

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 3. Signature details based on *Escherichia coli* RN587/1 reference

Region	Biomarker range in <i>Escherichia coli</i> RN587/1 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	Cytolethal distending toxin A/C family protein	205	100	2e-66	NH: CDtoxA, RICIN superfamily
B	0 Mbp to 1 Mbp	Type II restriction enzyme EcoRII	833	100	0.0	SH: EcoRII-C, EcoRII-C superfamily, NH: EcoRII_N, Bfil_C_EcoRII_N_B3 superfamily
		Modification methylase EcoRII	949	100	0.0	NH: dcm, DNA_methylase, Cyt_C5_DNA_methylase, Dcm, Cyt_C5_DNA_methylase superfamily, MD: PRK10458
		Outer membrane autotransporter barrel domain protein	214	100	3e-66	NH: PL1_Passenger_AT, PL1_Passenger_AT superfamily, NH: DUF4353, DUF4353 superfamily, MD: PHA03255
C	Around 1 Mbp	Hypothetical protein ECRN5871_0833	710	100	0.0	
		Hypothetical protein ECRN5871_0834	600	100	0.0	
		Hypothetical protein ECRN5871_0832	79.3	97	6e-18	
		Phage Tail Collar domain protein	528	100	0.0	NH: DUF3751, DUF3751 superfamily
		Caudovirales tail fiber assembly family protein	400	100	3e-136	SH: Caudo_TAP, Caudo_TAP superfamily
		Hypothetical protein ECRN5871_0827	442	100	6e-152	
		Phage protein	206	100	4e-67	NH: DUF2586, DUF2586 superfamily
		Pentapeptide repeats family protein	1,107	100	0.0	NH: SopA, SopA superfamily, NH: SopA_C, SopA_C superfamily, MD: Pentapeptide_4, PRK15377, COG1357
		hdmD	475	100	3e-167	
		Hypothetical protein ECRN5871_0812, [phage tail protein, P2 Phage tail completion protein R (GpR)]	316	100	1e-110	NH: P2_Phage_GpR, P2_Phage_GpR superfamily
		Phage protein	273	100	1e-95	NH: DUF2597, DUF2597 superfamily
		Hypothetical protein ECRN5871_0702	69.7	100	1e-16	
		Phage protein	375	100	2e-131	NH: DUF2586, DUF2586 superfamily, MD: PAT1, PHA03247
		Phage small terminase subunit	163	100	1e-50	NH: M, Phage_small_term superfamily
		Long tail fiber protein p37 domain protein	119	100	2e-35	
		Retron EC67 protein domain protein	850	100	0.0	SH: RT_Bac_retron_II, RT_like superfamily, MD: RVT_1
		Hypothetical protein ECRN5871_0823, [baseplate J-like protein]	814	100	0.0	NH: XkdT, Baseplate_J superfamily
		Hypothetical protein ECRN5871_0822, [PF10761 family protein]	77.8	100	6e-18	NH: DUF2590, DUF2590 superfamily
		Phage tail tape measure protein, TP901 family, core region	775	100	0.0	
		Hypothetical protein ECRN5871_0798	296	100	7e-101	
		Hypothetical protein ECRN5871_0799	276	100	4e-93	
		Regulatory protein CII	38.1	100	3e-04	
		Hypothetical protein ECRN5871_0797	119	100	6e-34	
		Hypothetical protein ECRN5871_0829	702	100	0.0	
		Phage portal protein, PBSX family	342	100	3e-119	NH: Q, portal_PBSX, Phage_portal, Phage_portal superfamily
		Hypothetical protein ECRN5871_0691	501	100	0.0	SH: DUF1311, DUF1311 superfamily, MD: PHA02067, LprI
		Terminase, ATPase subunit	605	100	0.0	MD: P, Terminase_6
		Protein rhsB	355	100	2e-113	NH: Rhs_assc_core, Rhs_assc_core superfamily
		Replication protein A	350	100	2e-118	NH: Phage_GPA, Phage_GPA superfamily
		Hypothetical protein ECRN5871_0824	273	100	5e-95	
Integrase	495	100	1e-177	NH: HP1_INT_C, DNA_BRE_C superfamily, MD: int, recomb_XerD, XerD,		
Scaffold domain protein	473	100	6e-171	NH: O, Phage_GPO, Phage_GPO superfamily, MD: Inca		
Repressor protein CI	338	100	4e-120	NH: Phage_CI_repr, Phage_CI_repr superfamily		
Hypothetical protein ECRN5871_0819	129	98	3e-40			

