scientific reports

Check for updates

OPEN

# Isolation, characterization, and receptor-binding protein specificity of phages PAS7, PAS59 and PAS61 infecting Shiga toxin-producing *Escherichia coli* O103 and O146

Célia Pas[1], Lars Fieseler[2], Joël F. Pothier[3] & Yves Briers[1]✉

**Shiga toxin-producing *Escherichia coli* (STEC) is a foodborne pathogen with 6,534 annual reported cases in the EU in 2021. This pathotype generally contains strains with smooth LPS with O-antigen serogroup O157 being the predominant serogroup in the US. However, non-O157 STEC serogroups are becoming increasingly prevalent. Here we announce the complete genomes of three newly isolated phages that infect STEC serogroups O103 and O146, namely Escherichia phages vB_EcoP_PAS7, vB_EcoP_PAS59 and vB_EcoP_PAS61. The genome sequences revealed that they belong to three distinct genera, namely the newly proposed genus *Cepavirus* within the *Slopekvirinae* subfamily, the genus *Suseptimavirus* and the genus *Uetakevirus*, respectively. We identified the tailspikes of phages PAS7 and PAS61 as a primary specificity determinant for the O-antigens O103 and O146, respectively, and predicted their active site in silico.**

**Keywords** Bacteriophage, Shiga toxin-producing *Escherichia coli* (STEC), Enterohemorrhagic *Escherichia coli* (EHEC), Phage isolation, O103, O146

Phages, viruses that infect bacteria, are featured by a variable specificity. While some phages have a relatively broad range infecting multiple or a majority of strains of a bacterial species, others are specific at the strain or serogroup level. Among others, particularly *Escherichia coli* infecting phages often exhibit such narrow-range specificity. *E. coli* has various pathotypes, each characterized by the presence of virulence factors and mechanisms that contribute to pathogenicity. Important pathotypes include enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC)/Shiga toxin-producing *E. coli* (STEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC) and diffusively adherent *E. coli* (DAEC)[1]. STEC are known for producing Shiga toxins, including STx1 and STx2, which can lead to the development of severe human diseases such as haemorrhagic colitis and haemolytic-uraemic syndrome (HUS)[2]. EHEC is an intestinal pathogenic *E. coli* and a subgroup of STEC able to cause additional symptoms such as bloody diarrhea[3,4]. EHEC/STEC strains are commonly associated with outbreaks of foodborne illnesses, particularly from contaminated beef and raw vegetables[5]. Among these pathotypes, the EHEC/STEC serotype O157:H7 is recognized as one of the most important zoonotic pathogens, especially in the United States[6]. However, many non-O157 STEC serogroups are causing foodborne infections world-wide, such as the STEC outbreak in Norway in 2006[7] and one in the United States in 2020[8], both strains exhibiting serogroup O103. The most prevalent serogroups in the US and the EU currently are O157, O26, O45, O103, O111, O121, O145 and O146[9–13]. This study describes the isolation and characterization of three highly specific phages infecting *E. coli* serogroups O103 (Escherichia phage vB_EcoP_PAS7) and O146 (Escherichia phages vB_EcoP_PAS59 and vB_EcoP_PAS61) with an additional focus on the encoded receptor-binding proteins (RBPs). A new phage genus *Cepavirus* was proposed with phage PAS7 as its reference phage. Treatment of STEC using antibiotics can often be harmful to the patients, as Shiga toxin production rises after stress induction resulting from the antibiotics[14]. The use of phages, however,

[1]Department of Biotechnology, Ghent University, Valentin Vaerwyckweg 1, Gent 9000, Belgium. [2]Food Microbiology Research Group, Institute of Food and Beverage Innovation, Zurich University of Applied Sciences (ZHAW), Einsiedlerstrasse 35, Wädenswil 8820, Switzerland. [3]Environmental Genomics and Systems Biology Research Group, Institute of Natural Resource Sciences, Zurich University of Applied Sciences (ZHAW), Einsiedlerstrasse 31, Wädenswil 8820, Switzerland. ✉email: yves.briers@ugent.be

has even been shown to decrease Shiga toxin production[15]. Therefore, isolating phages infecting different *E. coli* serogroups is of high importance for developing phage cocktails or phage-based alternatives for STEC treatments. Furthermore, these phage genomes contribute to a better understanding of the highly specific phage-host interactions, necessary to develop targeted antibacterials.

## Results

Seven collected bovine feces samples were screened for the presence of phages through enrichment of the samples with *E. coli* strains of serogroups O103 (strains 4215/4, NVH-848 and P11-2315) and O146 (A11-1675). Three distinct phages were successfully isolated from three out of seven collected samples, one infecting *E. coli* strain 4215/4 of serogroup O103 (phage PAS7) and two infecting *E. coli* strain A11-1675 of serogroup O146 (phages PAS59 and PAS61). Host range tests were conducted (Table 1) for all phages against a local *E. coli* strain collection (Centre for Food Safety and Quality Management of the Zurich University of Applied Sciences (ZHAW)) and one externally obtained strain (strain NVH-848), indicating that all three phages are infecting smooth *E. coli* strains with putative O serogroup specificity.

Phage PAS7 is a lytic phage with *E. coli* strain 4215/4 (serogroup O103) as its host and has a genome size of 42,514 bp. It has a highly variable plaque morphology surrounded by halos variable in size, ranging from 2 to 12 mm diameter (Fig. 1a). The closest relatives of phage PAS7 are Enterobacter phages ENC16 (GenBank accession: OL355133.1)[16] and phiKDA1 (GenBank accession: NC_027980.1) belonging to the *Koutsourovirus* phage genus of the *Slopekvirinae* subfamily (Fig. 2). Those relatives have an intergenomic similarity of 36.0 and 35.4% respectively, with phage PAS7 (Fig. 3). Since genomic similarity with the closest relative phage is lower than 70% a new genus *Cepavirus* within the *Slopekvirinae* subfamily was proposed to the International Committee on Taxonomy of Viruses (ICTV) subcommittee chair, with phage PAS7 as its representative. Phage PAS7 has a podo-like morphology, as confirmed by electron microscopy (Fig. 1b), similar to the confirmed morphology of its Shigella phage relative Buco (GenBank accession: MK562503.1; previously named HRP29) with which it shares 27.1%intergenomic similarity[17,18]. The size of the phage PAS7 virion was estimated at 67 nm. A high variability for the RBP gene was observed across the genomes of the members of the *Slopekvirinae* subfamily through MAUVE progressive alignment. This variability of the RBP within the genomes of phages PAS7 and phage relatives ENC16, phiKDA1, Buco and KP34 is visualized in Fig. 4 using clinker. Horizontal exchanges were detected for the RBP[19], with the PAS7 RBP (GenBank accession: WMX18832.1) showing similarity to the tailspikes of Escherichia phages TL-2011b (*Uetakevirus*) and Ro103C3Iw (*Kayfunavirus*) with 51 and 43% amino acid (aa) identity, respectively. These tailspikes are predicted to target the O103 antigen[19]. This corresponds to the serogroup of the host used for enrichment of PAS7 and indicates that the tailspike protein of phage PAS7 is likely binding/cleaving O103 antigen and responsible for the first phage-host interaction. Plating phage PAS7 on two other O103 strains resulted in semi-transparent zones of clearance instead of single plaques at phage concentrations higher than $10^6$ pfu/mL and no plaques with lower titers (Table 1), potentially indicating lysis from without (LO)[20]. The PAS7 RBP structure was predicted using AlphaFold2 (Fig. 5). Pairwise comparisons

| Host strain | Strain origin | Host serogroup | PAS7 | PAS59 | PAS61 |
|---|---|---|---|---|---|
| MG1655 | ZHAW | Rough | - | - | - |
| C600 | ZHAW | Rough | - | - | - |
| 5 | ZHAW, UZH | O26 | - | - | - |
| A11-2162 | ZHAW, UZH | O26:H31 | - | - | - |
| P11-2315 | ZHAW, UZH | O103 | -* | - | - |
| NVH-848 | NMBU | O103:H25 | -* | - | - |
| 4215/4 | ZHAW, UZH | O103 | + | - | - |
| 10-1120 | ZHAW, UZH | O111 | - | - | - |
| P10-1119 | ZHAW, UZH | O145 | - | - | - |
| A11-1581 | ZHAW, UZH | O145:H34 | - | - | - |
| A11-1675 | ZHAW, UZH | O146:H21 | - | + | + |
| 777/1 | ZHAW, UZH | O157 | - | - | - |
| 2905 | ZHAW, UZH | O157 | - | - | - |
| 264 | ZHAW, UZH | O157 | - | - | - |
| 584 | ZHAW, UZH | O157 | - | - | - |
| 332 | ZHAW, UZH | O157 | - | - | - |

**Table 1**. Host range testing of phages PAS7, PAS59 and PAS61. The three respective phages were spotted on lawns of bacteria of sixteen different *Escherichia coli* strains of the Centre for Food Safety and Quality Management of the Zurich University of Applied Sciences (ZHAW) strain collection, in collaboration with the Institute for Food Safety and Hygiene of the University of Zurich (UZH). Strain NVH-848 was additionally obtained from the Department of Paraclinical Sciences of the Norwegian University of Life Sciences (NMBU). A minus indicates the absence of bacterial killing and a plus indicates that the phage was able to infect the host, through identification of single plaques. An asterisk (*) indicates that zones of clearance were observed, without the presence of phage plaques in serial dilutions.

