

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 5. Signature details based on *Escherichia coli* STEC_C165_02 reference

Region	Biomarker range in <i>Escherichia coli</i> STEC_C165_02 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	Restriction endonuclease family protein	723	100	0.0	SH: COG4127, HsdM_N, NH: Mrr_cat, UPF0020, Mrr_cat superfamily, HsdM_N superfamily, MD: Mrr, HsdM, N6_Mtase
		TIR protein	411	100	2e-131	
		N-6 DNA Methylase family protein	974	100	0.0	SH: Methyltransf_26, AdoMet_Mtase superfamily, HsdM_N, HsdM_N superfamily, MD: N6_Mtase, hsdM
		Type I site-specific deoxyribonuclease, HsdR family protein	1,409	100	0.0	SH: HsdM_N, NH: UPF0020, HsdM_N superfamily, MD: HsdM, N6_Mtase, COG0610, hsdR, DEXDc
		Hypothetical protein ECSTECC16502_0289, [ABC transporter ATP-binding protein]	1,107	100	0.0	
		Hypothetical protein ECSTECC16502_0290	106	98	4e-27	
		Hypothetical protein ECSTECC16502_0291	74.3	100	4e-16	
		Type I restriction-modification system specificity determinant	1,176	100	0.0	NH: Methylase_S, Methylase_S superfamily, MD: N6_Mtase, HsdM, HsdS, PRK09737
		Putative membrane protein	364	100	6e-115	
		Type III restriction enzyme, res subunit	2,261	100	0.0	SH: HSDR_N, MD: COG0610, hsdR, hsdR
		Hypothetical protein ECSTECC16502_0339	1,450	100	0.0	SH: COG1479, DUF1524, DUF1524 superfamily, NH: DUF262, DUF262 superfamily, COG3586, COG3586 superfamily
		Metallo-beta-lactamase superfamily protein	569	100	0.0	SH: ElaC, NH: RNase_Z, PRK00055, Lactamase_B_2, Lactamase_B superfamily
		4-Hydroxyphenylacetate catabolism regulatory protein HpaA	618	100	0.0	MD: HpaA, AraC, HTH_ARAC, HTH_18, PRK09685
		4-Hydroxyphenylacetate permease	850	100	0.0	MFS superfamily, MD: HpaX, MFS_1, PRK11551, NarK
		Hypothetical protein ECSTECC16502_0280, [HNH endonuclease]	516	99	0.0	SH: HNH_2, HNHc superfamily, MD: COG3440
		Filamentation induced by cAMP protein Fic	138	100	2e-42	
		B	1 Mbp to 2 Mbp	Phage virion morphogenesis protein	70.5	100
Hypothetical protein ECSTECC16502_1311	133			98	1e-37	
AAA ATPase	795			100	0.0	NH: GP4d_helicase, RecA_like superfamily, MD: AAA_15, COG3950
Putative membrane protein	852			100	0.0	
Hypothetical protein ECSTECC16502_4950, [Rz1 lytic protein]	62.0			100	1e-12	NH: PRK14512, S14_ClpP_1, ClpP, clpP, ClpP_protease_like superfamily
Hypothetical protein ECSTECC16502_4968	442			100	4e-159	
Acetyltransferase family protein	394			100	4e-141	
Hypothetical protein ECSTECC16502_4986	51.2			100	2e-10	
Hypothetical protein ECSTECC16502_4996	466			100	2e-167	NH: DUF2829, DUF2829 superfamily
isaA	43.9			95	6e-06	
Hypothetical protein ECSTECC16502_1295, [acetyltransferase]	120			100	2e-35	
Integrase	208			100	3e-68	MD: int
Tail fiber	176			100	4e-55	
DNA-invertase	92.8			100	2e-23	NH: SR_ResInv, Resolvase, Ser_Recombinase superfamily, MD: PinR
Caudovirales tail fiber assembly family protein	338			100	6e-118	SH: Caudo_TAP, Caudo_TAP superfamily
Tail fiber domain protein	201			100	4e-65	

Supplementary Table 5. Signature details based on *Escherichia coli* STEC_C165_02 reference

