

## Supplementary Material

**Table S1. The ST131 clade A-unique genomic regions among 390 ST131 genomes.**

Region no.	Positions in SE15 <sup>a</sup>	Coding DNA sequences <sup>b</sup>	Similar segments in non-ST131 <i>E. coli</i> <sup>c</sup>
1	3657378–3659258	Hypothetical proteins	Yes
2	4051873–4054372	<i>yihW</i> , <i>yihV</i>	Yes
3	4069433–4076182	Hypothetical proteins	Yes
4 <sup>d</sup>	4371327–4373197	Hypothetical proteins	No

These segments were present in all of the 46 clade A genomes but were absent in the 344 non-clade A genomes.

<sup>a</sup> Clade A reference genome (Genbank accession no. GCA\_000010485.1).

<sup>b</sup> Genes were annotated using the RAST server (<http://rast.nmpdr.org/>).

<sup>c</sup> NCBI BLAST search (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) against nucleotide collection (nr/nt) was performed to identify similar regions in non-ST131 *E. coli* genomes.

<sup>d</sup> Primers of CladeAspe4-YF5 and CladeAspe4-YR5 (Table 1) for this region were designed to target nucleotide positions 4372231–4372937 of the SE15 genome.

**Table S2. The ST131 clade or subclade-unique SNPs among 390 ST131 genomes.**

Clade or subclade	or	SNP in target group	SNP in all others	SNP position in EC958	Gene	Locus tag	Nucleotide change	Amino acid change	Product
B (n=77) vs. non-B (n=313)	vs.	T	C	5078286	<i>prfC</i>	EC958_RS 25415	T564C	-	Peptide chain release factor
		G	T	5079073			G1351T	A451S	
		G <sup>a</sup>	A	5079108			G1386A	-	
		G <sup>a</sup>	A	5079114			G1392A	-	
C (n=267) vs. non-C (n=123)	vs.	T	A	3549826	<i>yqjA</i>	EC958_RS 17820	T551A	F184Y	Hypothetical protein
		A	G	3587935	<i>agaD</i>	EC958_RS 18040	A142G	N48D	PTS N-acetylgalactosamine transporter subunit IID
		A	C	4857366	<i>ybjT</i>	EC958_RS 24265	T366G	D122E	NAD(P)-dependent oxidoreductase
		C	T	4884400	<i>nrdD</i>	EC958_RS 24390	G752A	S251A	Anaerobic ribonucleoside triphosphate reductase
		G	A	4884915			C237T	-	
		T	G	4885585	<i>treC</i>	EC958_RS 24395	A1616C	N539T	Trehalose-6-phosphate hydrolase
		A	G	4889036	<i>treR</i>	EC958_RS 24405	T702C	-	Trehalose operon repressor
		C	T	4889063			G675A	-	
		C	T	4891435	<i>mgtA</i>	EC958_RS 24410	C1320T	-	Magnesium-translocating P-type ATPase
		G	A	4891756			G1641A	-	
		T <sup>a</sup>	C	4892699			T2584C	-	
		A	G	4892701			A2586G	-	
		A <sup>a</sup>	G	4892761			A2646G	-	
		T	C	4904571			<i>valS</i>	EC958_RS 24485	
		T	C	4904685	A1326G	-			
		A	C or T	4904703	T1308G/A	-			
C1 (n=114) vs. non-C1 (n=276)	vs.	T <sup>a</sup>	C	577077		EC958_RS 02710	T619C	-	Hypothetical protein

C1-M27 (n=29) vs. non-C1-M-27 (n=361)	A	C	74857	Non- CDS				
	A	G	249045	<i>dkgB</i>	EC958_RS 01160	G148A	A50T	2,5-diketo-D-gluconic acid reductase
	G	A	2538155	Non- CDS				
	T	C	2653122	Non- CDS				
	A <sup>a</sup>	T	3519471	<i>aer</i>	EC958_RS 17700	A1384T	C462S	Aerotaxis receptor
C2 (n=146) vs. non-C2 (n=244)	T <sup>a</sup>	C	2940033	<i>nrdI</i>	EC958_RS 14885	C121T	P41S	Ribonucleotide reductase stimulatory protein

