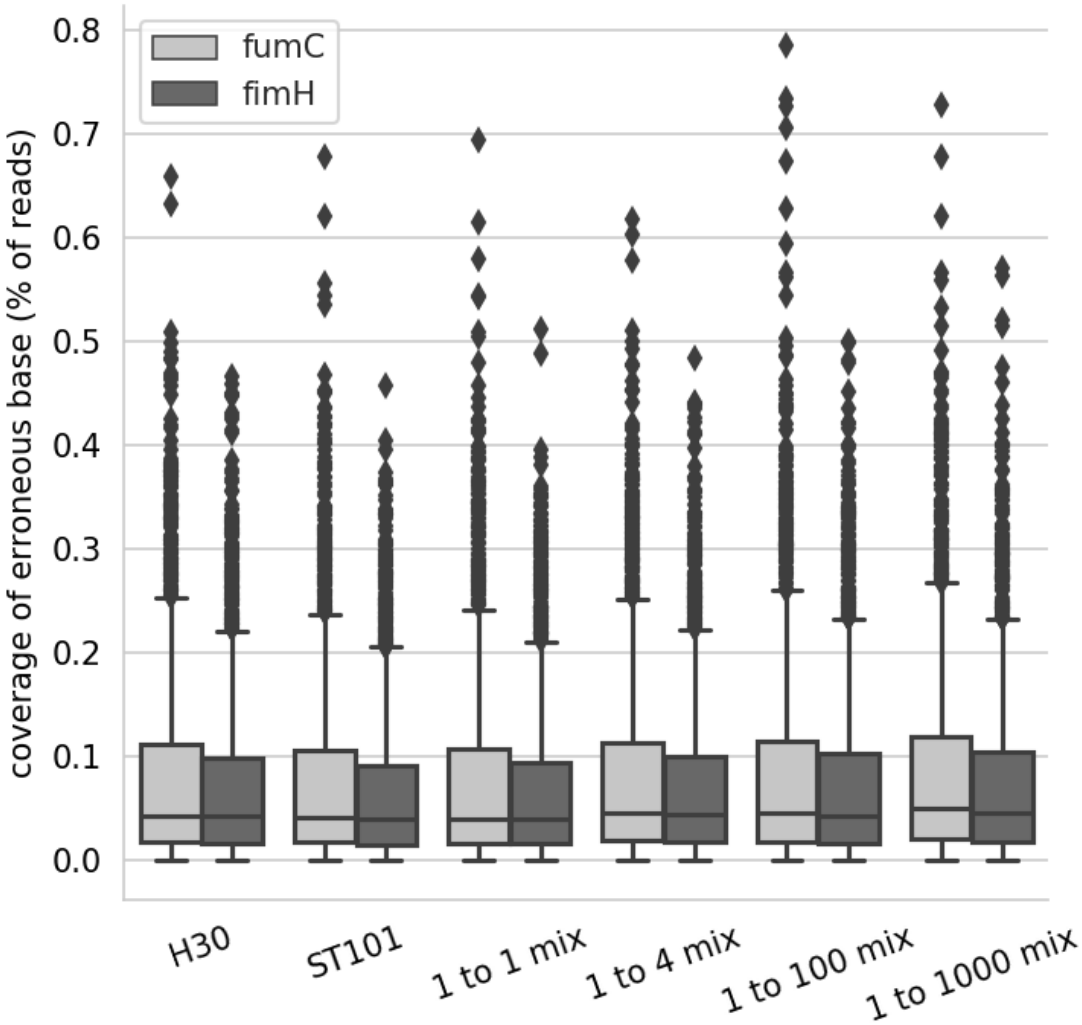
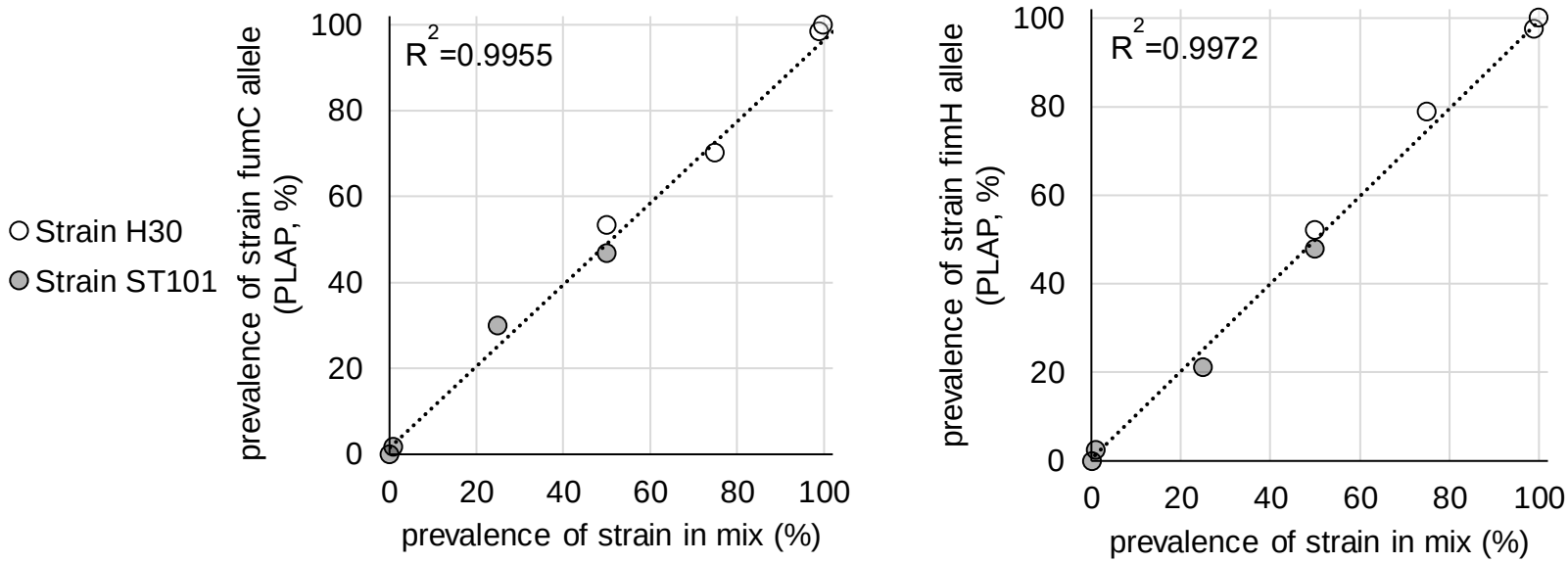
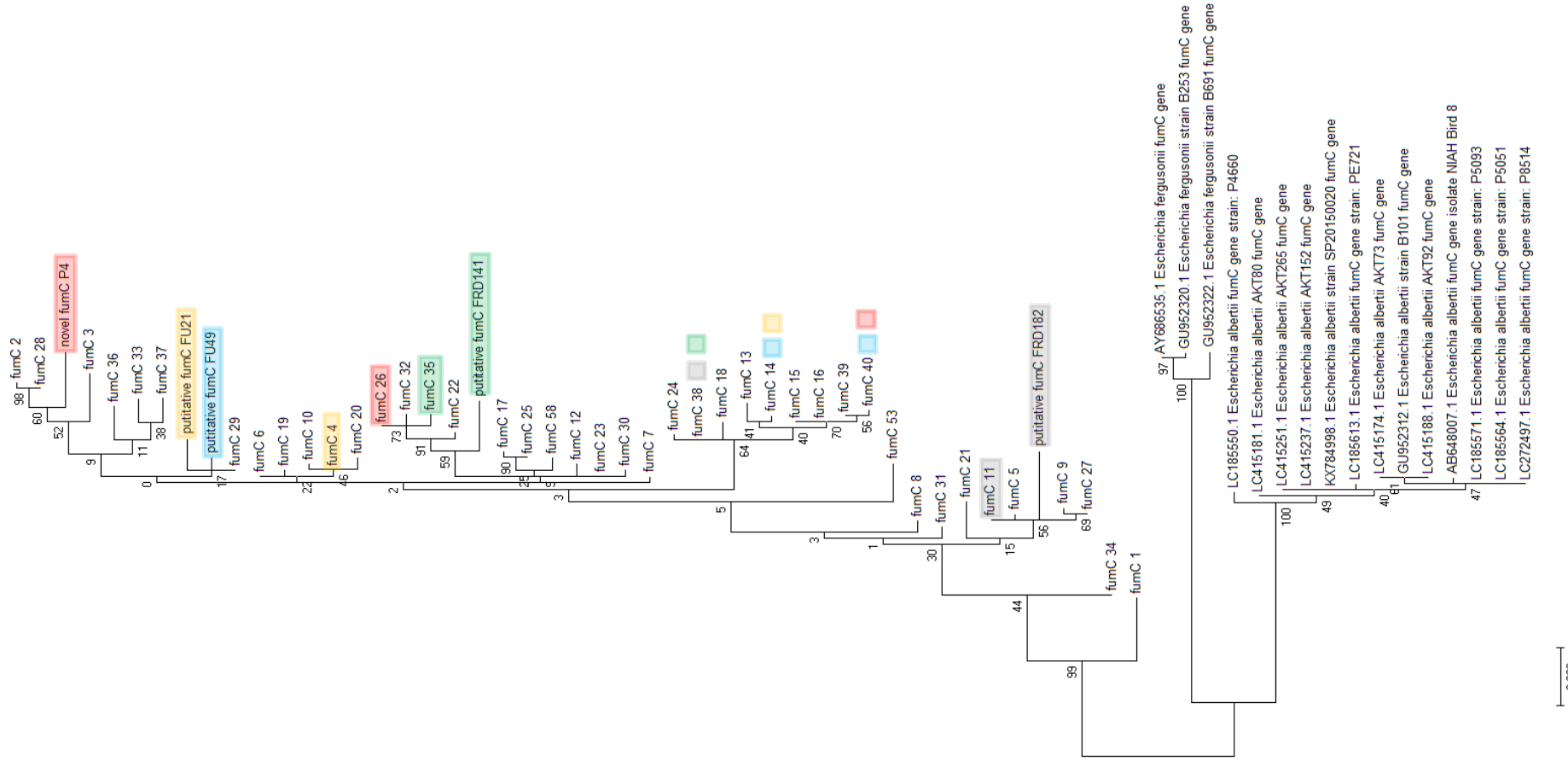