Supplementary Table 3. Signature details based on *Escherichia coli* RN587/1 reference

C	Around 1 Mbp	Hypothetical protein ECRN5871_0801	202	100	2e-68	NH: DUF3850, DUF3850 superfamily NH: Dam, PRK10904, MethyltransfD12, dam, MethyltransfD12 superfamily NH: PhageMin_Tail, MCP_signal superfamily MD: PRK14960 NH: Phage_holin_2, Phage_holin_2 superfamily NH: DUF2765, DUF2765 superfamily
		Retron EC67 DNA adenine methylase	204	100	3e-67	
		Phage tail tape measure protein, TP901 family, core region	200	100	2e-62	
		Replication protein A	201	100	2e-62	
		Phage major capsid protein, P2 family	48.9	100	1e-08	
		Holin	93.6	100	8e-27	
		Phage head completion protein family protein	98.2	100	4e-28	
		Phage protein	152	100	1e-48	
Hypothetical protein ECRN5871_0700	107	100	2e-32			
D	1 Mbp to 2 Mbp	Hypothetical protein ECRN5871_1072, [type III secretion system protein]	594	100	1e-174	MD: PRK15386 NH: Phage_CI_repr, Phage_CI_repr superfamily NH: DUF1482, DUF1482 superfamily NH: INT_Lambda_C, DNA_BRE_C superfamily
		Hypothetical protein ECRN5871_4139, [type III secretion system protein]	321	65	2e-89	
		Bacteriophage CI repressor helix-turn-helix domain protein	431	100	2e-134	
		Exonuclease family protein	676	100	0.0	
		Avirulence protein A domain protein	551	100	2e-173	
		Hypothetical protein ECRN5871_1040	185	100	8e-54	
		Hypothetical protein ECRN5871_1039	237	100	1e-71	
		Hypothetical protein ECRN5871_1038	155	100	2e-45	
		Eae-like protein	223	100	2e-66	
		Hypothetical protein ECRN5871_1027	224	100	2e-72	
Phage integrase family protein	227	100	2e-67			
E	1 Mbp to 2 Mbp	ea59 protein	1,043	100	0.0	SH: AAA_21, NH: COG3910, ABC_ATPase superfamily, MD: AAA_15 NH: INT_Lambda_C, DNA_BRE_C superfamily NH: NA37, PRK00378, NA37 superfamily NH: Phage_integ_N, Phage_integ_N superfamily
		Hypothetical protein ECRN5871_4153, [HNH endonuclease family protein]	576	100	0.0	
		Prophage lambda integrase	234	100	3e-68	
		Hypothetical protein ECRN5871_4172	829	100	0.0	
		Hypothetical protein ECRN5871_4173	301	100	5e-97	
		Hypothetical protein ECRN5871_4177, [Rz1 lytic protein]	91.3	98	3e-24	
		Prophage lambda integrase	104	100	1e-28	
F	1 Mbp to 2 Mbp	Reverse transcriptase	331	100	6e-107	NH: RT_Bac_retron_II, RT_like superfamily SH: AAA_23, NH: ABC_RecF, COG3910, AAA_21, ABC_ATPase superfamily, MD: COG3950, recF, recf, AAA_15 NH: TIGR02646, TIGR02646 superfamily NH: COG4688, COG4688 superfamily NH: PRK15131, PMI_typel, PLN02288, ABD superfamily, MD: ManA SH: Transposase_20, Transposase_20 superfamily, MD: COG3547 SH: Transposase_20, Transposase_20 superfamily, MD: COG3547 MD: pcrA, UvrD, uvrD, UvrD-helicase
		recF/RecN/SMC N terminal domain protein	1,100	100	0.0	
		Hypothetical protein ECRN5871_4523, [TIGR02646 family protein]	424	100	9e-142	
		Hypothetical protein ECRN5871_4556	775	100	0.0	
		Hypothetical protein ECRN5871_4549	91.7	98	1e-24	
		Transposase for insertion sequence element IS1111A	336	100	5e-118	
		Transposase IS116/IS110/IS902 family protein	338	100	1e-117	
Hypothetical protein ECRN5871_4552, [UvrD/REP helicase domain protein]	273	100	3e-91			
F	1 Mbp to 2 Mbp	tnpA	71.2	100	6e-18	NH: ABC_SMC_barmotin, AAA_23, ABC_ATPase superfamily,
		Hypothetical protein ECRN5871_4551, [chromosome segregation	206	100	6e-65	

