

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 1. Signature details based on *Escherichia coli* 53638 reference

Region	Biomarker range in <i>Escherichia coli</i> 53638 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	Invasion plasmid antigen Replicative DNA helicase homolog Replication protein Hypothetical protein Ec53638_0018 Hypothetical protein Ec53638_0019 Conserved hypothetical protein Putative bacteriophage protein Addiction module antitoxin, Axe family Plasmid stabilization system protein, RelE/ParE family Phage Head-Tail Attachment Putative DNA-packaging protein Bacteriophage CI repressor protein Putative tail fiber protein Conserved domain protein Hypothetical bacteriophage protein Putative tail component of prophage Putative tail fiber protein Exodeoxyribonuclease 8 Host-nuclease inhibitor protein Gam Gifsy-1 prophage VmtH Hypothetical protein Ec53638_0122, [gp41 domain protein] Putative phage protein Conserved hypothetical protein Putative Prophage Qin DNA packaging protein NU1 homolog Prophage minor tail protein Z (GPZ) Putative bacteriophage protein Phage antitermination Q type 1 family Conserved hypothetical protein	829 885 199 562 94.4 249 745 154 216 252 223 267 481 77.4 205 203 99.4 153 153 226 278 202 197 204 195 142 135 107	73 100 100 100 98 100 100 100 100 99 100 100 100 99 98 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100	0.0 0.0 6e-60 0.0 1e-24 1e-82 0.0 1e-47 3e-71 2e-85 5e-74 1e-91 6e-171 1e-19 7e-68 6e-67 3e-26 2e-44 2e-48 3e-70 3e-97 2e-68 5e-67 8e-71 8e-65 5e-44 5e-43 2e-31	SH: NEL, NEL superfamily, MD: PRK15387, COG4886, PLN00113 SH: DnaB, DnaB superfamily, NH: DnaB_C, RecA-like_NTPases superfamily, MD: phage_DnaB, PRK06904, DnaB, DnaB_C SH: HTH_XRE, NH: HTH_19, HTH_XRE superfamily, MD: PRK09706, HipB NH: Phage_attach, Gifsy-2 superfamily NH: Packaging_FI, Packaging_FI superfamily NH: Phage_CI_repr, Phage_CI_repr superfamily NH: phage_tail_N, phage_tail_N superfamily SH: ParE, NH: Plasmid_stabil, Plasmid_stabil superfamily MD: PRK09709 NH: Gam, Gam superfamily NH: tape_meas_lam_C, tape_meas_lam_C superfamily, NH: TMP_2, TMP_2 superfamily, MD: COG5281 NH: COG4220, Phage_Nu1, Phage_Nu1 superfamily NH: Minor_tail_Z, Minor_tail_Z superfamily NH: Phage_antitermQ, Phage_antitermQ superfamily NH: DUF1133, DUF1133 superfamily
B	0 Mbp to 1 Mbp	No significant results				
C	0 Mbp to 1 Mbp	Hypothetical protein Ec53638_3910	130	100	1e-41	
D	0 Mbp to 1 Mbp	No significant results				
E	1 Mbp to 2 Mbp	gp27 Putative DNA packaging protein Gp17 Tail fiber Phage protein Putative DNA packaging protein Gp17 Hypothetical protein Ec53638_1156, [membrane protein] DNA-binding protein RdgB Hypothetical protein Ec53638_1176 Phage tail sheath protein Baseplate Bcv gene product Conserved hypothetical protein	343 405 271 282 481 275 115 177 340 217 274 179	100 100 100 100 100 100 100 100 100 100 100 100	9e-119 3e-139 3e-91 5e-92 2e-171 1e-93 1e-32 1e-57 1e-116 1e-72 2e-93 2e-57	NH: DUF3486, DUF3486 superfamily NH: COG4373, Terminase_1 superfamily, MD: 17 NH: Tail_P2_I, gpI, Tail_P2_I superfamily NH: COG4373, Terminase_1 superfamily, MD: P, Terminase_6 NH: Mor, Mor superfamily NH: FI, COG3497, Phage_sheath_1 superfamily, NH: GATase1_DJ-1, GAT_1 superfamily NH: phage_P2_V, gpV, Phage_base_V superfamily
F	1 Mbp to 2 Mbp	Putative type I restriction-modification system, S subunit	352	100	9e-120	

Supplementary Table 1. Signature details based on *Escherichia coli* 53638 reference

