Supplementary Information

Evolution of extended-spectrum b-lactamase-producing ST131 *Escherichia coli* at a single hospital over 15 years

Shu-Ting Cho, Emma G. Mills, Marissa P. Griffith, Hayley R. Nordstrom, Christi L. McElheny, Lee H. Harrison, Yohei Doi, Daria Van Tyne

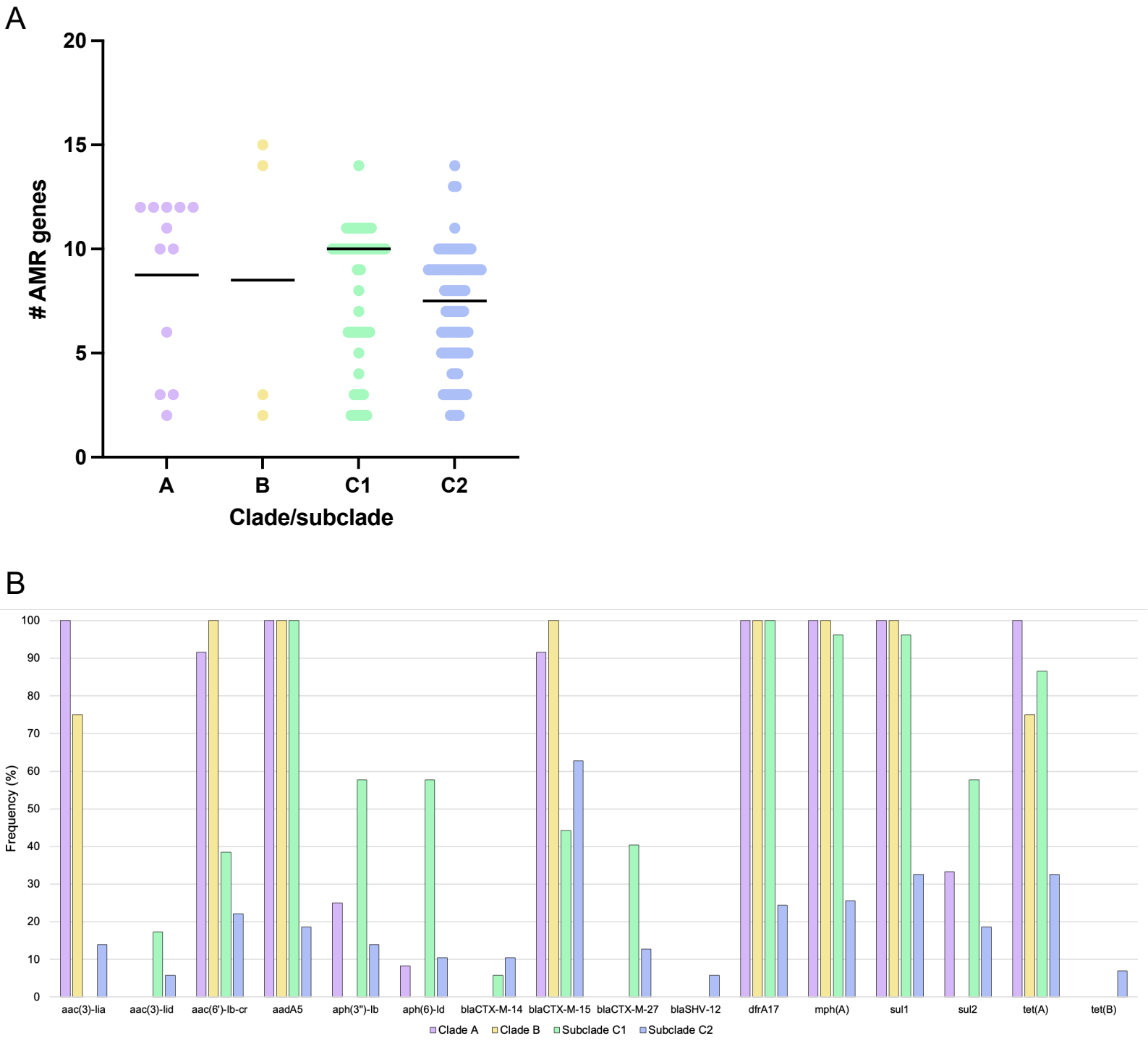

Figure S1. Differences in antimicrobial resistance (AMR) gene content between ST131 clades and subclades. (A) AMR gene abundance in isolates belonging to different ST131 clades and subclades. Each gene is counted once per genome. Horizontal lines show median values of AMR genes per genome in each isolate. AMR genes were identified by BLASTN to the ResFinder database. (B) Frequency of individual AMR genes among isolates in each clade or subclade. Genes with notable frequency differences between groups are shown. Complete data on AMR genes is provided in Table S2.

