1 **Complete genome sequence of C130_2, a novel Myovirus infecting pathogenic *Escherichia coli* and**

2 ***Shigella* strains**

3

4 Domonkos Sváb[1], Linda Falgenhauer [2], Manfred Rohde[3], Trinad Chakraborty[2] and

5 István Tóth[1]

6

7 [1]Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences,

8 Budapest, Hungary, [2] Institute of Medical Microbiology, Justus Liebig University Giessen and German Center

9 for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany, [3]Central Facility for

10 Microscopy, Helmholtz Centre for Infection Research, HZI, Braunschweig, Germany

11

12 **Abstract**

13 The genome sequence of a novel virulent bacteriophage termed C130_2 that is morphologically a member of the

14 family Myoviridiae, is reported. The 41,775 basepair double-stranded DNA genome of C130_2 encodes for 59

15 ORFs but exhibits overall low sequence homology to publicly available bacteriophage genomes. Phylogenetic

16 analyses indicates that C130_2 represents a new phage type. C130_2 propagated well on enterohemorrhagic

17 *Escherichia coli* (EHEC) O157:H7 and other pathogenic *E. coli* strains, as well as on strains comprising various

18 *Shigella* species.

19

20

21 Microbial resistance against antimicrobials is an increasing problem. In recent years, researchers' attention have

22 turned toward bacteriophages, the natural predators of bacteria, as alternative agents for use against bacteria,

23 either post-infection as phage therapy, or preventively as bio-control agents. (reviewed in [1] and [2]).

24 Several members of the Enterobacteriaceae bacterial family are important enteral pathogens; many of them cause

25 foodborne infections. In light of the serious problem of multidrug resistance, several members of the genus

26 Enterobacteriaceae, including *Escherichia coli* O157:H7, *Shigella* species and various *Salmonella* serovars have

27 been targeted in studies involving bacteriophages as biocontrol agents. Genomes of phages effective against *E.*

28 *coli* O157 strains, *Shigella* and *Salmonella* representing several groups from the order Caudovirales including

29 T4-like [3], T5-like [4], and rV5-like [5] bacteriophages have been reported. Studies testing the application of

30 phages towards *E. coli* O157:H7 on beef [6], cabbage [7] or on tomato surfaces [8], as well as their stability

31 under various conditions [8] were performed with promising results. Similar experiments were conducted with

32 phages against *Salmonella* on duck meat [9], as well as a phage cocktail against *Shigella* species in various

33 foodstuff [10].

34 In the current study, we present genomic characterisation of a new bacteriophage termed C130_2 isolated from

35 cheese. This phage exhibits broad host specificity and is quite unrelated to any previously characterised

36 bacteriophage. The phage was isolated from a cattle cheese sample from Ukraine in a project aiming to assess

37 the risk posed by illegally imported food in the EU [11]. The phage was isolated by applying the bacterium-free

38 supernatant of a pre-cultured food sample on layered soft agar plates containing *E. coli* K-12 C600 strain [4].

39 The host specificity of the isolated phage was investigated using a spot assay on various enterobacterial strains

40 (Supplementary Table 1).

41 The efficiency of plating (EOP) was determined by applying serial dilutions of phage suspensions employing

42 spot assays. The ratio of phage titre on the various enterobacterial strains (Supplementary Table 1) to the titre

43 measured on *E. coli* K-12 MG1655 was considered as the EOP of the phage on the given strain.

44 The morphology of the phage examined using transmission electron microscopy (TEM). C130_2 revealed a

45 Myoviridae morphology with an approx.75 x 78 nm icosahedral head and a 115 nm long contractile tail (Figure

46 1).

47 Phage DNA was isolated by the phenol-chloroform method [12] with the modifications outlined by Tóth et al

48 [13]. Genomic DNA sequencing libraries were prepared using the Nextera XT kit (Illumina, Eindhoven, NL).

49 Sequencing was performed using Nextseq Mid-output reagent kit v2 (2×150 bp) on an Illumina NextSeq 500.

50   Average read length was 233.39 nucleotides with an average coverage of 93.3%.-Assembly was performed with

51   Spades [14]. The genome was annotated using the RAST server [15]. A search for tRNAs was conducted with

52   tRNAScan-SE [16]. Homology searches were performed with the tools available on the NCBI website, protein

53   sequences of ORFs were investigated with PSI-BLAST, Prosite, and Uniprot databases. Protein masses were

54   predicted with ExPasy using an average resolution setting.