Supplementary Table 3. Signature details based on *Escherichia coli* RN587/1 reference

		protein SMC] mdaB domain protein Hypothetical protein ECRN5871_4526 Hypothetical protein ECRN5871_4525 Hypothetical protein ECRN5871_4553	172 137 103 58.9	100 100 100 61	4e-57 1e-40 3e-30 1e-13	MD: SMC_N, Smc, PRK14272, recf NH: Flavodoxin_2, PRK00871, FMN_red superfamily
G	1 Mbp to 2 Mbp	Peptidoglycan domain protein Tail fiber domain protein	374 131	100 100	2e-133 9e-41	NH: Glyco_hydro_108, Glyco_hydro_108 superfamily, NH: PG_binding_3, PG_binding_3 superfamily, MD: zliS
H	2 Mbp to 3 Mbp	Exonuclease family protein	349	100	3e-116	NH: dexA, DnaQ_like_exo superfamily, MD: PRK09709
I	2 Mbp to 3 Mbp	Non-LEE-encoded effector EspJ	111	100	6e-32	
J	2 Mbp to 3 Mbp	Phage protein Zinc-binding domain of primase-helicase family protein Hypothetical protein ECRN5871_3674 Hypothetical protein ECRN5871_3673 Hypothetical protein ECRN5871_3672 Prophage CP4-57 regulatory protein family protein Integrase Hypothetical protein ECRN5871_3699	268 1,586 432 271 216 151 833 343	100 100 100 100 99 100 100 100	3e-78 0.0 2e-140 2e-84 5e-65 2e-43 0.0 1e-120	NH: Phage_ASH, Phage_ASH superfamily SH: Prim_Zn_Ribbon, NH: Prim_Zn_Ribbon, Prim_Zn_Ribbon superfamily, NH: Toprim_3, Toprim superfamily, MD: COG4643 NH: Phage_Alpa, AlpA, Phage_Alpa superfamily SH: INT_P4, DNA_BRE_C superfamily, NH: DUF4102, DUF4102 superfamily, MD: PRK09692, XerC NH: DUF2544, DUF2544 superfamily
K	2 Mbp to 3 Mbp	Phage portal protein, PBSX family Terminase, ATPase subunit Presumed capsid-scaffolding protein Phage major capsid protein, P2 family Terminase, endonuclease subunit Head completion/stabilization protein Hypothetical protein ECRN5871_3522 Hypothetical protein ECRN5871_3519 Hypothetical protein ECRN5871_3516, [PF05449 family protein] Hypothetical protein ECRN5871_3515, [PF11860 family protein] P2 phage tail completion protein R family protein Phage virion morphogenesis protein hicB family protein Baseplate assembly protein V Baseplate assembly protein W Baseplate J-like family protein Phage tail protein I Phage tail fiber repeat family protein Tail fiber domain protein Phage Tail Collar domain protein	284 1,238 439 632 420 290 417 215 209 408 318 300 230 377 236 330 397 1,291 219 244	100 100 100 100 100 100 99 100 100 100 100 100 100 100 100 100 100 100 68 58	1e-86 0.0 7e-144 0.0 4e-138 2e-91 3e-137 3e-65 1e-62 2e-132 9e-101 3e-94 5e-70 5e-121 3e-72 5e-102 2e-127 0.0 7e-62 1e-69	NH: Q, COG5518, portal_PBSX, Phage_portal superfamily SH: Terminase_5, Terminase_5 superfamily, MD: P, Terminase_6, COG5484 NH: Phage_GPO, O, Phage_GPO superfamily NH: N, Phage_cap_P2, major_capsid_P2, Phage_cap_P2 superfamily NH: M, Phage_term_smal, Phage_term_smal superfamily NH: Phage_GPL, Phage_GPL superfamily NH: DUF754, DUF754 superfamily SH: DUF3380, DUF3380 superfamily NH: P2_Phage_GpR, P2_Phage_GpR superfamily SH: Phage_tail_S, NH: tail_comp_S, Phage_tail_S superfamily NH: HicB, HicB superfamily, MD: HicB NH: phage_P2_V, gpV, Phage_base_V, Phage_base_V superfamily NH: W, COG3628, GPW_gp25, GPW_gp25 superfamily NH: J, COG3948, Baseplate_J, Baseplate_J superfamily NH: gpl, tail_P2_I, tail_P2_I superfamily NH: DUF3751, DUF3751 superfamily, SH: Phage_fiber_2, Phage_fiber_2 superfamily, MD: COG5301 NH: DUF3751, DUF3751 superfamily NH: DUF3751, DUF3751 superfamily
K	2 Mbp to 3 Mbp	Hypothetical protein ECRN5871_3504,[tail fiber assembly protein]	470	100	6e-153	NH: DUF4376, DUF4376 superfamily, NH: Caudo_TAP, Caudo_TAP superfamily

