



Subtypes of tail spike proteins predicts the host range of *Ackermannviridae* phages



Anders Nørgaard Sørensen, Cedric Woudstra, Martine C. Holst Sørensen, Lone Brøndsted*

Department of Veterinary and Animal Sciences, University of Copenhagen, Stigbøjlen 4, 1870 Frederiksberg C, Denmark

ARTICLE INFO

Article history:

Received 22 June 2021
Received in revised form 19 August 2021
Accepted 19 August 2021
Available online 21 August 2021

Keywords:

Bacteriophage
Ackermannviridae family
Receptor-binding proteins
Tail spike proteins
Host range
O-antigen
Escherichia coli O:157
Salmonella

ABSTRACT

Phages belonging to the *Ackermannviridae* family encode up to four tail spike proteins (TSPs), each recognizing a specific receptor of their bacterial hosts. Here, we determined the TSPs diversity of 99 *Ackermannviridae* phages by performing a comprehensive *in silico* analysis. Based on sequence diversity, we assigned all TSPs into distinctive subtypes of TSP1, TSP2, TSP3 and TSP4, and found each TSP subtype to be specifically associated with the genera (*Kuttavirus*, *Agtrivirus*, *Limestonevirus*, *Taipeivirus*) of the *Ackermannviridae* family. Further analysis showed that the N-terminal XD1 and XD2 domains in TSP2 and TSP4, hinging the four TSPs together, are preserved. In contrast, the C-terminal receptor binding modules were only conserved within TSP subtypes, except for some *Kuttavirus* TSP1s and TSP3s that were similar to specific TSP4s. A conserved motif in TSP1, TSP3 and TSP4 of *Kuttavirus* phages may allow recombination between receptor binding modules, thus altering host recognition. The receptors for numerous uncharacterized phages expressing TSPs in the same subtypes were predicted using previous host range data. To validate our predictions, we experimentally determined the host recognition of three of the four TSPs expressed by *kuttavirus* S117. We confirmed that S117 TSP1 and TSP2 bind to their predicted host receptors, and identified the receptor for TSP3, which is shared by 51 other *Kuttavirus* phages. *Kuttavirus* phages were thus shown to encode a vast genetic diversity of potentially exchangeable TSPs influencing host recognition. Overall, our study demonstrates that comprehensive *in silico* and host range analysis of TSPs can predict host recognition of *Ackermannviridae* phages.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bacteriophages (phages) are viruses that specifically infect bacteria. As the first step in the infection cycle, tailed phages adsorb to the bacterial surface using their receptor-binding proteins (RBPs) located at the tip of the tail. Such RBPs often form long or short tail fibers or tail spike proteins (TSPs), allowing the phage to bind to specific receptors on the surface of the host bacterium. In contrast to tail fibers, TSPs often have enzymatic activity that bind and degrade polysaccharides structures such as lipopolysaccharides (LPS) or capsular polysaccharides (CPS) [1–3]. In the case of LPS, some TSPs recognize the core oligosaccharides, but more fre-

quently they bind to the external moiety of LPS; the O-antigen [4]. While the core oligosaccharides of LPS are conserved, the O-antigen consisting of a repetitive oligosaccharide of three to five sugars are highly diverse even within a bacterial species. For example, more than 185 different O-antigens have been identified in *E. coli* creating high surface diversity within this species [5]. As the binding of the TSP to the O-antigen is highly specific, phages encode highly diverse TSPs to match the bacterial O-antigen diversity [6].

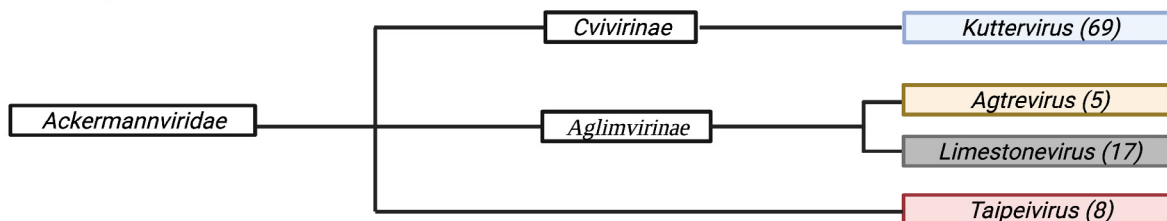
While most phages only express one RBP, some phages express several RBPs that recognize different receptors thereby allowing the phage to infect multiple hosts even if the surface is highly variable [7–9]. Phages previously belonging to the *Viunaviruses* but now part of the *Ackermannviridae* family are known to encode multiple TSPs and infect a wide range of Gram-negative bacteria [10]. At the time of writing, the *Ackermannviridae* family consists of two subfamilies; *Cvivirusinae* and *Aglimvirusinae*, as well as the *Taipeivirus* genus [11] (Fig. 1A). The *Kuttavirus* is the only genus of the *Cvivirusinae* subfamily and contains phages classified as either *Salmonella*

Abbreviations: ANI, Average nucleotide identity; CPS, Capsular polysaccharide; EOP, Efficiency of plating; LB, Luria-Bertani; LPS, Lipopolysaccharide; NCBI, National Center for Biotechnology Information; ORF, Open reading frame; PFU, Plaque formation unit; RBP, Receptor binding protein; TSP, Tail spike protein; VriC, Virulence-associated protein.

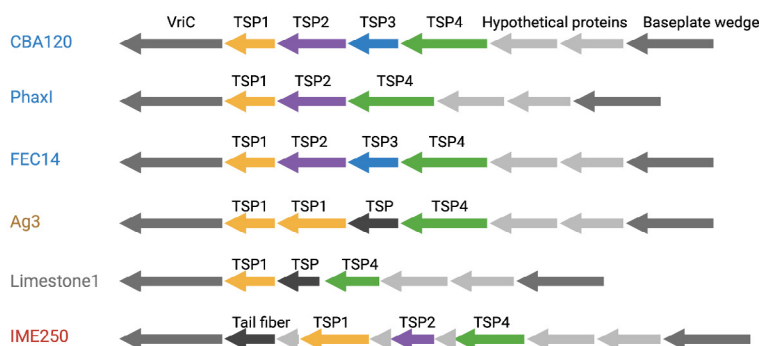
* Corresponding author.

E-mail address: lobr@sund.ku.dk (L. Brøndsted).

A) Taxonomy



B) TSP gene cluster



C) CBA120 receptor-binding complex

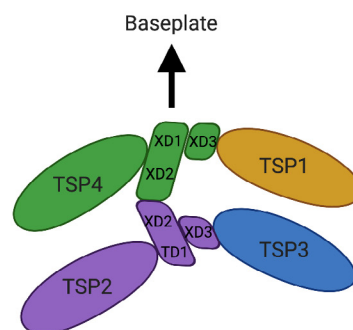


Fig. 1. Taxonomy of the Ackermannviridae family and the tail spike protein (TSP) gene clusters. A) Taxonomy of the Ackermannviridae family with the subfamilies and genera. The number of phages analyzed in this study within the respective genera are presented in the parentheses. B) The organization of the TSP gene clusters in different Ackermannviridae phages. Blue: Kuttervirus phages, olive-yellow: Agtrevirus phage, grey: Limestonevirus phage, red: Taipeivirus phages. TSP genes were colored according to annotation in GeneBank. TSP1: Yellow, TSP2: Purple, TSP3: Blue, TSP4: Green, Not specified: Black. B) Receptor-binding complex consisting of the four TSPs in kuttervirus CBA120 proposed by Plattner et al. Abbreviation: TSP, tail spike protein and VriC, Virulence-associated protein.

or *E. coli* phages [12,13]. The Aglimvirinae subfamily can be divided into the Agtrevirus and Limestonevirus genera that comprise phages infecting *Shigella*, *Enterobacter* and *Salmonella*, or *Dickeya*, respectively [14,15]. Finally, the Taipeivirus genus include phages that infects *Klebsiella*, *E. coli* and *Serratia* [1,16,17]. Most phages belonging to the Ackermannviridae family express four TSPs (TSP1 to 4), however, some phages of the Kuttervirus and Limestonevirus genera only contain three TSP genes (Fig. 1B) [1,14,15,18–20].

Even though the protein sequences of TSPs are highly diverse, their structure is conserved, forming a stable homotrimer in which each monomer displays right-handed β -helices. While the N-termini of TSPs, called the head-binding domain allows binding to the baseplate, the C-terminus contains a catalytic module domain that facilitate binding and degradation of the receptor as well as an intramolecular chaperone domain aiding in trimerization of the proteins [2,20–24]. A recent study of the four TSPs encoded by *E. coli* phage CBA120 belonging to the Kuttervirus genus showed that the N-termini of the four TSPs were distinct and facilitate interactions between the TSPs to form a receptor-binding complex [20]. In CBA120, the N-termini of all four TSPs contains similar tandem domains called TD1 and TD2 with the exception of TSP2 that only contain TD1. In addition, TSP2 and TSP4 contains two and three XD domains (XD1-3), respectively, upstream of the tandem repeats that are structurally similar to the first three domains found in Gp10 in *E. coli* phage T4. In phage T4, the *gp10* encodes a baseplate protein containing four domains that forms a complex with other baseplate proteins (Gp11 and Gp12), which are responsible for attachment of the short tail fibers (Gp12) to the baseplate [25]. In CBA120, Plattner et al. proposed that TSP4 interact with the baseplate through the first 80 amino acid resi-

dues and the XD1 domain. The XD2 of both TSP2 and TSP4 are believed to interact with each other, allowing interaction between the two proteins. TD1 of TSP1 and TSP3 interact with the XD3 of TSP4 and TSP2, respectively. All these interactions are believed to be necessary for forming the receptor-binding complex [20] (Fig. 1C).