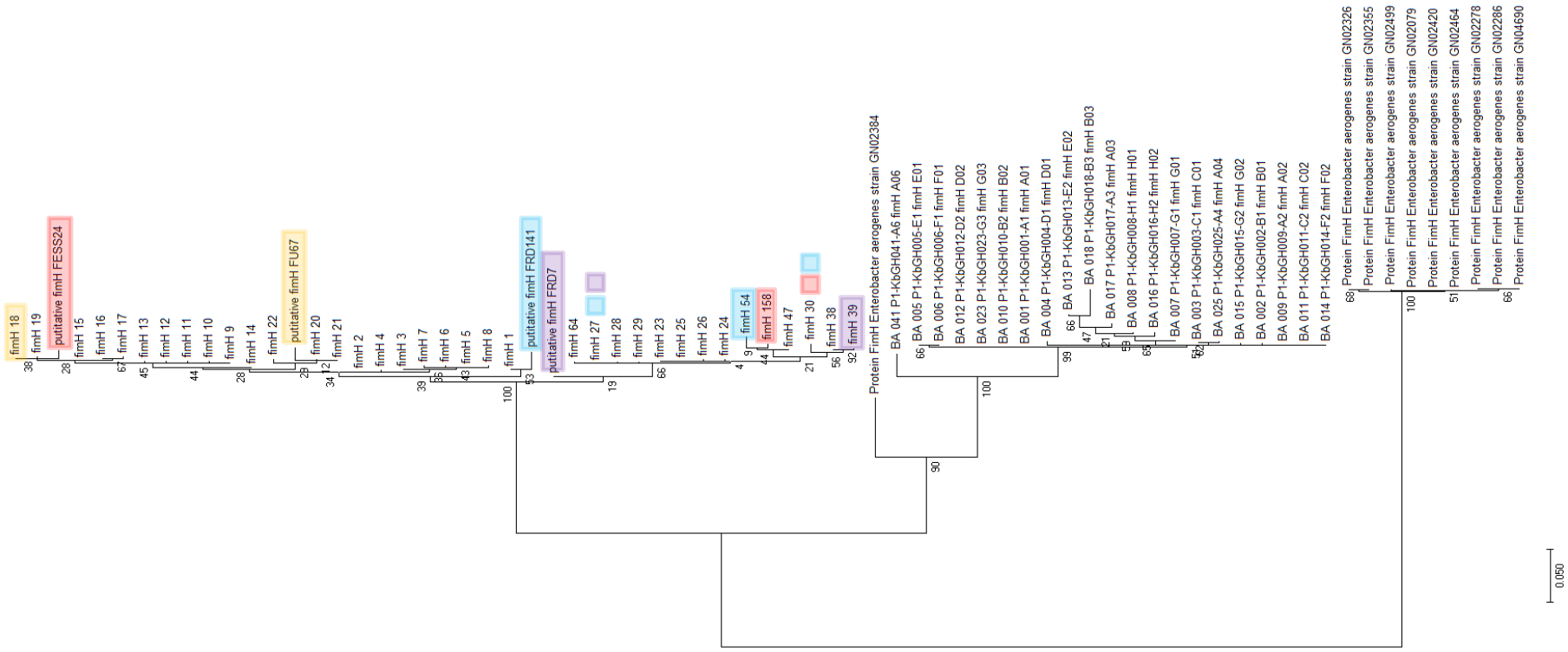
Supplemental Figure 1. Coverage of erroneous bases in *H*30-only, ST101-only, and mix sample sequencing. Coverage is expressed in percentage of total reads aligned to each gene.