G	1 Mbp to 2 Mbp	ISCps8, transposase Transposase	809 200	100 100	0.0 9e-68	SH: Transposase_20, Transposase_20 superfamily, MD: COG3547
H	1 Mbp to 2 Mbp	Cytolethal distending toxin A Cytolethal distending toxin B Cytolethal distending toxin subunit C Hypothetical protein Ec53638_1905 Putative phage protein	486 545 370 158 253	100 100 100 100 100	9e-166 0.0 3e-122 2e-50 2e-81	NH: CDtoxinA, RICIN superfamily NH: PRK15251, CdtB, EEP superfamily NH: CDtoxinA, RICIN superfamily
I	2 Mbp to 3 Mbp	ISCps8, transposase	809	100	0.0	SH: Transposase_20, Transposase_20 superfamily, MD: COG3547
J	2 Mbp to 3 Mbp	Transposase, IS1111 family ISAFe1, transposase	599 594	100 99	0.0 0.0	SH: Transposase_20, Transposase_20 superfamily, NH: DEDD_Tnp_IS110, DEDD_Tnp_IS110 superfamily, MD: COG3547
K	2 Mbp to 3 Mbp	No significant results				
L	2 Mbp to 3 Mbp	Transposase, IS1111 family ISAFe1, transposase	580 576	100 99	0.0 0.0	SH: Transposase_20, Transposase_20 superfamily, MD: COG3547 SH: Transposase_20, Transposase_20 superfamily, MD: COG3547
M	3 Mbp to 4 Mbp	Transposase, IS1111 family ISAFe1, transposase	604 600	100 99	0.0 0.0	SH: Transposase_20, Transposase_20 superfamily, MD: COG3547 SH: Transposase_20, Transposase_20 superfamily, MD: COG3547
N	3 Mbp to 4 Mbp	Gp33 TerL Putative phage-associated protein, HI1409 family Phage Mu protein F domain protein Phage protein gp13 Phage protein gp12 Conserved hypothetical protein Transglycosylase SLT domain protein Conserved hypothetical protein Phage P2 baseplate assembly protein gpV Putative bacteriophage protein Putative tail fiber protein Hypothetical bacteriophage protein Phage integrase Conserved hypothetical protein Putative tail component of prophage Invasion plasmid antigen, probably secreted by the Mxi-Spa machinery Dead box helicase Hypothetical protein Ec53638_3354 Hypothetical protein Ec53638_3355 Arc DNA binding domain protein Protein of unknown function Phage anti-repressor protein AntB Type I restriction-modification system specificity subunit Conserved hypothetical protein Phage regulatory protein Cro gpH Bcv gene product Tail fiber assembly protein Type I restriction-modification enzyme, R subunit Type I restriction-modification system, M subunit	736 1,016 530 288 711 221 888 451 456 761 586 410 1,079 1,628 247 690 65.5 557 162 198 471 534 348 102 95.5 163 101 77.4 210 148	100 100 100 100 100 100 100 100 100 100 100 99 99 100 64 100	0.0 0.0 1e-176 5e-89 0.0 7e-66 0.0 9e-146 8e-148 0.0 0.0 1e-131 0.0 0.0 0.0 0.0 4e-12 0.0 2e-48 4e-61 5e-168 0.0 6e-121 2e-29 1e-26 1e-49 5e-27 5e-19 7e-65 1e-44	NH: DUF1073, COG3567, Phage_portal superfamily, NH: phge_rel_HI1409, phge_rel_HI1409 superfamily, NH: COG3566, DUF2213, DUF2213 superfamily MD: COG2369 NH: DUF3383, DUF3383 superfamily SH: LT_GEWL, NH: SLT, Lysozyme_like superfamily, MD: MItE, mItD, PHA00368, NH: DUF2612, DUF2612 superfamily SH: INT_REC_C, DNA_BRE_C superfamily NH: COG4688, COG4688 superfamily NH: NEL, NEL superfamily, MD: PRK15387, COG4886, Golgin_A5 NH: Phage_pRha, Phage_pRha superfamily SH: DUF45, NH: COG1451, superfamily: DUF45 NH: COG3561, AntA, AntA superfamily NH: Methylase_S, Methylase_S superfamily, MD: HsdS NH: VapI, THE_XRE superfamily, MD: PRK09706 MD: N6_Mtase, HsdM, hsdM, rumA
O	3 Mbp to 4 Mbp	ISCps8, transposase	809	100	0.0	SH: Transposase_20, Transposase_20 superfamily, MD: COG3547