55   The genome sequence of bacteriophage C130_2 was deposited in GenBank and is available under accession no.

56   MH363708. The genome of bacteriophage C130_2 is a 41,775 bp long, linear double-stranded DNA, with a GC

57   content of 55.4%. The terminal repeats determined by a pile-up analysis of the raw reads by mapping of them to

58   the assembled phage genome using CLC genomic workbench (v. 9.5.4, Qiagen, Venlo, Netherlands), are 284

59   nucleotides in length, and located distally at the 5`and 3` ends of the genome from nucleotides 1-284 and 41,492-

60   41,775, respectively.

61   We identified a total of 59 potential protein-coding sequences (CDSs), but no tRNA genes. The list of ORFs

62   detected is provided in Supplementary Table 2. RAST- and PSI-BLAST- based annotations enabled assignment

63   of a function for 35 of 59 genes, with the remaining ORFs annotated as 'hypothetical', 'phage protein' or

64   'unknown' proteins. At the nucleotide –level, the genome does not show strong homology to any other

65   previously sequenced bacteriophage. Whole-genome based phylogenetic relations of phage C130_2 were

66   investigated with VICTOR [17]. This analysis placed IME_EC2 and vB_KpnS_IME279 as its closest neighbors,

67   albeit still too far apart to be considered as close relatives. At the same time it has shown that C130_2 indeed

68   represents a wholly new genotype within bacteriophages representing members of the order Caudovirales

69   (Figure 2).

70   Prosite search detected motifs in only 8 ORFs with the inclusion of high probability occurrence motifs. Four out

71   of these encode structural proteins, and the other four encode DNA modifying enzymes. In many cases the PSI-

72   BLAST and Uniprot searches indicated that the predicted proteins show homology to genes of Enterobacteria

73   phage IME_EC2 (GenBank KF591601.1; [18]) and *Klebsiella* phage vB_KpnS_IME279 (MF614100.1). For the

74   PSI-BLAST hits, CDSs exhibited an average coverage of 93.3% but with a low average homology of 47.4%.

75   The Uniprot hits showed 99% average coverage, and 76% average identity at the amino acid level (see

76   Supplementary Table 2 for details). Interestingly, these two phages are members of different families, as they

77   belong to Podoviridae and Siphoviridae, respectively. When studying the PSI-BLAST and Uniprot search

78   results, it should be noted that except for a major tail protein (locus 130-2_0057) the majority of ORFs bearing

79 similarity to corresponding ORFs in IME_EC2 or vB_KpnS_IME279 code for proteins associated with DNA

80 modification.

81 A blastN-based pairwise comparison analysis was performed for the three phages C130_2, IME_EC2 and

82 vB_KpnS_IME279 using Easyfig 2.1 [19] and visualized using Inkscape (Supplementary Figure 1). This

83 revealed only a few regions where similarity of C130_2 to either of the other phages approaches 75%, and

84 almost never exceeds 80%. The first of these regions contains two genes encoding putative proteins involved in

85 tail assembly (ORFs 14-15). The following region spans five genes encoding DNA modification enzymes,

86 (ORFs 28, 29, 31, 32, 33). These are followed by tail fiber, capsid, portal protein encoding genes and one that

87 encodes the terminase large subunit (Supplementary Table 3) The relatively conserved sequence of these genes

88 suggests their universal importance in the lifecycle of tailed bacteriophages. The order of these regions is the

89 same in all three of the phage genomes, suggesting that their overall genome organisation is colinear

90 (Supplementary Table 3). The rest of the C130_2 genome however, encodes for genes with as of yet unknown

91 functions, which have a low level of similarity (below 50%) to the other two phage genomes, indicating its

92 novelty.

93 C130_2 is capable of lysing *E. coli* K-12, EHEC O157:H7, enteropathogenic (EPEC), enteroinvasive (EIEC) and

94 *Shigella* strains with efficiency of plating (EOP) between approx.10^{-2} to 2×10^{-8} (Supplementary Table 1).

95 The fact that phage C130_2 lyses multiple *Shigella* strains is an important finding, as *Shigellae* are a leading

96 cause of bacillary dysentery [20]. Like other significant pathogens, antibiotic resistance is a rising menace

97 among *Shigella* strains [21], and promising experiments aiming the development of anti-*Shigella* phage cocktails

98 have been performed [10]. For foodborne pathogens, it is desirable that bacteriophages present in the same

99 foodstuff be considered as prime candidates in studies searching for biocontrol agents. Our study demonstrates

100 that so far completely uncharacterised bacteriophages potentially effective against and significant foodborne

101 pathogens are indeed present in the same foodstuff in which their hosts reside.