In CBA120, each of the four TSPs are able to recognize and bind different O-antigen moieties on the surface of *Salmonella* and *E. coli* [3,20]. Individual purified TSPs showed that while TSP1 binds and degrades the O:210-antigen on *Salmonella*, TSP2, TSP3 and TSP4 recognize O:157, O:77 and O:78 on *E. coli* strains, respectively [3,20]. While it is assumed that Ackermannviridae phages are able to recognize different receptors, like CBA120, most of these phages have only been assigned to one host, mainly because of lack of extended host range analysis. In addition, due to the presence of the multiple TSPs, it is not well understood which TSP is responsible for the recognition of the assigned hosts, even when an extended host range analysis has been performed [26–28]. While some studies of Ackermannviridae phages indicate that these phages encode very diverse sets of TSPs [29,30], other studies have demonstrated that some TSPs could be conserved within the different genera. An example is the TSP encoded by *orf169* of kuttervirus FEC14 which shows 99.5% identity to *orf211* encoding TSP2 in kuttervirus phage CBA120 [18]. Also, a recent study of limestonevirus phage PP35 showed that *orf156* was identical to several TSPs expressed by other Limestonevirus phages [31]. These studies suggest that TSPs in the Ackermannviridae family might be conserved, which could explain why different phages are able to infect the same hosts. Here we aim to reveal the diversity of TSPs encoded by Ackermannviridae phages by performing a comprehensive *in sil*-

ico analysis of these proteins in 99 phages. We found that different TSPs can be assigned into subtypes that correlate with the phage genus within the *Ackermannviridae* family. Furthermore, the TSP subtypes allowed us to predict host recognition of diverse TSPs. Finally, we experimentally determined the host recognition of three TSPs encoded by the *Salmonella* phage S117 (hereafter S117 or kuttervirus S117) belonging to the *Kuttervirus* genus and could confirm that phages expressing TSPs of the same subtypes recognize the same hosts. Our work demonstrates that by investigating TSP similarity and determining the host recognition of a few TSPs, it is possible to assign hosts for numerous uncharacterized phages.

2. Materials and methods

2.1. Bioinformatics analysis

To investigate the diversity of the TSPs in the *Ackermannviridae* family, all available genomes at the GenBank database were extracted (retrieved 24.11.2020). For the analysis, we searched for the well-conserved *virC* gene that is present upstream of the TSP gene cluster. We identified this in 99 phages and afterwards, we numbered the TSPs 1 to 4 according to their N-terminal sequence similarity of the four TSPs in *Kuttervirus* CBA120. All multiple amino acid alignments were done in CLC Workbench 21.0.3 (Qiagen Digital Insights, Aarhus, Denmark) with the default settings; Gap cost 10, gap extension cost 1, end gap cost: as any others and alignment mode: very accurate. Pairwise comparison of the alignment allowed for determination of the percent identity of the TSPs. TSPs with 75% or more identity were assigned into TSP subtypes. Each predicted TSP was verified using HHPRED [32]. Furthermore, to identify the structural domains in the N-termini we used one representative TSP of each subtype for the HHPRED analysis because of the high sequence similarity of the TSPs in each TSP subtype.

Whole genome alignment of the 99 *Ackermannviridae* phage genomes was done in CLC Workbench 21.0.3 (Qiagen Digital Insights, Aarhus, Denmark) with the default settings (minimum initial seed length: 15, allow mismatches in seeds: yes and minimum alignment block length: 100). Pairwise comparison of the analysis was conducted to create a heatmap displaying the average nucleotide identity (ANI) using default settings (table types: ANI, distance measure: euclidean distance and linkage criteria: complete linkage).

2.2. Appendices

Appendix A: overview of all phages analyzed with accession numbers, gene ID of the identified TSPs and the categorized TSP subtypes for each TSP. Appendix B-E: Overview of the TSP1 to TSP4 pair wise comparison and the number of phages expressing the specific TSP subtypes. Appendix F: overview of all bacterial strains used.

Mendeley data: <https://data.mendeley.com/datasets/k6pvnfdzch/2>.

2.3. Bacteria and phage

Salmonella strains for phage propagation, bacterial strains for host range analysis, strains used for TSP cloning and purification and *Salmonella* mutants for TSP3 receptor determination are listed in the Appendix F. *Salmonella* phage S117 (GenBank accession number MH370370.1) [33] and its four TSPs were used in the experimental work.

2.4. Phage propagation

Phage S117 was propagated on *S. Typhimurium* LT2c strain, which is cured from prophages *gifsy-1*, *gifsy-2*, *fels-1* and *fels-2*. First, a single colony from LT2c was inoculated into LB media (Lysogeny Broth, Merck, Darmstadt, Germany) and was grown to the exponential phase at 37 °C and 180 rpm. Then, a previous phage stock of S117 (4×10^{10} pfu/ml) was 10-fold diluted and the dilutions 10^{-3} to 10^{-5} were individually mixed with 100 μ L of the LT2c exponential cell culture. The cell-phage suspension was added to 3.3 mL molten top agar (LBov; LB broth with 0.6% Agar bacteriological no.1, Oxoid) and poured onto a LA plate (LB with 1.2% agar). After the plate was dried for 45 min in a laminar hood, the plate was incubated aerobically at 37 °C overnight. The next day, 5 mL of SM buffer (0.1 M NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 50 mM Tris-HCl, pH 7.5) was added to the plate with abundant plaques and the plate was incubated for 24 h at 4 °C and 50 rpm. The phage suspension in SM buffer was collected and centrifuged for 15 min at 11000 rpm. The supernatant was filtered twice with a 0.2 μ M filter and a 10-fold serial dilution of the new phage stock was prepared to determine the phage titer by phage plaque assay.

2.5. Phage plaque assay

To determine the phage titer, host range and to verify the receptor identified for for TSP3, 10-fold serial dilutions (10^{-1} to 10^{-8}) of the newly prepare S117 phage stock in SM buffer were prepared. Three times 10 μ L of each phage dilutions were spotted onto bacterial lawns and plates were dried for 45 min in a laminar hood. The plates were incubated overnight at 37 °C and the following day, plaques were counted and the plaque forming unit per ml (pfu per ml) was calculated. The efficiency of plating was calculated by comparing the pfu/mL of the S117 screened strains (Appendix F) with the propagated strain *Salmonella* Typhimurium LT2c.

2.6. DNA extraction and purification of phage DNA

The methods for DNA extraction and purification have been described elsewhere [33]. Briefly, the S117 phage stock was filtered three times with 0.2 μ M filters. RNase (10 μ g/mL) (Thermo Fischer Scientific, Waltham, MA, USA) and DNase (20 μ g/mL) (Thermo Fischer Scientific) was added to the lysate and incubated for 20 min at 37 °C in a thermoshaker (500 rpm, Eppendorf, Germany). After incubation, EDTA (pH8) (20 mM) (Thermo Fischer Scientific) and Proteinase K (50 μ g/mL) (Thermo Fischer Scientific) were added followed by incubation for 2 h at 56 °C. A sample was run on an 1% agarose gel to confirm the release of the DNA from the phage capsids. Phenol (Fluka), phenol–chloroform–isoamylalcohol (25:24:1, Ambion) and three rounds of chloroform–isoamylalcohol (24:1) were used for DNA purification. The DNA was precipitated with 0.1 vol 3 M sodium acetate (pH 5.5) (Thermo Fischer Scientific), Glycogen (0.05 μ g/ μ L) (Thermo Fischer Scientific) and 2.5x ice-cold 99% ethanol and incubated at –18 °C for 72 h. The precipitated DNA was centrifuged at 4 °C for 20 min at 12000 rpm and the pellet was washed three times with ice-cold 70% ethanol. Afterwards, the pellet was resuspended in 10 mM Tris-HCl (pH 8). The quality of the purified DNA was confirmed with Nanodrop (Thermo Fischer Scientific) and gel electrophoresis. The concentration was estimated with Qubit (Thermo Fischer Scientific).