Supplementary Table 1. Signature details based on *Escherichia coli* 53638 reference

P	3 Mbp to 4 Mbp	Invasion plasmid antigen	836	83	0.0	SH: NEL, NEL superfamily, MD: PRK15370
		Putative bacteriophage protein	528	100	0.0	
		Phage terminase large subunit	937	100	0.0	NH: phage_term_2, Terminase_3, Terminase_3 superfamily
		Putative bacteriophage protein	317	100	3e-100	NH: V, V superfamily
		Putative tail fiber protein	574	100	0.0	
		Hypothetical protein Ec53638_3420	77.0	100	6e-18	
		Putative tail component of prophage	239	99	3e-77	
		Putative bacteriophage protein	688	99	0.0	
		Hypothetical protein Ec53638_3782	478	100	1e-174	
		Conserved hypothetical protein	337	100	2e-118	NH: DedA, PRK10847, SNARE_assoc superfamily
		Bacteriophage lysis protein	52.8	100	2e-09	
		Exodeoxyribonuclease 8	162	100	7e-48	MD: PRK09709
		Host-nuclease inhibitor protein Gam	143	100	9e-45	NH: Gam, Gam superfamily
		gpH	180	100	6e-56	
		Bcv gene product	98.6	57	5e-26	
		Tail fiber assembly protein	102	100	3e-28	
		Hypothetical protein Ec53638_3785	277	100	1e-98	NH: DUF1627, DUF1627 superfamily
		Phage Mu protein F like protein	266	100	5e-93	NH: Phage_Mu_F, Phage_Mu_F superfamily, MD: COG2369
		Putative tail fiber protein	146	100	5e-46	NH: phage_tail_N, phage_tail_N superfamily
		Tail fiber assembly protein	137	100	5e-43	NH: Caudo_TAP, Caudo_TAP superfamily
Conserved hypothetical protein	130	100	2e-40	NH: DUF1133, DUF1133 superfamily		
Putative phage protein	141	100	2e-45			
Q	3 Mbp to 4 Mbp	Transcriptional regulator, AlpA family	114	98	1e-32	NH: Phage_AlpA, Phage_AlpA superfamily, MD: PRK09692
		Hypothetical protein Ec53638_3914	108	100	1e-30	
		Prophage integrase	256	100	4e-86	SH: DUF4102, DUF4102 superfamily, NH: INT_P4, DNA_BRE_C superfamily, MD: PRK09692
		Conserved hypothetical protein	379	100	2e-131	
		Hypothetical protein Ec53638_3910	119	100	4e-37	
R	3 Mbp to 4 Mbp	Putative tail component of prophage	80.1	100	1e-17	
		Invasion plasmid antigen	821	74	0.0	SH: NEL, NEL superfamily, MD: PRK15370, COG4886
		Hypothetical bacteriophage protein	412	100	2e-148	
		Putative tail fiber protein	666	100	0.0	
		Putative tail component of prophage	195	99	6e-60	
		Phage antitermination Q type 1 family	141	100	5e-44	NH: Phage_antitermQ, Phage_antitermQ superfamily
		Hypothetical protein Ec53638_4012	81.6	100	8e-22	
		ISCps8, transposase	809	100	0.0	SH: Transposase_20, Transposase_20 superfamily, MD: COG3547
		Hypothetical bacteriophage protein	393	100	6e-137	MD: PRK05643, DnaN
		Conserved hypothetical protein	496	100	3e-176	SH: ORF6N, ORF6N superfamily
		BRO family, N- domain protein	516	100	0.0	SH: Bro-N, NH: Bro-N, Bro-N superfamily, MD: COG3617
		Conserved domain protein	129	100	1e-40	NH: Phage_NinH, Phage_NinH superfamily
		Hypothetical protein Ec53638_4059	83.6	100	2e-23	
		Putative tail fiber protein	113	100	4e-32	NH: phage_tail_N, phage_tail_N superfamily
		Putative Prophage Qin DNA packaging protein NU1 homolog	134	100	6e-42	NH: COG4220, Phage_Nu1, Phage_Nu1 superfamily
		Hypothetical bacteriophage protein	96.3	100	2e-28	
S	4 Mbp to 5 Mbp	Putative membrane protein	944	100	0.0	NH: MATE_Wzx_like, MATE_Wzx_like superfamily, NH: Polysacc_synt, Polysacc_synt superfamily, MD: RfbX, spore_V_B
		Conserved hypothetical protein	909	100	0.0	NH: FrhB_FdhB_C, FrhB_FdhB_C superfamily, MD: PRK09326, FrhB
		Pyruvyl transferase	774	100	0.0	SH: PS_pyruv_trans, PS_pyruv_trans superfamily

