

## **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 4.** Signature details based on *Escherichia coli* STEC\_B2F1 reference

Region	Biomarker range in <i>Escherichia coli</i> STEC_B2F1 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	Hypothetical protein ECSTECB2F1_0150	80.1	100	1e-21	
		Hypothetical protein ECSTECB2F1_0149, [transposase]	82.4	100	5e-22	
B	0 Mbp to 1 Mbp	Protein 40A	582	100	0.0	
C	0 Mbp to 1 Mbp	Collagen triple helix repeat family protein	487	100	3e-166	SH: collagen, collagen superfamily
		Tail fiber assembly	144	100	6e-42	
		Hypothetical protein ECSTECB2F1_0901, [tail fiber assembly protein, caudovirales tail fiber assembly protein]	251	100	1e-79	
		Outer membrane protein lom	128	100	1e-39	NH: Ail_Lom, PRK15240, COG3637, OMP_b-brl superfamily
		Phage integrase family protein	56.2	100	1e-11	NH: phage_tail_N, phage_tail_N superfamily
Prophage tail fiber family protein	74.3	96	3e-19			
D	1 Mbp to 2 Mbp	Putative endopeptidase	63.9	100	2e-12	NH: DUF 262, DUF 262 superfamily
		Hypothetical protein ECSTECB2F1_1098	78.2	100	3e-18	
		Hypothetical protein ECSTECB2F1_1099	86.7	98	3e-21	
		Hypothetical protein ECSTECB2F1_1078, [PF03235 family protein]	726	100	0.0	
		Caudovirales tail fiber assembly family protein	125	100	5e-37	
		Prophage lambda integrase	275	100	2e-93	
		Prophage lambda integrase	139	100	6e-42	
Prophage lambda integrase	110	100	2e-31	NH: INT_Lambda_C, DNA_BRE_C superfamily NH: Phage_integ_N, Phage_integ_N superfamily		
E	1 Mbp to 2 Mbp	Transcriptional regulator, AraC family	497	100	3e-175	MD: PRK09940, COG4753
		Collagen triple helix repeat family protein	228	100	2e-70	
		Hypothetical protein ECSTECB2F1_1516, [tail fiber assembly protein, caudovirales tail fiber assembly protein]	221	100	4e-70	
		Hypothetical protein ECSTECB2F1_1255	89.4	100	1e-23	
		Hypothetical protein ECSTECB2F1_1256	188	100	5e-62	
		Host specificity protein J	308	100	1e-99	
		Antitermination protein Q	28.5	100	0.13	
		Hypothetical protein ECSTECB2F1_1251	207	100	6e-69	
		Hypothetical protein ECSTECB2F1_1263	197	100	2e-61	
		Hypothetical protein ECSTECB2F1_1515	60.1	100	9e-14	
		Helix-turn-helix family protein	130	100	6e-41	
		Outer membrane protein lom	127	100	6e-40	
		Prophage tail fiber family protein	82.8	96	2e-22	
		Collagen triple helix repeat family protein	85.9	100	4e-22	
F	1 Mbp to 2 Mbp	Phage integrase family protein	212	100	1e-69	NH: INT_Lambda_C, DNA_BRE_C superfamily
		Hypothetical protein ECSTECB2F1_1296	91.3	100	7e-26	
		Hypothetical protein ECSTECB2F1_1297	56.2	100	6e-12	
		Phage integrase family protein	141	100	5e-43	
G	1 Mbp to 2 Mbp	Hypothetical protein ECSTECB2F1_1685	902	100	0.0	
H	1 Mbp to 2 Mbp	BRO family, N-terminal domain protein	537	100	0.0	SH: Bro-N, NH: Bro-N, Bro-N superfamily, MD: COG3617 SH: HTH_36, HTH_36 superfamily, MD: PaaX_trns_reg SH: Phage_fiber_2, Phage_fiber_2 superfamily
		paaX-like family protein	271	100	3e-92	
		Phage tail fiber repeat family protein	355	100	2e-120	
		DNA-binding protein Roi	245	100	2e-83	
		DNA-damage-inducible protein I	59.7	100	6e-14	
		Hypothetical protein ECSTECB2F1_2003	69.7	100	4e-18	