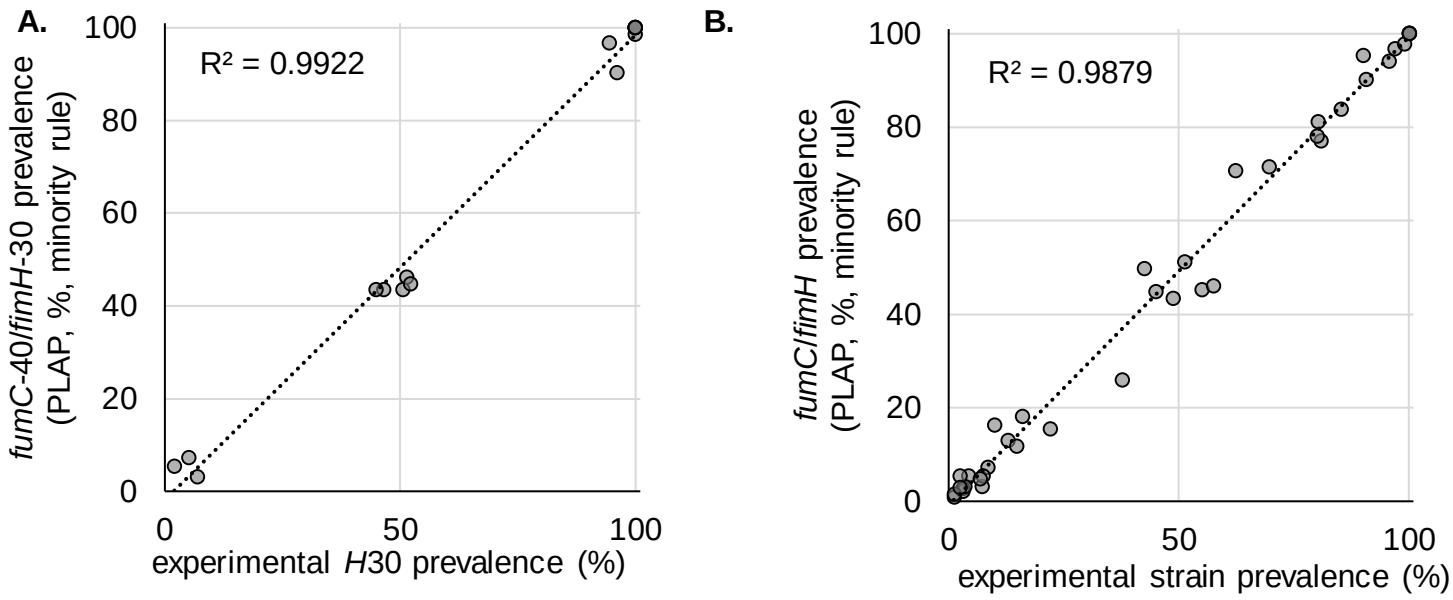
Supplemental Figure 2. Correlation between input and PLAP-derived (deep seq) prevalences of *fumC* and *fimH* alleles of *H*30 and ST101 in 1:1, 1:4, and 1:100 mixes.