2.7. TSP cloning and purification

The four TSPs of phage S117 were individually cloned into pET-28a (+) with a purification His-tag downstream of the TSP genes by

using In Vivo Assembly (IVA) cloning described previously [34]. Briefly, each *tsp* gene (TSP1 (2313 bp), TSP2 (2766 bp), TSP3 (2097 bp) and TSP4 (3504 bp)) were individually amplified with primers carrying overhang homologous to the pET-28a (+) in a PCR tube also containing the pET-28a (+) vector and primers for amplifying the vector. The homologous primer overhangs allow for homologous recombination upon transformation. The TSP genes were cloned into the multiple cloning site between the *Hin*CI and *Eco*52KI restriction sites. Phusion® High-Fidelity DNA polymerase (NEB®) was used to amplify the TSP genes and the pET-28a(+). PCR primers (Supplementary Table 1) were designed with SnapGene® and ordered from TAG Copenhagen A/S. The PCR reaction followed the manufactures description with 1 ng S117 DNA and 1 ng pET-28a(+) as templates with 0.1 μ M primers. Subsequently, 1 μ L FastDigest DpnI enzyme (Thermo Fisher Scientific) was added to the PCR sample and incubated for 15 min incubation at 37 °C to digest the methylated DNA template. The DpnI-digested PCR sample was propagated into Stellar™ Competent Cells (Takara Bio) using the manufactures protocol except for extending the heat shock to 60 sec. The TSP-inserted-plasmid (pET_TSP1 to 4) was isolated with GeneJET Plasmid Miniprep kit (Thermo Scientific) from an overnight culture. PCR on the isolated plasmid was used to validate the presence of the insert and successful plasmid constructions were confirmed by Sanger sequencing (Eurofins Genomics).

pET_TSP1 to 4 were transformed into electrocompetent *E. coli* BL21 cells (Agilent Technologies). To express the TSPs, a single colony of the strain containing each plasmid was inoculated into LB medium at 37 °C and 170 rpm. At optical density (OD_{600}) of 0.6 the culture was cooled on ice prior to addition of 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG) to induce the TSP expression. The culture was then incubated at 16 °C and 110 rpm for 16–18 h. The cells were harvest by centrifugation for 10 min at 12000 rpm and stored in the freezer until further processed. The pellet was resuspended in lysis buffer (0.5 M NaCl, 20 mM Na_2HPO_4 , 50 mM Imidazole, pH 7.4) and the cells disrupted on ice by sonication (Bandelin Sonopul HD 2070 homogeniser). The lysate was filtered twice with 0.22 μ M filters before the His-tagged TSP was purified using HisGraviTrap™ (GE Healthcare) according to the manufacture's description with elution buffer (0.5 M NaCl, 20 mM Na_2HPO_4 , 0.5 M Imidazole, pH 7.4). Amicon® Ultra-15 Centrifugal filter units with 50 kDa cut-off (Merck Milipore) was used to exchange the protein solutions to 20 mM HEPES (pH 7.4). Protein concentration was measured with Qubit™ Protein Assay Kit (Q33211) with a Qubit 2.0 Fluorometer (Invitrogen, Q32866).

2.8. Spot assay

To determine the host recognition for each of the four TSPs a spot assay was carried out. 100 μ L of a bacterial overnight culture was added to 4 mL top agar and poured onto a LB agar plate. Afterwards 1.5 μ g of the four TSPs was spotted onto the lawn and the protein buffer (20 mM HEPES) was used as a negative control. The plates were incubated overnight at 37 °C and the next day the plates were checked for the presence of a translucent clearing zone in the spots with purified TSP proteins indicating TSP degradation of the receptor.

2.9. TSP inhibition assay

To validate the spot assay, we determined if the TSPs were able to inhibit the infection of S117 on the respective hosts. *S. Minnesota* (JE02), *E. coli* O157:H7 (NTCT12900) and *S. Typhimurium* (LT2c) was used as strains for TSP1, TSP2 and TSP3 respectively. The bacteria were grown to OD_{600} of 0.3 and put on ice where the four TSPs (0.5 mg/mL) or nothing were added to 100 μ L of the bacteria. The TSP-bacteria suspension was preincubated at

37 °C for 20 min before being added to top-agar and poured onto a LB agar plate. After the bacterial lawn had dried, three times 10 μ L of a serial dilution (10^{-1} to 10^{-8}) of S117 was spotted onto the lawn. The plates were incubated overnight at 37 °C and the next day plaques were counted and the pfu/mL calculated. The efficiency of plating was calculated by comparing the pfu/mL of the S117 sensitive *Salmonella* Typhimurium LT2c mutant strains with the wild type.

2.10. Statistical analysis

All results were conducted in three independent experiments where the mean standard deviations are shown in the graphs. One-way ANOVA was used to access the statistical significance.

3. Results

3.1. TSPs subtypes are associated with phages genus

All classified *Ackermannviridae* phages encode multiple TSPs located in a gene cluster flanked by a conserved virulence associated protein gene (*vriC*) and a baseplate wedge gene (Fig. 1B). To investigate the diversity of the TSPs in the *Ackermannviridae* family, we extracted all available genomes (133) classified as such from NCBI and searched for the gene cluster using the conserved *vriC* gene as a reference. This approach allowed us to identify the *vriC* gene and the complete downstream TSP cluster in 99 genomes, whereas 34 genomes were discarded from further analysis. Of these, four phages contained the *vriC* gene, but the downstream genes could not be classified as full length TSPs, whereas 30 genomes were unclassified *Ackermannviridae* phages not encoding the *vriC-tsp* loci. We then performed homolog detection and structural prediction by HHPRED [32] of all individual TSP proteins in the 99 genomes to validate that the genes in fact encoded TSPs (results not shown). Among the 99 phages chosen for further analysis, 69 belonged to the *Kuttervirus* genus, five to the *Agtevirus* genus, 17 to the *Limestonevirus* genus and eight to the *Taipeivirus* genus. A detailed overview of all phages analysed as well as the protein ID of the defined TSPs when applicable are shown in Appendix A.

To investigate the conservation and similarity of the *Ackermannviridae* TSPs, we first assigned all TSPs as either TSP1, TSP2, TSP3 or TSP4 based on amino acid similarity to the N-termini of the four well-characterised TSPs of *Kuttervirus* phage CBA120. We identified: 97 TSP1s, 99 TSP2s, 78 TSP3s and 98 TSP4s (Appendix A). We then performed multiple protein alignment of all TSP1, TSP2, TSP3 and TSP4. This analysis showed that the similarity of the TSPs ranged from 5 to 100% identity, demonstrating that the TSPs in the *Ackermannviridae* family can be highly diverse (Appendix B-E). Interestingly, some phages encoded highly similar TSPs with 75–100% identity, yet these TSPs showed only limited similarity (7–30%) to other TSPs assigned as the same type. For instance, four phages encoded TSP1s that were 99–100% identical, whereas the remaining TSP1s only showed \leq 29% identity to these four TSPs (Appendix B). It was previously shown that TSPs sharing sequence similarity could be used to predict the receptor for the TSP [20,35]. For instance, TSP4 of *Kuttervirus* CBA120 were 74–79% identical to TSPs from three G7C-like phages. The authors showed that TSP4 of CBA120 recognize the same receptor as the three G7C-like phages [20]. Thus, we used this information to further assign the TSPs into subtypes using 75% identity as a cut-off value, however, the majority of the TSPs in the specific subtypes were generally more than 95% identical (Appendix B-E). Accordingly we could identify 33, 25, 16 and 23 distinctive subtypes of TSP1, TSP2, TSP3 and TSP4, respectively.

During our analysis, we observed that each TSP subtype could be associated with a specific phage genus with the only exception of subtype TSP4-14 that was encoded by phages belonging to both the *Limestonevirus* and *Agtrevirus* genera (Fig. 2). We found that *Kuttervirus* phages express a more diverse set of TSP1, TSP2 and TSP4 subtypes compared to the other phages. For instance, 21 of 32 TSP1 subtypes identified were expressed by *Kuttervirus* phages (Fig. 2). On the contrary, only six different TSP3 subtypes were identified for *Kuttervirus* phages as 52 of the 69 *Kuttervirus* phages express subtype TSP3-1. In contrast to the *Kuttervirus*, all 17 *Limestonevirus* phages analysed express TSP1 belonging to the same TSP subtype (TSP1-26). Furthermore, 15 out of 17 *Limestonevirus* phages express the subtypes TSP2-16 and TSP4-17, demonstrating that the TSPs within this group of phages are highly conserved. Interestingly, we did not identify a TSP3 in any of the *Limestonevirus* phage genomes. In the *Agtrevirus* genus, the five phages overall expressed more unique TSPs, as all of the five TSP2 and TSP3 grouped into separate subtypes. However, TSP1 of two phages grouped into TSP1-24, and four phages encoded TSP4-14, the same subtype as the *Limestonevirus* phages; RC-2014 and phiDP10.3 (Fig. 2). Similar to the *Agtrevirus* phages, the eight phages in the *Taipeivirus* genus also encode unique TSPs, as only one or a few phages encode TSPs of the same subtypes. Similar to the *Agtrevirus* phages, the eight phages in the *Taipeivirus* genus also encode unique TSPs, as only one or a few phages encode TSPs of the same subtypes.

The 69 of the 99 *Ackermannviridae* genomes analysed were *Kuttervirus* phages while only five, 17 and eight genomes belonged to *Agtrevirus* phages, *Limestonevirus* phages and *Taipeivirus* phages, respectively. Thus, we speculated that the larger diversity observed in the TSPs for the *Kuttervirus* and the highly similar TSPs found in *Limestonevirus* may result from the overall genomic diversity found within the representatives analysed for each genus. We therefore performed whole genome analysis of all 99 phages genomes and found that the *Limestonevirus* phages had a average nucleotide identity (ANI) between 96 and 100%, whereas the phages in the other genera were not as similar (*Kuttervirus* phages: ANI of 88–100%, *Agtrevirus* phages: ANI of 89–96% and *Taipeivirus*

phages: ANI of 86–97%) (Fig. S1). However, phages in the *Kuttervirus* genus that shared an ANI of 99–100% also expressed TSPs categorized into the same subtypes. Thus, as the genomes of the *Limestonevirus* phages analysed are highly conserved this may explain why these phages express similar TSPs compared to the other phage genera. Overall, we found that phages in the *Ackermannviridae* family express highly diverse TSPs that can be grouped into subtypes based on similarity. Furthermore, each TSP subtype were specifically associated with the phage genera.