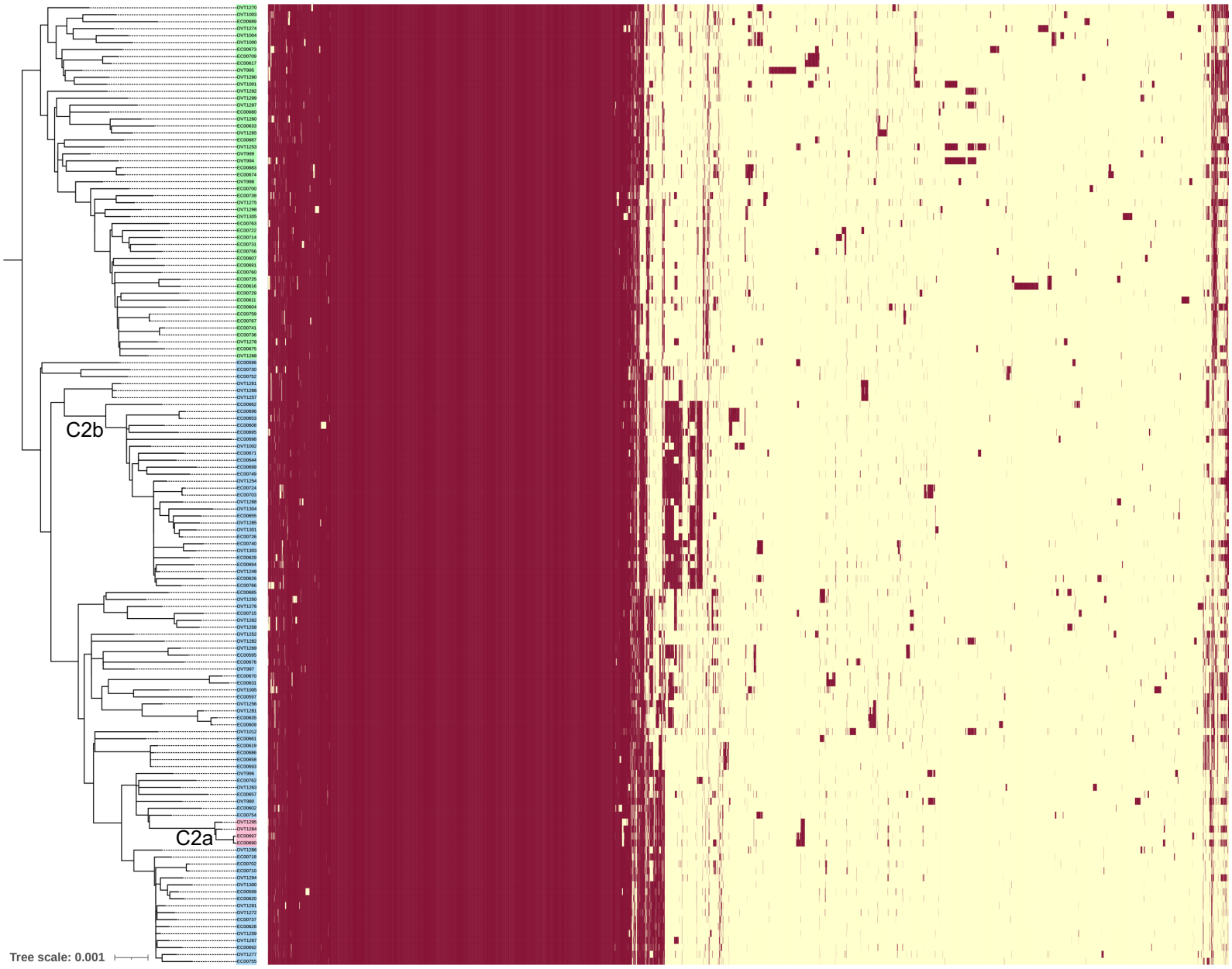

Figure S2. Pangenome analysis of 138 clade C ST131 *E. coli* isolates. Phylogenetic tree on the left was generated with RAxML using a core genome SNP alignment that was filtered to remove recombination with ClonalFrameML. The tree was midpoint rooted to separate subclades C1 (green shaded) and C2 (blue shaded). The heatmap on the right shows the pangenome matrix generated by Roary. Each column represents one gene group, and each row represent one genome. The presence or absence of a gene in a given genome is shown as red or yellow, respectively. Subgroups C2a and C2b are labeled below the corresponding branches on the phylogenetic tree.