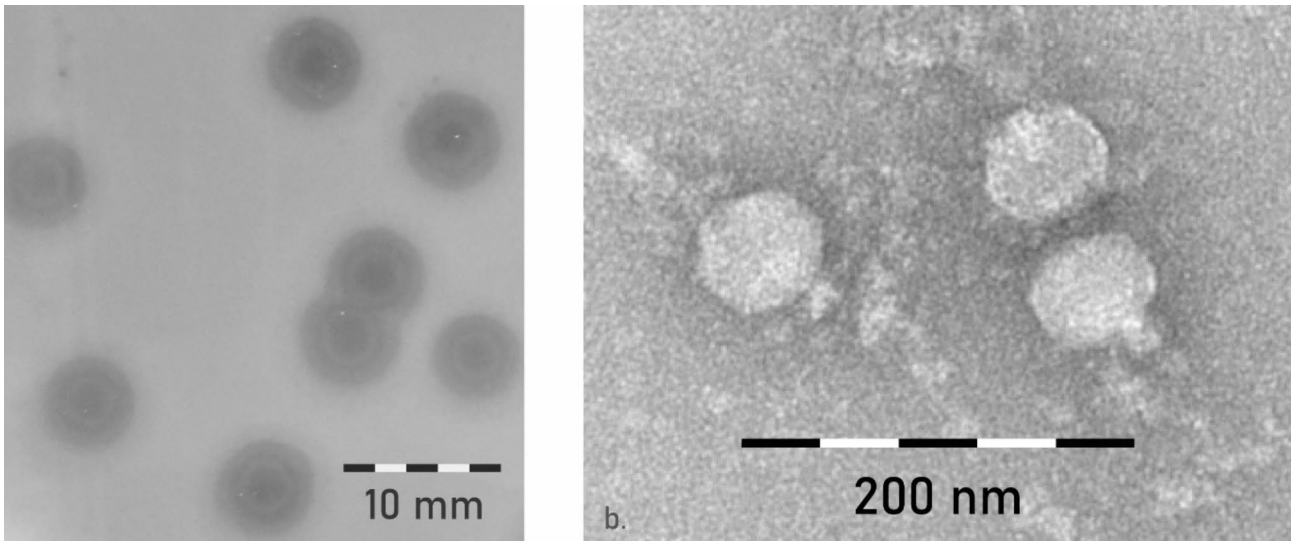


**Fig. 1**. (**a**) Plaque morphology of phage PAS7. Phage PAS7 has a multiple layered halo with a variable diameter between 2 and 9 mm. From the inside to the outside of the plaque, a clear plaque of 1–4 mm, a semi-transparent halo ≤ 1 mm, followed by a clear halo ≤ 0.5 mm and another semi-transparent halo of 1–2 mm. (**b**) Electron microscopy image of three phage PAS7 virions.

with the tailspikes of phages TL-2011b and Ro102C3Iw illustrate the high conservation of the receptor-binding domain (RBD), which is the C-terminal β-helix of the RBP, responsible for recognition of the O-antigen, while the N-terminal anchor domains are different. The latter are typically conserved at the taxonomical level and responsible for anchoring of the tailspike to the phage particle[19,21].

Phage PAS59 is a lytic phage with a genome size of 79,199 bp, infecting *E. coli* strain A11-1675 (serogroup O146:H21). Its closest relatives are Escherichia phages EK010[24] (GenBank accession: NC_070981.1) and SU7[25] (GenBank accession: NC_070980.1) with 86.0 and 85.5% intergenomic similarity, respectively (Fig. 3). Therefore, phage PAS59 belongs to the genus *Suseptimavirus* of the *Gordonclarkvirinae* subfamily, which has an uncommon, elongated head of morphotype C3[26]. Phage PAS59 has a plaque size of 1–2 mm without halo. In contrast to phages PAS7 and PAS61, phage PAS59 exhibits a RBP (GenBank accession: WMX18945.1) that is highly similar to those of its phage relatives, with minimum 52% amio acid similarity to the RBPs of Escherichia phages EK010, IME267 (GenBank accession: NC_070983.1), MLP3 (GenBank accession: OK148440.1), phiEco32 (GenBank accession: NC_010324.1), and SU7 (Fig. 4). Phages belonging to the *Gordonclarkvirinae* subfamily have a complex tail structure with long tail fibers composed of two separate proteins that initiate contact with a receptor on the outer cell membrane of their host, as can be observed in Escherichia phage SU10 (GenBank accession: NC_027395.1) belonging to the genus *Kuravirus* within the *Gordonclarkvirinae* subfamily[27]. In most cases, tailspikes are able to bind and cleave the O-antigen[28]. However, tail fibers were proven only to be involved in O-antigen binding[29]. The lack of a halo further supports the absence of enzymatic activity. Moreover, phage relatives EcoN5 (GenBank accession: NC_070986.1), EK010 and SU7 that have similar RBP sequences as phage PAS59 (68, 61 and 53% amino acid similarity, respectively) can infect various colistin-resistant *E. coli* strains and *E. coli* strains of different serogroups such as O6 for EcoN5, O157 for phage EK010 and O139 for phage SU7. This suggests that either (I) the RBP of phage PAS59 can bind various O-antigens, (II) the RBP of phage PAS59 is able to target an O-antigen serogroup different from those highly similar RBPs, or (III) the RBP targets a receptor different from the O-antigen.

Phage PAS61 is a temperate phage with a genome size of 39,287 bp. The closest relatives of phage PAS61 are phages phiV10[30] and phiv142-3[31] with 78.7 and 78.5% intergenomic similarity, respectively (Figs. 2, 3 and 4). Phage PAS61 has a 1 mm plaque size with a surrounding semi-transparent halo of 3–5 mm in diameter. Similar to the majority of *Uetakeviruses*, phage PAS61 presumably degrades the O-antigen with its tailspike[19]. For example, the tailspike of phage phiV10 cleaves the O157 antigen[32]. Furthermore, the tailspike of phage PAS61 (GenBank accession: WMX19002.1) shows similarity to two other phage RBPs. An amino acid similarity of 56.6% was found within its RBD to the tailspike of *Kagunavirus* 26 (GenBank accession: UAW06957.1), but the serogroup of its *E. coli* host is not known. Like *Uetakeviruses*, *Kagunaviruses* are known to encode tailspikes with O-antigen-degrading properties and a horizontal exchange of the C-terminal domain of tailspikes has been reported before between a *Uetakevirus* and a *Kagunavirus*[19]. Secondly, the tailspike of phage PAS61 also shows 81.4% aa simiarity over the complete tailspike of a *Uetakevirus* integrated within *Escherichia fergusonii* strain RHB38-C01 (GenBank accession: CP057104.1). Although the O-antigen of *E. fergusonii* can be different from that of *E. coli*, the serogroup of the strain was predicted to be serogroup O146, using the SerotypeFinder tool (with > 96.4% nucleotde (nt) identity, 100% coverage to the specific *wzx* and *wzy* genes resulting in serogroup O146)[33], which is in agreement with the serogroup of the observed host of phage PAS61 (Table 1). As the RBPs of *Uetakeviruses* have a high potential to be O-antigen specific and since two RBPs within the same RBP subtype hint towards O146 specificity, we hypothesize that the tailspike of phage PAS61 can bind and/or cleave the O146