		Recombination enhancement	298	99	3e-104	
		Phage Tail Collar domain protein	134	95	2e-41	NH: Collar, Collar superfamily
		DNA-invertase	28.9	100	0.061	
		Hypothetical protein ECSTECC16502_4984, [lysogeny establishment protein]	120	100	6e-37	
		Phage tail fiber repeat family protein	138	100	1e-41	SH: Phage_fiber_2, Phage_fiber_2 superfamily, MD: COG5301
		Hypothetical protein ECSTECC16502_1393	103	100	3e-31	
C	1 Mbp to 2 Mbp	Hypothetical protein ECSTECC16502_1803	264	100	4e-87	
		Hypothetical protein ECSTECC16502_1804	122	100	3e-34	
		Hypothetical protein ECSTECC16502_1805	213	100	1e-72	
		Hypothetical protein ECSTECC16502_1809	156	100	6e-49	
		Resolvase, N terminal domain protein	138	100	7e-41	
		Hypothetical protein ECSTECC16502_1807	126	100	5e-40	
D	2 Mbp to 3 Mbp	Hypothetical protein ECSTECC16502_2561	194	100	3e-62	
		Phage holin, lambda family	68.9	100	8e-16	
		Peptidoglycan domain protein	374	100	3e-133	NH: Glyco_hydro_108, Glyco_hydro_108 superfamily, NH: PG_binding_3, PG_binding_3 superfamily, MD: zliS
		Hypothetical protein ECSTECC16502_2576	188	100	5e-62	
		Hypothetical protein ECSTECC16502_2577	74.7	100	3e-18	
		gp41 domain protein	312	99	2e-110	
		Hypothetical protein ECSTECC16502_2586	32.0	100	0.006	
		Hypothetical protein ECSTECC16502_2559	75.9	97	1e-19	
		Hypothetical protein ECSTECC16502_2587, [DNA-binding protein]	168	100	1e-55	NH: PHA00675, PHA00675 superfamily
		Hypothetical protein ECSTECC16502_2588	133	100	8e-41	NH: DUF1627, DUF1627 superfamily
E	2 Mbp to 3 Mbp	Prophage CP4-57 integrase	108	100	1e-27	NH: INT_P4,DNA_BRE_C superfamily, MD: PRK09692
		Hypothetical protein ECSTECC16502_2827	168	100	6e-53	
		Hypothetical protein ECSTECC16502_2828	99.4	100	5e-27	
		Prophage CP4-57 regulatory protein family protein	136	99	1e-40	NH: Phage_Alpa, Alpa, Phage_Alpa superfamily
		Terminase small subunit	374	100	2e-132	SH: COG4220, NH: Phage_Nu1, Phage_Nu1 superfamily
		Hypothetical protein ECSTECC16502_2842	101	100	5e-28	
		Major head protein	288	100	8e-97	NH: Phage_cap_E, Phage_cap_E superfamily
		Prophage CP4-57 integrase	237	100	6e-77	NH: INT_P4,DNA_BRE_C superfamily, SH: DUF4102, DUF4102 superfamily, MD: PRK09692
		Hypothetical protein ECSTECC16502_2843	128	100	2e-40	
F	3 Mbp to 4 Mbp	DNA topoisomerase IV, A subunit	177	100	6e-54	NH: TOP4c, TOP4c superfamily, MD: parC_Gneg, PRK05561, GyrA, TOP4c, DNA_topoisolV, 52, PLN03128
G	4 Mbp to 5 Mbp	Cytolethal distending toxin A/C family protein	456	100	3e-154	NH: CDtoxinA, RICIN superfamily
		Endonuclease/Exonuclease/phosphatase family protein	546	100	0.0	NH: PRK15251, CdtB, EEP superfamily
		Hypothetical protein ECSTECC16502_4757	189	99	3e-62	
H	4.5 Mbp to end of the genome	upf89.5	194	100	4e-59	
		Putative lipoprotein	185	100	2e-57	
		humD	122	100	3e-34	NH: PRK10276, peptidase_S24_S26 superfamily, MD: LexA
		Hot protein	87.0	98	1e-21	SH: DNA_pol3_theta, NH: PRK10969, DNA_pol3_theta superfamily
		Hypothetical protein ECSTECC16502_1327	96.7	98	3e-25	
		Terminase B protein domain protein	565	100	0.0	

Supplementary Table 5. Signature details based on *Escherichia coli* STEC_C165_02 reference

		Putative membrane protein	80.9	100	6e-21	NH: Phage_ASH, Phage_ASH superfamily
		Hypothetical protein ECSTECC16502_1333	155	100	3e-49	
		Terminase B protein	283	100	1e-94	
		VRR-NUC domain protein	165	100	2e-53	
		Hypothetical protein ECSTECC16502_4927	57.4	100	2e-13	