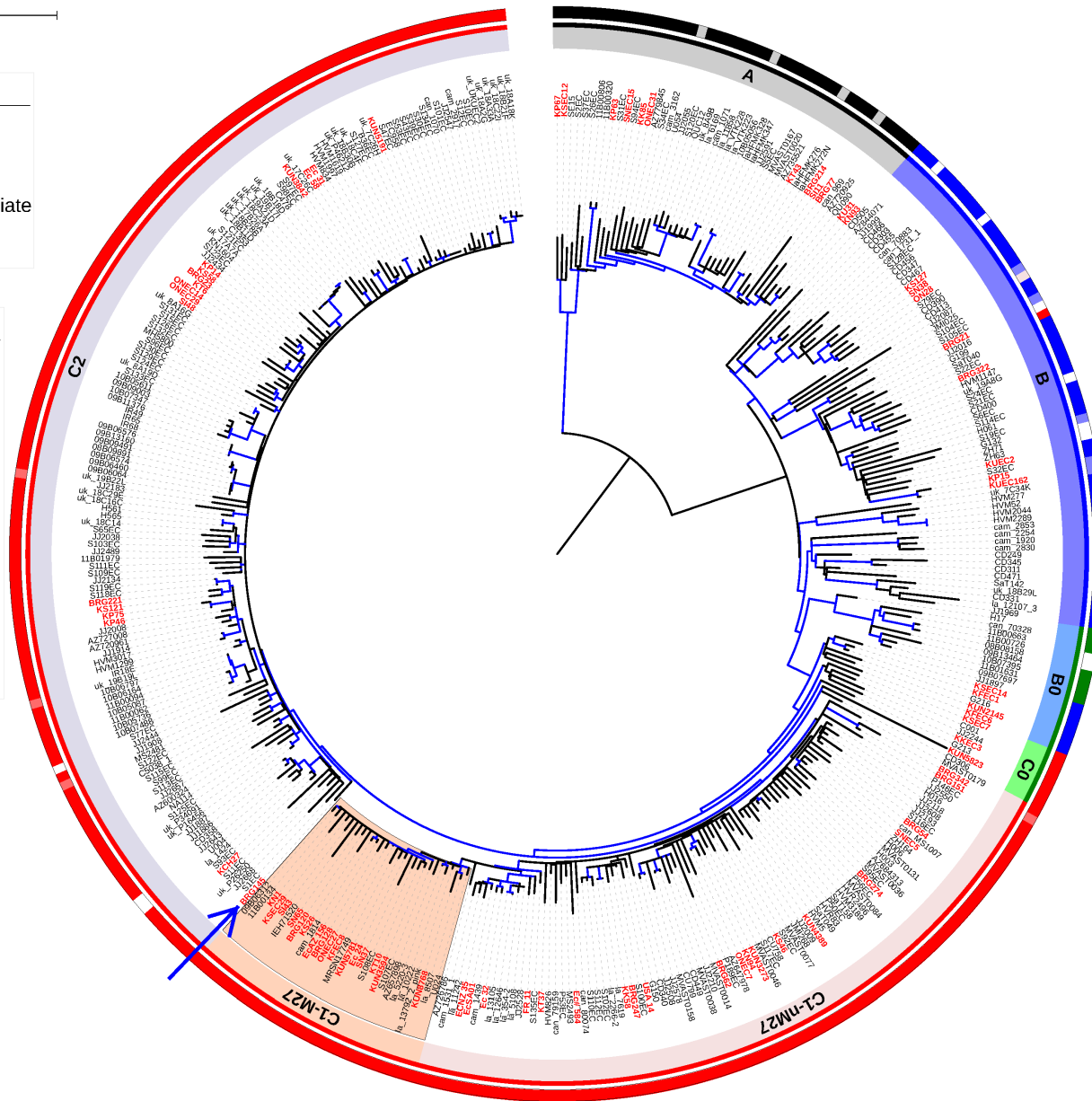
CDS, coding DNA sequence.

