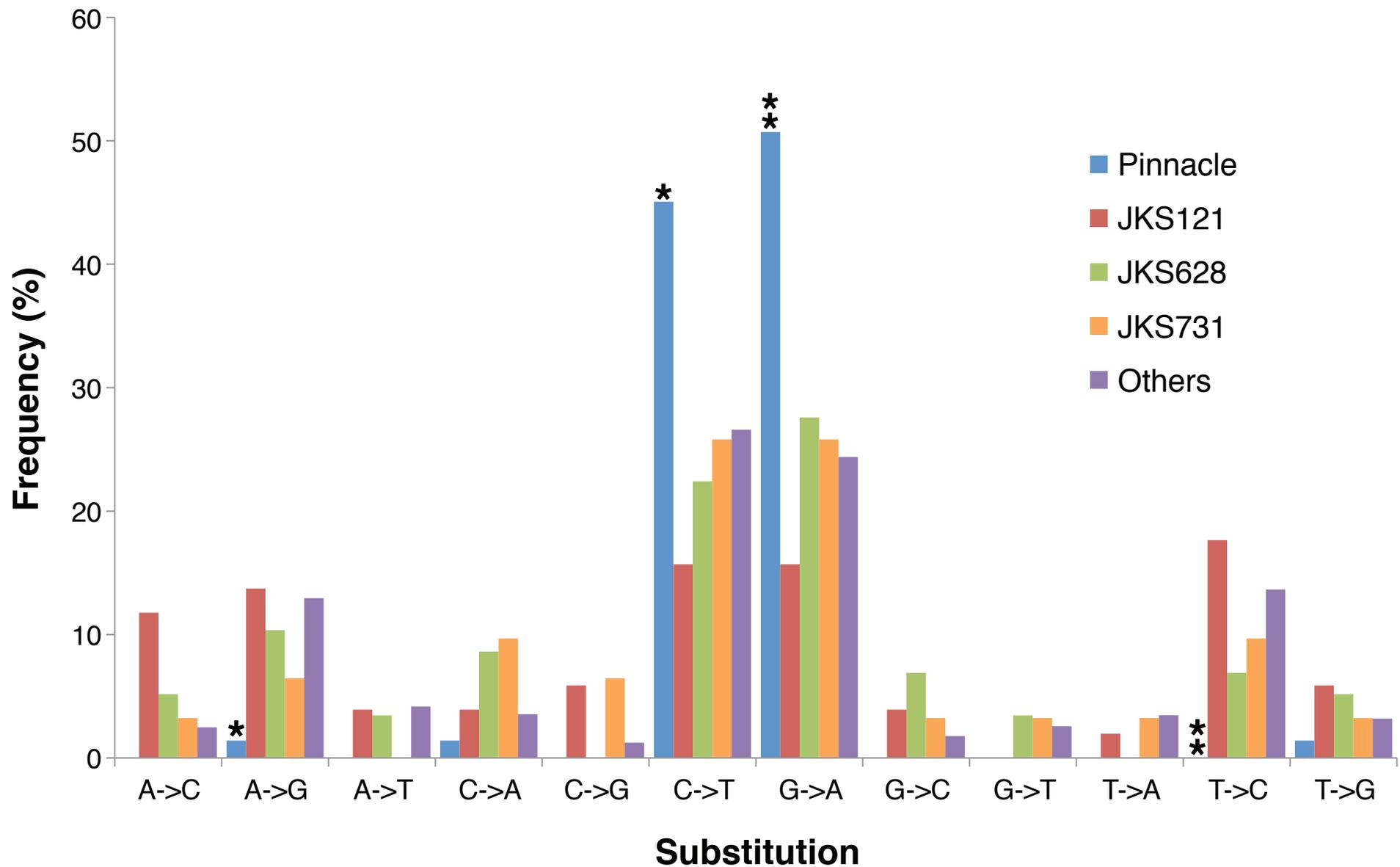
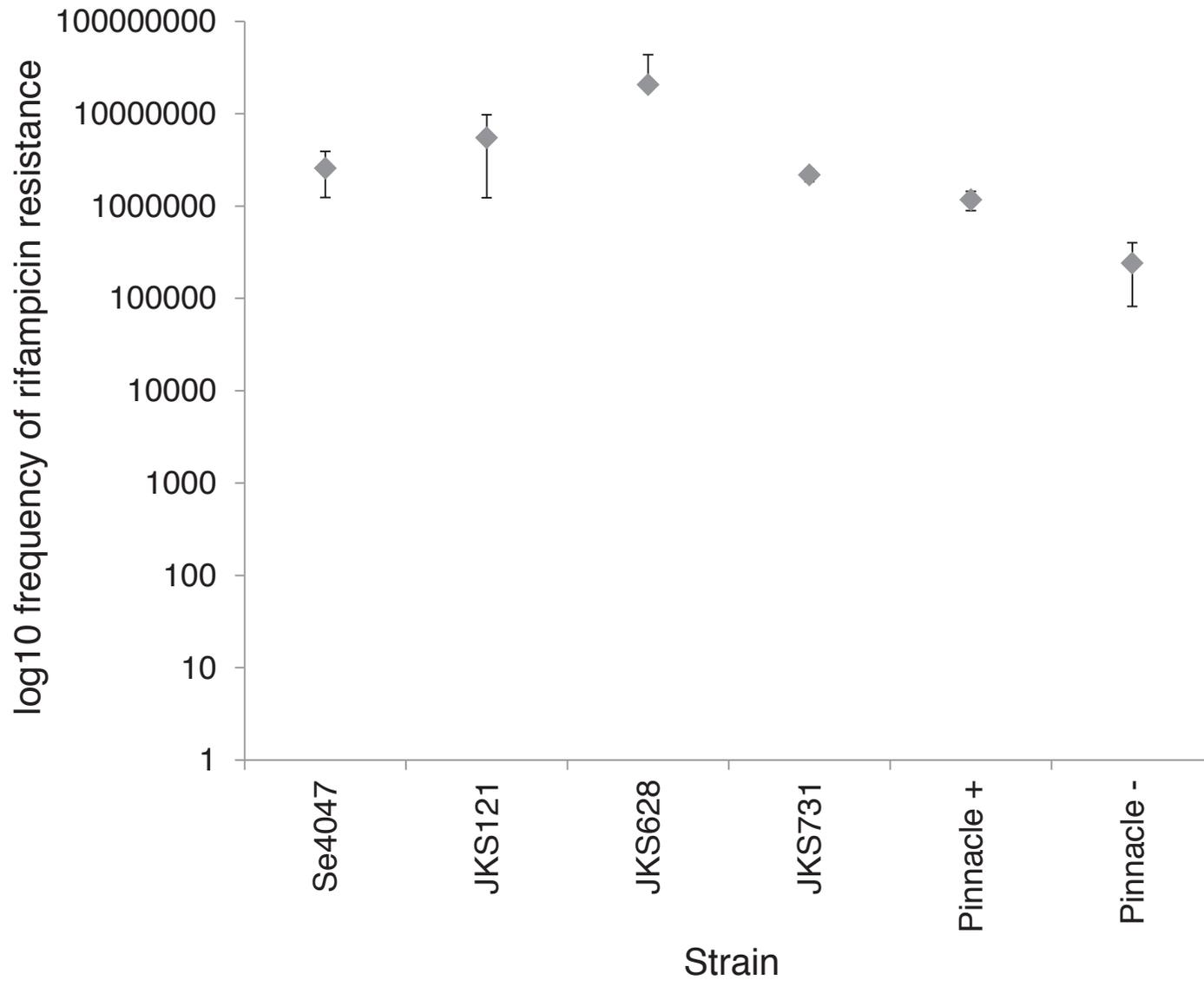


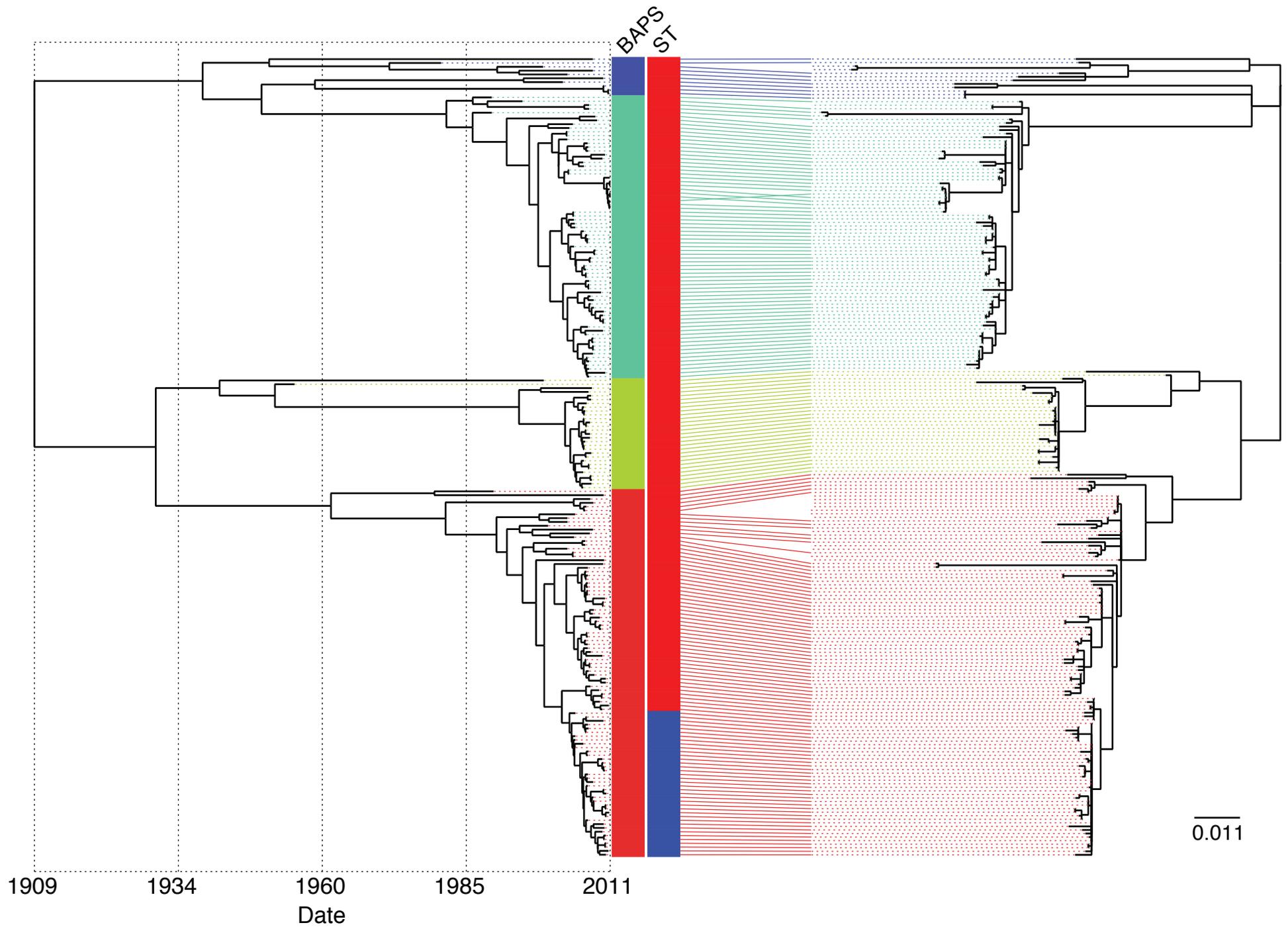
**Supplementary Figure 1.** Representation of predicted homologously-recombined regions. The left panel represents the ML phylogeny of *S. equi*, with BAPs cluster and MLST type shown in columns adjacent to the tree, as in Figure 1. The right panel represents regions identified as exhibiting significantly raised SNP density, indicative of import of variation en masse via homologous recombination or regions under high selective pressures. Above the panel is a representation of the genome annotation of *S. equi* Se4047.