Supplementary Table 3. Signature details based on *Escherichia coli* RN587/1 reference

		Enoyl-CoA hydratase/carnithine racemase-like protein	321	100	4e-101	SH: Dam, NH: dam, PRK10904, MethyltransfD12,
		D12 class N6 adenine-specific DNA methyltransferase family protein	528	100	7e-174	MethyltransfD12 superfamily
		Major tail sheath protein	546	100	2e-178	NH: FI, COG3497, Phage_sheath_1 superfamily
		Phage major tail tube protein	328	100	5e-103	NH: FI, Phage_tube, COG3498, tail_tube, Phage_tube superfamily
		Phage tail tape measure protein, TP901 family, core region	1,687	100	0.0	NH: tape_meas_TP901, PhageMin_Tail, MCP_signal superfamily, MD: COG5283, NH: PHA01399, PHA01399 superfamily
		Phage P2 GpU family protein	287	100	7e-89	NH: COG3499, Phage_P2_GpU, Phage_P2_GpU superfamily
		Phage late control gene D family protein	727	100	0.0	NH: D, COG3500, Phage_GPD, Phage_GPD superfamily
		Caspase domain protein	671	100	0.0	
		Hypothetical protein ECRN5871_3490	1,759	100	0.0	
		Hypothetical protein ECRN5871_3489	259	100	0.0	
		Integrase	511	100	2e-169	SH: INT_P4, DNA_BRE_C superfamily, NH: DUF4102, DUF4102 superfamily, MD: PRK09692
		Phage portal protein, PBSX family	320	100	2e-106	NH: Q, portal_PBSX, COG5518, Phage_portal superfamily
		Hypothetical protein ECRN5871_3481, [RatA]	202	100	5e-63	MD: PRK15316
		Aldose 1-epimerase family protein	139	100	1e-42	NH: Aldose_epim_Ec_YphB, GalM, Aldose_epim superfamily
L	3 Mbp to 4 Mbp	Hypothetical protein ECRN5871_3053	1,435	100	0.0	MD: AAA_13, COG4694, SMC_prok_B
		Hypothetical protein ECRN5871_3052	380	100	1e-123	
		Hypothetical protein ECRN5871_3051, [nucleotidyl transferase, PF08843 family]	679	100	0.0	NH: COG2253, DUF1814 superfamily
		Hypothetical protein ECRN5871_3050, [ATPase, tellurite resistance protein TerB]	1,535	100	0.0	SH: TerB-N, TerB-N superfamily, SH: TerB-C, TerB-C superfamily, NH: terB, terB_like superfamily
		Biotin carboxylase	619	100	0.0	NH: DUF2791, AAA superfamily
		DEAD/DEAH box helicase family protein	988	100	0.0	SH: DEAD, NH: DEADc, DEXDc superfamily, NH: HELICc, HELICc superfamily, NH: HELICc, Helicase_C, Helicase_C superfamily, MD: Lhr, PRK13767, DEXH_lig_assoc, DEXDc, PTZ00110, PLN00206
		Hypothetical protein ECRN5871_3043, [ATP-dependent helicase]	407	100	2e-130	
		Hypothetical protein ECRN5871_3042, [PF09983 domain protein]	607	100	0.0	
		Hypothetical protein ECRN5871_3041	777	100	0.0	
		Hypothetical protein ECRN5871_3040	352	100	6e-111	
		Hypothetical protein ECRN5871_3039, [PF12128 family protein]	2,302	100	0.0	SH: Tropomyosin_1, Cep57_CLD, ApoLP_III_like superfamily, NH: SPEC, SPEC superfamily, NH: iSH2_P13K_IA_R, iSH2_P13K_IA_R superfamily, MD: DUF3584, Smc, PRK03918, SMC_prok_B, AIP3
		DNA-binding prophage protein	597	100	0.0	SH: Integrase_1, Integrase_1 superfamily, NH: Phage_int_SAM_2, Phage_int_SAM_2 superfamily
		DNA-binding protein Ner	145	100	5e-42	NH: Nlp, HTH_35, PRK10344, HTH_35 superfamily
		Integrase domain protein	450	100	6e-161	NH: INT_P4, Phage_integrase, DNA_BRE_C superfamily, MD: PRK09692, XerC, SH: DUF4102, DUF4102 superfamily
M	3 Mbp to 4 Mbp	Death on curing protein	255	100	4e-87	NH: Doc, DOC_P1, Fic superfamily
N	4 Mbp to 5 Mbp	espA-like secreted family protein	303	100	3e-104	NH: EspA, EspA superfamily
		Secretion system effector C (SseC) like family protein	147	100	1e-41	NH: COG5613, COG5613 superfamily

Supplementary Table 3. Signature details based on *Escherichia coli* RN587/1 reference

