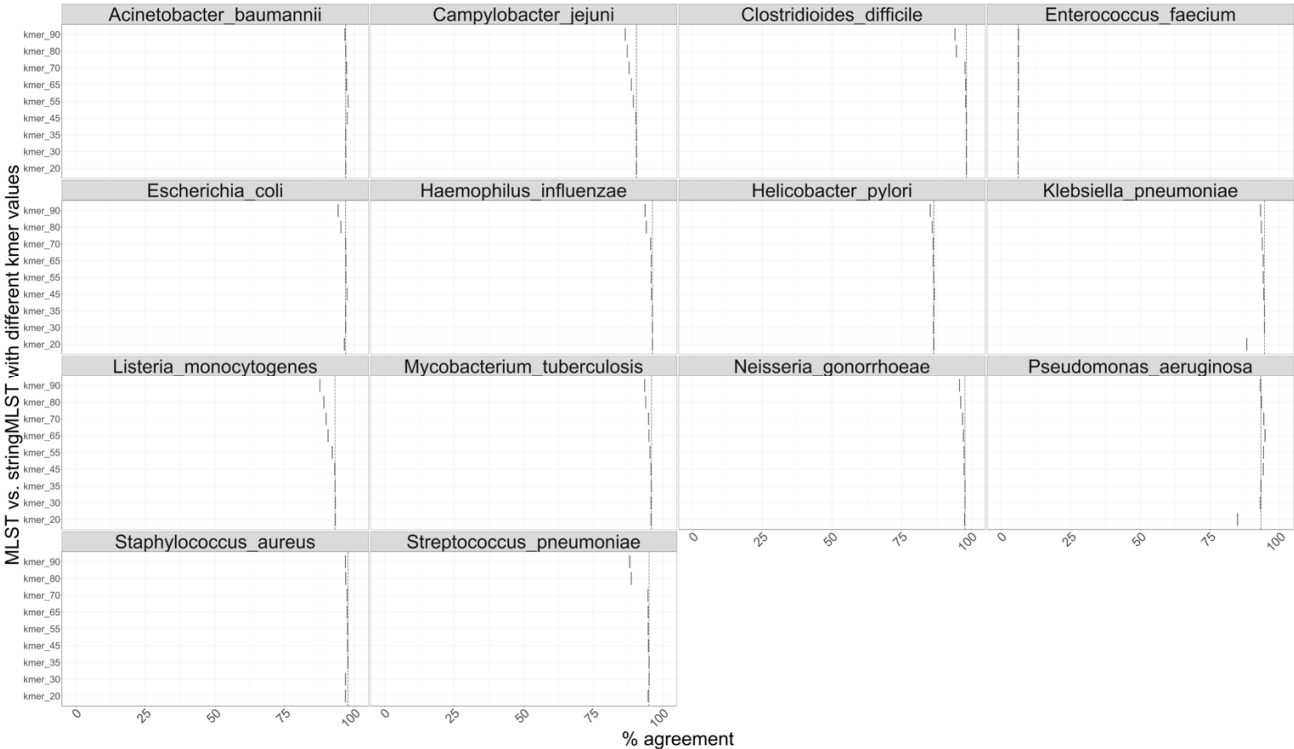
D.



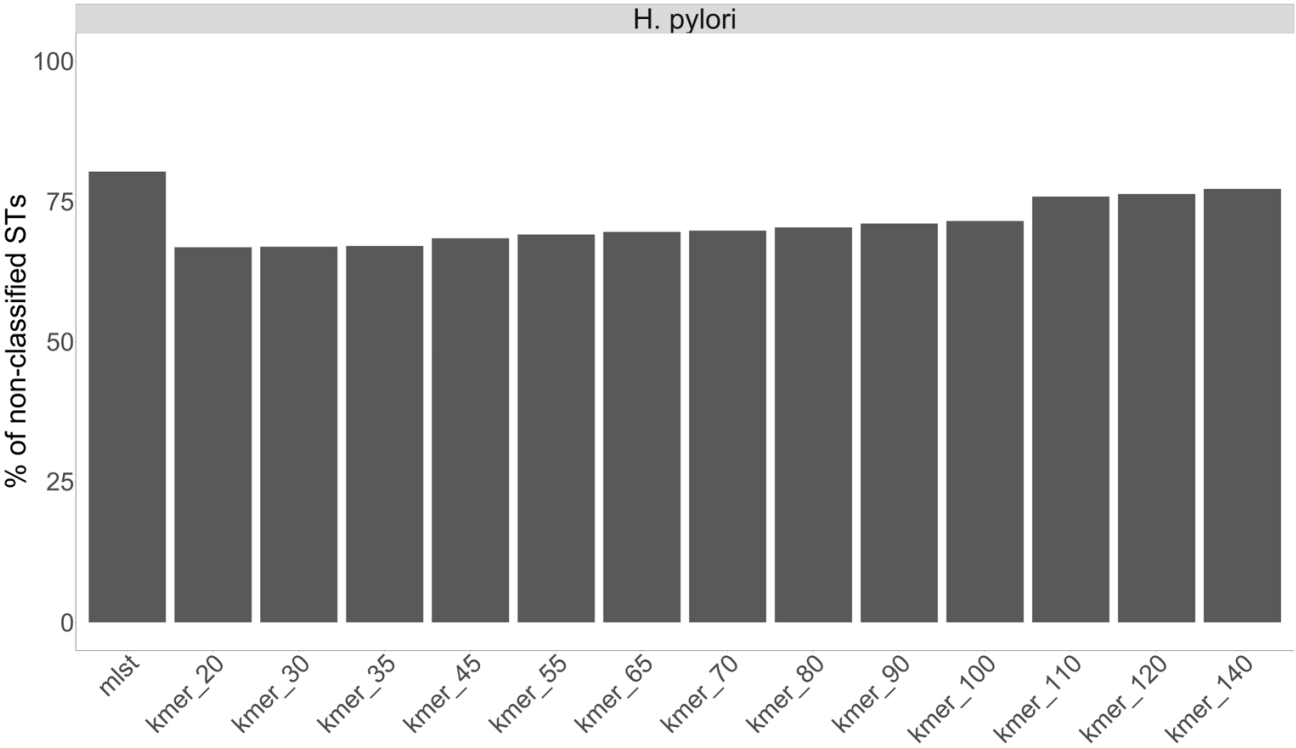
E.