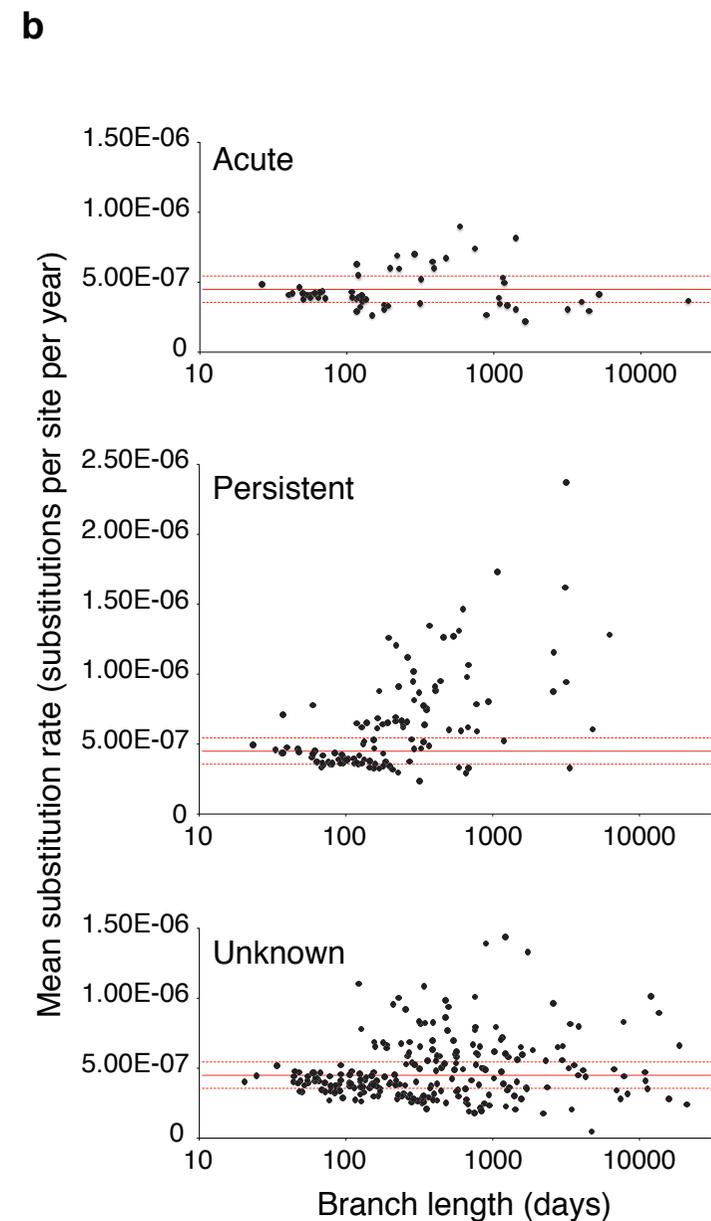
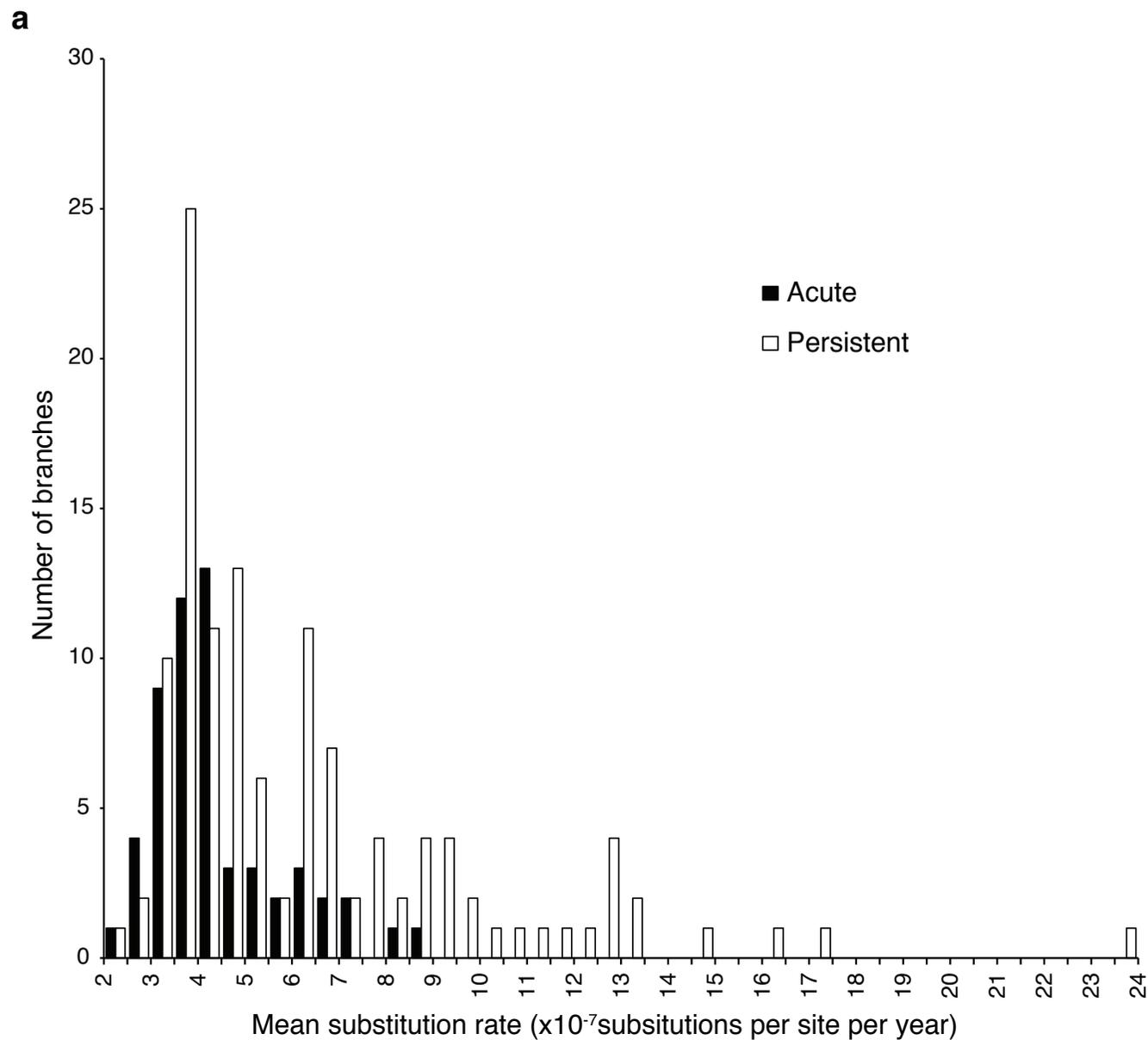
**Supplementary Figure 2.** Mutation spectra associated with branches on the tree leading to the outliers in the root-to-tip analysis (Supplementary Fig. 1) and all other branches. \* indicates significant difference to 'others' at the 0.1 level, while \*\* indicates significance at the 0.05 level. Colors correspond to colors in Supplementary Figure 1.



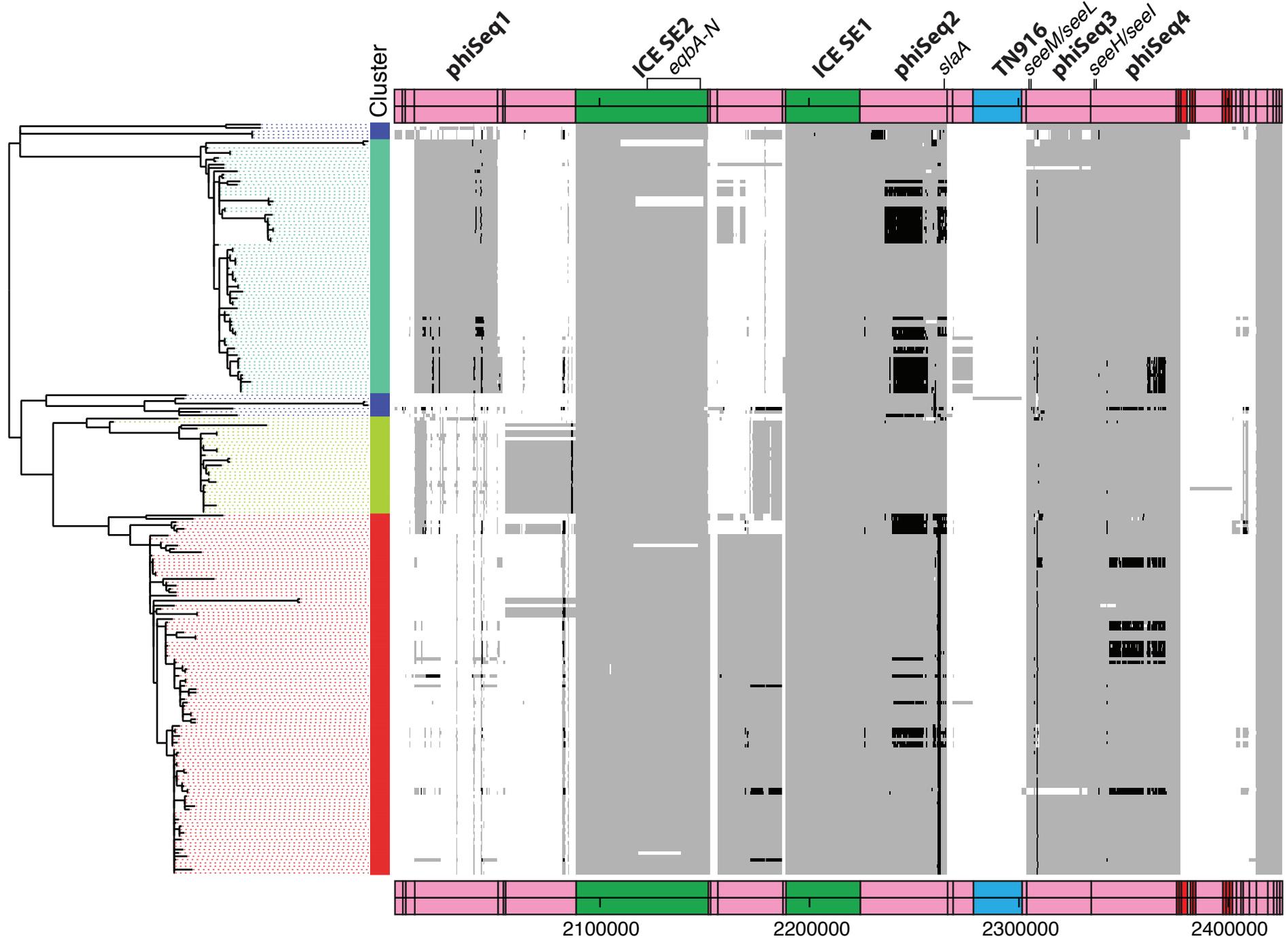
**Supplementary Figure 3.** Mean resistance frequency of long-branch isolates and the reference *Se4047* in vitro to rifampicin, Values represent the means of three independent experiments conducted in triplicate. Error bars indicate 95% confidence intervals.



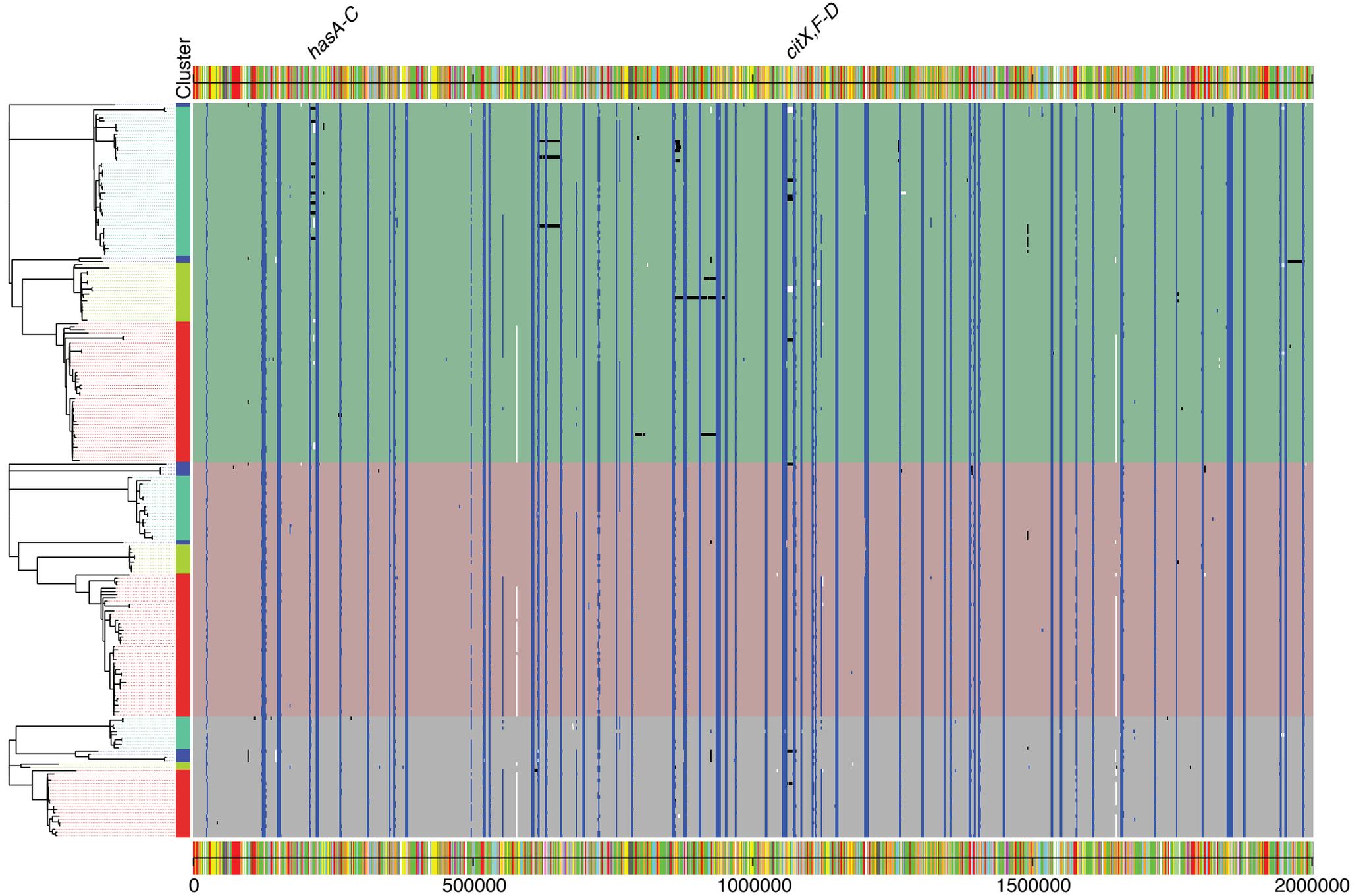
**Supplementary Figure 4.