Supplementary Information:

Genome Analysis of Bacteriophage (U1G) of *Schitoviridae*, host receptor prediction using machine learning tools and its evaluation to mitigate colistin resistant clinical isolate of *Escherichia coli in vitro* and *in vivo*

Niranjana Sri Sundaramoorthy[a,b], Vineetha KU[a] Jean Sophy Roy[a], , Sneha Srinath[c], Santhosh Kumar S[c], Suma Mohan S.*[c] and Saisubramanian Nagarajan[a,d]*

[a] Center for Research on Infectious Diseases (CRID), School of Chemical and Biotechnology, SASTRA Deemed University, Tamil Nadu, India

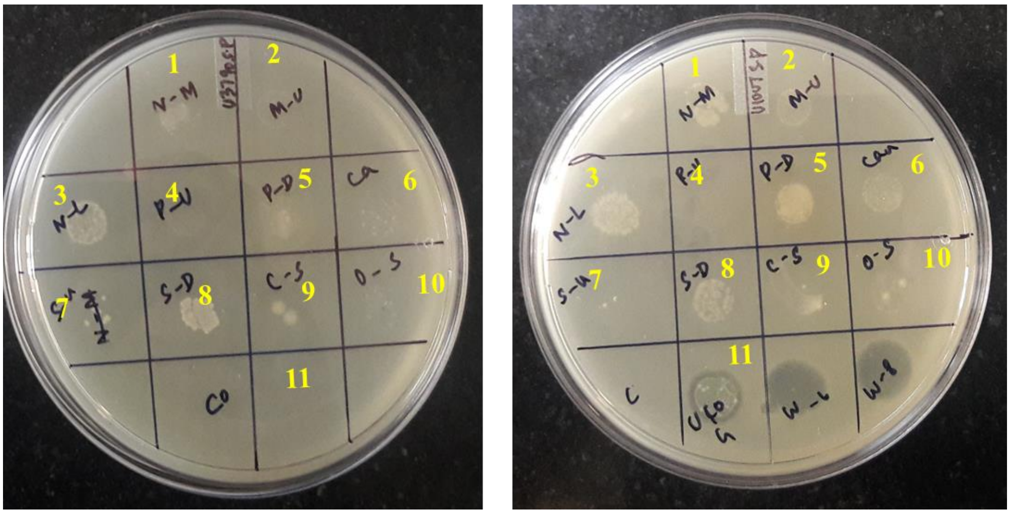
[b] Translational Health Sciences Technology Institute, Faridabad

[c] Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA Deemed University, Tamil Nadu, India

[d] Antimicrobial Resistance Lab, ASK-I-312, School of Chemical and Biotechnology, SASTRA Deemed University, Tamil Nadu, India

*Jointly Communicated
Communicating Authors
Dr. Saisubramanian Nagarajan
Mail id: sai@scbt.sastra.edu
Dr. Suma Mohan
Mail id: sumamohan@scbt.sastra.edu

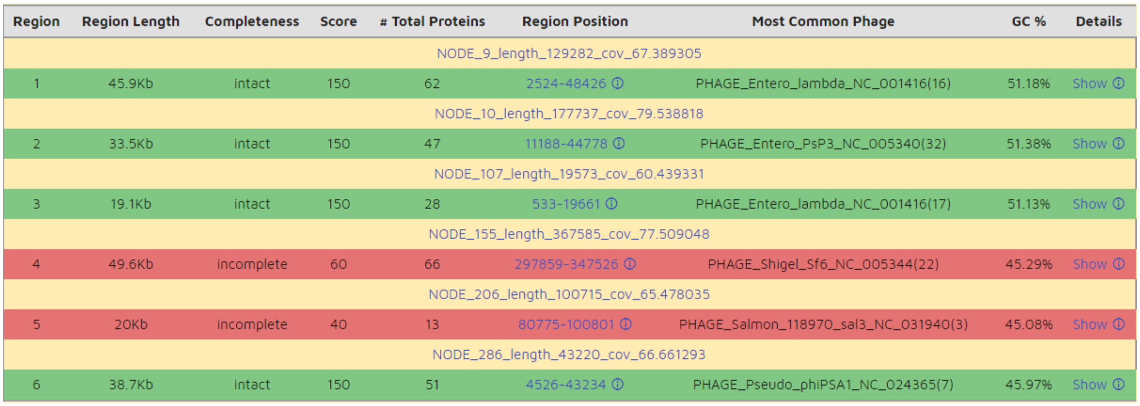
Supplementary Figure 1: Spot assay showing the presence of bacteriophages specific to U1007. 1- Water sample from Madurai Pond 1, 2- Cow's urine sample from Madurai, 3- Water sample from Madurai pond 2, 4 – Urine sample from cow shed at Perambalur, 5 – Dung sample from cow shed at Perambalur, 6 – Water sample from Cauvery river, 7 – Urine sample from SASTRA cow shed, 8 – Dung sample from SASTRA cow shed, 9 – Soil sample from a corn farm, 10 – Soil sample from a onion farm, 11 – Water sample from Ganges river



