Check for updates

GENOME NOTE

Isolation and sequencing of three RB49-like bacteriophages infecting O antigen-producing *E. coli* strains [version 1; peer review: 1 approved, 1 approved with reservations]

Alexander Efimov¹, Eugene Kulikov ¹, Alla Golomidova¹, Ilya Belalov¹, Vladislav Babenko², Andrey Letarov¹

¹Microbial viruses lab, FRC Biotechnology, Russian Academy of Sciences, Moscow, Russian Federation

²Genomic research and computational biology lab, FSCC of physico-chemical medicine, Federal Medical-Biological Agency, Moscow, Russian Federation

V1 First published: 03 Nov 2021, 10:1113
<https://doi.org/10.12688/f1000research.74169.1>

Latest published: 03 Nov 2021, 10:1113
<https://doi.org/10.12688/f1000research.74169.1>

Abstract

E. coli strains 4s, F5 and F17, whose O antigens are structurally characterized and shown to effectively shield the cell surface from bacteriophage attack, were used as the hosts to isolate novel RB49-like bacteriophages. Three novel phage isolates were obtained, and their genomes were sequenced and annotated. Despite high overall identity levels of these genomic sequences, the variants of large distal tail fiber subunit, orthologous to the bacteriophage T2 long tail receptor recognition protein gp38, were unique for each phage, suggesting their role in host range determination. The annotated genomes are available via NCBI Genbank, acc. numbers MZ504876-MZ504878.

Keywords

Bacteriophage genomes, RB49-like bacteriophages, coliphages, O antigens, receptor recognition proteins.

Open Peer Review

Approval Status

	1	2
version 1 03 Nov 2021	view	view

1. **Marcin Los** , University of Gdańsk, Gdansk, Poland
2. **Mikael Skurnik** , University of Helsinki, Helsinki, Finland

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Pathogens** gateway.

This article is included in the **Genomics and Genetics** gateway.

This article is included in the **Cell & Molecular Biology** gateway.

Corresponding authors: Alexander Efimov (efsasha@bk.ru), Eugene Kulikov (eumenius@gmail.com), Andrey Letarov (letarov@gmail.com)

Author roles: **Efimov A:** Conceptualization, Formal Analysis, Investigation; **Kulikov E:** Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Golomidova A:** Conceptualization, Formal Analysis, Investigation, Methodology, Validation; **Belalov I:** Resources, Software, Supervision, Validation, Visualization; **Babenko V:** Investigation, Software, Validation; **Letarov A:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The research was funded by the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement № 075-15-2020-907 of 16.11.2020 on providing a grant in the form of subsidies from the Federal budget of Russian Federation to FRC Biotechnology RAS. The grant was provided for state support for the creation and development of a World-class Scientific Center “Agrotechnologies for Future”.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Efimov A *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Efimov A, Kulikov E, Golomidova A *et al.* **Isolation and sequencing of three RB49-like bacteriophages infecting O antigen-producing *E. coli* strains [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2021, **10**:1113 <https://doi.org/10.12688/f1000research.74169.1>

First published: 03 Nov 2021, **10**:1113 <https://doi.org/10.12688/f1000research.74169.1>

Introduction

The threat of rapidly spreading microbial resistance to antibacterial drugs has revitalized the interest to the usage of bacteriophages as biocontrol agents for agriculture (Allen *et al.* 2013), aquaculture (D'Accolti *et al.* 2021), phage therapy in humans and animals (Abedon 2018, Clokie and Mary Ann Liebert Inc. 2020) and other applications.

One of the main limitations of phage-based biocontrol technologies is narrow host range of most of bacteriophages. T-even-related bacteriophages, including RB49-like viruses, are widely spread in nature and feature significant sequence flexibility in the receptor recognition proteins (RBPs), coupled with an high overall homology of the other regions in their genomes (Comeau *et al.* 2007). This makes them attractive platforms for developing systems of artificial host range management.

The paradigm of the host cell recognition by T-even-like phages was studied in great detail on the bacteriophage T4 model system. The infection of the host cell by T4 is initiated by binding of the tips of the phage long tail fibers (LTFs) to their cognate receptor molecules (OmpC) (Hu *et al.* 2015). This initial binding triggers the baseplate rearrangement and deployment of the short tail fibers (Taylor *et al.* 2016) that bind to different receptors tightly fixing the virion on the cell surface to prevent it from been pushed out by the contraction of the phage tail.