** Tanglegram showing concordance in BEAST (left) and ML (right) tree topologies, but not branch lengths. Branch lengths in the Bayesian phylogeny produced with BEAST represent time, while those in the ML tree represent genetic diversity. Dates are shown beneath the BEAST tree, and BAPs cluster and MLST type in columns adjacent to the tree.



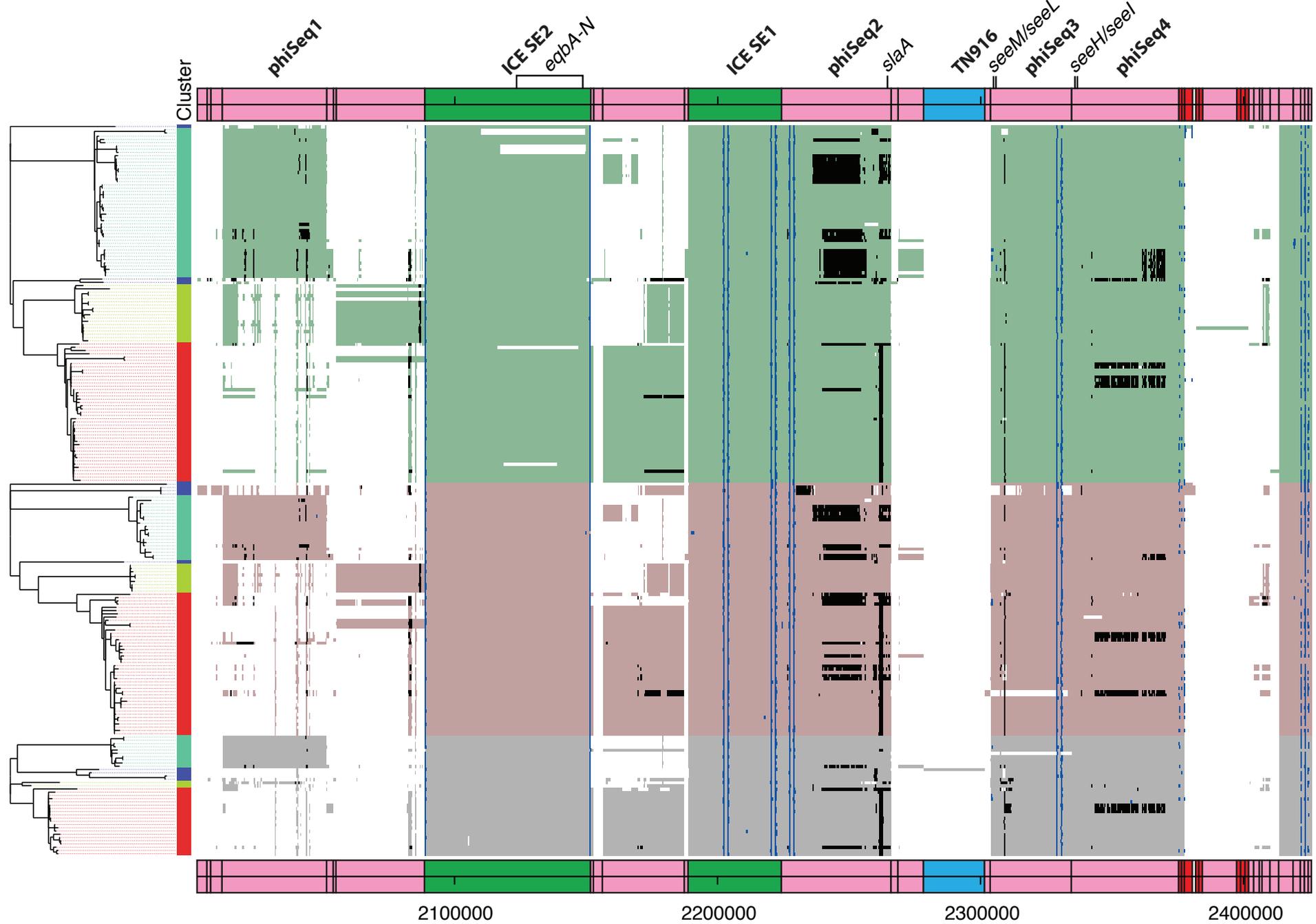
**Supplementary Figure 5.** Variation in substitution rates between acute and persistent isolates. A) Histogram showing the frequency of branches subtending acute and persistent isolates with different estimated mutation rates. B) Scatter plots of branch length vs mean estimated substitution rate for branches subtending acute, persistent and other (unknown) isolates. The unbroken and dashed red lines in each plot indicate the mean and 95% HPD estimates for the entire data from BEAST.



**Supplementary Figure 6.** Coverage of the accessory genome across the species. The left panel shows the ML tree, with BAPs cluster in a column to the right, as in Figure 1. The right-hand panel shows coverage