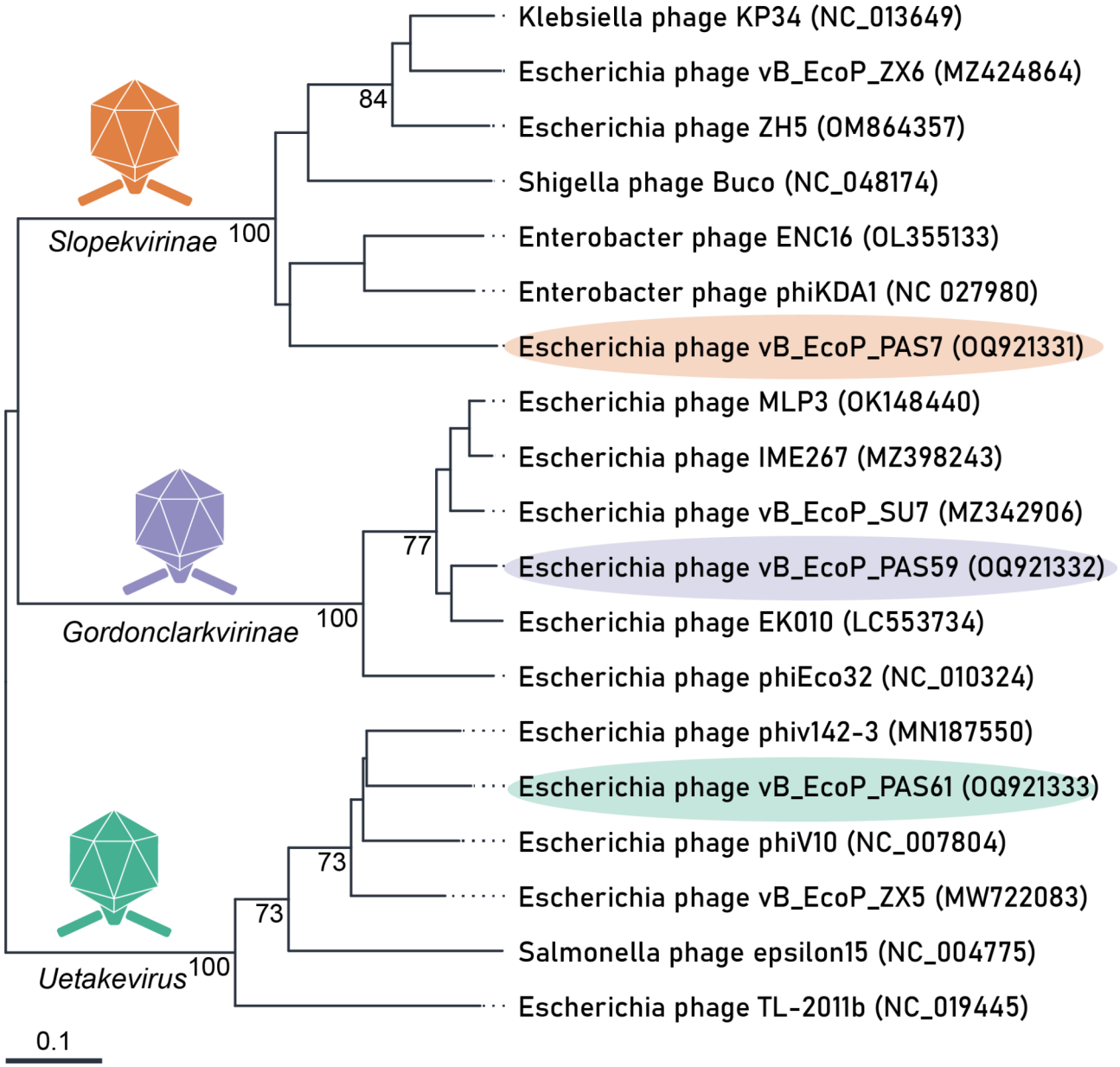


**Fig. 2**. Phylogenetic tree of isolated phages PAS7, PAS59 and PAS61 and their phage relatives using VICTOR phylogeny. The numbers below the branches are GBDP pseudo-bootstrap support values from 100 replications. The branch lengths of the resulting VICTOR trees are scaled in terms of the distance formula D6[22]. The GenBank accession numbers of all phage genomes are listed between parentheses.

antigen. In summary, both phages PAS7 and PAS61 are predicted to target the same serogroups as taxonomically unrelated phages, illustrating evolution through horizontal gene transfer.

Potential active sites of the RBPs of phages PAS7 and PAS61 were identified in silico. Therefore, the protein data bank (PDB) was searched for structurally similar RBPs using DALI. For both RBPs, we found various similarities to phage tailspikes in the PDB and selected those with highest similarities (DALI Z scores > 30 or RMSD scores < 4) (Supplementary Table S1). The RBP of phage PAS7 shares structure similarity with the Acinetobacter phage AS12 RBP (PDB code: 6eu4), the second RBP of phage CBA120 (TSP2; PDB code: 5w6p) and the RBP of Escherichia phage Sf6 (PDB code: 2vbk). DALI also showed similarity of the PAS61 RBP to the RBP of phage Sf6 (PDB code: 2vbk) and the third and first RBP of phage CBA120 (TSP3 and TSP1; PDB codes: 6nw9 and 4ojl). Most of these structurally similar RBPs have glycosidase catalytic activity and intersubunit active sites[34,35]. The glycosidase catalytic machinery generally employs the carboxyl groups of two neighbouring Asp/Glu amino acid residues, consistent with an acid/base mechanism[36]. By calculating the surface charge potential, negatively charged pockets were found along the groove formed at the subunit interface for both the PAS7 and PAS61 RBPs (Fig. 6a, c). In the located pocket of the phage PAS7 RBP, localization of the conserved Asp/Glu amino acid residues revealed two neighbouring clusters of acid residues, D166*, E177, E268 on the one hand,