The bioinformatic and experimental evidence suggest that in other T-even-like phages the outline of the early stages of infection is essentially the same as in T4, however in most of these viruses (including RB49 and related phages) the recognition of LTF receptors is performed not by the C-terminal domain of the trimeric gp37 protein forming the distal moiety of the LTF, but by a tiny monomeric protein gp38 attached to the end of the gp37 trimer (Trojet *et al.* 2011, Hyman and van Raaij 2018).

Interestingly, the above-described host recognition scenario was mainly evaluated by studying the phage T4 infection of the laboratory *E. coli* strains that are depleted of the O antigen biosynthesis, such are the derivatives of K-12 or B strains (Tetart *et al.* 2001). At the same time evidence accumulates that many types of *E. coli* O antigen serve as an effective shields restricting the access of the phage receptor-recognition structures to the receptors, located at the outer membrane surface (Peng *et al.* 2007, Golomidova *et al.* 2021) (Broeker and Barbirz 2017). The penetration through this shield requires specific recognition and, in some cases, enzymatic degradation of the O polysaccharide, by the phage RBPs (Letarov and Kulikov 2017, Kulikov *et al.* 2019). In order to evaluate the mechanisms employed by T-even-like bacteriophages to attack the O antigen-producing *E. coli* strains it is necessary to obtain such phages infective against the host strains in which the protective features of the O antigens were studied in some detail. We previously investigated the structure and protectivity of the O antigens of several wild type isolates of *E. coli*, including the strains 4s (Knirel *et al.* 2015, Kulikov *et al.* 2017), F5 (Golomidova *et al.* 2019), and F17 (Knirel *et al.* 2019). Here we report the isolation and sequencing of three RB49-like bacteriophages able to infect these strains. Novel viruses with high degree of overall similarity but featuring different RBP variants serve as useful model systems for in-depth analysis of T-even phage strategies used to circumvent O antigen protection of the host cell, efficiently shielded otherwise.

Methods

Three previously characterized *E. coli* strains featuring different O antigen types were used for bacteriophage isolation. These were: O28 strain F5 (Golomidova *et al.* 2019), *E. coli* F17, producing a novel O antigen type (Knirel *et al.* 2019) and O22 strain 4s (Knirel *et al.* 2015). To prepare concentrated phage stocks the the relevant host culture was propagated at 37°C with agitation up to OD600 = 0.2, and inoculated with phage at multiplicity of infection (moi) = 0.1. The culture was incubated under the same conditions until the lysis was observed, then 0.05 volume of chlorophorm was added and the lysates were cleared by centrifugation at 10 000 g for 10 min.

The bacteriophages were isolated by direct plating of the commercial therapeutical bacteriophage preparation “Intesti-bakteriofag” (Mikrogen, Russia) that is marketed as a broad range therapeutic phage cocktail (active against *Shigella flexneri*, *S. sonnei*, *S. paratyphi A*, *S. paratyphi B*, *Salmonella typhimurium*, *S. infantis*, *S. choleraesuis*, *S. oranienburg*, *S. enteritidis*, *E. coli*, *P. vulgaris*, *P. mirabilis*, *Enterococcus sp.*, *Staphylococcus*, *P. aeruginosa*) (<https://www.mikrogen.ru/products/bakteriofagi/intesti-bakteriofag/>). This preparation was titrated to single plaques by the conventional double-layer plating technique using three above-mentiones *E. coli* strains as the hosts. All the methods for phage isolation were as described in Kulikova *et al.* (2007). PCR assay with RB49 g38 specific primers cttgctggatgag ccaattcgccggagcgtctaaagagattcattatagca-gp38C(RB49)-F and ccttttaagggtatttattcacattacgcgcgagcaccgtaga-gp38C(RB49)-R was used to pick RB49-like phage candidates for further study.

Transmission electron microscopy study was performed using uranyl acetate negative contrast as described in (Kulikov *et al.* 2014).