of the accessory genome in each isolate. To the top and bottom of the right panel are representations of the assembled accessory contigs from isolates in the study. The contig color gives an indication of the content of the contig. Pink = bacteriophage, green = integrative and conjugative element (ICE), blue = transposon, red = IS element. Contigs present in the reference genome are labeled in bold above the panel, along with the location of some important virulence genes. For each isolate in the tree, regions are colored gray if they were present in single copy, and black if they were in multiple copy. i.e. duplications. Missing regions are in white.



**Supplementary Figure 7.** Core genome insertions, duplications and IS elements in persistent, acute and other isolates. Three ML phylogenetic trees are shown in the left panel, created from only persistent isolates (top), only acute isolates (middle) and other isolates (bottom). For each, the column to the right of the tree indicates the BAPs cluster of the isolates in the species phylogeny in Figure 1. In all cases, the topologies of the individual trees are consistent with the tree in Figure 1. The right-hand panel shows coverage of the core genome in each isolate. To the top and bottom are a representation of the annotation of the core genome, with the has and cit loci labeled. Regions of single copy coverage are colored green in persistent isolates, red in acute isolates and gray in others. Regions of duplication are colored black, and deletions are white. IS element insertion locations are shown in blue.



**Supplementary Figure 8.** Accessory genome insertions, duplications and IS elements in persistent, acute and other isolates. Three ML phylogenetic trees are shown in the left panel, created from only persistent isolates (top), only acute isolates (middle) and other isolates (bottom). For each, the column to the right of the tree indicates the BAPs cluster of the isolates in the species phylogeny in Figure 1. In all cases, the topologies of the individual trees are consistent with the tree in Figure 1. The right-hand panel shows coverage of the accessory genome in each isolate. To the top and bottom of the right panel are representations of the assembled accessory contigs from isolates in the study. The contig color gives an indication of the content of the contig. Pink = bacteriophage, green = integrative and conjugative element (ICE), blue = transposon, red = IS element. Contigs present in the reference genome are labeled in bold above the panel, along with the location of some important virulence genes. Regions of single copy coverage are colored green in persistent isolates, red in acute isolates and gray in others. Regions of duplication are colored black, and deletions are white. IS element insertion locations are shown in blue.