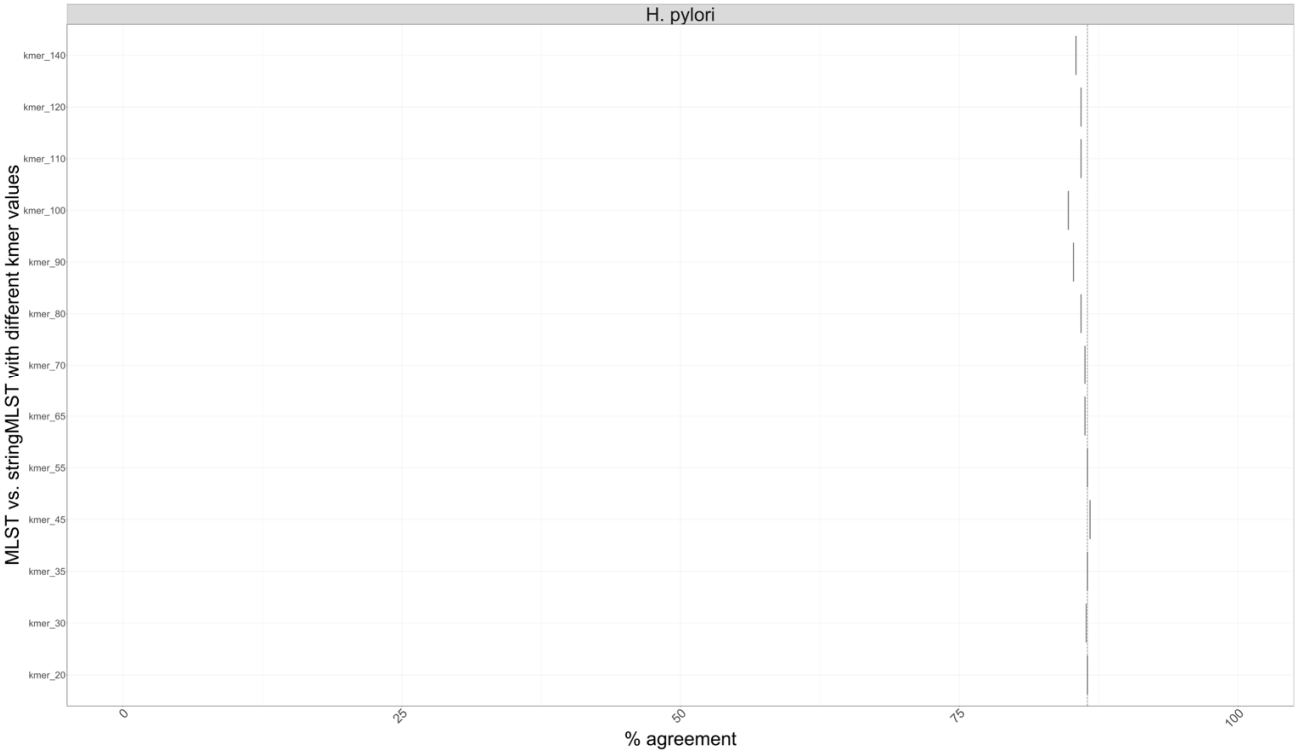
F.