To extract the DNA the lysates were treated with DNase (0.01 mg mL^{-1}) for 30 min at room temperature and the phages were collected by ultracentrifugation (1 h, 75000 g). Viral genomic DNA was extracted from the pellets with CTAB (cetyltrimethylammonium bromide) as described in (Kulikov *et al.* 2020) using. DNA quality and quantity was assessed by agarose gel analysis and Qubit dsDNA HS fluorometer assay (Qubit, USA). The phage genomic DNA libraries were prepared and sequenced using an Ion Torrent sequencer system (Applied Biosystems, USA) with 400-fold coverage and a median read length of 185 bp. The raw reads from the run were then combined and filtered using the spectral alignment error correction tool SAET 3 (Pevzner *et al.* 2001). This yielded 157,407-163,568 reads per library with average 190-fold coverage. Primary assembly was conducted with Newbler version 3.0 (Roche Diagnostics, USA), resulting in a single contigs.

Annotation was performed using Prokka (Seemann 2014) with further manual curation. Potential open reading frames (ORFs) were detected by use of SnapGene (GSL Biotech, USA) and subsequently analyzed with HMMER, HHPRED (Soding *et al.* 2005) (MPI Bioinformatics Toolkit), BLAST (Altschul *et al.* 1990), tRNAscan-SE (Lowe and Eddy 1997).

A subset of 25 highly similar RB49-like phages using with blastN search with RB49 genome sequence as a query was retrieved from GenBank. The thresholds used were 93% of the query coverage and the average nucleotide identity over aligned segments 92%. Full genome nucleotide alignment and phylogenetic tree reconstruction were performed with NGPhylogeny.fr web services. The genome of RB49 and the bacteriophage genomes obtained in this work were aligned and visualized with Clinker v0.0.21 and clustermap.js (Gilchrist and Chooi 2021).

Results

The direct plating of the commercial therapeutical phage mixture on the lawns of the *E. coli* strains 4s, F5 and F17 yielded phage plaques. PCR screening with the primers hybridizing to bacteriophage RB49 g38 sequence was performed and about half of the plaques were found positive. One PCR-positive phage isolate per host strain was selected and purified by repeated single plaque isolation. The obtained bacteriophage strains were named as follows: F5-derived phage – Cognac49, F17-derived phage – Whisky49 and 4s-derived phage – Brandy49. The concentrated phage stocks were grown and analyzed by transmission electron microscopy. The bacteriophages were morphologically undistinguishable (Figure 1) having typical appearance of T-even-like viruses (Ackermann and Krisch 1997, Comeau *et al.* 2007) that was compatible with their expected relatedness to the phage RB49.

The genomic DNA was extracted from the phage stocks and sequenced using the Ion Torrent technology. The assembly procedure yielded a single contig for each of the bacteriophages with the coverage level > 190. Since T-even related phages feature unlimited circular permutations (Comeau *et al.* 2007), the start of each genomes was set arbitrary according to the standard representation of the RB49 genome. The main features of the genomes assembled are listed in the Table 1.

Gene functions were predicted based on the existing annotations of RB49-like phages producing the best blastN hits for each of the genomes for Brandy49: Phi1 (nucleotide identity 95.77%), vB_EcoM_011D4 (95.57%), kaaroo (96.86%);

Figure 1. Virion morphology of the phages Cognac49, Whisky49 and Brandy49 as revealed by transmission electron microscopy (1% uranyl acetate in methanol as a negative stain).

Table 1. Genomic features of the bacteriophages under study.

Phage	Genome length, bp	Number of ORFs annotated	Average GC content, %	GenBank accession number
Brandy49	169,233	279	41	MZ504876
Cognac49	164,033	271	40	MZ504877
Whisky49	165,617	278	40	MZ504878

Figure 2. Schematic alignment of the genomes of the phages RB49, Cognac49, Whisky49 and Brandy49. The grayscale indicates the level of the amino acid identity. The ORF colors reflect the groups of homology as revealed by Clinker software.

for Cognac49: Sf20 (98.04%), RB49 (97.59%), KFS-EC (97.79%); for Whisky49: vB_EcoM_011D4 (97.76%), Sf20 (98.38%), GEC-3S (98.17%); query cover >95% for all comparisons. The annotations were then manually curated.

Scan by tRNAscan-SE detected no tRNAs in phage genomes.

The comparison of three genomes to the bacteriophage RB49 (Figure 2) revealed that 42 ORFs shared identical aa sequences in all four genomes, another 140 ORFs were also very similar with the aa identity ≥ 0.90 . The phages share also 71 more distantly related ORFs with aa identity levels ≥ 0.50 but ≤ 0.90 . This result indicates that the isolated bacteriophages are closely related to each other.