3.2. The N-terminal XD domains of TSP4 are conserved within *Ackermannviridae* phages

It was previously suggested that the XD domains located in the N-termini of TSP2 and TSP4 are essential for forming the branched receptor-binding protein complex as observed in phage CBA120 [20]. To investigate if such receptor-binding complex are preserved in *Ackermannviridae* phages, we analyzed the N-termini of all TSP subtypes for homology detection and structure prediction using HHPRED. In phage CBA120 TSP2 and TSP4, the XD domains are located within the first 300 and 500 residues, respectively, thus we used these regions for our analysis [20].

For all TSP2 subtypes, we found high similarity to the N-terminus of TSP2 of CBA120 (PDB ID 6W4Q) with a probability ranging from 98.16 to 100% and e-values spanning from $4e^{-11}$ to $7.7e^{-151}$ (Table 1). For some TSP2 subtypes, only the first 85 residues were similar to TSP2 of CBA120 (Table 1), which correlates with the proposed XD2 domain that spans the first 88 residues in TSP2 of CBA120. The XD domains of TSP2 of phage CBA120 were identified based on structural similarity to domain two and three in Gp10 of phage T4 (PDB ID 5XH2). We did, however, not detect such similarity in any of the TSP2 subtypes. Yet, all TSP2 subtypes were structurally similar to TSP2 of CBA120, suggesting an overall similar fold of the N-termini of TSP2 subtypes in *Ackermannviridae* phages. Analysis of the TSP4 subtypes showed structural similarity to the Gp10 protein (PDB ID 5XH2) encoded by phage T4 as previously observed for TSP4 of CBA120 [20] (Table 2). In addition, we also found similarity to Gp9 of T4 (PDB ID 12SE) from the first resi-

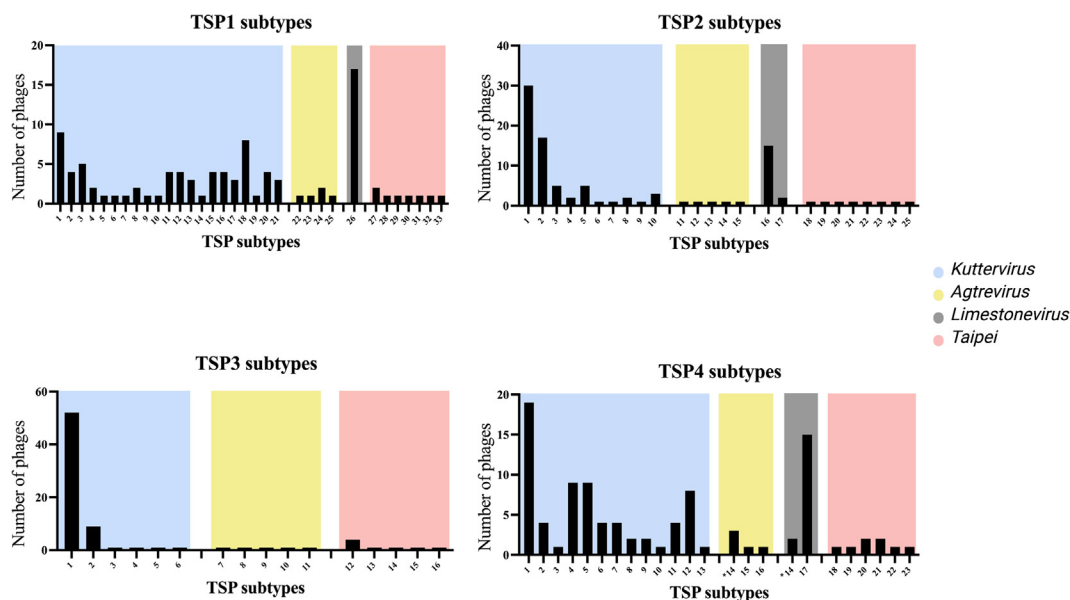


Fig. 2. TSP subtypes in the *Ackermannviridae* family correlate with the phage genus. TSP genes were identified in the phage genomes and grouped into TSP1, TSP2, TSP3 and TSP4 and aligned. TSPs that were 75% or more identical were grouped into subtypes. The number of phages grouped into the individual TSP subtypes are illustrated on the y-axis. Blue: TSP subtypes expressed by *Kuttervirus* phages, Olive-yellow: TSP subtypes expressed by *Agtrevirus* phages, Grey: TSP subtypes expressed by *Limestonevirus* phages and Red: TSP subtypes expressed by *Taipeivirus* phages. Asterisk: TSPs in the TSP4-14 subtype are expressed by *Agtrevirus* and *Limestonevirus* phages. The figure was generated using Prism 9.

Table 1
HHPRED analysis overview of the N-terminal structural similarity in the TSP2 subtypes.

TSP2 subtypes representatives ^a	Position of similarity between TSP compared to the position of the HHPRED hit	PDB ID	probability	e-value
TSP2-1 S117	1-300 aa vs. 1-249 aa of CBA120 TSP2	6W4Q	100%	1.8E-32
TSP2-2 Matapan	1-269 aa vs. 1-269 aa of CBA120 TSP2	6W4Q	99.94%	1.6E-25
TSP2-3 Sh19	1-269 aa vs. 1-269 aa of CBA120 TSP2	6W4Q	99.9%	1.4E-22
TSP2-4 SS3	1-297 aa vs. 1-292 aa of CBA120 TSP2	6W4Q	99.93%	4.5E-24
TSP2-5 SeTs-2	1-273 aa vs. 1-255 aa of CBA120 TSP2	6W4Q	99.96%	1.5E-27
TSP2-6 SJ2	1-295 aa vs. 1-295 aa of CBA120 TSP2	6W4Q	99.96%	4.6E-27
TSP2-7 ViI	1-265 aa vs. 1-269 aa of CBA120 TSP2	6W4Q	98.69%	4.1E-7
TSP2-8 Mutine	1-299 aa vs. 1-358 aa of CBA120 TSP2	6W4Q	99.95%	4.1E-26
TSP2-9 LPST94	1-266 aa vs. 1-266 aa of CBA120 TSP2	6W4Q	99.95%	5E-27
TSP2-10 Det7	1-85 aa vs. 1-88 aa of CBA120 TSP	6W4Q	99.62%	7.1E-14
TSP2-11 P46FS4	1-85 aa vs. 1-86 aa of CBA120 TSP2	6W4Q	99.44%	1.3E-11
TSP2-12 AG3	1-82 aa vs. 1-83 aa of CBA120 TSP2	6W4Q	99.46%	3.7E-12
TSP2-13 MK-13	1-86 aa vs. 1-87 aa of CBA120 TSP2	6W4Q	99.63%	1.4E-13
TSP2-14 SKML39	1-267 aa vs. 1-297 aa of CBA120 TSP2	6W4Q	99.87%	1.3E-19
TSP2-15 PhiEM4	1-300 aa vs. 1-300 aa of CBA120 TSP2	6W4Q	100%	6.7E-42
TSP2-16 Limestonevirus	1-176 aa vs. 1-188 aa of CBA120 TSP2	6W4Q	99.25%	6E-11
TSP2-17 RC-2014	1-248 aa vs. 1-244 aa of CBA120 TSP2	6W4Q	99.84%	1.6E-19
TSP2-18 0507KN21	6-161 aa vs. 1-155 aa of CBA120 TSP2	6W4Q	99.29%	2.7E-10
TSP2-19 May	1-290 aa vs. 1-358 aa of CBA120 TSP2	6W4Q	99.29%	7.5E-24
TSP2-20 KpS110	1-300 aa vs. 1-357 aa of CBA120 TSP2	6W4Q	99.97%	1.9E-29
TSP2-21 Menlow	1-84 aa vs. 1-84 aa of CBA120 TSP2	6W4Q	99.67%	3.8E-15
TSP2-22 Magnus	1-202 aa vs. 1-194 aa of CBA120 TSP2	6W4Q	99.93%	4E-11
TSP2-23 KWBSE43-6	1-300 aa vs. 1-358 aa of CBA120 TSP2	6W4Q	99.97%	3.2E-29
TSP2-24 UPM2146	6-129 aa vs. 1-124 aa of CBA120 TSP2	6W4Q	99.84%	4.5E-19
TSP2-25 IME250	1-114 aa vs. 1-114 aa of CBA120 TSP2	6W4Q	99.84%	6.2E-19

^a Subtypes expressed by phages belonging to *Kuttervirus* (blue), *Agtevirus* (yellow), *Limestonevirus* (grey), *Taipeivirus* (red).

due to residue 217 with a general higher probability than T4 Gp10 (Table 2). Yet, it was previously shown that Gp10 and Gp9 have similar folding in the N-termini, which explains our findings [25]. While all TSP4s have structural similarity to the first two domains in T4 Gp10, hence XD1 and XD2 domains, there were no structural similarity to domain 3 (XD3) of Gp10 in some TSP4s (Table 2). We did not find any distinctive structural domains in the N-termini of TSP1 and TSP3. Overall, our analysis showed that the structural domains XD1 and XD2 proposed to be necessary for hinging the receptor-binding complex together are preserved in all *Ackerman-nviridae* phages, whereas the XD3 domain interacting with TSP1 and TSP3 are not always preserved.