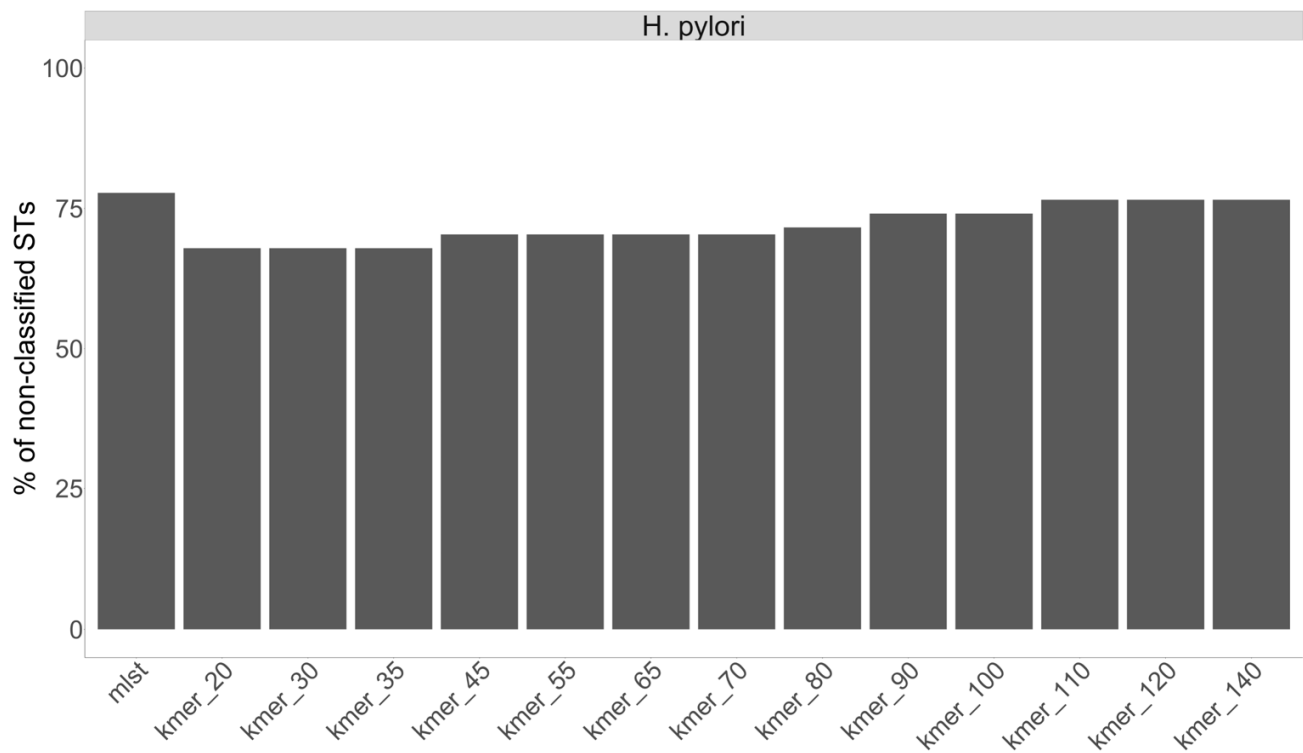
G.



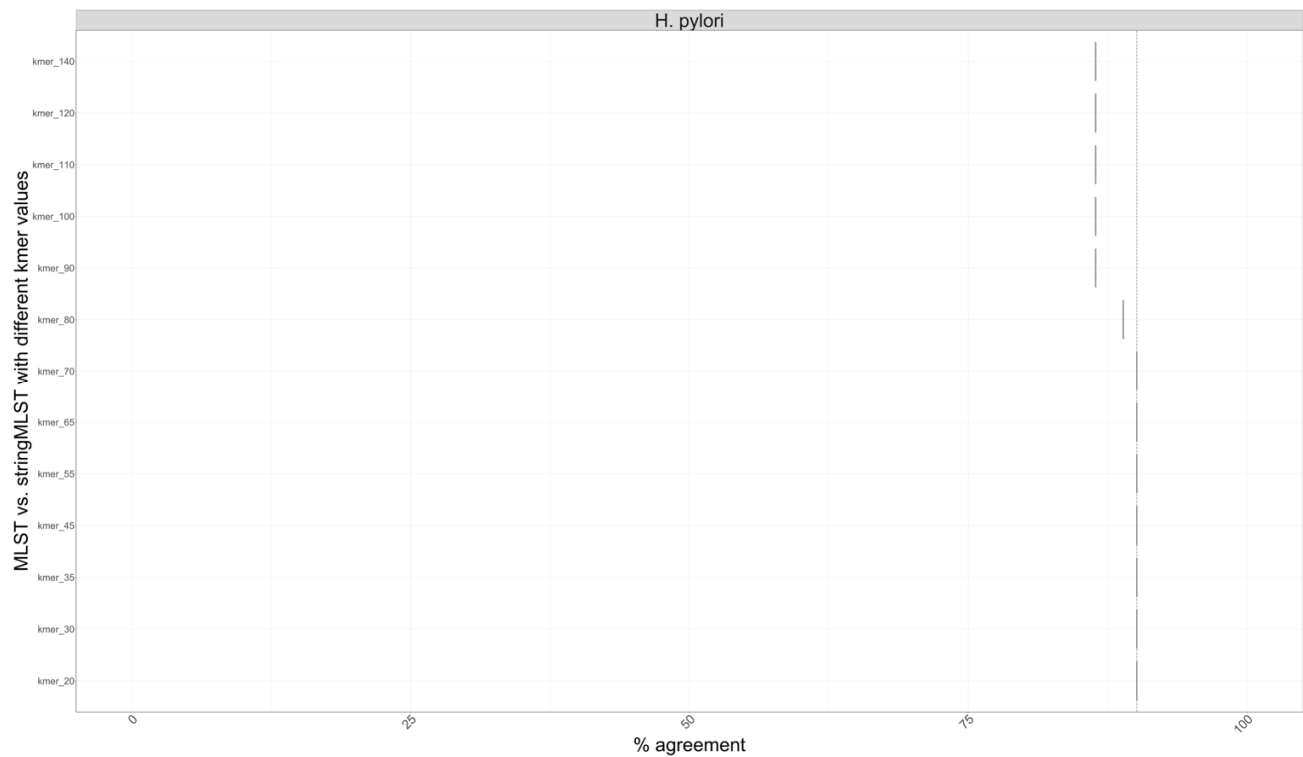
H.