<sup>a</sup> SNP used in the PCR assay.

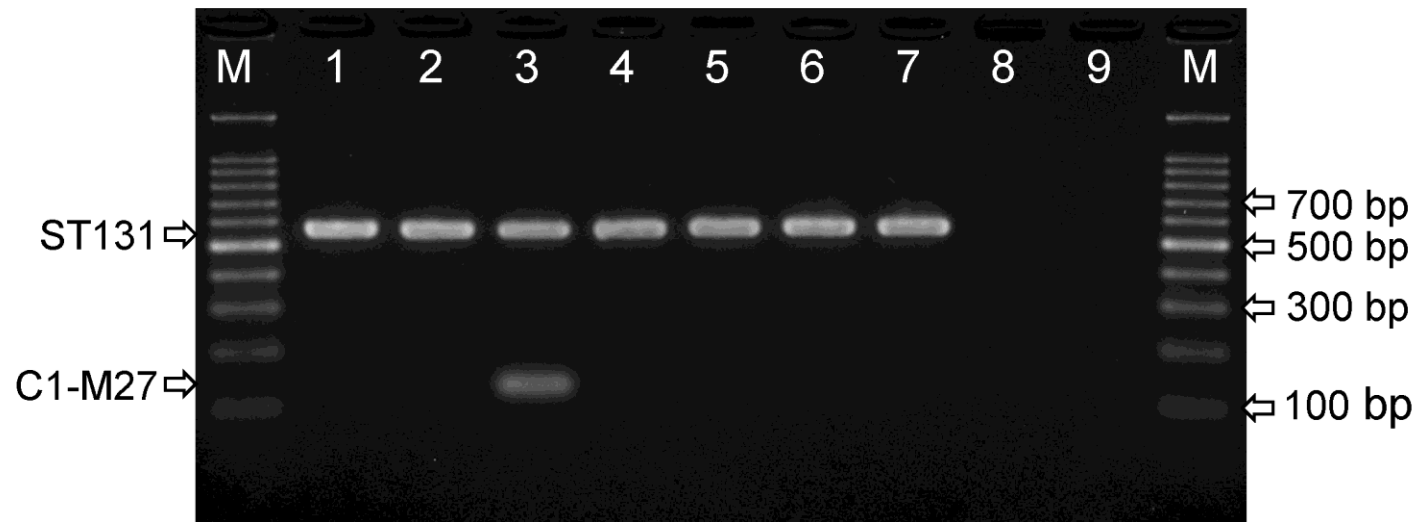
Tree scale: 0.01

Clade	
■	A
■	B
■	Intermediate
■	C

fimH	
●	41
○	41 group
●	22
○	22 group
●	30
○	30 group
●	27
○	31
○	Others



**Figure S1. Recombination-free core SNP-based phylogeny of 390 ST131 genomes.** This maximum-likelihood phylogenetic tree is rooted by using the clade A isolates. A total of 6,667 core SNP sites were used after excluding the 18,725 SNP sites that were located within the 1,953,372 bp recombination region. Branches that had bootstrap support > 90% from 100 replicates are highlighted in blue.



**Figure S2. Agarose gel electrophoresis of the C1-M27 subclade PCR amplicons.** M, 100-bp DNA ladder; 1, strain SNEC15, clade A; 2, strain KFEC6, clade B; 3, strain KFEC8, subclade C1-M27; 4, strain SNEC5, subclade C1-nM27 (C1-non-M27); 5, strain ONEC14, clade C2; 6, strain KSEC7, subclade C0; 7, strain BRG210, subclade I1; 8, strain BRG28, non-ST131; 9, no template control.