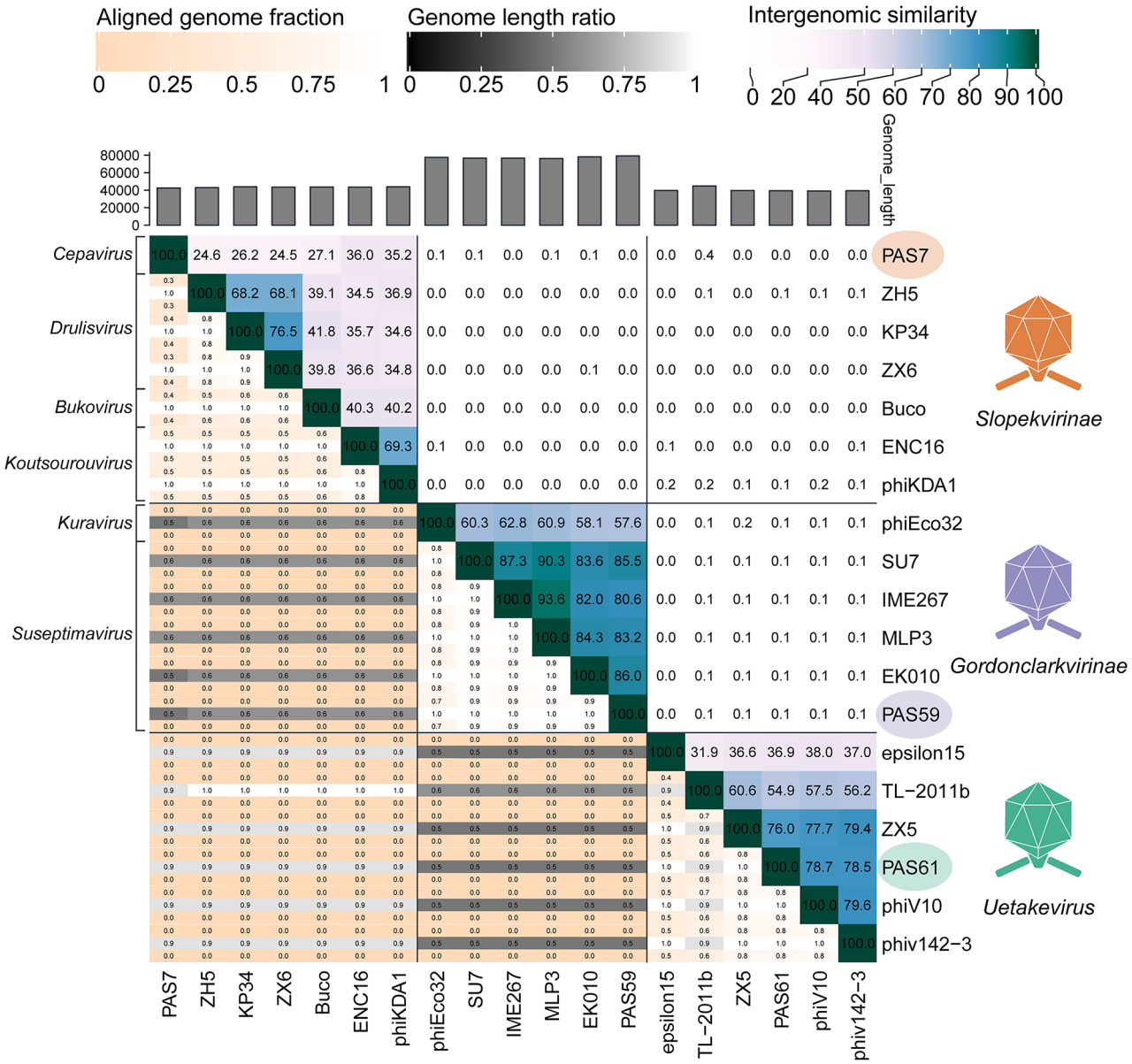


**Fig. 3**. Phage intergenomic similarity between phages PAS7, PAS59 and PAS61 and their phage relatives using the VIRIDIC phage genome comparison tool. Per square, the intergenomic similarity percentages are displayed in the right upper half along with the corresponding scale on top. In the left lower half, alignment indicators are shown. Three different values are shown for each phage genome comparison. From top to bottom: aligned fraction genome 1 (for the genome found in this row), genome length ratio (for the two genomes in this pair) and aligned fraction genome 2 (for the genome found in this column). The darker colours emphasize low values, indicating genome pairs with only a small fraction of the genome aligned (orange to white colour gradient for low to high values), or with a high difference in genome length (black to white colour gradient for high to low values)[23]. For each phage, its genome length in bp is indicated on top as a bar plot.

and E347, E348, E377 and D407 on the other hand (Fig. 6, b) (*adjacent chain). These residues are conserved across the RBD homologs of phages TL-2011b and Ro103C3lw, except for E348 and D166* (Supplementary Figure S1). Specifically, residue E348 is not conserved in the RBP of RBD homolog phage TL-2011b and residue D166* is not conserved in RBD homolog phage Ro103C3lw. However, in the latter, an alternative Asp residue is present 10.4 Å away from its E177 residue homolog, potentially replacing the function of residue D166*. Similarly, glycosidase activity at a negatively charged pocket was predicted for the phage PAS61 RBP, with the putative Asp/Glu amino acid residues E404*, E437*, D447 on the one hand, and E491*, E521, D541 and D573 on the other hand, which constitute two neighbouring clusters. All seven potential active site residues are conserved across the RBD homologs of RHB38-C01_prophage and phage 26 (Fig. 6, d; Supplementary Figure S2). The average atomic distance between the carboxylic oxygen atoms of two neighbouring Glu/Asp residues in a glycosidase hydrolysis reaction is approximately 5-5.5 Å for retaining glycosidases and around 9.0-9.5 Å

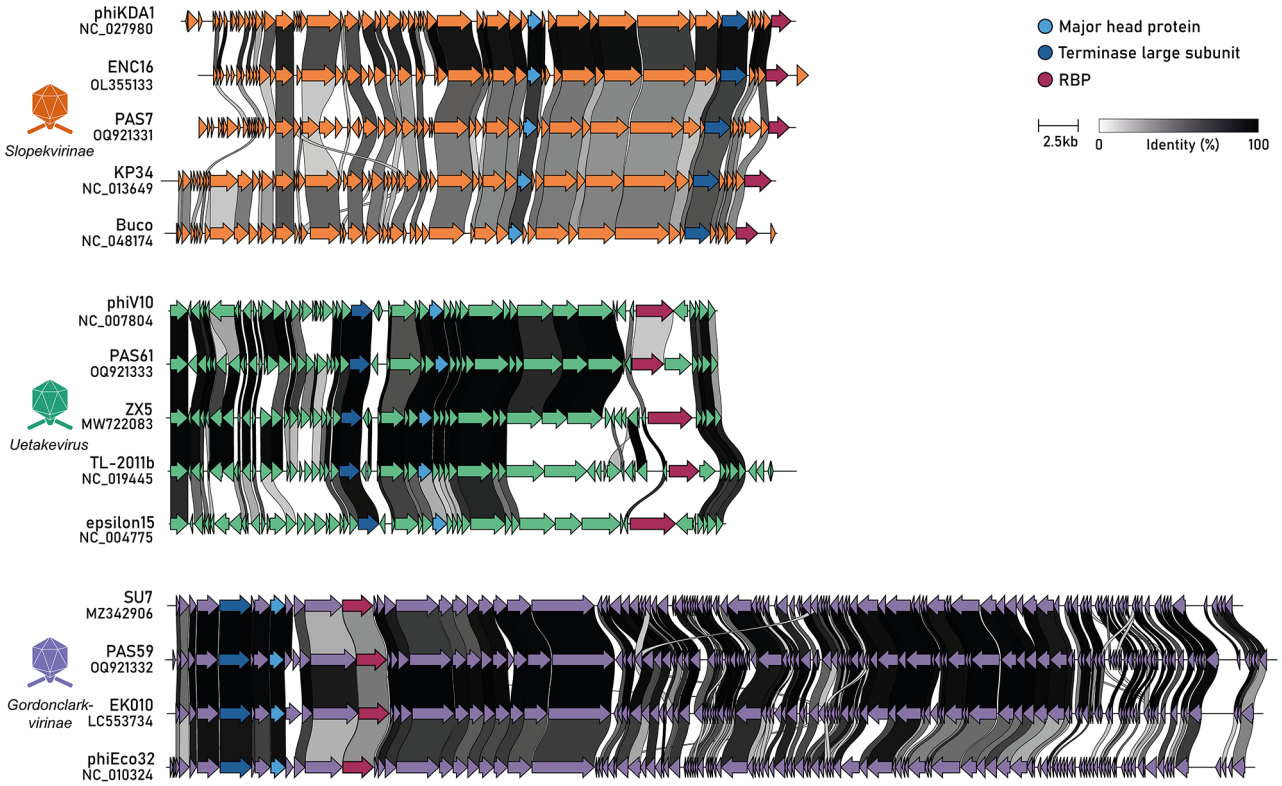


**Fig. 4**. Phage genome comparison of isolated phages PAS7, PAS61 and PAS59 and phages sharing genomic similarity. The three different taxonomic groups to which the isolated phages belong are indicated using three different colours (*Slopekvirinae*, orange; *Uetakevirus*, green; *Gordonclarkvirinae*, purple). Two genes encoding conserved phage proteins, namely the major head protein and the terminase large subunit are indicated for orientation purposes (light blue and dark blue respectively). The receptor-binding proteins (RBPs) of the phages (tailspikes for phages belonging to the *Slopekvirinae* subfamily and *Uetakevirus* genus, and the long tail fiber tip for phages belonging to the *Gordonclarkvirinae* subfamily) are shown in red. Other genes are shown as orange, green or purple arrows according to their taxonomic group. Homology between the genes is indicated with a light grey to black colour gradient, from low to high % nucleotide identity.

for inverting glycosidases[37,38]. Based on conservation and atomic distance, we highlight residue pairs E177 and E268 (8.7 Å), and E377 and E407 (5.0 Å) as potential active residue candidate pairs for the RBP of phage PAS7. Residue pairs E404* and D447 (9.8 Å), and E491* and E521 (5.4 Å) are plausible active residue pairs for the RBP of phage PAS61. However, AlphaFold2 structure predictions are error prone. While the negatively charged pocket is prominent, the exact location of the candidate pairs might diverge from their effective location. Further investigation with single and double mutants is needed to pin down the exact residues taking part in the enzymatic reaction.