102 Whole genome sequencing of new bacteriophages may reveal hitherto unknown genes regulating host

103 specificity, as well as those that play key roles in lysis and survival. Detailed knowledge of the host spectrum,

104 stability and efficiency of different phages and the associated genes could help in assembling more effective

105 phage cocktails or even the generation of specifically modified phages to be applied against different arrays of

106 pathogenic bacteria.

107

115

References

2. Abedon ST (2017) Bacteriophage clinical use as antibacterial "drugs": Utility and Precedent. Microbiol Spectr 5: . doi: 10.1128/microbiolspec.BAD-0003-2016

3. Hong Y, Pan Y, Harman NJ, Ebner PD (2014) Complete genome sequences of two *Escherichia coli* O157:H7 phages effective in limiting contamination of food products. Genome Announc 2: . doi: 10.1128/genomeA.00519-14

4. Sváb D, Falgenhauer L, Rohde M, et al (2018) Identification and characterization of T5-like bacteriophages representing two novel subgroups from food products. Front Microbiol 9:202 . doi: 10.3389/fmicb.2018.00202

5. Kim M, Heu S, Ryu S (2014) Complete genome sequence of enterobacteria phage 4MG, a new member of the subgroup "PVP-SE1-like phage" of the "rV5-like viruses." Arch Virol 159:3137–3140 . doi: 10.1007/s00705-014-2140-1

6. Liu H, Niu YD, Meng R, et al (2015) Control of *Escherichia coli* O157 on beef at 37, 22 and 4°C by T5-, T1-, T4-and O1-like bacteriophages. Food Microbiol 51:69–73 . doi: 10.1016/j.fm.2015.05.001

7. Lee H, Ku H-J, Lee D-H, et al (2016) Characterization and genomic study of the novel bacteriophage HY01 infecting both *Escherichia coli* O157:H7 and *Shigella flexneri*: Potential as a biocontrol agent in food. PloS One 11:e0168985 . doi: 10.1371/journal.pone.0168985

8. Ramirez K, Cazarez-Montoya C, Lopez-Moreno HS, Castro-Del Campo N (2018) Bacteriophage cocktail for biocontrol of *Escherichia coli* O157:H7: Stability and potential allergenicity study. PloS One 13:e0195023 . doi: 10.1371/journal.pone.0195023

9. Wang C, Chen Q, Zhang C, et al (2017) Characterization of a broad host-spectrum virulent *Salmonella* bacteriophage fmb-p1 and its application on duck meat. Virus Res 236:14–23 . doi: 10.1016/j.virusres.2017.05.001

10. Soffer N, Woolston J, Li M, et al (2017) Bacteriophage preparation lytic for *Shigella* significantly reduces *Shigella sonnei* contamination in various foods. PloS One 12:e0175256 . doi: 10.1371/journal.pone.0175256

11. Nagy B, Szmolka A, Smole Možina S, et al (2015) Virulence and antimicrobial resistance determinants of verotoxigenic *Escherichia coli* (VTEC) and of multidrug-resistant *E. coli* from foods of animal origin

illegally imported to the EU by flight passengers. Int J Food Microbiol 209:52–59 . doi: 10.1016/j.ijfoodmicro.2015.06.026

12. Sambrook J, Maniatis T, Fritsch EF, Laboratory CSH (1987) Molecular cloning : a laboratory manual, 2nd ed. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press

13. Tóth I, Sváb D, Bálint B, et al (2016) Comparative analysis of the Shiga toxin converting bacteriophage first detected in *Shigella sonnei*. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 37:150–157 . doi: 10.1016/j.meegid.2015.11.022

14. Bankevich A, Nurk S, Antipov D, et al (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477 . doi: 10.1089/cmb.2012.0021

15. Overbeek R, Olson R, Pusch GD, et al (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42:D206-214 . doi: 10.1093/nar/gkt1226

16. Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54-57 . doi: 10.1093/nar/gkw413

17. Meier-Kolthoff JP, Göker M (2017) VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33:3396–3404 . doi: 10.1093/bioinformatics/btx440

18. Hua Y, An X, Pei G, et al (2014) Characterization of the morphology and genome of an *Escherichia coli* podovirus. Arch Virol 159:3249–3256 . doi: 10.1007/s00705-014-2189-x

19. Sullivan MJ, Petty NK, Beatson SA (2011) Easyfig: a genome comparison visualizer. Bioinforma Oxf Engl 27:1009–10 . doi: 10.1093/bioinformatics/btr039

20. Anderson M, Sansonetti PJ, Marteyn BS (2016) Shigella diversity and changing landscape: Insights for the twenty-first century. Front Cell Infect Microbiol 6:45 . doi: 10.3389/fcimb.2016.00045