A) U3790

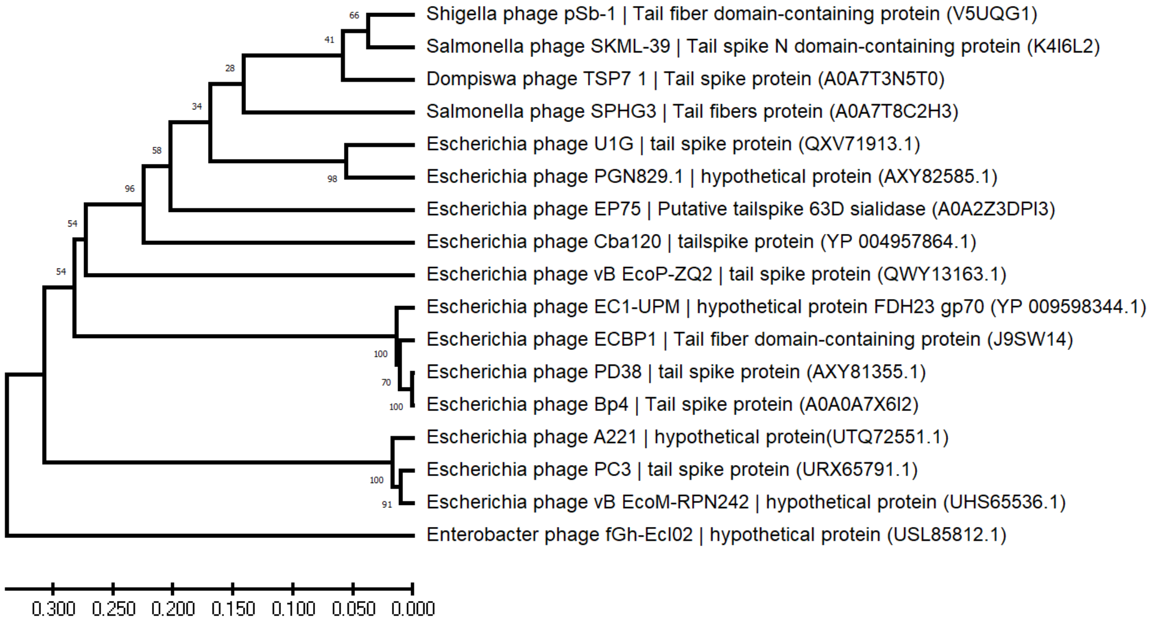
B) U1007

Supplementary Figure 2: Presence of four intact prophages in U3790 genome was revealed by PHASTER