3.3. Conserved receptor binding domains are found in TSP1, TSP3 and TSP4 subtypes suggesting recombination events

While the N-terminal regions of each of TSP1, TSP2, TSP3 and TSP4 were highly conserved, the C-terminal regions were highly diverse, even among TSP subtypes (Fig. 2 and Appendix B-E). However, when we aligned all TSP1, TSP2, TSP3 and TSP4, we observed that among *Kuttervirus* phages, some TSP1 and TSP3 amino acid sequences were similar to TSP4 sequences except for the XD domains found in TSP4 subtypes (Figs. 3 and S1). More precisely, sequence similarity was observed between TSP1-1 and TSP4-7 subtypes (Fig. 3A), TSP3-3 and TSP4-2 subtypes (Fig. S2A), and TSP3-1 and TSP4-8 subtypes (Fig. S2B). The similarity starts at amino acid position 65 for the TSP1s and TSP3s and position 391 for the TSP4s (Fig. 3B). When we aligned the sequences from position 65 and 391, respectively, TSPs belonging to subtype TSP1-1 and TSP4-7

were 92–100% identical, TSPs belonging to TSP3-3 and TSP4-2 were 93–100% identical, and TSPs belonging to TSP3-1 and TSP4-8 subtypes were 91–100% identical (results not shown). The results suggest that the same receptor binding module can be located on different TSP types.

It has previously been suggested that the β -helix, hence the receptor binding module, can be swapped among TSPs thereby changing the host range of the phage [2,36]. When using the structural domains defined in the CBA120 TSP1 as a reference, we identified a conserved sequence motif; GTTAVSL at the distal part of the TD1 domains of subtypes in the TSP1-1, TSP4-7, TSP3-3, TSP4-2, TSP3-1, TSP4-8 subtypes showing similarity across TSPs (Fig. 3B). We therefore speculated if the end of the TD1 domain could form a conserved site, allowing recombination between the TSPs. To further investigate if this region is conserved in all TSP1, TSP2, TSP3 and TSP4s, we extracted the sequences spanning the TD1 domain from all TSP subtypes (region 1–96 aa in TSP1 and TSP3, region 157–257 aa in TSP2 and region 344–421 aa in TSP4 [20]) and aligned them. Firstly, across phage genera only little sequence similarity of the TD1 domains of TSPs was observed (results not shown). Secondly, the sequence spanning the TD1 domain in *Kuttervirus* TSP2 subtypes were not similar to the TD1 domain in any other *Kuttervirus* TSP subtypes (results not shown). However, when we aligned sequences spanning the TD1 domains of all *Kuttervirus* TSP1, TSP3 and TSP4 subtypes, we found an overall conservation, where the GTTAVSL motif was present in many of the TSP subtypes with the exception of TSP4-13 where only limited similarity was found (Fig. S3). These observations further suggest that a conserved site in the TD1 domains of TSP1, TSP3 and TSP4 in

Table 2
HHPRED analysis overview of the N-terminal structural similarity in the TSP4 subtypes.

TSP4 subtypes representatives ^a	Position of similarity	HHPRED hit	PDB ID	probability	e-value	Preserved XD domains
TSP4-1 S117	79-249 aa vs. 46-247 aa in T4 Gp10	5XH2	98.21%	6.2E-5	XD1-XD2	
	80-176 aa vs. 65-164 aa in T4 Gp9	12SE	99.44%	8.7E-12		
TSP4-2 CBA120	82-329 aa vs. 49-383 aa in T4 Gp10	5XH2	98.67%	5.4E-6	XD1-XD2- XD3	
	82-197 aa vs. 67-185 aa in T4 Gp9	12SE	99.5%	5.4E-12		
TSP4-3 Heyday	77-252 aa vs. 44-250 aa in T4 Gp10	5XH2	99.87%	1.5E-21	XD1-XD2	
	73-233 aa vs. 58-207 aa in T4 Gp9	12SE	99.74%	3E-17		
TSP4-4 GG32	82-330 aa vs. 49-383 aa in T4 Gp10	5XH2	98.56%	1.5E-5	XD1-XD2- XD3	
	82-176 aa vs. 67-164 aa in T4 Gp9	12SE	99.41%	3.9E-12		
TSP4-5 Sh19	82-329 aa vs. 49-383 aa in T4 Gp10	5XH2	98.77%	2.2E-6	XD1-XD2- XD3	
	82-176 aa vs. 67-164 aa in T4 Gp9	12SE	99.45%	3E-12		
TSP4-6 LPST94	82-329 aa vs. 49-383 aa in T4 Gp10	5XH2	99.15%	1.8E-6	XD1-XD2- XD3	
	82-209 aa vs. 67-198 aa in T4 Gp9	12SE	99.47%	6.7E-12		
TSP4-7 FEC14	55-248 aa vs. 25-246 aa in T4 Gp10	5XH2	99.94%	6E-26	XD1-XD2	
	40-221 aa vs. 9-215 aa in T4 Gp9	12SE	99.85%	1.1E-20		
TSP4-8 SE-14	82-329 aa vs. 49-383 aa in T4 Gp10	5XH2	98.65%	6E-6	XD1-XD2- XD3	
	82-197 aa vs. 67-185 aa in T4 Gp9	12SE	99.46%	1.2E-11		
TSP4-9 Det7	82-329 aa vs. 49-383 aa in T4 Gp10	5XH2	98.75%	3.7E-6	XD1-XD2- XD3	
	82-197 aa vs. 67-185 aa in T4 Gp9	12SE	99.47%	7.7E-12		
TSP4-10 Mutine	49-249 aa vs. 19-247 aa in T4 Gp10	5XH2	99.66%	3.9E-15	XD1-XD2	
	50-177 aa vs. 25-165 aa in T4 Gp9	12SE	99.78%	9.2E-18		
TSP4-11 SJ3	82-329 aa vs. 49-383 aa in T4 Gp10	5XH2	98.63%	1.1E-5	XD1-XD2- XD3	
	82-197 aa vs. 67-185 aa in T4 Gp9	12SE	99.48%	6.4E-12		
TSP4-12 Mooltan	79-250 aa vs. 46-248 aa in T4 Gp10	5XH2	99.91%	2.5E-22	XD1-XD2	
	78-216 aa vs. 63-207 aa in T4 Gp9	12SE	99.74%	1.8E-17		
TSP4-13 Bering	79-250 aa vs. 46-248 aa in T4 Gp10	5XH2	99.91%	2E-22	XD1-XD2	
	78-216 aa vs. 63-207 aa in T4 Gp9	12SE	99.74%	2.4E-17		
TSP4-14 AG3	44-251 aa vs. 14-249 aa in T4 Gp10	5HX2	100%	7.8E-32	XD1-XD2	
	41-216 aa vs. 16-207 aa in T4 Gp9	12SE	99.92%	2.24E-24		
TSP4-15 MK-13	44-250 aa vs. 11-248 aa in T4 Gp10	5XH2	100%	1E-31	XD1-XD2	
	43-223 aa vs. 15-217 aa in T4 Gp9	12SE	99.95%	5.1E-27		
TSP4-16 PhiEM4	45-250 aa vs. 13-248 aa in T4 Gp10	5XH2	100%	8E-32	XD1-XD2	
	43-223 aa vs. 15-217 aa in T4 Gp9	12SE	99.94%	9.6E-26		
TSP4-17 Limestone	44-247 aa vs. 10-249 aa in T4 Gp10	5XH2	99.97%	1.6E-30	XD1-XD2	
	40-309 aa vs. 14-207 aa in T4 Gp9	12SE	99.93%	1.4E-25		
TSP4-18 KpS110	83-251 aa vs. 48-247 aa in T4 Gp10	5XH2	99.86%	2.3E-20	XD1-XD2	
	81-184 aa vs. 64-207 aa in T4 Gp9	12SE	99.69%	6.6E-16		
TSP4-19 0507KN21	84-251 aa vs. 60-247 aa in T4 Gp10	5XH2	99.87%	1.4E-20	XD1-XD2	
	93-218 aa vs. 77-207 aa in T4 Gp9	12SE	99.67%	5E-15		
TSP4-20 UMP2146	84-332 aa vs. 49-383 aa in T4 Gp10	5XH2	98.52%	2.3E-5	XD1-XD2- XD3	
	83-178 aa vs. 66-164 aa in T4 Gp9	12SE	99.32%	2.5E-11		
TSP4-21 Menlow	83-251 aa vs. 48-247 aa in T4 Gp10	5XH2	99.84%	3.7E-19	XD1-XD2	
	82-218 aa vs. 65-207 aa in T4 Gp9	12SE	99.65%	2.3E-15		
TSP4-22 May	83-252 aa vs. 48-248 aa in T4 Gp10	5XH2	99.75%	2.9E-17	XD1-XD2	
	81-218 aa vs. 64-207 aa in T4 Gp9	12SE	99.64%	2.4E-15		
TSP4-23 IME250	92-251 aa vs. 59-207 aa in T4 Gp10	5XH2	99.89%	3.6E-21	XD1-XD2	
	83-218 aa vs. 66-207 aa in T4 Gp9	12SE	99.68%	3.9E-16		

^a Subtypes expressed by phages belonging to *Kuttervirus* (Blue), *Agtrevirus* (yellow), *Limestonevirus* (grey), *Taipeivirus* (red).