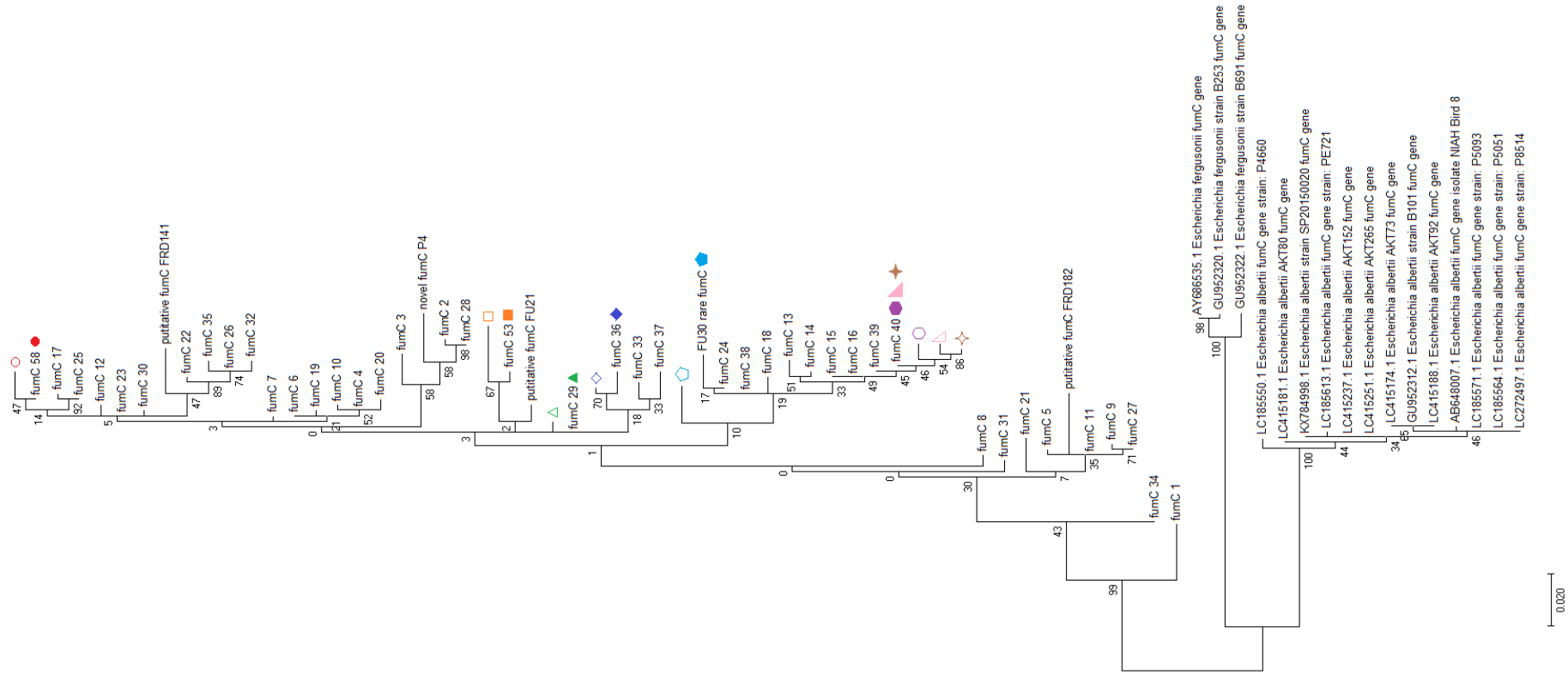
Supplemental Figure 3. Phylogenetic relationships between predicted novel *fumC* alleles and known *E. coli fumC* alleles. *Escherichia fergusonii* and *albertii fumC* alleles also presented for outgroup reference. Alleles not labelled with a species are known *E. coli* alleles or putative novel alleles. Alleles found in the sample as the novel allele are highlighted in the same color as the novel allele to show distance between predicted novel alleles and other *fumC* alleles present in the sample. Alleles present in multiple different samples are marked with the appropriate colors next to the allele name.