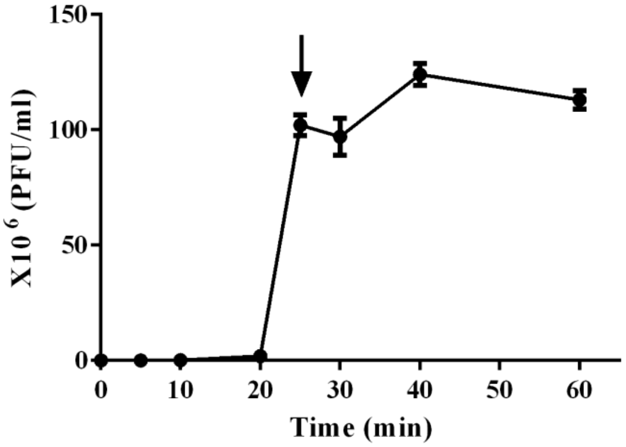
Region	Region Length	Completeness	Score	# Total Proteins	Region Position	Most Common Phage	GC %	Details
NODE_9_length_129282_cov_67.389305								
1	45.9Kb	intact	150	62	2524-48426 ⓘ	PHAGE_Entero_lambda_NC_001416(16)	51.18%	Show ⓘ
NODE_10_length_177737_cov_79.538818								
2	33.5Kb	intact	150	47	11188-44778 ⓘ	PHAGE_Entero_PsP3_NC_005340(32)	51.38%	Show ⓘ
NODE_107_length_19573_cov_60.439331								
3	19.1Kb	intact	150	28	533-19661 ⓘ	PHAGE_Entero_lambda_NC_001416(17)	51.13%	Show ⓘ
NODE_155_length_367585_cov_77.509048								
4	49.6Kb	incomplete	60	66	297859-347526 ⓘ	PHAGE_Shigel_Sf6_NC_005344(22)	45.29%	Show ⓘ
NODE_206_length_100715_cov_65.478035								
5	20Kb	incomplete	40	13	80775-100801 ⓘ	PHAGE_Salmon_118970_sal3_NC_031940(3)	45.08%	Show ⓘ
NODE_286_length_43220_cov_66.661293								
6	38.7Kb	intact	150	51	4526-43234 ⓘ	PHAGE_Pseudo_phiPSA1_NC_024365(7)	45.97%	Show ⓘ

Supplementary Figure 3: Phylogenetic tree of the tailspike protein Phylogenetic tree of the tailspike protein constructed with the UPGMA method with 1000 bootstrap values using MEGA11. The homologs of the tail spike protein of U1G from other bacteriophages were obtained using BLASTp search.

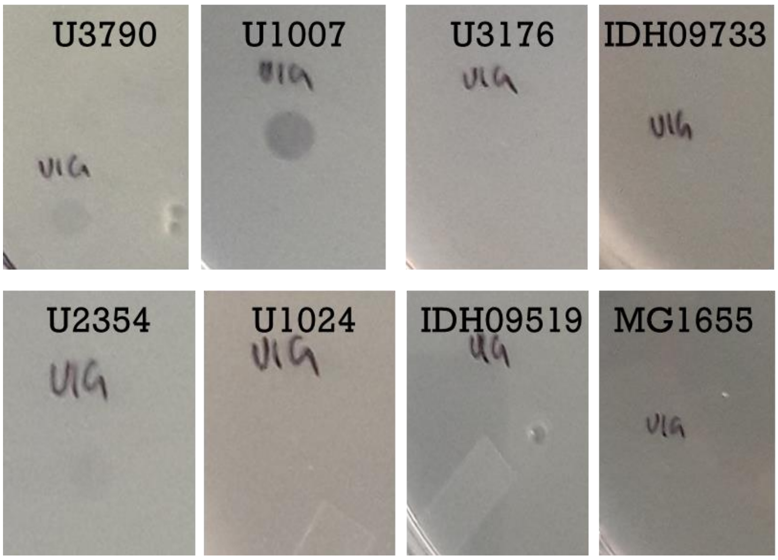


Supplementary Figure 4: Genome alignment of the close homologs of Escherichia phage U1G with Mauve