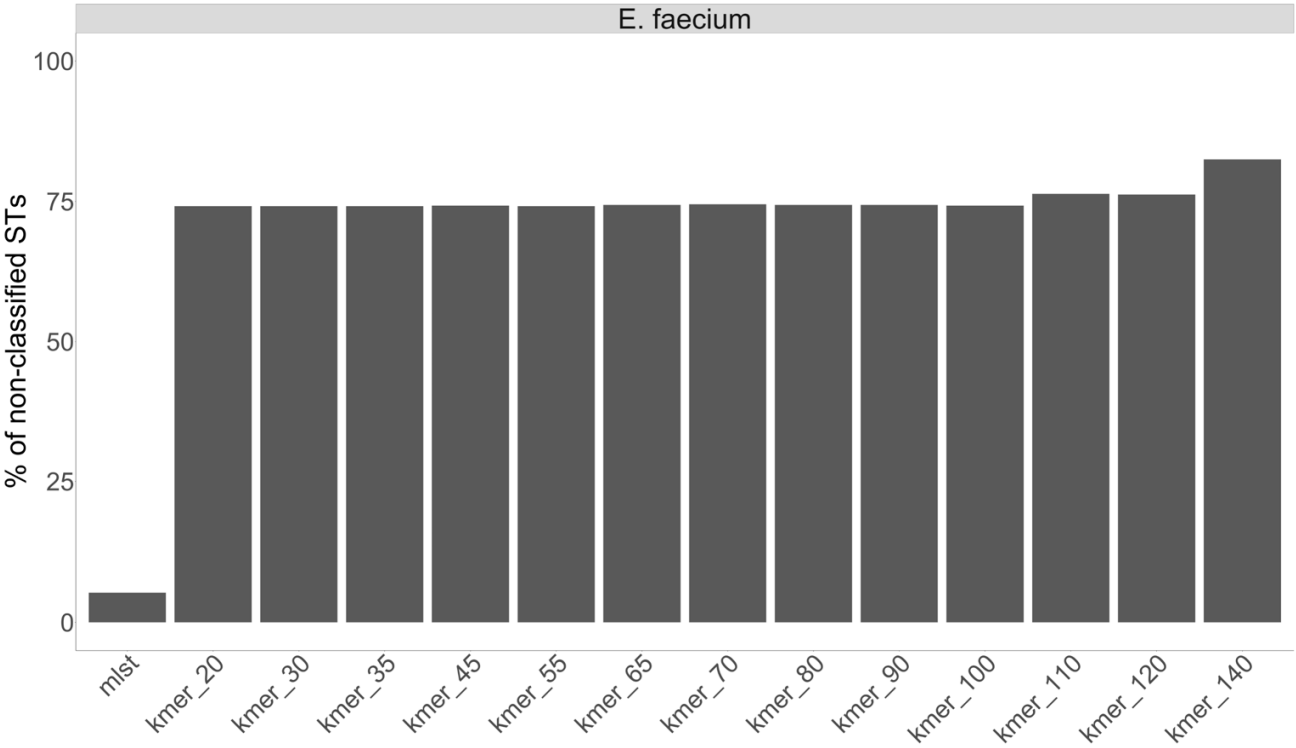
I.