Supplementary Table 1. Signature details based on *Escherichia coli* 53638 reference

		Glycosyl transferase domain protein, group 2 family	202	100	1e-57	SH: Glyco_tranf_GTA_type, NH: Glycos_transf_2, PLN02726, PRK10018, Glyco_tranf_GTA_type superfamily, MD: WcaA, PRK10073, Glyco_tranf_2_3, PgaC_IcaA, PTZ00260
		Glycosyltransferase, sugar-binding region containing DXD motif	539	99	1e-177	MD: Caps_synth, OCH1
		Putative membrane protein Cps23Fh	680	100	0.0	SH: EpsG, 7TMR_DISM_7TM superfamily
			695	100	0.0	SH: Glyco_tranf_GTA_type, NH: Glycos_transf_2, PRK10018, PLN02726, Glyco_tranf_GTA_type superfamily, MD: WcaA, PRK10073, glyc2_xrt_Gpos1, PTZ00260, Glyco_tranf_2_3
		UDP-galactopyranose mutase	758	99	0.0	SH: GLF, GLF superfamily, SH: NAD_binding_8, NAD_binding_8 superfamily, MD: Gif, UDP-GALP_mutase, PRK07208
		WfbU	620	99	0.0	NH: GT_2_like_b, Glycos_transf_2, Glyco_tranf_GTA_type superfamily, MD: COG1216, WcaA
		ISCps8, transposase	810	100	0.0	SH: Transposase_20, Transposase_20 superfamily, MD: COG3547
T	4 Mbp to 5 Mbp	gpH	705	100	0.0	SH: DUF3751, DUF3751 superfamily
		Putative phage gene	811	100	0.0	
		Conserved hypothetical protein	228	100	4e-69	NH: DUF2590, DUF2590 superfamily
		Phage tail tape measure protein, family	1,292	100	0.0	SH: PhageMin_Tail, MCP_signal superfamily
		Conserved hypothetical protein	185	100	2e-54	SH: DUF2765, DUF2765 superfamily
		Bacteriophage lysis protein	238	100	5e-72	SH: Phage_lysis, Phage_lysis superfamily
		Phage lysozyme	306	100	7e-96	NH: endolysin_autolysin, COG3772, Phage_lysozyme , Lysozyme_like superfamily
		Phage holin, lambda family	171	100	5e-49	
		Tail tube	311	100	2e-97	NH: DUF2597, DUF2597 superfamily
		Tail sheath	774	100	0.0	NH: DUF2586, DUF2586 superfamily
		Putative phage gene	498	100	2e-162	NH: P2_Phage_GpR, P2_Phage_GpR superfamily
		Conserved hypothetical protein	324	100	9e-102	
		Phage head completion protein (GPL)	321	100	9e-101	NH: Phage_GPL, Phage_GPL superfamily
		Putative repressor protein Cl	373	100	1e-118	
		Putative DNA-binding protein Ner	170	100	3e-49	NH: Nlp, HTH_35, PRK10344, HTH_35 superfamily
		Phage transposase	579	100	0.0	NH: HTH_Tnp_Mu_1, HTH_Tnp_Mu_1 superfamily,
		Packaging protein	450	100	7e-145	NH: M, Phage_term_smal, Phage_term_smal superfamily
		Phage major capsid protein, P2 family	730	100	0.0	NH: N, major_capsid_P2, Phage_cap_P2, Phage_cap_P2 superfamily
		Scaffold	666	100	0.0	NH: O, Phage_GPO, Phage_GPO superfamily
		Terminase	1,232	100	0.0	NH: Terminase_5, Terminase_5 superfamily, MD: P, Terminase_6, COG5484
		Phage portal protein, pbsx family	674	100	0.0	NH: Q, portal_PBSX, COG5518, Phage_portal, Phage_portal superfamily
		Phage transcriptional activator, Ogr/delta	186	100	1e-54	
		Hypothetical protein Ec53638_4365	176	100	6e-51	
		Hypothetical protein Ec53638_4366	162	100	3e-46	
		Putative phage replication protein	254	100	4e-70	
		Hypothetical protein Ec53638_4288	102	98	3e-26	
		Putative phage gene	1,133	100	0.0	
		Conserved hypothetical protein	217	100	9e-66	NH: DUF4406, DUF4406 superfamily
		DNA adenine methylase	641	100	0.0	NH: dam, PRK10904, Dam, MethyltransfD12, MethyltransfD12 superfamily
T	4 Mbp to 5 Mbp	Hypothetical protein Ec53638_4374	161	100	2e-46	
		Hypothetical protein Ec53638_4375	372	100	6e-121	
		Hypothetical protein Ec53638_4376	293	100	2e-92	
		Hypothetical protein Ec53638_4378	171	100	7e-50	

