

SUPPLEMENTARY INFORMATION

***In Silico* Signature Prediction Modeling in Cytolethal Distending
Toxin-Producing *Escherichia coli* Strains**

Maryam Javadi, Mana Oloomi*, Saeid Bouzari

Department of Molecular Biology, Pasteur Institute of Iran, Tehran 13164, Iran

Supplementary Table 2. Signature details based on *Escherichia coli* IHE3034 reference

Region	Biomarker range in <i>Escherichia coli</i> IHE3034 genome between	Protein obtained by NCBI Blastx [identical protein in other <i>Escherichia coli</i> strains]	Blast score	Blast Identity (%)	Blast E-value	Putative conserved domains non-specific hits (NH), specific hits (SH), multi domains (MD)
A	0 Mbp to 1 Mbp	Tail fiber protein Phage tail tape measure protein	559 135	100 100	0.0 4e-39	
B	2 Mbp to 3 Mbp	Cytolethal distending toxin, subunit C Cytolethal distending toxin, subunit B Cytolethal distending toxin, subunit A Exonuclease family protein Hypothetical protein ECOK1_2135, [conserved domain protein] Hypothetical protein ECOK1_2134 Hypothetical protein ECOK1_2122, [membrane protein] Putative transcriptional regulator DicA157 Antitermination protein	382 520 483 350 120 147 164 370 98.6	100 100 99 100 100 100 100 100 100	1e-127 1e-179 5e-166 1e-116 4e-36 3e-47 2e-52 3e-106 2e-27	NH: CDtoxinA, RICIN superfamily NH: PRK15251, CdtB, EEP superfamily NH: CDtoxinA, RICIN superfamily NH: dexA, DnaQ-like-exo superfamily
C	2 Mbp to 3 Mbp	Integrase/recombinase, phage integrase family Putative regulatory protein Cox Hypothetical protein ECOK1_2557, [membrane protein] hypothetical protein ECOK1_2558 Hypothetical protein ECOK1_2602, [type VI secretion protein] Hypothetical protein ECOK1_2601, [double zinc ribbon family protein] Hypothetical protein ECOK1_2600 Holin, lambda family Hypothetical protein ECOK1_2581	693 196 131 83.2 585 194 675 58.9 375	100 100 100 98 100 100 100 100 100	0.0 1e-59 1e-36 9e-20 0.0 4e-60 0.0 3e-12 1e-133	NH: HP1_INT_C, Phage_integrase, DNA_BRE_C superfamily MD: int, XerD, recomb_XerD, xerD NH: Phage_Cox, Phage_Cox superfamily NH: VI_minor_1, DUF 3121 superfamily NH: Glyco_hydro_108, Glyco_hydro_108 superfamily NH: PG_binding_3, PG_binding_3 superfamily MD: zliS
D	2 Mbp to 3 Mbp	Hypothetical protein ECOK1_2812 Hypothetical protein ECOK1_2814 Hypothetical protein ECOK1_2815 Hypothetical protein ECOK1_2816 Hypothetical protein ECOK1_2809 Enterohemolysin 1	519 2,219 1,837 103 222 317	100 100 99 100 100 100	8e-169 0.0 0.0 6e-25 2e-72 2e-110	NH: V, V superfamily MD: PRK03918, COG1340, SMC_N, SMC_prok_B
E	4 Mbp to 5 Mbp	Site-specific recombinase, phage integrase family Protein cII Hypothetical protein ECOK1_4790, [beta family protein, Enterobacteria phage P4]	223 531 700	100 100 100	6e-65 0.0 0.0	NH: INT_P4, DNA_BRE_C superfamily, MD: PRK09692 NH: Beta_protein, Beta_protein superfamily
F	5 Mbp to end of the genome	Hypothetical protein ECOK1_4914	133	100	9e-42	