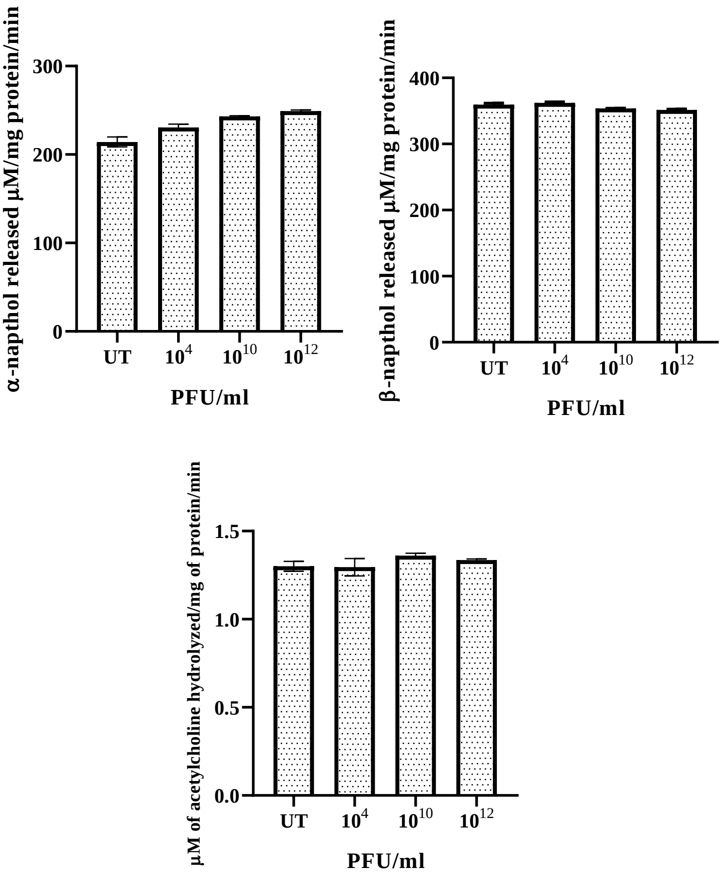
Supplementary Figure 5: One-step growth curve of U1G



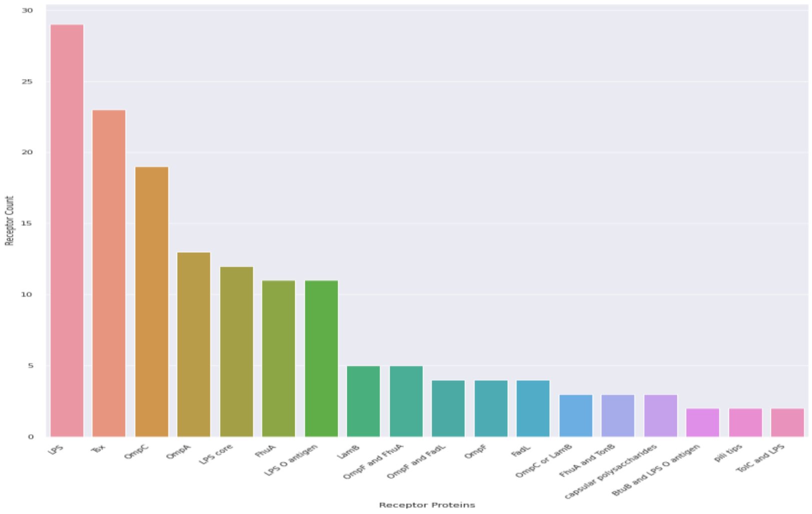
Supplementary Figure 6. U1G is highly specific to U1007, with a slight lysis against U3790.



Supplementary Figure 7: U1G was non-toxic to zebrafish at different concentrations as evidenced by brain and liver enzyme profiles**.**



Supplementary figure 8: The receptor label distribution in the dataset with multiple entries.



Supplementary Table 1: Features of U1G genome annotated by RAST and PHASTER

Start	Stop	RAST	PHASTER
170	493	Phage protein	hypothetical protein
584	943	hypothetical protein	hypothetical protein
997	1203	hypothetical protein	Major Capsid protein
1330	1713	Phage integrase	hypothetical protein
1857	2156	hypothetical protein	hypothetical protein
2159	2377	hypothetical protein	-
2374	2547	hypothetical protein	-
2571	3044	Phage HNH homing endonuclease (ACLAME 27)	gp59
3049	3180	hypothetical protein	hypothetical protein
3314	3622	hypothetical protein	hypothetical protein
3619	3840	hypothetical protein	hypothetical protein
3837	4094	hypothetical protein	hypothetical protein
4091	4411	Phage antirepressor protein	hypothetical protein
4414	4596	hypothetical protein	hypothetical protein
4593	4973	hypothetical protein	hypothetical protein
5027	5845	hypothetical protein	RNA polymerase RNAP1
5835	5996	hypothetical protein	hypothetical protein
6043	7260	hypothetical protein	T7-like RNA polymerase
7318	7500	hypothetical protein	-
7575	8408	Phage fibritin (wac) protein	Multi domain protein
8474	8788	Superinfection exclusion protein (Protein gp17)	Capsid decoration protein
8893	9081	hypothetical protein	hypothetical protein
9078	9338	hypothetical protein	hypothetical protein
9335	9520	hypothetical protein	hypothetical protein
9521	9847	hypothetical protein	hypothetical protein
9882	10202	Phage protein	hypothetical protein
10313	10882	HNH homing endonuclease # Phage intron	Putative homing endonuclease HNH family
11142	11816	Phage protein	hypothetical protein