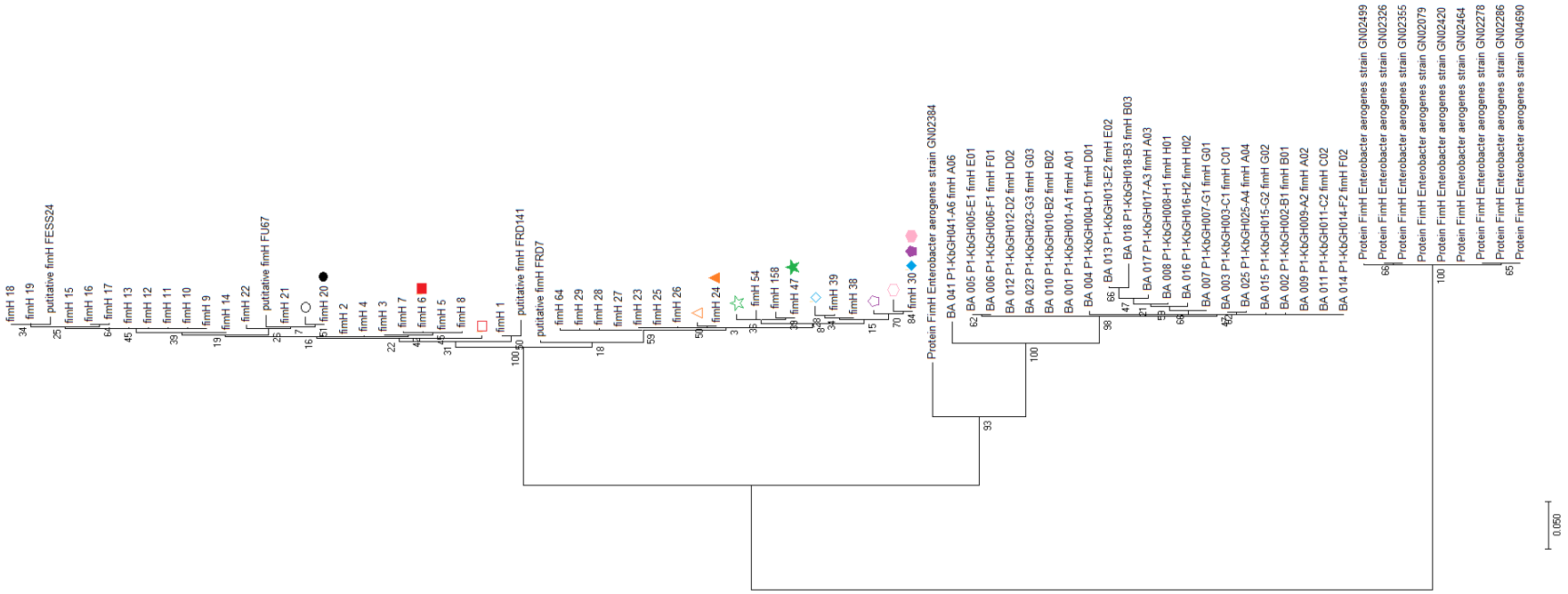
Supplemental Figure 4. Phylogenetic relationships between predicted novel *fimH* alleles and known *E. coli fimH* alleles. *Klebsiella pneumoniae* and *Enterobacter aerogenes fimH* alleles also presented for outgroup reference. Alleles not labelled with a species are known *E. coli* alleles or putative novel alleles. Alleles found in the sample as the novel allele are highlighted in the same color as the novel allele to show distance between predicted novel alleles and other *fimH* alleles present in the sample. Alleles present in multiple different samples are marked with the appropriate colors next to the allele name.