Supplementary Table 1. Signature details based on *Escherichia coli* 53638 reference

		Hypothetical protein Ec53638_4377	71.2	100	3e-15	
		Hypothetical protein Ec53638_4380	133	100	8e-37	
		Integrase for prophage CP-933T	711	100	0.0	NH: HP1_INT_C, Phage_integrase, DNA_BRE_C superfamily, MD: int, recomb_XerC, XerD, xerC
		Hypothetical bacteriophage protein	393	100	6e-137	DM: PRK05643, DnaN
		Conserved hypothetical protein	496	100	2e-176	NH: ORF6N, ORF6N superfamily
		Putative tail fiber protein	586	100	0.0	NH: phage_tail_N, phage_tail_N superfamily
		Putative tail component of prophage CP-933K	202	100	3e-63	
		Flagellin FliC	469	100	8e-164	SH: FliC, FliC superfamily, MD: PRK08026
		Putative DNA-packaging protein	223	100	5e-74	NH: Packaging_FI, Packaging_FI superfamily
		Phage Head-Tail Attachment	252	99	2e-85	NH: Phage_attach, Gifsy-2 superfamily
		Hypothetical protein Ec53638_4317	387	100	1e-137	
		Phage transposase	488	100	3e-170	SH: rve, rve superfamily
		PaaX family protein	266	99	3e-90	SH: HTH_36, HTH_36 superfamily, MD: PaaX_trns_reg
		Tail fiber protein	273	100	4e-91	
		Bcv gene product	262	96	7e-89	
		Putative phage replication protein	149	100	2e-43	
		Gifsy-1 prophage VmtH	229	100	1e-71	NH: tape_meas_lam_C, tape_meas_lam_C superfamily, MD: COG5281
		Putative phage regulatory protein, Rha family	204	100	4e-66	
		DNA transposition protein	203	100	2e-66	NH: Phage-MuB_C, Phage-MuB_C superfamily
		Gifsy-1 prophage VmtH	163	100	3e-48	SH: TMP_2, TMP_2 superfamily, MD: COG5281
		Prophage minor tail protein Z (GPZ)	195	100	7e-65	NH: Minor_tail_Z, Minor_tail_Z superfamily
		Putative Prophage Qin DNA packaging protein NU1 homolog	134	100	6e-42	NH: COG4220, Phage_Nu1, Phage_Nu1 superfamily
		Hypothetical protein Ec53638_4368, [inositol Monophosphatase 1 (IMPase 1) (IMP 1)(or 4)]	139	100	2e-44	
		Protein gp42	140	100	5e-41	
		Bacteriophage Mu Gam like protein	122	100	4e-37	NH: COG4396, Phage_Mu_Gam, Phage_Mu_Gam superfamily
U	5 Mbp to end of the genome	Putative bacteriophage protein	761	100	0.0	
		Putative tail fiber protein	575	100	0.0	
		Hypothetical bacteriophage protein	410	100	6e-135	NH: DUF2612, DUF2612 superfamily
		Phage P2 baseplate assembly protein gpV	456	100	5e-152	
		Invasion plasmid antigen	1,126	100	0.0	SH: TTSSLRR, TTSSLRR superfamily, SH: NEL, NEL superfamily, MD: PRK15370, COG4886
		Putative tail component of prophage CP-933K	246	99	6e-80	NH: DUF4376, DUF4376 superfamily
		Tail fiber assembly protein	137	100	7e-43	NH: Caudo_TAP, Caudo_TAP superfamily
		gp33 TerL	142	100	3e-42	
		Putative tail fiber protein	146	100	4e-44	NH: phage_tail_N, phage_tail_N superfamily
		Conserved hypothetical protein	236	100	3e-70	