## Discussion

To summarize, three phages were retrieved from seven bovine environmental samples, of which the RBPs of phages PAS7 and PAS61 are predicted to specifically bind and/or cleave to the O-antigen serogroup of their *E. coli* host. Additionally, phage PAS7 has low sequence homology to any other genome within NCBI, resulting in the proposal of a new phage genus *Cepavirus*. We highlight the importance of identifying the RBPs of phages and studying their targeting receptor. Additionally, this work illustrates the need to use emerging STEC serogroups for phage isolation purposes, as using new hosts for phage isolation can lead to further discoveries of new phage taxonomic groups and further unraveling of phage-host interactions.

Unravelling phage-host relationships is crucial for selecting the right phage in phage therapy or for the creation of new phage-derived antimicrobial solutions. For example, the RBPs of phages can be used to change the host range of other phages through phage engineering[39]. Additionally, a capsule degrading RBP has been shown to make multidrug-resistant bacteria susceptible to serum-mediated killing, highlighting their potential for use in combined treatments[40]. RBPs can also be used to develop highly specific antimicrobials, such as tailocins or innolysins[32,41,42]. New STEC-targeting phages or phage-derived products can benefit patients in phage therapy. Unlike most traditional therapies using antibiotics[15], Shiga toxin production is not exacerbated by phages. Additionally, these phages could be applied in a preventive manner such as in food processing, given that EHEC/STEC are foodborne pathogens[14]. In future phage-host specificity prediction tools based on machine learning could aid us in the search for specific phages or their RBPs[43,44]. However, the effectiveness and accuracy of such

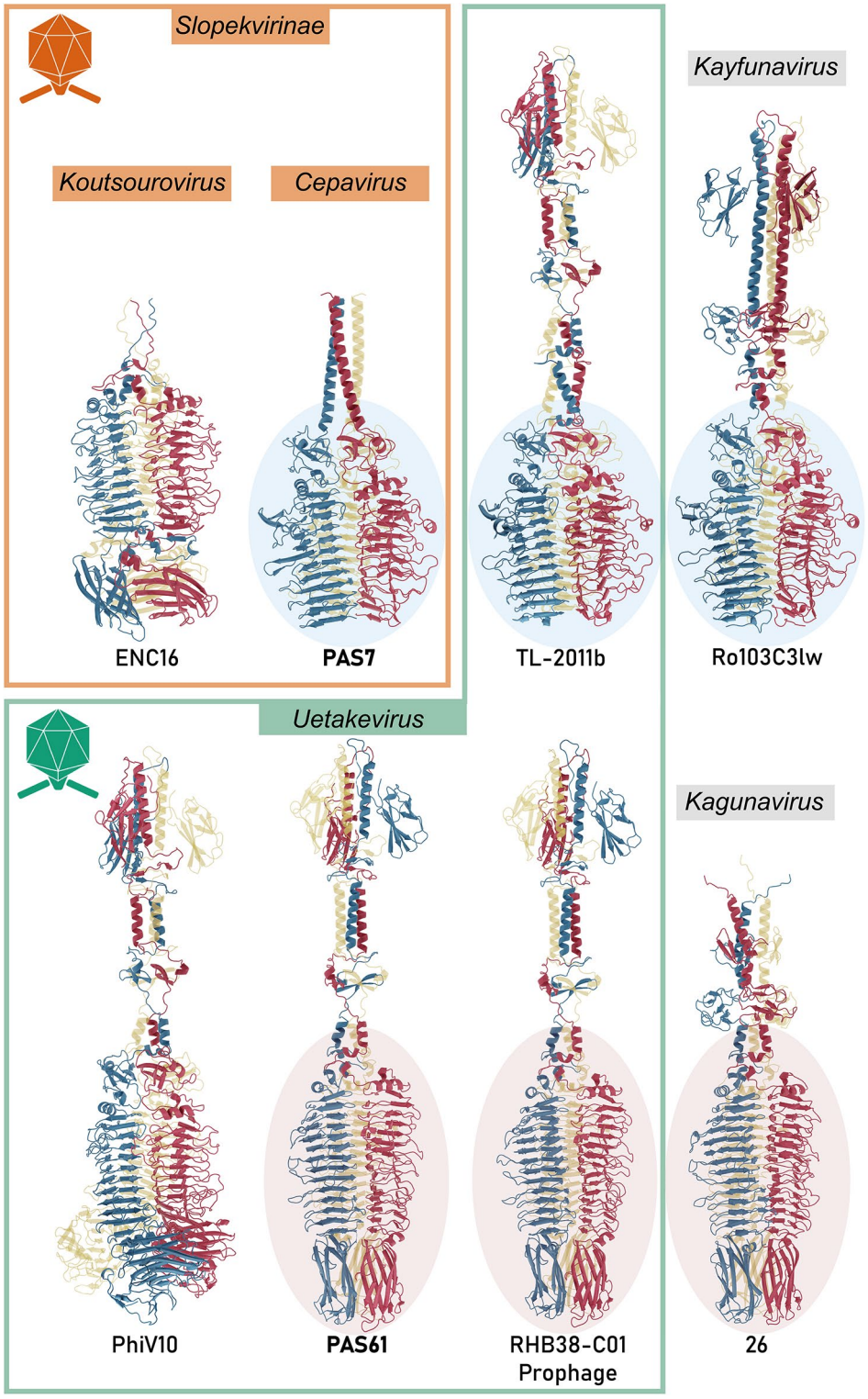


**Fig. 5**. Visualization of the receptor-binding protein (RBP) structures of isolated phages PAS7 and PAS61 (indicated in bold) using AlphaFold v2.3.1. The predicted phage RBP structures of the phages with the highest intergenomic similarities with phages PAS7 and PAS61. The RBPs of phages ENC16 and phiV10 are shown, respectively. These comparisons illustrate the variation of the C-terminal domains of the RBP structure within taxonomic phage groups, whereas the N-terminal anchor is conserved within all *Uetakeviruses*. Additionally, RBPs with amino acid similarity to the receptor-binding domains of phages PAS7 and PAS61 are shown to highlight the horizontal gene transfer of the C-terminal domain of the RBPs across variable taxonomic phage groups (indicated in light blue and red coloured ovals).

