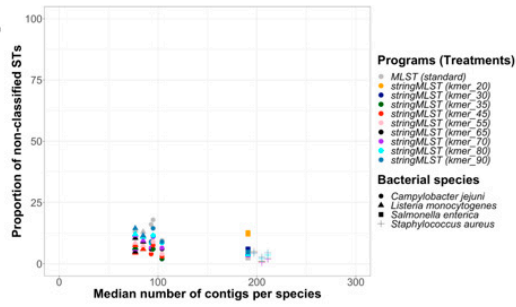
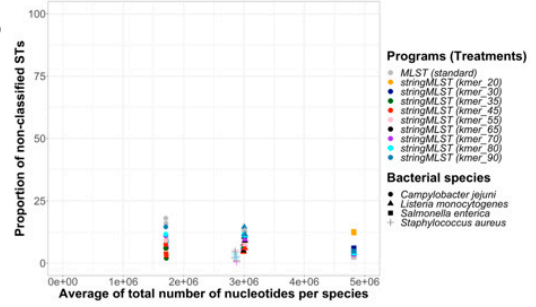


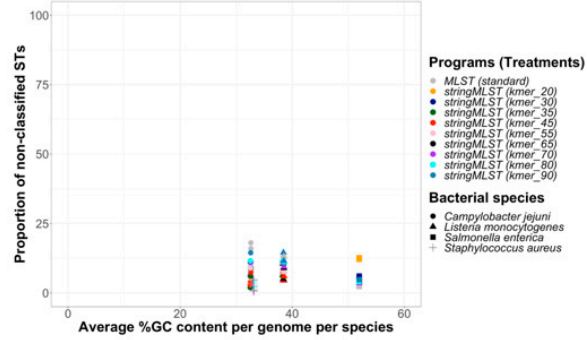
G.



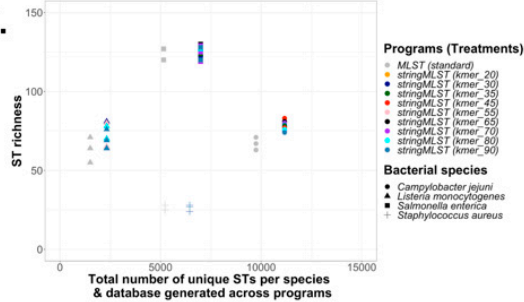
H.



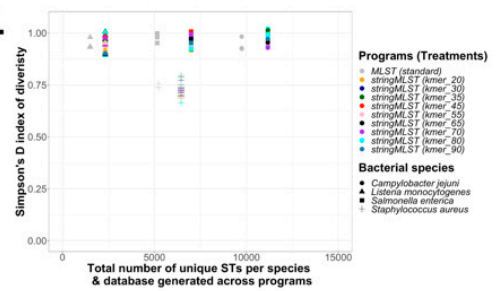
I.



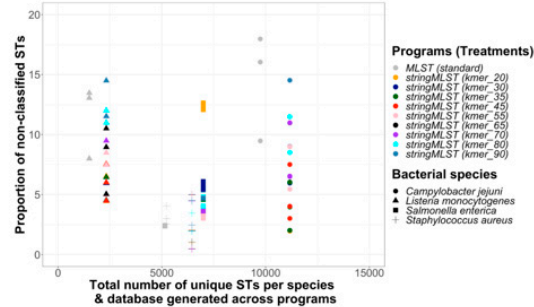
J.



K.



L.



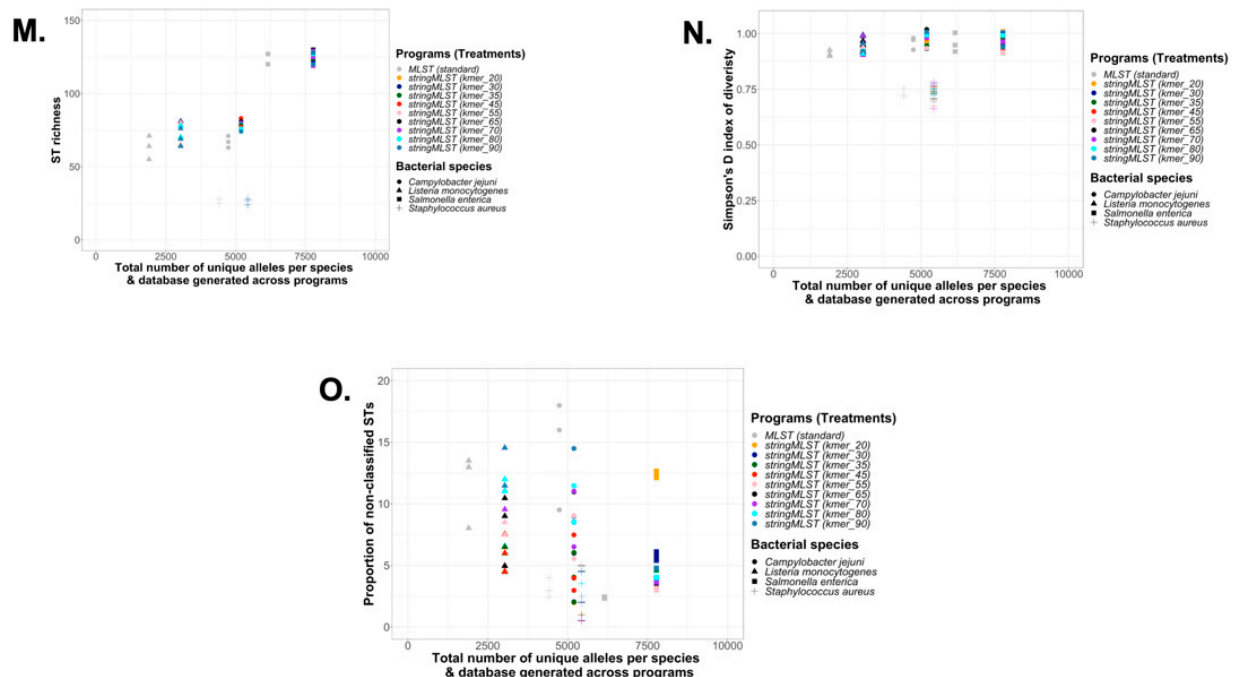


Figure S11. Bivariate association between accuracy-based classification metrics for ST calls and genome-intrinsic and –extrinsic variables.

Plot A-C depict groupings generated based on the relationship between ST richness (Richness) vs. the median number of contigs per species, or the average of total number of nucleotides per species, or the average GC% content per genome per species, respectively. Plot D-F depict groupings generated based on the relationship between the Simpson's D index of diversity vs. the median number of contigs per species, or the average of total number of nucleotides per species, or the average GC% content per genome per species, respectively. Plot G-I depict groupings generated based on the relationship between the proportion of non-classified STs vs. the median number of contigs per species, or the average of total number of nucleotides per species, or the average GC% content per genome per species, respectively. Plot J-L depict groupings generated based on the relationship between the total number of unique STs per species (generated across databases per program) vs. ST richness, or the Simpson's D index of diversity, or the proportion of non-classified STs, respectively. Plot M-O depict groupings generated based on the relationship between the total number of unique alleles per species (per database generated across programs) vs. ST richness, or the Simpson's D index of diversity, or the proportion of non-classified STs, respectively.