About 56 ORFs making the difference between these three phages (Figure 2), mostly encode for hypothetical proteins with unknown function. Not all RB49 ORFs or their homologues are present in all phages. A trace of recombination event was detected in genome of Cognac49 phage. An integration of foreign genetic material occurred within DNA polymerase gene, which is essential for phage viability. This insertion contains an 825 bp ORF predicted to be an HNH homing nuclease, often present within self-splicing introns. Rfam search (Kalvari *et al.* 2021) yielded no significant hits, while the analysis of putative secondary structures of corresponding RNA revealed a complex pattern of the stem-loop structures suggesting that insertion could be a bacterial type I intron containing homing nuclease. Ten ORFs with unknown function present in RB49, Brandy49 and Whisky49 are missing in genome of Cognac49, spanning the region homologous to NP_891764.1-NP_891775.1 in RB49.

Figure 3. Phylogeny of whole genomic nucleotide sequences of a subset of RB49-like bacteriophages.

High similarity of the studied phages to RB49 allows us to classify them as follows: Realm: Duplodnaviria, Kingdom: Heunggongvirae, Phylum: Uroviricota, Class: Caudoviricetes, Order: Caudovirales, Family: Myoviridae, Subfamily: Tevenvirinae, Genus: Krischvirus; *Escherichia* virus RB49 type viruses, just as expected from g38 PCR and TEM results. The genomes of 25 bacteriophages closely related to RB49 were extracted from GenBank. Overall nucleotide phylogeny tree of this subset of the phage genomes with the addition of three genomes sequenced during this work is shown on the Figure 3.

Despite their close relatedness, the isolated bacteriophages showed marked differences in their host specificity. When they were cross-tested for the infectivity on all three host strains used, all three phages were found infective towards F5 strain, however, phage Whisky49 could form plaque on both *E. coli* F5 and F17 strains, while Brandy49 had the broadest host range infecting all three host strains (been the only one forming plaques on the *E. coli* 4s lawns). In all the cases the efficiency of plating was similar that observed on the respective isolation host strain.

The specificity of host cell recognition by most of the T-even-like phages is determined by two proteins – the LTF adhesin gp38 and by the short tail fiber protein gp12 (PMID: 21746838). The comparison of gp12 aa sequences from the bacteriophages under the study revealed them to be nearly identical to gp12 from RB49 phage (data not shown). At the same time the sequences of gp38 were found more diverged (Figure 4). Gp38 from the phage Brandy49 was more distant from the consensus sequence compared to other two viruses (Figure 4) that is correlated with a broader host range of this phage. Some of the aa polymorphisms were observed within the loops located between polyglycine stretches and predicted to be responsible for interactions with the receptor (Trojet *et al.* 2011, Dunne *et al.* 2018). Therefore, differences in the particular regions of gp38 sequence may explain different host specificities of these bacteriophages.

The characterization of three new viruses that been closely related to each other show distinct patterns of the infectivity towards *E. coli* strain producing different types of the O antigen provides a useful model system for in-depth analysis of the strategies employed by T-even like phages to infect the host cells featuring effective non-specific protection of the outer membrane surfaces by this polysaccharide structure. The close relatedness of these phages to each other will

Figure 4. Alignment of the aa sequences of gp38 LTF receptor recognition proteins. The aa residues that differ from the consensus are highlighted in red. Blue bars indicate polyglycine stretches and yellow bars the loops according to the structural data on phage S16 gp38 (Dunne *et al.* 2018).

facilitate the comparison of the results of experiments aiming at deciphering fine mechanisms employed by RB49-like phages to penetrate the O antigen shields of different *E. coli* strains.