the *Kuttervirus* phages may be involved in swapping the receptor binding regions of these TSPs, thus providing a mechanism for alternate host recognition. Overall, our results demonstrate

sequence similarity in the β -helix and C-terminal domains between some TSP1, TSP3 and TSP4 subtypes in the *Kuttervirus* genus. Furthermore, we identified a conserved site in the TD1

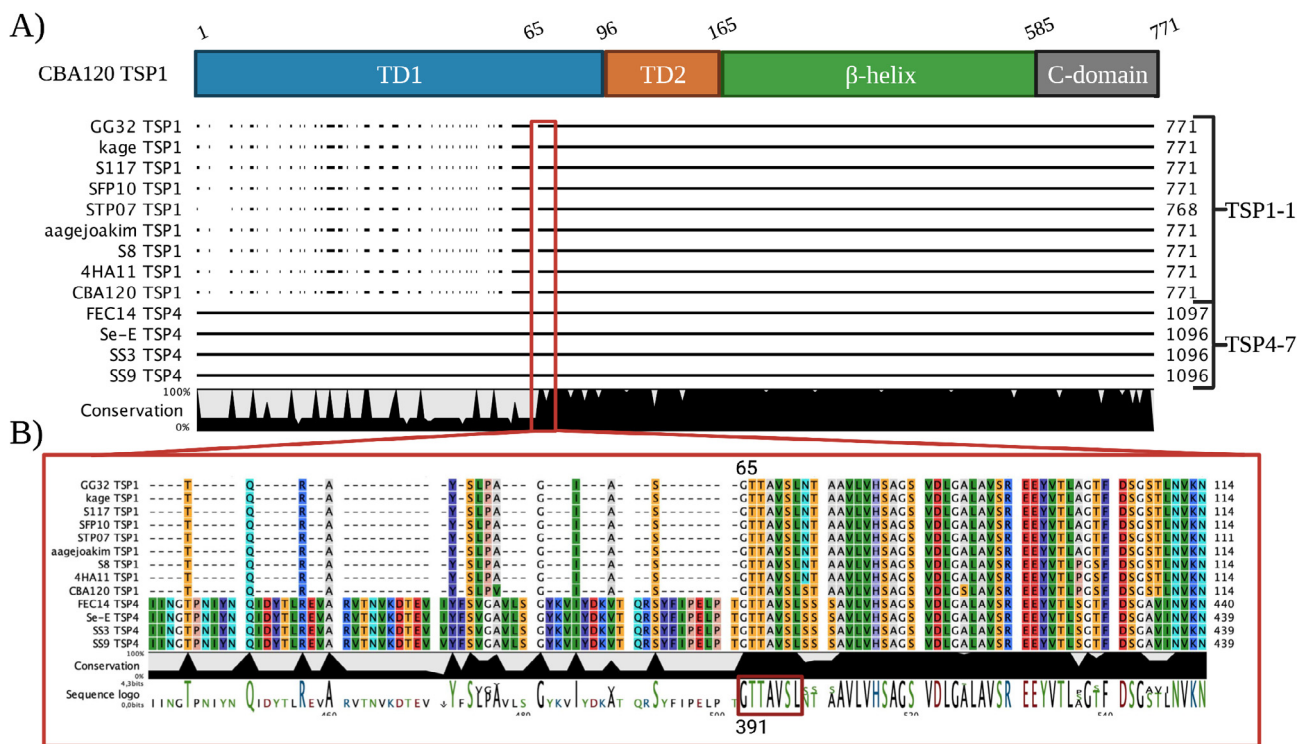


Fig. 3. The receptor binding domains can be swapped between TSPs in the *Kuttervirus* phages. A) All TSP1s belonging to subtype 1–1 were aligned with all TSP4-7. B) Zoom in on the start positions of similarity of the TSPs. The alignment shows a conserved sequence motif that starts at the same amino acid position 65 and 391 for TSP1s and TSP4s, respectively. The structural domains of CBA120 TSP1 was used as a reference.

domains of these TSPs, which potentially could promote recombination events among TSPs in this genus.

3.4. TSP subtypes can be used to predict hosts bacteria of *Ackermannviridae* phages

Our *in silico* analysis showed that the majority of the *Ackermannviridae* phages encode four TSPs, which most likely recognize different bacterial receptors, suggesting that phages in this family have a wide host range. In general, phages belonging to the four *Ackermannviridae* genera have been shown to infect diverse Gram-negative bacteria [1,14,33,37]. However, for most of the phages, a detailed host range analysis has not been conducted. In addition, the respective hosts for each of the TSPs have only been shown for *kuttervirus* CBA120 [20]. To investigate if we could associate TSP subtypes with host recognition, we used published data of TSPs already characterized in terms of receptor recognition.

TSP1 of *kuttervirus* CBA120 was assigned to the TSP1-1 subtype and a total of 9 phages encode this subtype (Fig. 2). TSP1 of CBA120 recognize the O:21O-antigen of *Salmonella* Minnesota, thus we predict that all phages expressing the TSP1-1 subtype also recognize the O:21 antigen (Table 3). In addition, our bioinformatic analysis showed that TSP1-1 was almost identical to TSP4-7 except for the N-terminal region containing the XD domains promoting hinging of the receptor binding complex together (Fig. 3). It is therefore likely that the four phages encoding TSP4-7 subtypes recognize *Salmonella enterica* O:21 serotypes as well (Table 3). TSP2 of CBA120, belonging to the TSP2-1 subtype, binds and degrades the O:157O-antigen on Shiga toxin (Stx)-producing *Escherichia coli* (STEC) strains, suggesting that all 32 phages expressing TSP2-1 s recognize the O:157O-antigen [20]. This is also supported by the fact that 11 out of the 32 phages expressing TSP2s belonging to the TSP2-1 subtype are already known to infect O:157 *E. coli* strains [12,18,19,30,33,38–41]. The last two TSPs of CBA120, TSP3 (TSP3-

4) and TSP4 (TSP4-2), recognize *E. coli* O:77 and O:78, respectively. While CBA120 is the only phage that express a TSP3-4 subtype, CBA120 and 3 other *Kuttervirus* phages express TSP4-2 and thus may recognize *E. coli* O:78. In addition, we showed that the C-terminal domains of TSP4-2 and TSP3-3 subtypes are almost identical, suggesting that TSPs belonging to the TSP3-3 subtype also recognize *E. coli* O:78 (Fig. 4 and Table 1).

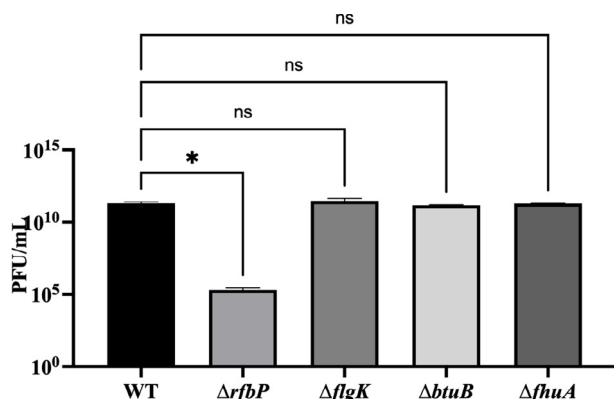
The receptors of TSP2 (TSP2-10) and TSP3 (TSP3-2) of *Kuttervirus* Det7 have previously also been determined [42,43]. TSP2 of Det7 binds to the O:30-antigen from *S. Anatum* [42]. Thus, we predict that the two phages; SenM-2 and Barely that also express a TSP2-10 subtype also recognize *S. Anatum*, which indeed has been shown for phage SenM-2 [43]. TSP3of phage Det7 binds and degrades the O:40-antigen of *S. Typhimurium* [21,44]. A total of nine phages express the TSP3-2 subtype, thus we predict that these nine phages bind to *S. Typhimurium* O:4. Indeed, *S. Typhimurium* has already been identified as a host of four of these nine phages (EP75, Mutine, PM10 and SenM-2) [29,43,45,46]. A recent study of the four TSPs of EP75 showed that TSP1 in the TSP1-3 subtype recognize the O:18A O-antigen of *E. coli* [46]. It is, therefore, likely that the four other phages expressing this subtype recognize the O:18A O-antigen (Table 3).

The TSP1-26 of limestonevirus phage PP35 recognize the O-antigen; (→2)-β-D-6-deoxy-D-altrose-1) →) of its hosts *Dickeya solani* and *Lelliottia* sp. F154 [31]. Our analysis revealed that all *Limestonevirus* phages express TSP1-26, thus the results suggest that all phages in the *Limestonevirus* genus can bind and degrade this LPS. This is in accordance with the known hosts of almost all *Limestonevirus* phages [14,31,47–53]. Finally, TSP2 (TSP2-18) of taipeivirus 0507KN21 recognize the KN2 capsular polysaccharide [17]. This phage is the only *Ackermannviridae* phage expressing a TSP2 of that subtype (Table 3). Overall, we use our *in silico* analysis and already published experimental data to associate the presence of specific TSPs with potential hosts of phages belonging to the

Table 3
Host range predictions of Ackermannviridae phages based on the TSP subtypes.