21. Nüesch-Inderbinen M, Heini N, Zurfluh K, et al (2016) *Shigella* antimicrobial drug resistance mechanisms, 2004-2014. Emerg Infect Dis 22:1083–1085 . doi: 10.3201/eid2206.152088

22. Blattner FR (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1462 . doi: 10.1126/science.277.5331.1453

23. Hayashi T (2001) Complete genome sequence of enterohemorrhagic *Eschelichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res 8:11–22 . doi: 10.1093/dnares/8.1.11

24. Perna NT, Plunkett G, Burland V, et al (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529–533 . doi: 10.1038/35054089

25. Marchès O, Ledger TN, Boury M, et al (2003) Enteropathogenic and enterohaemorrhagic *Escherichia coli* deliver a novel effector called Cif, which blocks cell cycle G2/M transition. Mol Microbiol 50:1553–1567

26. Sváb D, Bálint B, Maróti G, Tóth I (2016) Cytolethal distending toxin producing *Escherichia coli* O157:H43 strain T22 represents a novel evolutionary lineage within the O157 serogroup. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. doi: 10.1016/j.meegid.2016.11.003

27. Iguchi A, Thomson NR, Ogura Y, et al (2009) Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J Bacteriol 191:347–354 . doi: 10.1128/JB.01238-08

28. Hochhut B, Wilde C, Balling G, et al (2006) Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. Mol Microbiol 61:584–595 . doi: 10.1111/j.1365-2958.2006.05255.x

184   29.   Moriel DG, Bertoldi I, Spagnuolo A, et al (2010) Identification of protective and broadly conserved
185   vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. Proc Natl Acad Sci U S
186   A 107:9072–9077 . doi: 10.1073/pnas.0915077107

187   30.   Tóth I, Nougayrède J-P, Dobrindt U, et al (2009) Cytolethal distending toxin type I and type IV genes are
188   framed with lambdoid prophage genes in extraintestinal pathogenic *Escherichia coli*. Infect Immun
189   77:492–500 . doi: 10.1128/IAI.00962-08

190   31.   Allué-Guardia A, García-Aljaro C, Muniesa M (2011) Bacteriophage-encoding cytolethal distending toxin
191   type V gene induced from nonclinical *Escherichia coli* isolates. Infect Immun 79:3262–3272 . doi:
192   10.1128/IAI.05071-11

193   32.   Sváb D, Bálint B, Vásárhelyi B, et al (2017) Comparative genomic and phylogenetic analysis of a Shiga
194   toxin producing *Shigella sonnei* (STSS) strain. Front Cell Infect Microbiol 7:229 . doi:
195   10.3389/fcimb.2017.00229

196   33.   Petty NK, Bulgin R, Crepin VF, et al (2010) The *Citrobacter rodentium* genome sequence reveals
197   convergent evolution with human pathogenic *Escherichia coli*. J Bacteriol 192:525–538 . doi:
198   10.1128/JB.01144-09

199

200

201 **Legend to the figures**

202 **Figure 1. Transmission electron micrograph of bacteriophage 130/2 showing Myoviridae morphology with**

203 **contracted (A) and non-contracted (B) tail structure.**

204

205 **Figure 2. Whole-genome based phylogeny of bacteriophage C130_2 prepared with VICTOR, comparing it**

206 **to representative members of Caudovirales, as well as bacteriophages IME_EC2 and vB_KpnS_IME279.**

207 The GenBank accession numbers of phage genomes and type designations of the phages are indicated next to the

208 branches. In the case of IME_EC2 and vB_KpnS_IME279, the phage families are indicated.

209

210 **Supplementary Table 1.**

211 Host spectrum and efficiency of plating (EOP) of bacteriophage C130_2. EOP values are given relative to the

212 titre measured on *E. coli* K-12 MG1655 strain.

213 **Supplementary Table 2.**

214 List of ORFs of phage C130_2 with assigned functions and protein homology searches. Prosite search was

215 performed including motifs with high probability occurrences, Uniprot search was performed with narrowing

216 down to viral proteins.

217   **Supplementary Table 3.**

218   List of ORFs of phages C130_2, IME_EC2 and vB_KpnS_IME279, with corresponding ORFs above 75%

219   similarity highlighted in blue.

220   **Supplementary Figure 1.**

221   BLAST-based comparison of the whole genomes of bacteriophages C130_2, IME_EC2 and vB_KpnS_IME279.

222   Orange arrows represent genes, numbers on C130_2 genes correspond to ORF numbers in Supplementary Tables

223   2 and 3. Regions showing >50% similarity are interconnected with grey lines.