11824	12327	hypothetical protein	hypothetical protein
12324	12998	hypothetical protein	hypothetical protein
12998	13504	Phage protein	deoxycytidine triphosphate deaminase
13514	13723	hypothetical protein	hypothetical protein
13797	14039	hypothetical protein	-
14055	14387	hypothetical protein	hypothetical protein
14554	14997	hypothetical protein	hypothetical protein
14997	15941	Phage protein	hypothetical protein
16009	16221	hypothetical protein	hypothetical protein
16214	16543	Phage protein	hypothetical protein
16536	16730	hypothetical protein	hypothetical protein
16774	19341	Phage rIIA lysis inhibitor	rIIA lysis inhibitor
19346	21406	hypothetical protein	rIIB protector from prophage-induced early lysis
21469	21864	hypothetical protein	hypothetical protein
21911	22045	hypothetical protein	hypothetical protein
22108	22233	Phage protein	hypothetical protein
22725	24077	Phage DNA helicase	Putative HNH homing endonuclease
24088	24618	hypothetical protein	hypothetical protein
24628	26442	DNA polymerase I (EC 2.7.7.7), phage-associated	DNA polymerase I
26439	27095	Phage protein (ACLAME 141)	hypothetical protein
27076	28047	DNA polymerase I (EC 2.7.7.7), phage-associated	DNA polymerase I
28044	28349	hypothetical protein	hypothetical protein
28349	28822	hypothetical protein	hypothetical protein
28822	29799	hypothetical protein	HNH endonuclease
29796	30233	Phage-associated homing endonuclease	-
30276	32426	Phage-associated DNA primase	hypothetical protein
32483	32884	Phage protein	hypothetical protein
32950	33234	Phage protein	hypothetical protein
33430	34071	hypothetical protein	Single-stranded DNA binding protein
34071	34625	hypothetical protein	hypothetical protein

34627	35067	hypothetical protein	hypothetical protein
35299	35412	hypothetical protein	hypothetical protein
36836	37216	hypothetical protein	hypothetical protein
48080	37326	DNA polymerase, phage-associated	Gram positive anchor domain protein
49890	48175	hypothetical protein	Putative structural protein
50138	49884	hypothetical protein	66 kDa protein
50594	50151	hypothetical protein	16.5 kDa protein
53262	50608	hypothetical protein	hypothetical protein
54100	53264	hypothetical protein	hypothetical protein
54809	54177	hypothetical protein	hypothetical protein
56085	54883	hypothetical protein	Major coat protein
57322	56102	Phage tape measure protein	hypothetical protein
57686	57342	Phage protein	hypothetical protein
59970	57700	Phage portal (connector) protein	94 kDa protein
60479	59973	hypothetical protein	hypothetical protein
60891	60460	Phage protein	hypothetical protein
61094	60996	hypothetical protein	-
61657	61328	Phage tail length tape-measure protein T	hypothetical protein
62449	61916	hypothetical protein	hypothetical protein
63285	62485	hypothetical protein	hypothetical protein
63800	63477	Phage tailspike protein	hypothetical protein
65447	64050	hypothetical protein	Tail fibre protein
67642	65492	hypothetical protein	Tail fibre protein
68349	67639	hypothetical protein	30 kDa protein
69945	68356	Phage terminase, large subunit	hypothetical protein
70627	69938	hypothetical protein	hypothetical protein
70811	71089	hypothetical protein	-
71654	71968	hypothetical protein	hypothetical protein
71972	72292	hypothetical protein	hypothetical protein
72292	72576	hypothetical protein	hypothetical protein
72573	72887	hypothetical protein	hypothetical protein
72877	73215	hypothetical protein	hypothetical protein

Supplementary table 2: The dataset with RBP and host receptor information used for building the ML models.

https://docs.google.com/spreadsheets/d/13BCVyxdjq-PVnH3MuY0T2RAiINW6wk5L/edit#gid=230086930