Identified/predicted host	TSP subtype (s)	Phages in total	Ackermannviridae phages expressing TSPs in the same TSP subtype
S. Ruiru and S. Minnesota O:21	TSP1-1	9	GG32, Kage, S117, SFP10, STP07, Aagejoakim, S8, 4HA11 and CBA120 [20]
	TSP4-7	4	FEC14, SS3, SS9 and Se-E
<i>E. coli</i> O:157	TSP2-1	32	Rabagast, Moki, Heyday, SeHz-1, S118, S115, Se-U, Pa-sanjiao, S8, Se-D, Se-B, BSP101, Kage, SFP10, STP07, Se_EM1, GG32, 4HA11, Se-E, FEC14, S117, EP75, SJ3, PhaxI, FSL_SP-063, FSL_SP-029, STML-13-1, ECML-4, Aagejoakim and CBA120 [20]
S. Typhimurium, S. Derby, S. 4.12:i:-, S. 4.5.12:i:-, S. Enteritidis, S. Goettingen O:4 and O:9	TSP3-1	52	TSP3-1:PS5, Marshall, Mooltan, STML-13-1, Matapan, Maynard, FSL-SP-029, FSL-SP-063, SJ2, Sal157lw, Bering, ISTEP3, Se-H, BSP101, Se-I, Se-J, Se-N, Se-S, SP1, Dinky, SenASZ3, Sh19, Moki, S117, Pa-sanjiao, S115, S118, Se-U, SFP10, Aagejoakim, Kage, Rabagast, 4HA11, SeHz-1, Se-E, Heyday, S8, GG32, STP07, Se-F, Se-G, SenALZ1, SeTs-2, Se-B, Se-D, Pertopsoe, SS3, SeSz-1, Se_AO1, Se_EM3, Se_EM1, and Se_EM4
<i>E. coli</i> O:77	TSP4-8	2	SE14 and mane
<i>E. coli</i> O:78	TSP3-4	1	CBA120 [20]
	TSP4-2	4	BSP101, S8, PhaxI and CBA120 [20]
	TSP3-3	1	FEC14
S. Typhimurium O:4	TSP3-2	9	SJ3, EP75, LPSTP4, Mutine, PM10, Chennai, SenM-2 and barely and Det7 [21]
S. Anatum O:3	TSP2-10	3	SenM-2 and barely and Det7 [42]
<i>E. coli</i> O:18A	TSP1-3	5	SeTs-2, Se-F, Se-G and SenALZ1 and EP75 [46]
<i>Dickeya Solani</i> and <i>Lelliottia</i> F154	TSP1-26	17	PhiDP23.1, Coodle, XF4, PhiD3, Kamild, Limestone1, JA15, PhiDP10.3, RC-2014, Ds16CZ, Ds9CZ, Ds5CZ, Ds3CZ, Ds23CZ, Ds20CZ and Ds25CZ and PP35 [31]
<i>Klebsiella</i> KN2 capsular	TSP2-18	1	0507-KN2-1 [17]

S117 on *Salmonella* Typhimurium mutants



LT2C strains	Translucent zone by TSP3	EOP (±SD)
WT	+	1 (±0.29)
ΔrfbP	-	1.02E-06 (±0.35)
ΔflgK	+	1.48 (±0.47)
ΔbtuB	+	0.72 (±0.1)
ΔfhuA	+	0.98 (±0.01)

Fig. 4. TSP3-1 was not able to produce a translucent zone on an LPS mutant (ΔrfbP). Phage S117 and TSP3-1 were spotted on *Salmonella*. Typhimurium (LT2c) mutant strains lacking known phage receptors; O-Ag (rfbP), flagella (flgK), ferrichrome transporter (fhuA) and vitamin B₁₂ transporter (btuB). The experiment was carried out in triplicates and the error bars represent the standard deviation. The graph was generated in prism9 where the p-values were calculated using the ordinary one-way ANOVA. +: appearance of a translucent zone, -: no zone.

Ackermannviridae family. Revealing the TSP homology between phages in the Ackermannviridae family, allowed us to assign the receptors for many phages by only knowing the receptor recognition for a few TSPs.

3.5. Three TSPs of Kuttervirus phage S117 recognize specific hosts expressing different O-antigens

To further validate our host range predictions, we expressed and purified the four TSPs of kuttervirus S117 phage from our collection [33]. From an extensive host range analysis, we previously showed that S117 infects *E. coli* O:157 as well as several *Salmonella enterica* serotypes [33]. Our bioinformatic analysis assigned the four TSPs of phage S117 into the following subtypes: TSP1-1 (9 phages), TSP2-1 (32 phages), TSP3-1 (52 phages) and TSP4-1 (18 phages) and as indicated in brackets these subtypes are conserved in many other Kuttervirus phages (Fig. 2). Furthermore, the *in silico* analysis predicts that TSP1 and TSP2 of phage S117 target O:21 of *Salmonella* and O:157 of *E. coli*, respectively. Purified TSPs at 95° for 10 min ran as monomers during SDS-PAGE analysis, whereas sizes corresponding to TSP trimerization could be detected on the gels without heating, suggesting that the TSPs were correctly folded and thereby enzymatically functional (Fig. S4). These properties are a common characteristic for TSPs [54]. TSPs are known for their enzymatic activity towards their receptors, and when spotted on a bacterial lawn, a translucent zone appears as the TSPs degrade the polysaccharide receptor [17,55,56]. Purified TSPs as well as phage S117 were spotted on 65 *Salmonella enterica* strains, 76 *E. coli* strains and 7 other gram-negative strains to identify the host of each of the TSPs (Table 4). This allowed us to identify the TSP responsible for binding to a specific host and identified the O-antigen recognized by TSP1, TSP2 and TSP3, while host recognition for TSP4 could not be found. Our analysis showed that TSP1 was able to bind to different *Salmonella enterica* serovars belonging to O:21 serotype (Table 4), whereas TSP2 was indeed able to recognize *E. coli* O:157 strains (Table 4), in agreement with our *in silico* predictions (Table 3).

TSP3 from phage S117 belonging to subtype TSP3-1, showed the broadest host range as it bound to *Salmonella enterica* serotypes expressing O-antigens of both type O:4 and O:9. The backbone sugar residues in the O:4 and O:9 serogroups are identical (→2)-α-D-Man-(1 → 4)-α-L-Rha-(1 → 3)-α-D-Gal-(1 →), only the branched sugars are different, where tyvelose residues are present in the O:9 group and abequose residues are found in the O:4 group on the mannose sugar residue [57]. The similarity of the O-antigens could explain why TSP3 was able to form translucent clear-

Table 4

Summary of the host range analysis of S117 and its four TSPs. Successful phage infection or detection of a translucent zone are indicated in bold. A detailed description of the strains used can be found in Appendix F.

Genus	Species or serovar	O-antigen	Infected or translucent zone/Tested strains				
			S117	TSP1-1	TSP2-1	TSP3-1	TSP4-1
<i>Salmonella</i>	Derby	O:4	2/17	0/17	0/17	2/17	0/17
<i>Salmonella</i>	Typhimurium	O:4	6/8	0/8	0/8	6/8	0/8
<i>Salmonella</i>	4.12:i:-	O:4	2/4	0/4	0/4	2/4	0/4
<i>Salmonella</i>	4.5.12:i:-	O:4	5/6	0/6	0/6	5/6	0/6
<i>Salmonella</i>	Bradenburg	O:4	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Bradford	O:4	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Dublin	O:9	4/4	0/4	0/4	4/4	0/4
<i>Salmonella</i>	Enteritidis	O:9	3/3	0/3	0/3	3/3	0/3
<i>Salmonella</i>	Infantis	O:7	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Seftenberg	O:3	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Adelaide	O:35	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Wesleco	O:42	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Montevideo	O:54	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Tanger	O:13	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Cerro	O:18	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Basel	O:58	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Anatum	O:3,10	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Eilbeck	O:61	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Worthington	O:13	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Onderstepoort	O:6	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Deversoir	O:45	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Telaviv	O:28	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Munester	O:3	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Aberdeen	O:11	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Inverness	O:38	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Bergen	O:47	0/1	0/1	0/1	0/1	0/1
<i>Salmonella</i>	Ruiru	O:21	1/1	1/1	0/1	0/1	0/1
<i>Salmonella</i>	Minnesota	O:21	2/2	2/2	0/2	0/2	0/2
<i>Escherichia coli</i>	coli	O:157	3/3	0/3	3/3	0/3	0/3
<i>Escherichia coli</i>	coli (K12)		1/1*	0/1	0/1	0/1	0/1
<i>Escherichia coli</i>	coli (ECOR)		1/72*	0/72	0/72	0/72	0/72
<i>Shigella</i>	sonnei		1/1*	0/1	0/1	0/1	0/1
<i>Acinetobacter</i>	baumanni		0/1	0/1	0/1	0/1	0/1
<i>Klebsiella</i>	spp.		0/3	0/3	0/3	0/3	0/3
<i>Enterobacter</i>	aerogenes		0/1	0/1	0/1	0/1	0/1
<i>Cronobacter</i>	sakazakii		0/1	0/1	0/1	0/1	0/1
<i>Citrobacter</i>	freundii		0/1	0/1	0/1	0/1	0/1

*) Highly reduced infection (EOP of 10^{-6}).