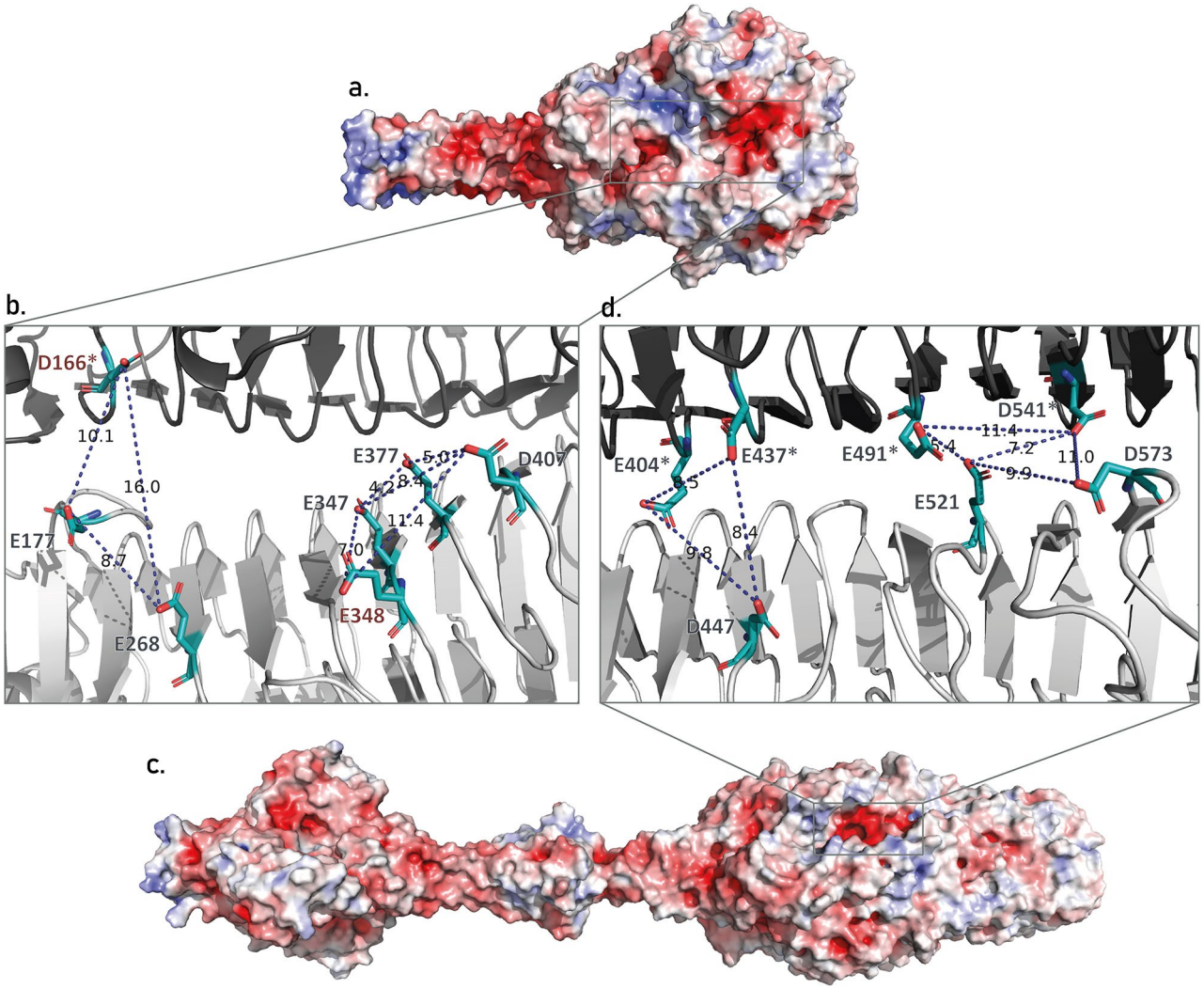


**Fig. 6**. In silico predicted catalytic pockets of the receptor-binding proteins (RBPs) of phages PAS7 and PAS61 with the Asp/Glu plausible active site residues. (**a**) The surface charge potential of the RBP of PAS7 with blue and red colours representing the positive and negative charge, respectively. (**b**) Cartoon representation of the predicted catalytic pocket with potential active site Asp/Glu residues shown as sticks. Residues of the adjacent subunit (dark grey) are indicated with an asterisk. The distances between the carboxyl groups of neighbouring residues are shown. Residues which are not conserved across the RBD homologs of the RBP targeting the same serogroup are labelled in brown (panel **b**, residues D166* and E348). (**c**) Surface charge potential of the RBP of phage PAS61. (**d**) Predicted catalytic pocket with plausible active site Asp/Glu residues of the RBP of phage PAS61.

prediction tools heavily rely on the availability of comprehensive data on phage-host relationships. Therefore, continuous efforts in phage isolation and the collection of phage host range data are crucial to advancing the development and refinement of these prediction tools[15].

## Methods

**Media, bacterial strains, and strain serogroup confirmation.** SM buffer was used for all phage solutions (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-Cl, pH 7.5). Bacterial strains were grown in lysogeny broth (LB; 10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl). Agar was added at a concentration of 1.5% for plating and 0.5% for top agar layers. The serogroups of *E. coli* strains 4215/4 (serogroup O103) and A11-1675 (serotype O146:H21) were confirmed using an O-antigen PCR test using primers (Table 2) available in the Bacterial Surface Polysaccharide Genes Database (BSPdb) targeting O103 and O146 *wzx* genes[33]. Serogroup analysis of *E. coli* genomes from online databases was performed using the SerotypeFinder tool (release date: 28 January 2019)[45].

**Phage isolation.** Seven calf/cow feces samples were collected from local farms in Wädenswil, Switzerland. Samples (approximately 20 g) were diluted with SM buffer to 50 mL, mixed and centrifuged at 7,000 × *g* for 10 min. To 35 mL of the supernatant, 10 mL of overnight incubated (> 18 h) cultures of rough *E. coli* K12-like

| Primer name | Primer sequence (5'-3') |
|---|---|
| O103wzx_Fw | TAAGTACGGGGGTGCTTTTT |
| O103wzx_Rv | AAGCTCCCGAGCACGTATAA |
| O146wzx_Fw | CGCCACAATTACCATGGGA |
| O146wzx_Rv | CCCCTCCAGGCAAAATTACA |

**Table 2**. Primers used for serogroup verification of *Escherichia coli* strains 4215/4 and A11-1675, targeting the *wzx* genes of the bacteria. These genes are specific for the synthesis of the O103 and the O146 O-antigen, respectively.

strains (5 mL of strains C600 and MG1655) and 5 mL of 10x concentrated SM buffer (pH 8) were added, to remove O-antigen aspecific phages. This mixture was incubated at 21 °C for 20 min and then centrifuged for 2 min at 10,000 × *g*. Remaining phages were enriched from the supernatant with bacterial *E. coli* strains 4215/4 (serogroup O103) and A11-1675 (serotype O146:H21) as hosts. To 15 mL of sample supernatant, 25 mL double concentrated LB medium, 5 mL bacterial culture and 5 mL 10x concentrated SM buffer (pH 8) were added, and the mixture was then incubated for 3 h at 37 °C and 160 rpm. Enriched mixtures were centrifuged for 2 min at 10,000 × *g* and the supernatant was further filter sterilized using 0.45 µm syringe filters. Phages were then plated using the double agar overlay method with an LB agar top layer, supplemented with 200 µL of 18 h incubated host culture and 10 µL phage solution. Dilutions $10^{0}$, $10^{-2}$, $10^{-4}$, $10^{-5}$ and $10^{-6}$ of the phage solution were plated for each sample. Single phage isolation was ensured by picking phage plaques and repeating the double agar overlay method three consecutive times.

**Phage propagation, purification, and host range.** Phages were first propagated using the double agar overlay method at the phage titer resulting in semi-confluently lysed plates ('web' plate morphology). Ten plates of this titer were washed off with 5 mL SM buffer for at least 5 h at 80 rpm. Next, NaCl was added to the phage mixture, to a final concentration of 0.5 M and incubated for 30 min at around 21 °C and mixed regularly. After 10 min centrifugation at 10,000 × *g*, the supernatant was transferred into a new container and 10% (*w*/*v*) PEG8000 was added and mixed until dissolved. This mixture was incubated at least 18 h at 4 °C on ice. The next day, a 15 min centrifugation step at 10,000 × *g* and 4 °C was performed and the phage-PEG pellet was resuspended in 21 mL SM buffer. A spotting assay was performed to verify the titer of the propagated phage mixture. A titer of at least $10^{12}$ PFU/mL was preferred. Consecutively, a CsCl density gradient purification was performed with 20 mL of the phage solution and centrifuged for 2 h at 50,000 × *g* at 10 °C[46]. After at least 12 h of dialysis of the phage solution in SM buffer at 4 °C, the titer of the concentrated phage solution was determined using a double agar layer spotting assay, with 200 µL of > 18 h incubated host cultures in the top agar layer and 5 µL spots of dilutions dilution 1–$10^{-8}$ of the phage solution. The host range of the phages was determined using the double agar layer spotting assay on 16 *E. coli* strains with various lipo-polysaccharide (LPS) structures (Table 1).

**Phage DNA extraction and genome sequencing.** The DNA was extracted from purified phage solution using a phenol-chloroform method. First, 1 U/mL DNase I and RNase A were added to the purified phage solution, mixed and incubated at 37 °C for 15 min. Next, EDTA (pH 8), Proteinase K and SDS were added to final concentrations of 20 mM, 50 µg/mL and 0.5% (*w*/*v*), respectively, and incubated for one hour at 56 °C. One volume of phenol-chloroform-isoamylalcohol (PCI) mixture (25:24:1) was then added and well mixed, after which a 10-min centrifugation step (15,000 × *g*) was performed and the upper watery phase was retained. The addition and centrifugation were performed three times, the last round using chloroform-isoamylalcohol (49:1). After a final step of adding two and a half volumes of ice-cold ethanol and 0.1 volumes of NaOAc (3 M, pH 5.2), visible DNA clumps were extracted using a pipette tip and dissolved in 50 µL UltraPure water. The DNA was quantified using the Quant-iT PicoGreen double-stranded DNA (dsDNA) quantification assay (Thermo Fisher Scientific, Waltham, MA) and a Synergy HT microplate reader (BioTek, Winooski, VT, USA). The phage genomic DNA was prepared using a Nextera XT DNA library prep kit (Illumina, San Diego, CA, USA) followed by 2 × 150-bp paired-end reads using a MiSeq sequencing reagent kit v2 (Illumina) according to the manufacturer's instructions and assembled using MIRA (version 4.0.2) on a HPC at ZHAW. Genome sequences of the isolated Escherichia phages vB_EcoP_PAS7, vB_EcoP_PAS59 and vB_EcoP_PAS61 can be accessed through the GenBank database with the respective accessions OQ921331, OQ921332 and OQ921333.