Data availability

Repository: NCBI Nucleotide database

Accession number: MZ504876 for Brandy49

Root URL: <https://www.ncbi.nlm.nih.gov/nuccore>

Accession number URL: <https://www.ncbi.nlm.nih.gov/nuccore/MZ504876>

Accession number: MZ504877 for Cognac49

Root URL: <https://www.ncbi.nlm.nih.gov/nuccore>

Accession number URL: <https://www.ncbi.nlm.nih.gov/nuccore/MZ504877>

Accession number: MZ504878 for Whisky49

Root URL: <https://www.ncbi.nlm.nih.gov/nuccore>

Accession number URL: <https://www.ncbi.nlm.nih.gov/nuccore/MZ504878>

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

References

- Abedon ST: **Phage Therapy: Various Perspectives on How to Improve the Art.** *Host-Pathogen Interactions: Methods and Protocols*. 2018; **1734**: 113–127.
[Publisher Full Text](#)
- Ackermann HW, Krusch HM: **A catalogue of T4-type bacteriophages.** *Arch. Virol.* 1997; **142**(12): 2329–2345.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Allen HK, Levine UY, Looft T, et al.: **Treatment, promotion, commotion: antibiotic alternatives in food-producing animals.** *Trends Microbiol.* 2013; **21**(3): 114–119.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Altschul SF, Gish W, Miller W, et al.: **Basic local alignment search tool.** *J. Mol. Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Broeker NK, Barbirz S: **Not a barrier but a key: How bacteriophages exploit host's O-antigen as an essential receptor to initiate infection.** *Mol. Microbiol.* 2017; **105**(3): 353–357.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Clokie MRJ, Mary Ann Liebert Inc: *PHAGE: therapy, applications, and research*. New Rochelle, NY: Mary Ann Liebert, Inc; 2020.
- Comeau AM, Bertrand C, Letarov A, et al.: **Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery.** *Virology*. 2007; **362**(2): 384–396.
[PubMed Abstract](#) | [Publisher Full Text](#)
- D'Accolti M, Soffritti I, Mazzacane S, et al.: **Bacteriophages as a Potential 360-Degree Pathogen Control Strategy.** *Microorganisms*. 2021; **9**(2):
[Publisher Full Text](#)
- Dunne M, Denyes JM, Arndt H, et al.: **Salmonella Phage S16 Tail Fiber Adhesin Features a Rare Polyglycine Rich Domain for Host Recognition.** *Structure*. 2018; **26**(12): 1573–1582.e4.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Golomidova AK, Efimov AD, Kulikov EE, et al.: **O antigen restricts lysogenization of non-O157 Escherichia coli strains by Stx-converting bacteriophage phi24B.** *Sci. Rep.* 2021; **11**(1): 3035.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Golomidova AK, Naumenko OI, Senchenkova SN, et al.: **The O-polysaccharide of Escherichia coli F5, which is structurally related to that of E. coli O28ab, provides cells only weak protection against bacteriophage attack.** *Arch. Virol.* 2019; **164**(11): 2783–2787.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hu B, Margolin W, Molineux IJ, et al.: **Structural remodeling of bacteriophage T4 and host membranes during infection initiation.** *Proc. Natl. Acad. Sci. U. S. A.* 2015; **112**(35): E4919–E4928.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hyman P, van Raaij M: **Bacteriophage T4 long tail fiber domains.** *Biophys. Rev.* 2018; **10**(2): 463–471.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kalvari I, Nawrocki EP, Ontiveros-Palacios N, et al.: **Rfam 14: expanded coverage of metagenomic, viral and microRNA families.** *Nucleic Acids Res.* 2021; **49**(D1): D192–D200.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Knirel YA, Ivanov PA, Senchenkova SN, et al.: **Structure and gene cluster of the O antigen of Escherichia coli F17, a candidate for a new O-serogroup.** *Int. J. Biol. Macromol.* 2019; **124**: 389–395.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Knirel YA, Prokhorov NS, Shashkov AS, et al.: **Variations in O-antigen biosynthesis and O-acetylation associated with altered phage sensitivity in Escherichia coli 4s.** *J. Bacteriol.* 2015; **197**(5): 905–912.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kulikov E, Golomidova A, Babenko V, et al.: **A Simple Method for Extraction of the Horse Feces Virome DNA, Suitable for Oxford Nanopore Sequencing.** *Microbiology*. 2020; **89**(2): 246–249.
[Publisher Full Text](#)
- Kulikov EE, Golomidova AK, Letarova MA, et al.: **Genomic sequencing and biological characteristics of a novel Escherichia coli bacteriophage 9g, a putative representative of a new Siphoviridae genus.** *Viruses*. 2014; **6**(12): 5077–5092.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kulikov EE, Golomidova AK, Prokhorov NS, et al.: **High-throughput LPS profiling as a tool for revealing of bacteriophage infection strategies.**

Sci. Rep. 2019; **9**(1): 2958.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kulikov EE, Majewska J, Prokhorov NS, *et al.*: **Effect of O-acetylation of O antigen of Escherichia coli lipopolysaccharide on the nonspecific barrier function of the outer membrane.** *Microbiology.* 2017; **86**(3): 310–316.