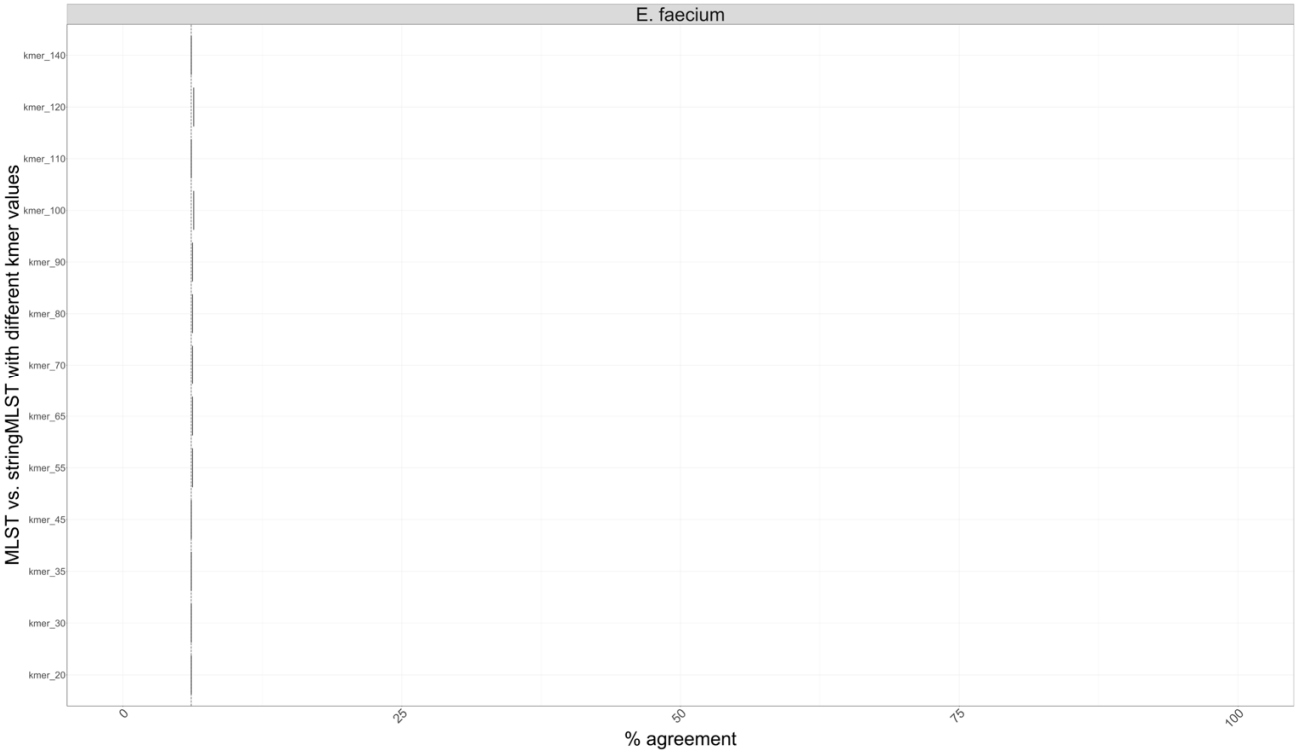
J.



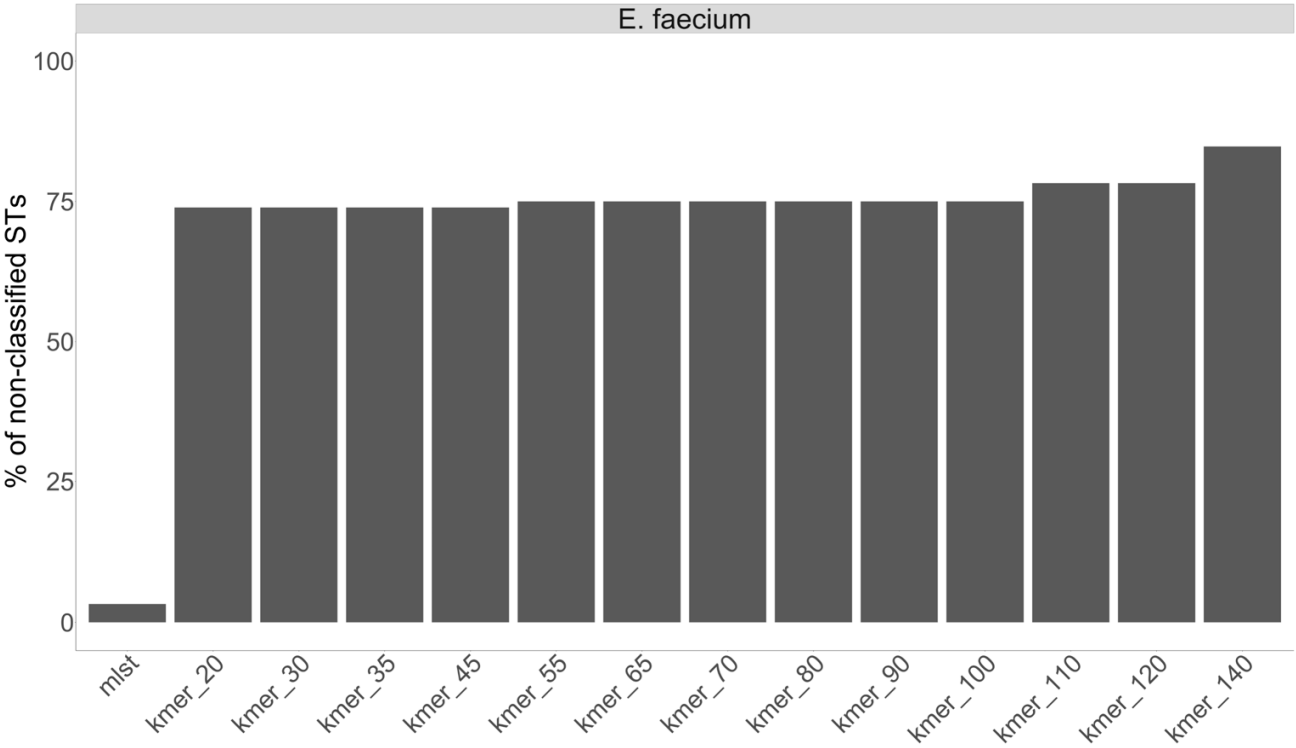
K.