**Phage genome annotation, phylogeny and taxonomy.** The KBase online platform (kbase.us)[47] was used to annotate the genome with RASTtk version 1.073[48]. MAUVE progressive alignment version 20.150.226[49] of the isolated phage genomes with phage relatives was used to manually annotate the RBPs such as tailspikes of the isolated phages. The virulent or temperate nature of the phages was verified using PhageAI v1.0.2[50]. BLASTn version 2.15.0 was used to identify the closest relatives of the phages. A phylogenetic tree was built using the VICTOR phylogeny tool (release date 26 October 2017)[22] and processed using iTOL v6 to obtain Fig. 2[51]. The VIRIDIC v1.1 tool[23] was used to calculate the intergenomic similarity of the phage genomes with their closest relatives, as displayed in Fig. 3. Genome visualization, as shown in Fig. 4, was performed using clinker v0.0.28[52]. All figures were edited using Adobe Illustrator version 26.5.1.

**Tailspike structure prediction, structure similarity and in silico prediction of catalytic domains.** To investigate the presence of a catalytic domain within the tailspikes, the sequences of the RBPs of phages PAS7, TL-2011b and Ro103C3lw, and PAS61, 26 and RHB38-C01_prophage were aligned using MUSCLE v3.8.31[53]. These alignments were visualized using SnapGene Viewer version 6.0.5 (Supplementary Figures S1 and S2). AlphaFold2 (v2.3.1; multimer, maximum recycles = 30) was used on the HPC-UGent to predict the homotrimeric quaternary structures of the RBPs[54]. The reliability of the AF predictions was assessed by the Local

Distance Difference Test (LDDT) score reported for each structure (Supplementary Figure S3-S10). Structures were further processed and RMSD scores and surface charge potential were calculated using PyMol version 2.5.2. For calculating the RMSD scores, the estimated receptor-binding domains (RBDs) of the RBPs were compared. RBDs were estimated to start at amino acid positions 1, 1, 206, 227, 216, 206, 213 and 134 for the respective RBPs of phages PAS7 (WMX18832.1), ENC16 (UIW12862.1), TL-2011b (YP_007001999.1), Ro103C3lw (QDH94159.1), PAS61 (WMX19002.1), PhiV10 (YP_512279.1), RHB38-C01_prophage (WP_249418615.1) and 26 (UAW06957.1). The estimate anchor domain of the RBP of phage PAS61 was removed to focus on the RBD. DaliLite v5 software (http://ekhidna2.biocenter.helsinki.fi/) was used to analyse structure similarities of the RBPs with other proteins in the Protein Data Bank (PDB). Figures were further optimised using Blender, version 2.93.3, Adobe Illustrator version 26.5.1 and Adobe InDesign version 16.4.3.

**Electron microscopy**. Prior to electron microscopy, phage PAS7 was purified using CsCl (Sigma Aldrich, Switzerland) gradient purification[55]. Samples were negatively stained with 2% (*w*/*v*) uranyl acetate (Sigma Aldrich, Switzerland) and visualized using a TEM Morgagni 268 (FEI, NL) operated at 100 kV in bright field mode. Micrographs were recorded using a CCD-camera (Keen View, Olympus SIS GmbH, D). Images were taken at Scope M, ETH Zurich.

## Data availability

The datasets supporting the conclusions of this article are included within the article and its supplementary materials. Complete genome sequences of the isolated Escherichia phages vB_EcoP_PAS7, vB_EcoP_PAS59 and vB_EcoP_PAS61 are deposited in the GenBank database with the respective GenBank accessions OQ921331, OQ921332 and OQ921333. In this work, these phages are referred to as phages PAS7, PAS59 and PAS61, respectively.

Received: 17 July 2024; Accepted: 22 October 2024
Published online: 30 October 2024

## References

1. Darwish, W. S., Fathy, W., Eldin, S., Eldesoky, K. I. & Prevalence Molecular characterization and antibiotic susceptibility of *Escherichia coli* isolated from duck meat and giblets. *J. Food Saf.* **35**, 410–415 (2015).
2. Krü, A. & Lucchesi, P. M. A. Shiga toxins and stx phages: highly diverse entities. *Microbiology*. **161**, 451–462 (2015).
3. Caprioli, A., Scavia, G. & Morabito, S. Public Health Microbiology of Shiga Toxin-Producing Escherichia coli. *Microbiol. Spectr.* **2**, 0014–2013 (2014).
4. Kim, J. S., Lee, M. S. & Kim, J. H. Recent Updates on Outbreaks of Shiga Toxin-Producing Escherichia coli and Its Potential Reservoirs. *Front. Cell. Infect. Microbiol* https://doi.org/10.3389/fcimb.2020.00273 (2020).
5. Rúgeles, L. C., Bai, J., Martínez, A. J. & Vanegas, M. C. Gómez-Duarte, O. G. Molecular characterization of diarrheagenic *Escherichia coli* strains from stools samples and food products in Colombia. *Int. J. Food Microbiol.* **138**, 282–286 (2010).
6. Pakbin, B., Brück, W. M. & Rossen, J. W. A. Virulence factors of enteric pathogenic *Escherichia coli*: a review. *Int. J. Mol. Sci.* **22**, 9922. https://doi.org/10.3390/ijms22189922 (2021).
7. Schimmer, B. et al. Outbreak of haemolytic uraemic syndrome in Norway caused by stx2-positive *Escherichia coli* O103:H25 traced to cured mutton sausages. *BMC Infect. Dis.* **8**, 1–10 (2008).
8. CDC. Outbreak of *E. coli* Infections Linked to Clover Sprouts. (2020). https://www.cdc.gov/ecoli/2020/o103h2-02-20/index.html
9. Vishram, B. et al. The emerging importance of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England. *J. Med. Microbiol.* **70**, 001375. https://doi.org/10.1099/jmm.0.001375 (2021).
10. Long, J. et al. Large-scale phylogenetic analysis reveals a New Genetic Clade among *Escherichia coli* O26 strains. *Microbiol. Spectr.* **10**, 02525–02521. https://doi.org/10.1128/spectrum.02525-21 (2022).
11. Dewsbury, D. M. A. et al. Summer and Winter Prevalence of Shiga Toxin-Producing *Escherichia coli* (STEC) O26, O45, O103, O111, O121, O145, and O157 in feces of feedlot cattle. *Foodborne Pathog Dis.* **12**, 726–732 (2015).
12. Cernicchiaro, N. et al. Associations between season, Processing Plant, and hide cleanliness scores with prevalence and concentration of Major Shiga Toxin-Producing *Escherichia coli* on beef cattle hides. *Foodborne Pathog Dis.* **17**, 611–619 (2020).
13. ECDC. *STEC Infection Annual Epidemiological Report for 2021 Key Facts*. (2022). https://www.ecdc.europa.eu/en/publications-data/stec-infection-annual-epidemiological-report-2021
14. Tarr, P. I. & Freedman, S. B. Why antibiotics should not be used to treat shiga toxin-producing *Escherichia coli* infections. *Curr. Opin. Gastroenterol.* **38**, 30–38 (2022).
15. Howard-Varona, C. et al. Fighting Fire with Fire: Phage Potential for the Treatment of *E. coli* O157 Infection. *Antibiotics* https://doi.org/10.3390/antibiotics7040101 (2018).
16. Finney, A. G. et al. Isolation and characterization of Lytic bacteriophages Targeting Diverse *Enterobacter* spp. Clinical isolates. *PHAGE: Therapy Appl. Res.* **3**, 50–58 (2022).
17. Doore, S. M. et al. A cornucopia of Shigella phages from the Cornhusker State. *Virology*. **538**, 45–52. https://doi.org/10.1016/j.virol.2019.09.007 (2019).
18. Subramanian, S., Bergland Drarvik, S. M., Tinney, K. R. & Parent, K. N. Cryo-EM structure of a *Shigella* podophage reveals a hybrid tail and novel decoration proteins. *Structure*. **32**, 24–34e4 (2024).
19. Pas, C., Latka, A., Fieseler, L. & Briers, Y. Phage tailspike modularity and horizontal gene transfer reveals specificity towards *E. Coli* O-antigen serogroups. *Virol. J.* **20**, 1–19 (2023).
20. Abedon, S. T. Lysis from without. *Bacteriophage*. **1**, 46–49. https://doi.org/10.4161/bact.1.1.13980 (2011).
21. Latka, A., Leiman, P. G., Drulis-Kawa, Z. & Briers, Y. Modeling the Architecture of Depolymerase-Containing Receptor Binding Proteins in Klebsiella Phages. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2019.02649 (2019).
22. Meier-Kolthoff, J. P. & Göker, M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics*. **33**, 3396–3404 (2017).
23. Moraru, C., Varsani, A. & Kropinski, A. M. VIRIDIC—A Novel Tool to calculate the intergenomic similarities of Prokaryote-infecting viruses. *Viruses*. **12**, 1268. https://doi.org/10.3390/v12111268 (2020).
24. Shahin, K. et al. Bio-control of O157:H7, and colistin-resistant MCR-1-positive *Escherichia coli* using a new designed broad host range phage cocktail. *LWT-Food Sci. Technol.* **154**, 112836 (2022).
25. Koonjan, S., Cooper, C. J. & Nilsson, A. S. Complete genome sequence of vb_ecop_su7, a podoviridae coliphage with the rare c3 morphotype. *Microorganisms*. **9**, 1576 (2021).
26. Batinovic, S., Fujii, Y. & Nittami, T. Expansion of Kuravirus-like phage sequences within the past Decade, including Escherichia Phage YF01 from Japan, prompt the creation of three New Genera. *Viruses*. **15**, 506. https://doi.org/10.3390/v15020506 (2023).