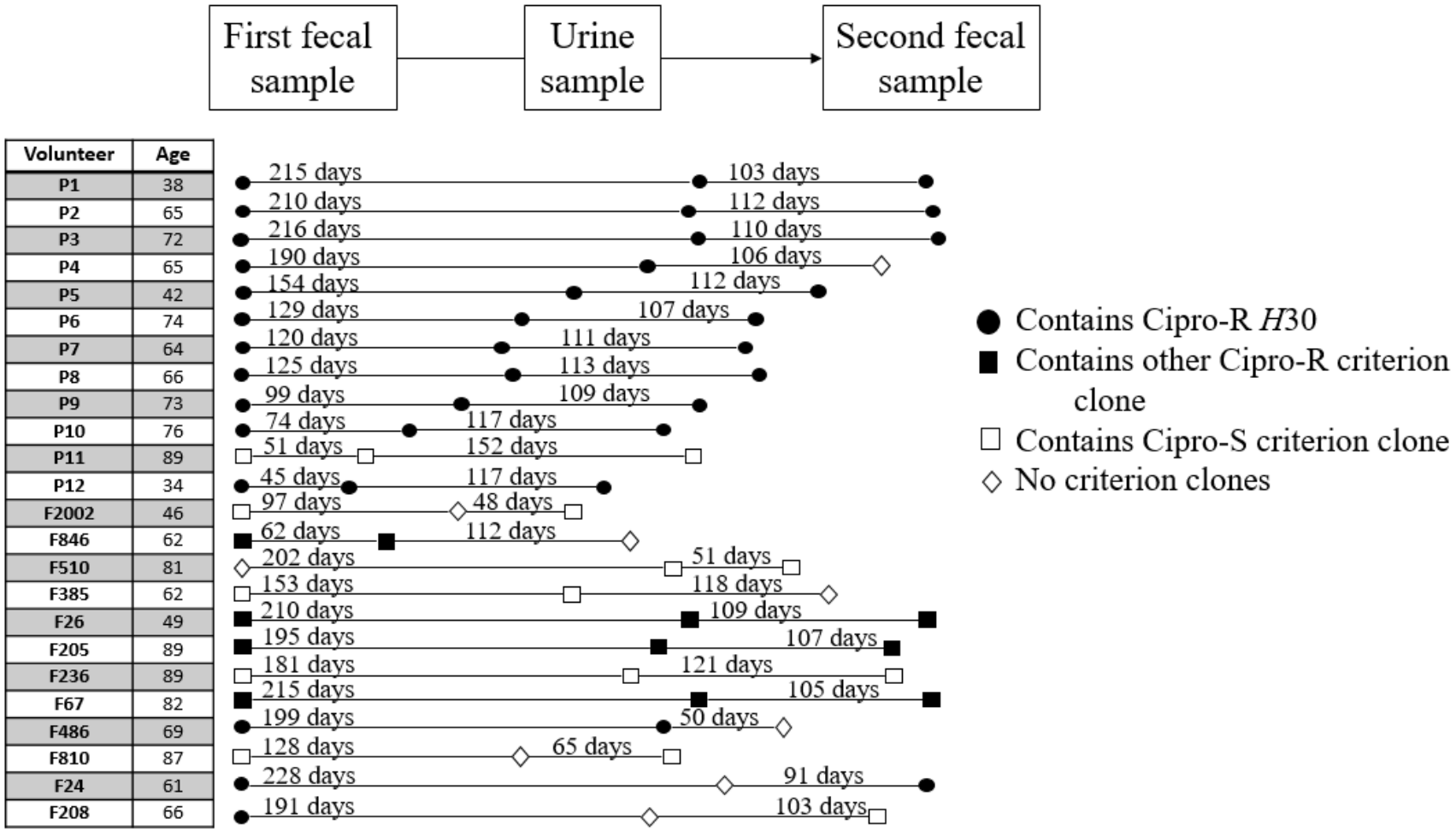
Supplemental Figure 5. A. Comparison of actual *H*30 load in *H*30-containing fecal samples to PLAP-predicted *fumC*-40/*fimH*-30 prevalences with minority rule correction (i.e. the smaller prevalence of the two was used). Prevalence of *fumC*-40/*fimH*-30 is expressed as percentage of all *E. coli* in each sample. *H*30 load is expressed as ratio of *H*30 (ciprofloxacin-resistant) single colonies to all plated *E. coli* single colonies in percent. **B.** PLAP-predicted allele prevalence (with minority rule correction) compared to experimental allele prevalence as determined by surveying at least 40 single colonies per sample.