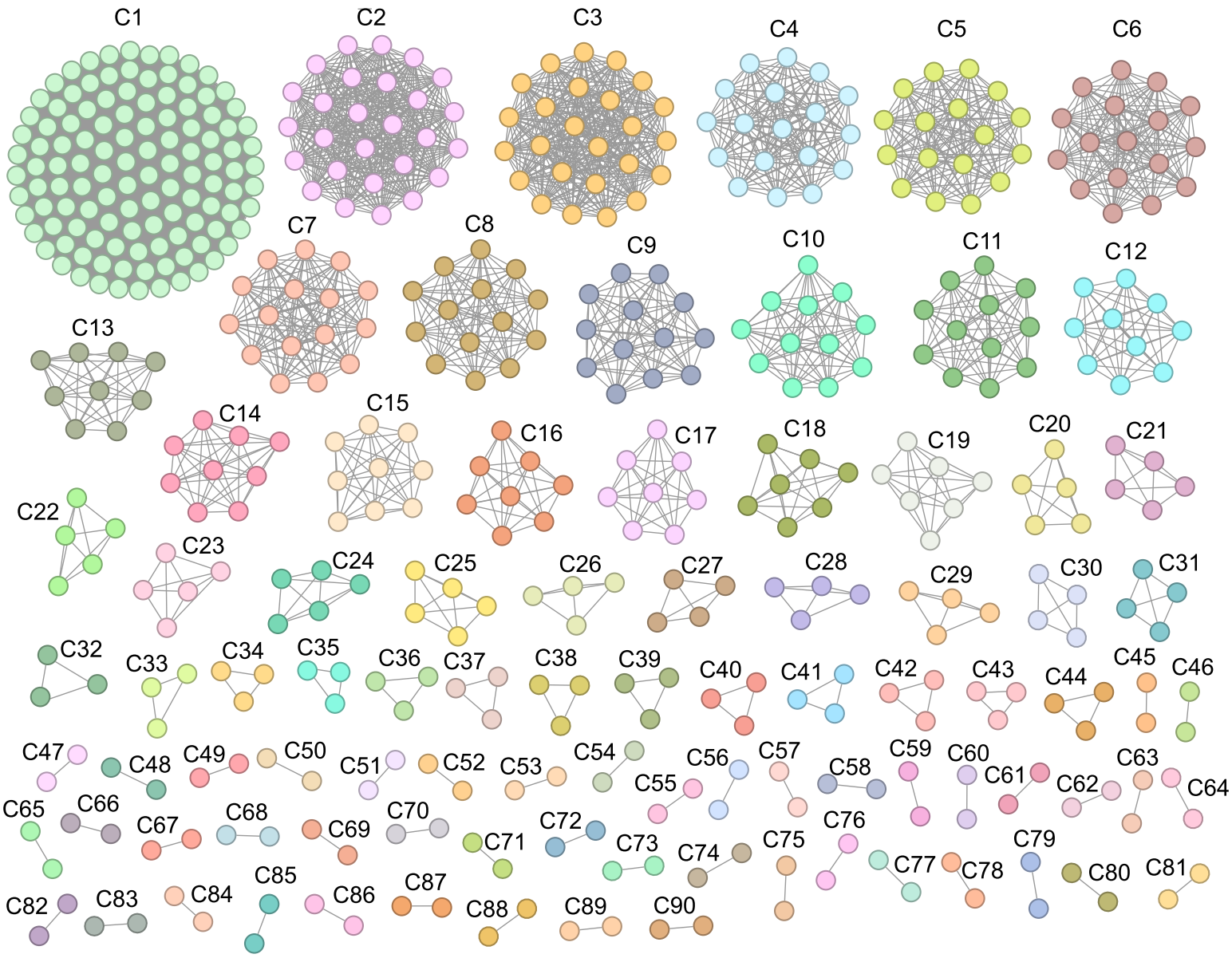

Figure S3. Prophage clusters of high genetic similarity among 154 ST131 *E. coli* isolates. Prophages were identified using PHASTEST. Sequences of prophages predicted to be intact or questionable were extracted and compared to one another via all-by-all BLASTN analysis. Prophages that had ≥90% nucleotide similarity and ≥90% sequence coverage were grouped together into clusters, with cluster number noted by C#.