O	4 Mbp to 5 Mbp	No significant results				
P	4 Mbp to 5 Mbp	Anaerobic C4-dicarboxylate transporter dcuA Fumarylacetoacetate (FAA) hydrolase family protein Dihydrodipicolinate synthetase family protein Transcriptional regulator, LacI family Sugar (and other) transporter family protein Porin B	721 610 600 466 941 837	100 100 100 100 100 100	0.0 0.0 0.0 2e-152 0.0 0.0	NH: Dcu, PRK09412, DcuB, DcuA_DcuB, DcuA_DcuB superfamily SH: FAA_hydrolase, NH: PRK10691, HpaG-N-term, FAA_hydrolase superfamily, MD: MhpD, HpaG-C-term, PRK15203 SH: DHDPS-like, NH: DapA, PRK03170, DHDPS, dapA, PLN02417, Aldolase_Class_I superfamily SH: HTH_LacI, NH: HTH_LacI, HTH_LacI superfamily, NH: PBP1_LacI_like_7, Periplasmic_Binding_Protein_Type_I superfamily, NH: Peripla_BP_3, MD: PurR, PRK10703, ccpA, Peripla_BP_1 SH: MFS, MFS superfamily, NH: 2_A_01_02, PRK11195 superfamily, MD: 2A0115, PRK11551, Sugar_tr, ProP, synapt_SV2 NH: OprB, OprB superfamily
Q	4 Mbp to 5 Mbp	Type I restriction enzyme specificity protein Type I restriction-modification system, M subunit Type I site-specific deoxyribonuclease, HsdR family	213 256 139	100 100 100	2e-69 7e-82 2e-40	NH: Methylase_S, Methylase_S superfamily, MD: HsdS, sufB MD: HsdM, hsdM, N6_Mtase SH: HSDR_N, HSDR_N superfamily, MD: COG0610, hsdR
R	4.5 Mbp to end of the genome	Hypothetical protein ECRN5871_0098 Hypothetical protein ECRN5871_0099 Hypothetical protein ECRN5871_0100 sel1 repeat family protein Hypothetical protein ECRN5871_0138, [secretion protein EspT] Hypothetical protein ECRN5871_0139 Hypothetical protein ECRN5871_0129 Hypothetical protein ECRN5871_0128, [secretion protein EspM] Transposase IS116/IS110/IS902 family protein Hypothetical protein ECRN5871_0087 Cysteine protease domain, YopT-type domain protein Hypothetical protein ECRN5871_0025, [N-acetyltransferase GCN5] Hypothetical protein ECRN5871_0026 Hypothetical protein ECRN5871_0137 Superoxide dismutase [Cu-Zn] 1 Hypothetical protein ECRN5871_0110 Transposase for insertion sequence element IS1111A Retron EC67 protein domain protein Phage portal protein, PBSX family yadA Hypothetical protein ECRN5871_0131	1,119 1,435 814 1,603 406 546 435 103 233 390 1,030 332 205 399 344 161 292 257 95.9 316 114	100 100 99 100 100 99 100 100 100 100 100 100 100 100 100 86 100 100 100 100	0.0 0.0 0.0 0.0 4e-136 0.0 1e-149 7e-27 2e-72 4e-134 0.0 2e-114 1e-65 4e-142 2e-121 8e-51 3e-101 4e-88 3e-24 3e-111 4e-33	NH: DLP_2, Dynamin_N, Ras_like_GTPase superfamily NH: DLP_2, PRK09866, Ras_like_GTPase superfamily NH: DLP_2, Ras_like_GTPase superfamily NH: IpaB_EvcA, IpaB_EvcA superfamily NH: DUF1076, DUF1076 superfamily SH: IpaB_EvcA, IpaB_EvcA superfamily, NH: sifB, sif superfamily SH: Transposase_20, Transposase_20 superfamily, MD: COG3547 NH: DUF3491, DUF3491 superfamily SH: COG4453, DUF 1778 NH: DUF1076, DUF1076 superfamily NH: PRK15388, SodC, Cu-Zn_Superoxide_Dismutase, Sod_Cu, PLN02386, Cu-Zn_Superoxide_Dismutase superfamily NH: DEDD_Tnp_IS110, DEDD_Tnp_IS110 superfamily, SH: Transposase_20, Transposase_20 superfamily, MD: COG3547 NH: Q, Phage_portal superfamily MD: PTZ00102
R	4.5 Mbp to end of the genome	Terminase large subunit Terminase small subunit Serine protease eatA	226 206 345	100 100 100	5e-71 7e-68 8e-112	NH: COG5525, Terminase_GpA, Terminase_GpA superfamily SH: COG4220, NH: Phage_Nu1, Phage_Nu1 superfamily, MD: PLN03114 MD: Peptidase_S6

Supplementary Table 3. Signature details based on *Escherichia coli* RN587/1 reference

	ST44 protein	133	100	4e-42	NH: Transposase_mut, Transposase_mut superfamily, MD: COG3328
	ygeA	162	100	5e-53	
	Baseplate assembly protein V	140	100	8e-44	NH: gpV, phage_P2_V, Phage_base_V, Phage_base_V superfamily
	Transposase, IS605 OrfB family	62.4	100	8e-15	
	Putative transposase	136	100	7e-44	
	Cysteine protease domain, YopT-type domain protein	92.0	100	1e-23	