27. Šiborová, M. et al. Tail proteins of phage SU10 reorganize into the nozzle for genome delivery. *Nat. Commun.* https://doi.org/10.1038/s41467-022-33305-w (2022).
28. Broeker, N. K. & Barbirz, S. Not a barrier but a key: how bacteriophages exploit host's O-antigen as an essential receptor to initiate infection. *Mol. Microbiol.* **105**, 353–357 (2017).
29. Efimov, A. D. et al. RB49-like bacteriophages recognize O antigens as one of the alternative primary receptors. *Int. J. Mol. Sci.* **23**, 11329. https://doi.org/10.3390/ijms231911329 (2022).
30. Perry, L. L. et al. Sequence analysis of *Escherichia coli* O157:H7 bacteriophage PhiV10 and identification of a phage-encoded immunity protein that modifies the O157 antigen. *FEMS Microbiol. Lett.* **292**, 182–186 (2009).
31. Li, D. et al. Prophage phiv142-3 enhances the colonization and resistance to environmental stresses of avian pathogenic *Escherichia coli*. *Vet. Microbiol.* **218**, 70–77 (2018).
32. Scholl, D. et al. An engineered R-type pyocin is a highly specific and sensitive bactericidal agent for the food-borne pathogen *Escherichia coli* O157:H7. *Antimicrob. Agents Chemother.* **53**, 3074–3080 (2009).
33. Liu, B. et al. Structure and genetics of *Escherichia coli* O antigens. *FEMS Microbiol. Rev.* **44**, 655–683 (2020).
34. Greenfield, J. et al. Structure and tailspike glycosidase machinery of ORF212 from *E. Coli* O157:H7 phage CBA120 (TSP3). *Sci. Rep.* **9**, 1–11 (2019).
35. Müller, J. J. et al. An Intersubunit Active Site between Supercoiled Parallel β Helices in the Trimeric Tailspike Endorhamnosidase of Shigella flexneri Phage Sf6. *Structure* **16**, 766–775 (2008).
36. Maciejewska, B. et al. Klebsiella phage KP34gp57 capsular depolymerase structure and function: from a serendipitous finding to the design of active mini-enzymes against K. pneumoniae. *mBio*. **14**, e0132923. https://doi.org/10.1128/mbio.01329-23 (2023).
37. Steinbacher, S. et al. Crystal structure of phage P22 tailspike protein complexed with Salmonella sp. O-antigen receptors. *Proc. Natl. Acad. Sci. USA*. **93**, 10584 (1996).
38. McCarter, J. D. & Stephen Withers, G. Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885–892 (1994).
39. Latka, A. et al. Engineering the modular receptor-binding proteins of Klebsiella Phages switches their Capsule Serotype specificity. *mBio*. **12**, e00455–e00421. https://doi.org/10.1128/mBio.00455-21 (2021).
40. Oliveira, H. et al. A tailspike with exopolysaccharide depolymerase activity from a new Providencia stuartii phage makes multidrug-resistant bacteria susceptible to serum-mediated killing. *Appl. Environ. Microbiol.* **86** (20), e00073. https://doi.org/10.1128/AEM.00073-20 (2020).
41. Scholl, D., Gebhart, D., Williams, S. R., Bates, A. & Mandrell, R. Genome sequence of *E. Coli* O104:H4 leads to rapid development of a targeted antimicrobial agent against this emerging pathogen. *PLoS One*. **7**, e33637. https://doi.org/10.1371/journal.pone.0033637 (2012).
42. Zampara, A. et al. Exploiting phage receptor binding proteins to enable endolysins to kill Gram-negative bacteria. *Sci. Rep* **10**, 1–12 (2020).
43. Boeckaerts, D. et al. Predicting bacteriophage hosts based on sequences of annotated receptor-binding proteins. *Sci. Rep.* **11**, 1467 (2021).
44. Boeckaerts, D. et al. Prediction of Klebsiella phage-host specificity at the strain level. *Nature Comm* **15**, 1–10 (2024).
45. Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M. & Scheutz, F. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* **53**, 2410–2426 (2015).
46. Boulanger, P. Purification of bacteriophages and SDS-PAGE analysis of phage structural proteins from ghost particles. in Bacteriophages: Methods and Protocols, Volume 2 Molecular and Applied Aspects (eds Clokie, M. R. J. & Kropinski, A. M.) 227–238 (Humana, Totowa, NJ, https://doi.org/10.1007/978-1-60327-565-1_13. (2009).
47. Arkin, A. P. et al. KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).
48. Brettin, T. et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 1–6 (2015).
49. Darling, A. C. E., Mau, B., Blattner, F. R., Perna, N. T. & Mauve Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394 (2004).
50. Tynecki, P. et al. PhageAI-Bacteriophage Life Cycle Recognition with Machine Learning and Natural Language Processing. *Preprint at.* https://doi.org/10.1101/2020.07.11.198606v1.full (2020). https://www.biorxiv.org/content/
51. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296. https://doi.org/10.1093/nar/gkab301 (2021).
52. Gilchrist, C. L. M. & Chooi, Y. H. Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics*. **37**, 2473–2475 (2021).
53. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792 (2004).
54. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).
55. Sambrook, J., Russel, D., Fritsch, E. F. & Maniatis, T. & others. *Molecular Cloning: A Laboratory Manual*.Cold Spring Harbor Laboratory press, (2001).


## Acknowledgements

We would like to thank Prof. Toril Lindbäck of the Department of Paraclinical Sciences at the University of Life Sciences (NMBU) in Oslo, Norway for providing us with *E. coli* strain NVH-848. Next, we are grateful to Dr. Roger Stephan of the Institute for Food Safety and Hygiene of the University of Zurich, Switzerland, who provided most of the bacterial *E. coli* host strains. We gratefully acknowledge the help of Stephan Handschin, for performing electron microscopy at the Scientific Center for Optical and Electron Microscopy (ScopeM) of the Federal Institute of Technology (ETH, Zurich). Additionally, we would like to thank Nicola Rhyner (ZHAW) for his assistance during genome sequencing. We are grateful to the HPC team of the School for Life Sciences and Facility Management at ZHAW for computer resources and support. J.F.P. acknowledges support from the Department of Life Sciences and Facility Management of the Zurich University of Applied Sciences (ZHAW) in Wädenswil. C.P. received a fellowship from the Research Foundation Flanders (FWO, 1S79422N).

## Author contributions

C.P.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Software, Data curation, writing - Original Draft, Visualization. L.F.: Conceptualization, Methodology, Project administration, Resources, Supervision, Investigation, Writing - Review & Editing. J.F.P.: Investigation, Data curation, Formal analysis, Resources, Software. Y.B.: Conceptualization, Methodology, Project administration, Supervision, Investigation, Writing - Review & Editing.

## Declarations

### Competing interests
The authors declare no competing interests.

### Additional information
**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-77463-x.

**Correspondence** and requests for materials should be addressed to Y.B.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.