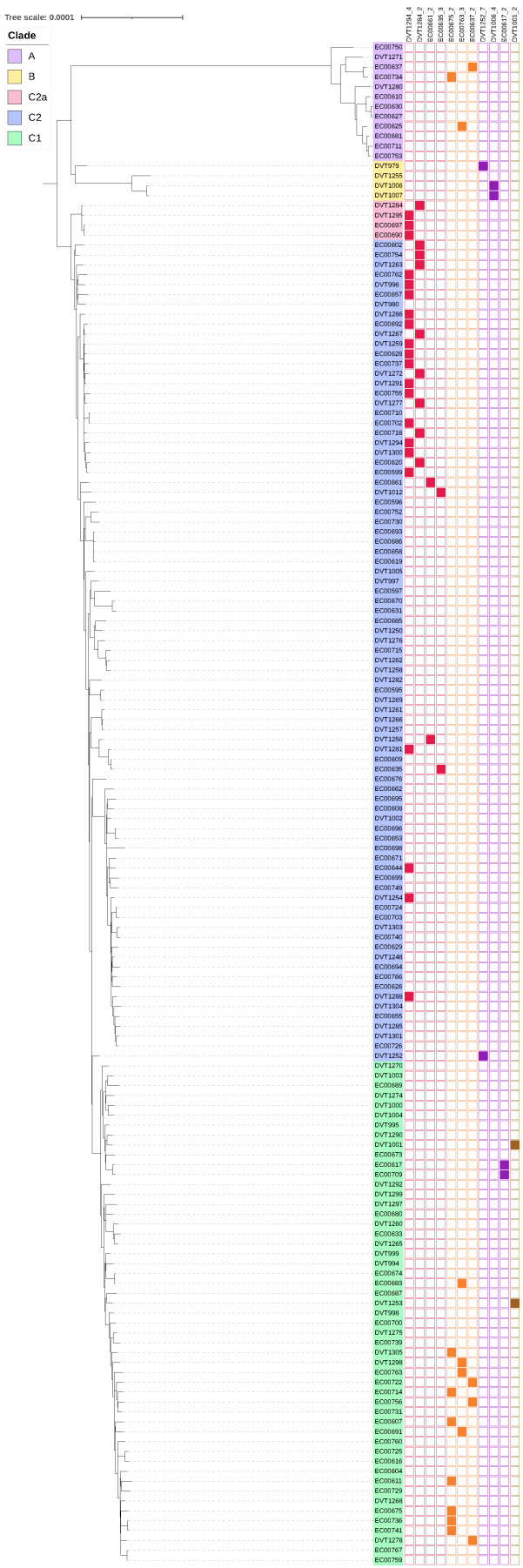

Figure S4. Distribution of ESBL-encoding plasmids among ST131 *E. coli* isolates. The core genome phylogeny is annotated with the presence of 11 ESBL-encoding reference plasmids detected in more than one genome in the dataset. Plasmids DVT1294_4 (Unknown Rep/Inc Type), DVT1284_2 (IncFIA, IncFIC, IncFII), EC00661_2 (IncFIA, IncFIC ,IncFII), and EC00635_3 (Unknown Rep/Inc Type) harbor *bla*$_{\text{CTX-M-15}}$ (red); plasmids EC00675_2 (IncFIA ,IncFIA, IncFIB), EC00763_3 (Unknown Rep/Inc Type), and EC00637_2 (IncFIA, IncFIA, IncFIB, rep_cluster_2131) harbor *bla*$_{\text{CTX-M-27}}$ (orange); plasmids DVT1252_7 (Unknown Rep/Inc Type), DVT1006_4 (Unknown Rep/Inc Type), and EC00617_2 (IncFIA,IncFIA, IncFIB, IncN ,rep_cluster_2131) harbor *bla*$_{\text{SHV-12}}$ (purple); and plasmid DVT1001_2 harbors *bla*$_{\text{CTX-M-2}}$ (brown).