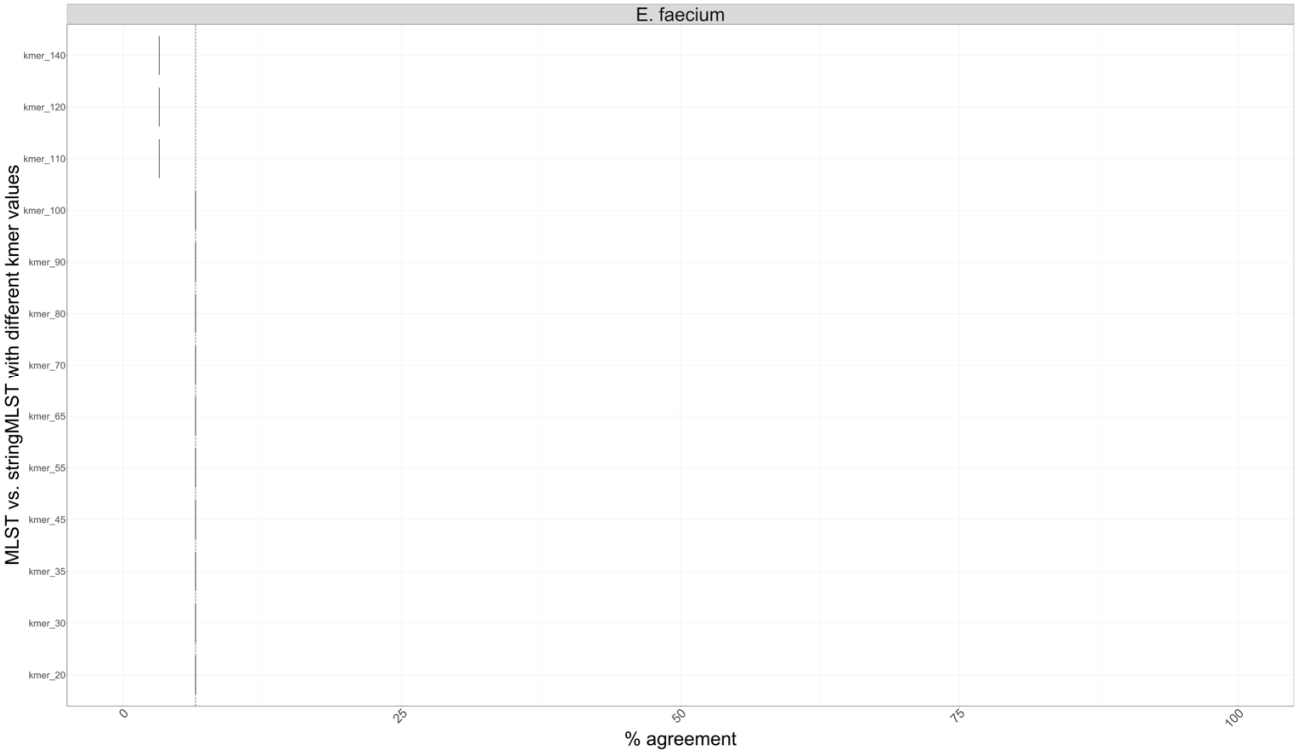
L.