[Publisher Full Text](#)

Kulikova EE, Isaeva AS, Rotkina AS, *et al.*: **Diversity and dynamics of bacteriophages in horse feces.** *Mikrobiologiya.* 2007; **76**(2): 271–278.

[PubMed Abstract](#) | [Publisher Full Text](#)

Letarov AV, Kulikov EE: **Adsorption of Bacteriophages on Bacterial Cells.** *Biochemistry (Mosc).* 2017; **82**(13): 1632–1658.

[PubMed Abstract](#) | [Publisher Full Text](#)

Lowe TM, Eddy SR: **tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res.* 1997; **25**(5): 955–964.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Peng D, Hu WG, Choudhury BP, *et al.*: **Role of different moieties from the lipooligosaccharide molecule in biological activities of the Moraxella catarrhalis outer membrane.** *FEBS J.* 2007; **274**(20): 5350–5359.

[Publisher Full Text](#)

Pevzner PA, Tang H, Waterman MS: **An Eulerian path approach to DNA fragment assembly.** *Proc. Natl. Acad. Sci. U. S. A.* 2001; **98**(17): 9748–9753.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Seemann T: **Prokka: rapid prokaryotic genome annotation.** *Bioinformatics.* 2014; **30**(14): 2068–2069.

[PubMed Abstract](#) | [Publisher Full Text](#)

Soding J, Biegert A, Lupas AN: **The HHpred interactive server for protein homology detection and structure prediction.** *Nucleic Acids Res.* 2005; **33**(15980461): W244–W248.

[Publisher Full Text](#)

Taylor NM, Prokhorov NS, Guerrero-Ferreira RC, *et al.*: **Structure of the T4 baseplate and its function in triggering sheath contraction.** *Nature.* 2016; **533**(7603): 346–352.

[Publisher Full Text](#)

Tetart F, Desplats C, Kutateladze M, *et al.*: **Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages.** *J. Bacteriol.* 2001; **183**(1): 358–366.

[Publisher Full Text](#)

Trojet SN, Caumont-Sarcos A, Perrody E, *et al.*: **The gp38 adhesins of the T4 superfamily: a complex modular determinant of the phage's host specificity.** *Genome Biol. Evol.* 2011; **3**: 674–686.

[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 25 May 2022

<https://doi.org/10.5256/f1000research.77892.r138503>

© 2022 Skurnik M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mikael Skurnik

Department of Bacteriology and Immunology, University of Helsinki, Helsinki, Finland

The manuscript of Efimov *et al.* describes three new phages isolated against three *E. coli* strains expressing structurally different LPS O-antigens. While I find the manuscript interesting, I have a few major and many minor comments that I think the authors could consider:

Major comments

- It would be nice to have a picture of the O-unit structures of the three LPSs, just to see how they differ structurally.
- p.4: "Potential open reading frames (ORFs) were detected by use of SnapGene (GSL Biotech, USA) and subsequently analyzed with HMMER, HHPRED (Soding *et al.* 2005) (MPI Bioinformatics Toolkit), BLAST (Altschul *et al.* 1990), tRNAscan-SE (Lowe and Eddy 1997)." When describing the annotation of the genomes, I personally prefer using "predicted genes" or "predicted gene products" instead of ORFs. The genomes are full of ORFs in all six reading frames but only a small fraction of them contain potential genes. In addition, the predicted gene products were not analyzed with tRNAscan-SE, but the phage genomes were.
- The authors should discuss what role the different O-antigens play, if any, for the three phages. Is the LTF receptor an OMP, or is it an LPS structure? If it is OmpC or OmpF (as we identified for three *Yersinia* phages¹), for example, would it be possible to compare the sequences of these proteins of the three *E. coli* strains?

Minor comments

Page 3

- Chlorophorm ==> chloroform
- OD600 ==> OD₆₀₀
- Check the usage of italics for bacterial names, specifically for *E. coli*.