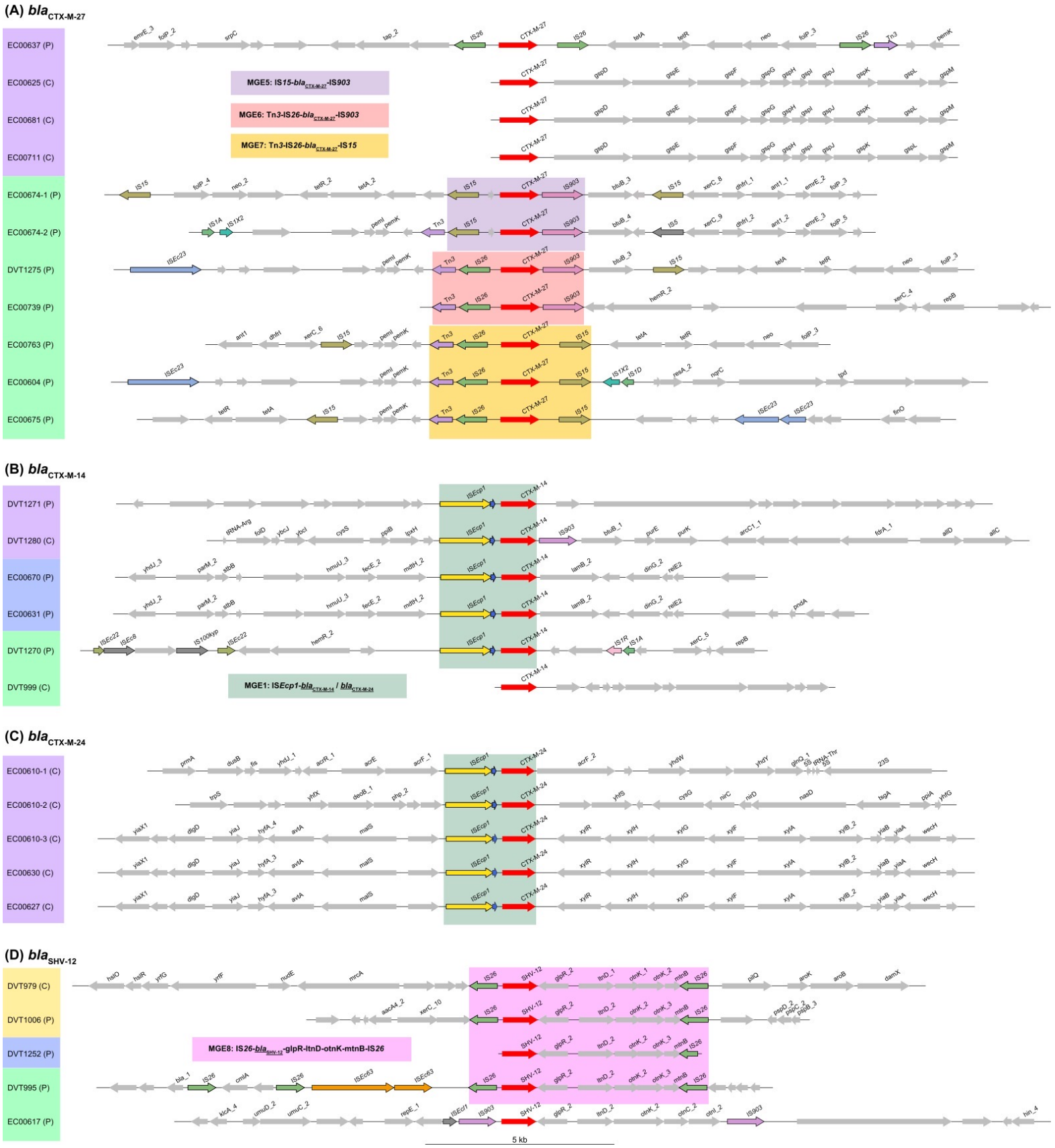

Figure S5. Regions flanking ESBL genes among ST131 *E. coli* isolates. (A-D) Genomic context of different ESBL-carrying MGEs is shown. Isolate names are shaded based on their phylogenetic clade assignments (clade A=purple; subclade C1=green; subclade C2=blue; clade B=yellow). The genomic context of each sequence is indicated (C=chromosome, P=plasmid) and ESBL genes are colored red. Genes were annotated with Prokka, and genes with predicted functions are labeled. Genes associated with MGEs and transposases are highlighted with black outlines and are colored if found in more than one region. Regions that were used for MGE classification are shaded in each panel.