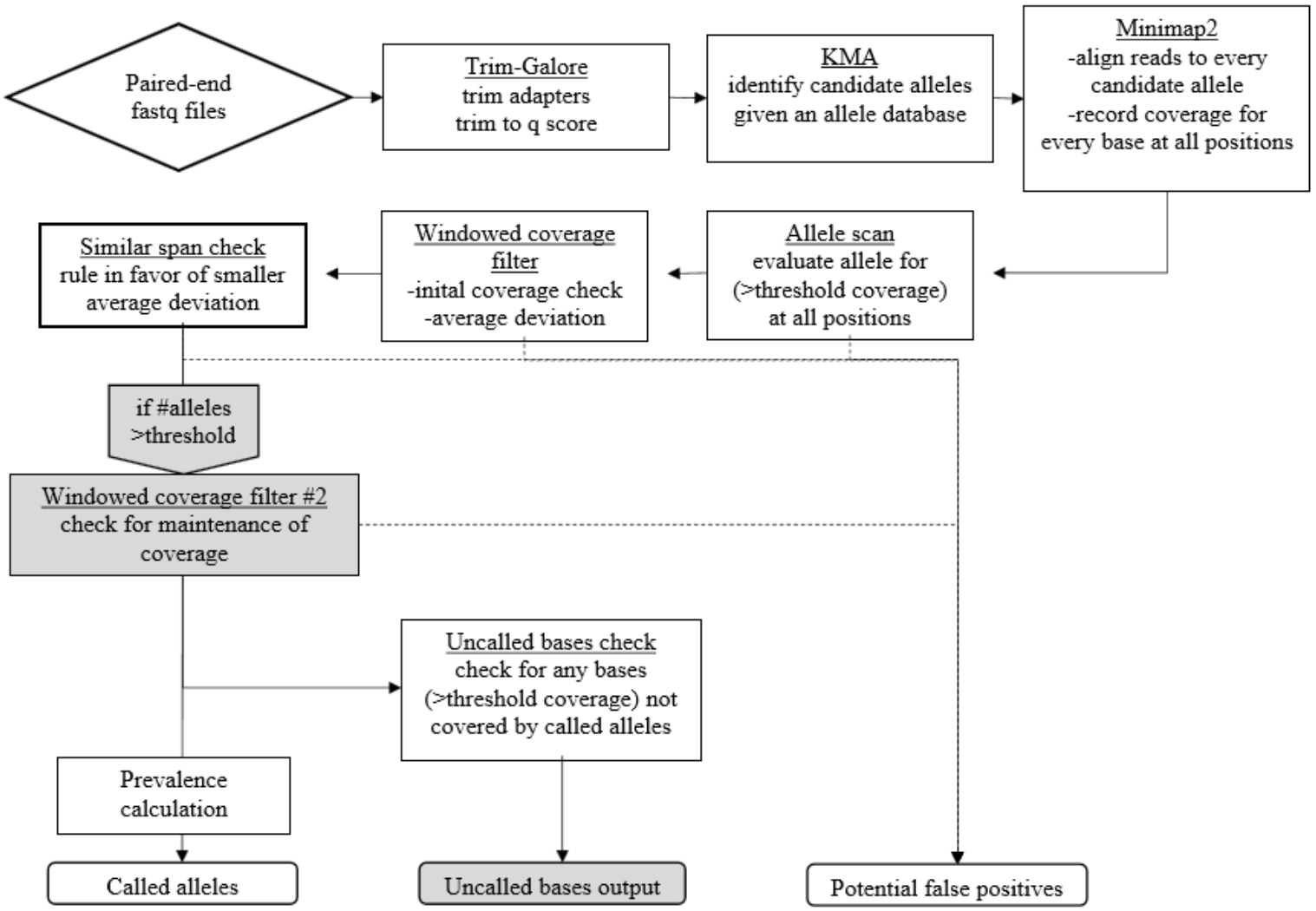
Supplemental Figure 6. Putative rare novel *fumC* alleles identified by lowering the error threshold from 0.8% to 0.5%, marked in open shapes. Known alleles from the same sample as the rare novel allele are marked in filled-in shapes of the same type and color. FumC-40 was present in 3 different samples and therefore is marked by 3 different shapes.



Supplemental Figure 7. Putative rare novel *fimH* alleles identified by lowering the error threshold from 0.8% to 0.5%, marked in open shapes. Known alleles from the same sample as the rare novel allele are marked in filled-in shapes of the same type and color. FimH-30 was present in 3 different samples and therefore is marked by 3 different shapes.



Supplemental Figure 8. Sampling of volunteer sample sets. Length of segments is proportional to number of days between samples.



Supplemental Figure 9. PLAP algorithm workflow. Algorithms previously developed by other groups include Trim-Galore, KMA, Minimap2. Not pictured but used during windowed coverage checks is SAMtools.