Page 4

- The sentence, "*Viral genomic DNA was extracted from the pellets with CTAB (cetyltrimethylammonium bromide) as described in (Kulikov et al. 2020) using.*" is not finished.
- The sentence, "*A subset of 25 highly similar RB49-like phages using with blastN search with RB49 genome sequence as a query was retrieved from GenBank.*" needs attention. Perhaps like this: "A subset of 25 highly similar RB49-like phages identified with blastN search using the RB49 genome sequence as a query was retrieved from GenBank."
- Also, the next sentence is clumsy: "*The thresholds used were 93% of the query coverage and the average nucleotide identity over aligned segments 92%.*" Perhaps like this: "A threshold of 93% was used for the query coverage, and 92% for the average nucleotide identity over the aligned segments."
- "*Gene functions were predicted based on...*", perhaps like this: "The functions of the gene products were predicted based on..."

Page 5

- Table 1 column: "*Number of ORFs annotated*" ==> Number of predicted genes.
- Figure 2 caption: "*The ORF colors*"==> The colors of the predicted genes.

Page 6

- query cover ==> query coverage
- "*Scan by tRNAscan-SE detected no tRNAs in phage genomes.*" ==> tRNAscan-SE detected no tRNAs encoding genes in phage genomes.
- This text is problematic: "*The comparison of three genomes to the bacteriophage RB49 (Figure 2) revealed that 42 ORFs shared identical aa sequences in all four genomes, another 140 ORFs were also very similar with the aa identity ≥ 0.90 . The phages share also 71 more distantly related ORFs with aa identity levels ≥ 0.50 but ≤ 0.90 . This result indicates that the isolated bacteriophages are closely related to each other.*" If you compare genomes, you use nucleotide sequences, if you compare predicted gene products, you use amino acid sequences. ORFs ==> predicted gene products. aa identity ==> aa sequence identity. Perhaps it would be more clear to use identity percentages instead of a 0-1 scale.
- "*About 56 ORFs making the difference between these three phages (Figure 2), mostly encode...*" ==> "The three phages differed based on the presence or absence of altogether 56 predicted genes (Figure 2), most of which encode..."

Page 7

- were extracted from GenBank ==> were downloaded from GenBank
- efficiency of plating was similar that observed ==> efficiency of plating was similar to that observed
- Some of the aa polymorphisms ==> Some of the aa sequence polymorphisms
- The sentence, "*The characterization of three new viruses that been closely related to each other show distinct patterns of the infectivity towards E. coli strain producing different types of the O*"

antigen provides a useful model system for in-depth analysis of the strategies employed by T-even like phages to infect the host cells featuring effective non-specific protection of the outer membrane surfaces by this polysaccharide structure." is far too long and confusing. Better to split it into 2-3 shorter sentences.

- Figure 4 caption. "*Blue bars indicate polyglycine stretches and yellow bars the loops*". The colors are mixed up: blue indicates the loops and yellow the polyglycines.

References

1. Salem M, Pajunen M, Jun J, Skurnik M: T4-like Bacteriophages Isolated from Pig Stools Infect *Yersinia pseudotuberculosis* and *Yersinia pestis* Using LPS and OmpF as Receptors. *Viruses*. 2021; **13** (2). [Publisher Full Text](#)

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular microbiology, bacterial genetics, bacteriophages, phage therapy

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 16 March 2022

<https://doi.org/10.5256/f1000research.77892.r114759>

© 2022 Los M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Marcin Los

Department of Molecular Biology, University of Gdańsk, Gdansk, Poland

The paper describes the sequencing and analysis of three T4-like phages capable of degradation of O-antigens of *E. coli*. The degradation of O-antigen and use of surface receptors by phage may play a crucial role in the host range of T4-like phages and their ability to be potential components of phage cocktails for phage therapy. The comparison of the sequences of isolated phages shed a light on their potential mechanisms involved in the broadening of host range among this group of phages.

The manuscript is well written and the data is clear. Also, the interpretation of the data seems to be justified. I have a request for minor revision - since Brandy49 had the broadest host range, but also it contained additional genes not present in other phages studied, can you please explain if the function of these additional genes can be predicted and if these genes can play a role in broader host range?

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bacteriophage biology, phage-bacterial host interaction, prophage induction

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research