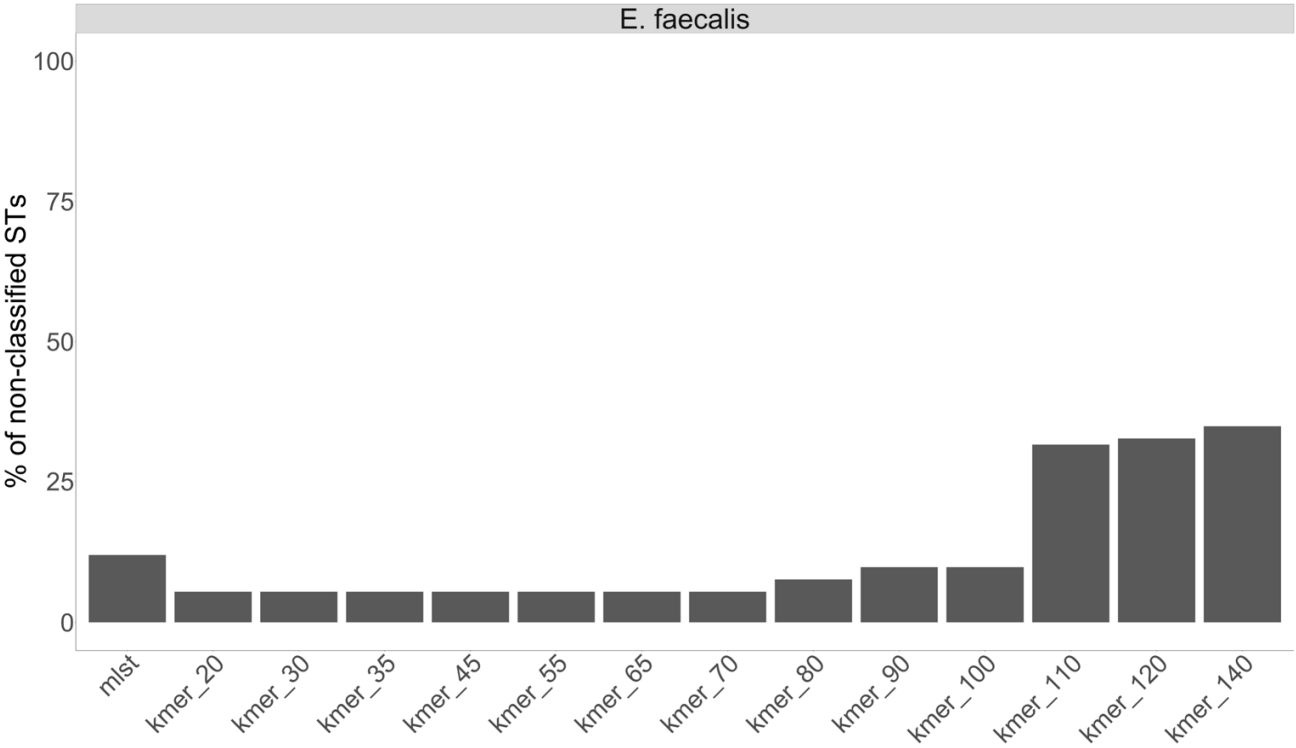
M.



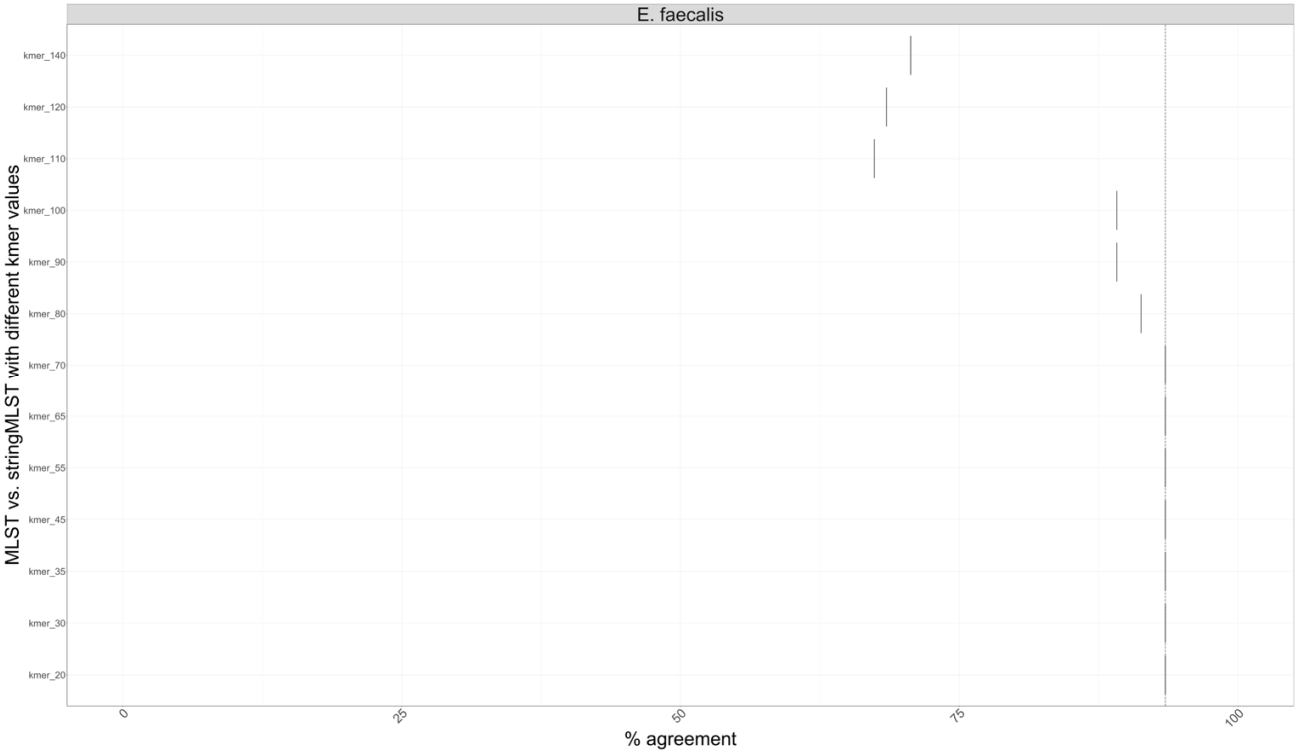
N.