**Supplementary Table 4.** Signature details based on *Escherichia coli* STEC\_B2F1 reference

		outer membrane protein lom	136	100	2e-42	NH: Ail_Lom, PRK15240, COG3637, OMP_b-brl superfamily
I	2 Mbp to 3 Mbp	Flagellin	229	100	4e-74	
J	2 Mbp to 3 Mbp	Putative membrane protein	632	100	0.0	Glyco_transf_GTA type superfamily
		Glycosyl transferase family 2 family protein	572	100	0.0	NH: Glycos_transf_2, CESA_like, Glyco_transf_GTA type superfamily, MD: PRK10073
		Hypothetical protein ECSTECB2F1_2214	328	99	8e-103	
		Polysaccharide biosynthesis family protein	764	100	0.0	SH: MATE_like_10, NH: PRK15099, MATE_like superfamily, MD: RfbX
		Erythromycin biosynthesis sensory transduction protein eryC1	753	100	0.0	NH: AHBA_syn, DegT_DnrJ_EryC1, WecE, PRK11658, NHT_00031, AAT_I superfamily, MD: PRK15407, c
		wbtB	164	99	4e-47	
		Putative teichuronic acid biosynthesis glycosyltransferase tuaG	525	100	2e-170	SH: Glycos_transf_2, GT_2_like_d, PRK10018, PLN02726, Glyco_transf_GTA type superfamily, MD: Glyco_tranf_2_3, WcaA, PRK10073, PTZ00260
		Glucose-1-phosphate thymidyltransferase	556	100	0.0	NH: G1P_TT_short, NTP_transferase, GalU, galU, PRK10122, Glyco_transf_GTA type superfamily, MD: rmlA, PRK15480, RfbA
K	2 Mbp to 3 Mbp	Hypothetical protein ECSTECB2F1_2378	99.8	98	1e-25	
		Hypothetical protein ECSTECB2F1_2379	94.7	100	6e-24	
		clp protease family protein	1,288	100	0.0	SH: S14_ClpP_1, NH: ClpP, CLP_protease, clpP, Clp_protease_like superfamily, MD: SDH_sah, SppA
		Bacteriophage P4 DNA primease	203	100	8e-65	
		Hypothetical protein ECSTECB2F1_2371	132	100	3e-40	
L	2 Mbp to 3 Mbp	Hemagglutination activity domain protein	1,036	99	0.0	SH: Fil_haemagg_2, fil_hemag_20aa, Fil_haemagg_2 superfamily, MD: FhaB, PRK15319, Hia, PRK12688
		Hypothetical protein ECSTECB2F1_3213	105	99	3e-26	
		Sulfatase family protein	769	99	0.0	NH: Sulfatase, Sulfatase superfamily, MD: AsIA, PRK13759, chol_sulfatase
		Hypothetical protein ECSTECB2F1_3190, [arylsulfatase]	261	100	3e-81	NH: Sulfatase, Sulfatase superfamily, MD: AsIA
		Outer membrane porin protein LC	611	100	0.0	NH: PRK10554, Porin_1, OmpC, gram_neg_porins, OM_channels superfamily
		RTX toxin acyltransferase family protein	361	100	1e-118	NH: Haemagg_act, SH: Haemagg_act, Haemagg_act superfamily
		Hemolysin secretion/activation protein ShIB/FhaC/HecB family pr	746	100	0.0	NH: HlyC, SH: HlyC, HlyC superfamily, MD: FhaC, ShIB
		Hypothetical protein ECSTECB2F1_3193	130	100	9e-37	NH: PRK09750, DUF1187, DUF1187 superfamily
		Transcriptional regulator, AraC family	377	100	9e-128	MD: HTH_ARAC, PRK09940, AraC
		Serine protease eatA	1,121	100	0.0	MD: Peptidase_S6, AidA
		Hypothetical protein ECSTECB2F1_3192, [membrane protein]	60.5	100	9e-13	
		Hypothetical protein ECSTECB2F1_3199	89.4	100	5e-24	
		Hypothetical protein ECSTECB2F1_3200	133	100	1e-40	
		Neurotensin receptor R8	133	100	2e-40	
		tonB-dependent vitamin B12 receptor	268	100	9e-88	NH: ligand_gated_channel, Plug, Plug superfamily, OM_channels superfamily, MD: BtuB, TonB-B12
		Acetyl-CoA acetyltransferase	44.7	100	6e-07	NH: Thiolase_C, thiolase, Thiolase_C superfamily, cond-enzymes superfamily, MD: PRK05790, PaaJ, AcCoA-C-

**Supplementary Table 4.** Signature details based on *Escherichia coli* STEC\_B2F1 reference

		Hypothetical protein ECSTECB2F1_3178 ompA-like transmembrane domain protein	76.6 135	97 100	9e-20 2e-41	Actrans NH: COG3637, OMP_b-brl, OMP_b-brl superfamily
M	2 Mbp to 3 Mbp	Hypothetical protein ECSTECB2F1_3480	453	100	6e-161	
N	3 Mbp to 4 Mbp	Replication gene A protein Cytolethal distending toxin C Endonuclease/Exonuclease/phosphatase family protein Cytolethal distending toxin A/C family protein Caudovirales tail fiber assembly family protein Hypothetical protein ECSTECB2F1_4300 Phage tail fiber repeat family protein	69.7 158 545 486 46.6 696 257	100 100 100 100 100 100 100	6e-13 6e-46 0.0 9e-166 6e-07 0.0 5e-83	NH: CDtoxinA, RICIN superfamily NH: PRK15251, CdtB, EEP superfamily NH: CDtoxinA, RICIN superfamily
O	4 Mbp to end of the genome	Hypothetical protein ECSTECB2F1_4680 Putative membrane protein Hypothetical protein ECSTECB2F1_4682 Type I restriction modification DNA specificity domain protein Type I site-specific deoxyribonuclease, HsdR family	564 267 70.9 279 142	100 100 100 100 100	0.0 1e-85 1e-15 5e-96 2e-41	NH: ResIII, DEXDc superfamily, MD: hsdR, COG0610
P	4 Mbp to end of the genome	Host specificity protein J Hypothetical protein ECSTECB2F1_1326, [hok/gef family protein] Prophage tail fiber family protein Major tail protein V Hypothetical protein ECSTECB2F1_1281 Hemin receptor domain protein Minor tail protein M Phage tail tape measure protein, lambda family Phage terminase large subunit family protein	374 63.2 89.7 337 130 40.0 144 67.0 140	100 100 100 100 100 99 100 100	7e-119 2e-13 3e-22 6e-118 9e-41 9e-06 4e-46 1e-14 6e-42	SH: DUF3672, DUF3672 superfamily NH: phage_tail_N, phage_tail_N superfamily SH: BID_2, Big_2 superfamily, MD: COG5492 NH: COG4718, Phage_min_tail, Phage_min_tail superfamily NH: COG5525, Terminase_GpA superfamily