O.



P.





**Figure S16.** Comparison of the statistical performance metrics between mlst and stringMLST with different k-mer lengths across *Salmonella enterica* serovars and other phylogenetic divergent bacterial pathogens.

For the twenty-three *Salmonella enterica* serovars (*S. Agona*, *S. Anatum*, *S. Braenderup*, *S. Derby*, *S. Dublin*, *S. Enteritidis*, *S. Hadar*, *S. Heidelberg*, *S. Infantis*, *S. Javiana*, *S. Johannesburg*, *S. Kentucky*, *S. Mbandaka*, *S. Montevideo*, *S. Muenchen*, *S. Newport*, *S. Oranienburg*, *S. Poona*, *S. Saintpaul*, *S. Schwarzengrund*, *S. Senftenberg*, *S. Thompson*, and *S. Typhimurium*), we randomly chose and downloaded 100 paired-end reads from NCBI-SRA. Next, for each dataset we ran mlst and stringMLST with k-mer lengths ranging from 20, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90. For the fourteen bacterial pathogens (*Acinetobacter baumannii*, *Clostridioides difficile*, *Enterococcus faecium*, *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Klebsiella pneumoniae*, *Mycobacterium tuberculosis*, *Neisseria gonorrhoeae*, *Pseudomonas aeruginosa*, *Streptococcus pneumoniae*, *Campylobacter jejuni*, *Listeria monocytogenes*, and *Staphylococcus aureus*), we randomly chose and downloaded 1,000 paired-end reads from NCBI-SRA. Next, for each dataset we ran mlst and stringMLST with k-mer lengths ranging from 20, 30, 35, 45, 55, 65, 70, 80, 90. (A) Percentage of non-classified STs (ST calls that returned missing/blank values) across twenty-three *S. enterica* serovars for mlst and stringMLST using an array of k-mer lengths (20, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90); (B) Percentage of agreement (concordance) between programs (“good” or “bad” ST calls that matched between mlst and stringMLST for different k-mer lengths) across all twenty-three *S. enterica* serovars; (C) Percentage of non-classified STs for *S. Saintpaul* using mlst and longer k-mer lengths with stringMLST; (D) Percentage of agreement between programs for *S. Saintpaul* using mlst and longer k-mer lengths with stringMLST; (E) Percentage of non-classified STs (ST calls that returned missing/blank values) across fourteen phylogenetic divergent bacterial pathogens for mlst and stringMLST with range of k-mer lengths (20, 30, 35, 45, 55, 65, 70, 80, 90); (F) Percentage of agreement between programs (“good” or “bad” ST calls that matched between mlst and stringMLST for different k-mer lengths) across all fourteen phylogenetic divergent bacterial pathogens; (G) Percentage of non-classified STs for *H. pylori* (dataset of 1,000 genomes) using mlst and longer k-mer lengths with stringMLST; (H) Percentage of agreement between programs for *H. pylori* (dataset of 1,000 genomes) using mlst and lost k-mer lengths with stringMLST; (I) Percentage of non-classified STs for *H. pylori* (dataset of 100 genomes) using mlst and longer k-mer lengths with stringMLST; (J) Percentage of agreement between programs for *H. pylori* (dataset of 100 genomes) using mlst and lost k-mer lengths with stringMLST; (K) Percentage of non-classified STs for *E. faecium* (dataset of 1,000 genomes) using mlst and longer k-mer lengths with

stringMLST; (L) Percentage of agreement (concordance) between programs for *E. faecium* (dataset of 1,000 genomes) using mlst and longer k-mer lengths with stringMLST; (M) Percentage of non-classified STs for *E. faecium* (dataset of 100 genomes) using mlst and longer k-mer lengths with stringMLST; (N) Percentage of agreement (concordance) between programs for *E. faecium* (dataset of 100 genomes) using mlst and longer k-mer lengths with stringMLST; (O) Percentage of non-classified STs for *Enterococcus faecalis* (dataset of 100 genomes) using mlst and longer k-mer lengths with stringMLST; (P) Percentage of agreement (concordance) between programs for *E. faecalis* (dataset of 100 genomes) using mlst and longer k-mer lengths with stringMLST.