ing zones on to strains expressing both serogroups, similar to phages P22, KB1 and P27 which also recognize both O-antigens [6,58]. However, TSP3 of S117 did not form translucent zones on all strains belonging to O:4 and O:9 serogroups (Table 4), suggesting that some strains may have modified O-antigens, like acetylation which have been implicated in phage resistance in *Salmonella* Typhimurium [59]. To validate that the O-antigen indeed is the receptor for TSP3-1, we spotted TSP3 and 10-fold serial dilutions of phage S117 on lawns of *S. Typhimurium* (O:4) mutants strains lacking known phage receptors; O-Antigen ($\Delta rfbP$), flagella ($\Delta flgK$), ferrichrome transporter ($\Delta fhuA$) and vitamin B₁₂ transporter ($\Delta btuB$) (Fig. 5). As expected, TSP3 did not produce any translucent zones on the O-antigen mutant ($\Delta rfbP$) and the efficiency of plating (EOP of 10^{-6}) of S117 dropped significantly (Fig. 4). On the contrary, TSP3 was able to form translucent zones on the other mutants and the plaque formation of the phage was comparable to the wild type (Fig. 4), demonstrating that O-antigen of *Salmonella* indeed is the receptor for TSP3-1. Our *in silico* analysis showed that 51 phages also express TSP3s belonging to the TSP3-1 subtype (Fig. 2), thus these phages are likely to recognize O:4 and O:9 O-antigens on *Salmonella enterica* subspecies. Indeed, it has been shown that many of these phages are able to infect *Salmonella enterica* belonging to both O:4 and O:9 serogroups [30,33,38,60–62]. In addition, based on the high similarity of the catalytic domains between TSP3-1 and TSP4-8 subtypes, we predict that TSPs in the TSP4-8 subtype also recognize O:4 and O:9

serogroups which have also been shown for phage SE14 [26] (Fig. 3 and Table 3).

To further validate the host recognition of TSP1, TSP2 and TSP3 of phage S117, we performed a phage inhibition assay. Here each of the TSPs were preincubated with their respective bacterial hosts prior to testing infectivity of phage S117 using a double layer plaque assay. Indeed, each of the TSPs were able to completely inhibit the infection of phage S117 (10^5 PFU/mL) of their respectively hosts, whereas no effect was observed when the TSPs were preincubated with a non-host (Fig. 5). In summary, by using purified TSPs we could confirm the hosts predicted for TSP1 and TSP2 of phage S117 and identify a new host target for TSP3. Furthermore, identifying the host target of TSP3 encoded by phage S117 allowed us to predict putative hosts for 52 phages encoding TSPs in the TSP3-1 subtype. For all TSPs host recognition could be associated with specific O-antigen of the LPS.

4. Discussion

The *Ackermannviridae* is a novel family within the *Caudovirales* order named after the renowned Professor Hans-Wolfgang Ackermann. The family consist of the four genera; *Kuttervirus*, *Agrevirus*, *Limestonevirus* and *Taipeivirus*, and contains phages infecting a broad range of Gram-negative bacteria including many pathogenic bacteria like *E. coli*, *Salmonella*, *Shigella* and *Klebsiella* [13,16,33,37].

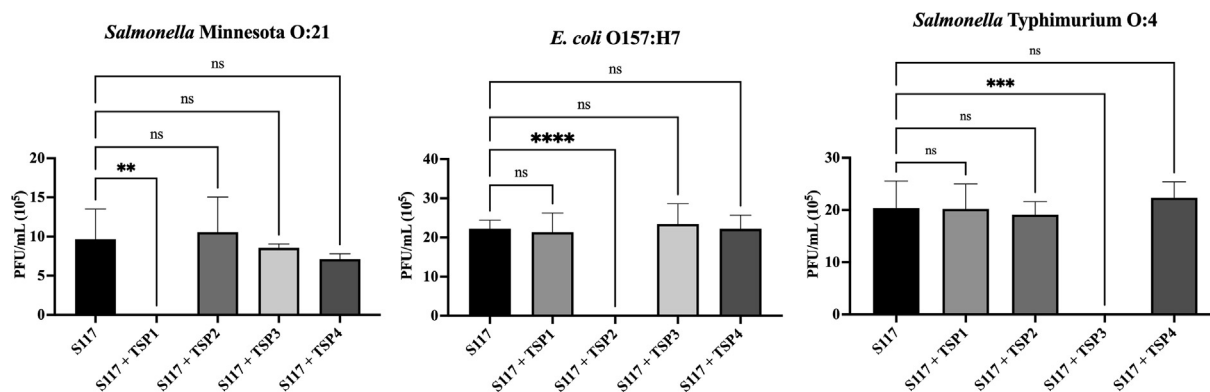


Fig. 5. TSPs can inhibit the infection of phage S117 of their respective hosts. The infectivity of phage S117 in the presence of TSP1-1, TSP2-1, TSP3-1, TSP4-1 or nothing mixed with the bacterial hosts were measured using a double layered plaque assay. 0.5 mg/mL of the TSPs were able to block the infection of phage S117 on their respective hosts when a phage titer of 10^5 PFU/mL was used. The experiment was carried out in triplicates and the error bars represent the standard deviation. The graphs were generated in prism9 where the p-values were calculated using the ordinary one-way ANOVA.

Ackermannviridae phages express a receptor-binding complex consisting of up to four TSPs recognizing different receptors on their bacterial host. For instance, the four TSPs of kuttervirus CBA120 recognize different O-antigens on *Salmonella* and *E. coli* [20]. However, only very few phages and their individual TSPs have been characterized in detail for their host range. In this study, we revealed the diversity of TSPs in the *Ackermannviridae* family through comprehensive *in silico* analysis. Based on TSP similarity we divided them into subtypes, allowing us to predict the host range of numerous uncharacterized *Ackermannviridae* phages by knowing the host recognition of a few TSPs.

Our comprehensive *in silico* analysis revealed a large number of distinct subtypes within each of the four TSPs encoded by phages belonging to the *Ackermannviridae* family. The high sequence conservation found within each subtype suggests similar receptor-binding specificity, which we validated using published data and biological experiments. On the other hand, the limited similarity at the C-terminal receptor binding modules between subtypes, suggest that each subtype may recognize different receptors. Based on the total number of subtypes each carrying diverse C-termini, potentially up to 97 different receptors may be recognized by these phages, thus displaying a vast pool of diverse RBPs for host binding within the *Ackermannviridae* family. In accordance, *Ackermannviridae* phages binds to highly variable polysaccharides on the surface of their bacterial host like the LPS or CPS [17,20,21,31]. For instance, 185 and 46 O-antigens are found in *E. coli* and *Salmonella*, respectively, and 79 K-antigens are found in *Klebsiella* [5,63,64]. A large study of the P22-like phages (*Lederbergvirus* genus) also showed that the phages express highly diverse receptor binding modules to match the diverse O-antigens on the bacterial surface [6]. For instance, the phages HK620, P22 and Sf6 all belonging to the *Lederbergvirus* genus express structural similar TSPs, but with no sequence similarity in the receptor binding module. Each of the TSPs recognize different O-antigens found on the surface of *Salmonella*, *E. coli* and *Shigella* [2,6]. Barbirz et al. suggested that TSPs share a common ancestor, which during the course of evolution have mutated and thereby lost sequence similarity in the receptor binding module leading to different receptor recognition, while still preserving TSP structural folds and functionality [2]. Thus, it seems likely that TSPs in the *Ackermannviridae* family have evolved to match the diverse O-antigens or K-antigens on their bacterial hosts, explaining the diversity of the TSPs in the family.

The four TSPs of kuttervirus CBA120 have previously been detailed characterized [20] and using their distinct N-terminal sequences we could assign all identified TSPs into TSP1, TSP2, TSP3 and TSP4. Further analysis of the conserved N-termini

showed that the XD domains identified in TSP2 and TSP4 of CBA120 are preserved in all TSP2 and TSP4 in the *Ackermannviridae* family. These XD domains are structurally similar to the T4 Gp10 protein that facilitate interactions between proteins composing the distal tail fiber complex in phage T4 [25]. Interestingly, Gp10-like domains are not unique for *Ackermannviridae* phages, as they have been identified in many TSPs or proteins associated with TSPs, like in *E. coli* phages G7C and *Salmonella* phage SP6 [65,66]. A recent study of TSPs in *Klebsiella* phages found multiple TSPs that contain T4 Gp10-like domains that were all located at the N-terminus [9]. As these Gp10-like domains allow complex formation hinging multiple TSPs together, they are commonly found in phages expressing multiple TSPs including phages of the *Ackermannviridae* family. Thus, there may be an evolutionary relationship of Gp10-like domains with phages expressing multiple TSPs.

When comparing all TSPs, we showed that some *Kuttervirus* TSP1 and TSP3 were similar to specific TSP4s, except for the N-terminus of the TSP4. This suggests that the receptor binding modules may be exchanged between TSPs thereby alternating their host recognition. We identified a conserved sequence motif (GTTAVSL) in TSP1, TSP3 and TSP4 expressed by *Kuttervirus* phages, suggesting a site for recombination. Exchange of domains through horizontal gene transfer is a known strategy for phages to alternate their host range [36,67]. For instance, the TSP of *Lederbergvirus* CUS-3 is similar to the C-terminus of TSP of kayfunavirus K1F [68]. Furthermore, the receptor binding module of TSP3 of *Kuttervirus* Det7 is similar to the TSP of phage P22, proposing an exchange between even distantly related phages [21]. Thus, *Ackermannviridae* phages may alternate their host recognition by exchanging the receptor binding module between TSPs within each genus or acquiring new ones by exchanging TSPs with distant related phages thereby further expanding the vast diversity of the TSPs. Finally, the sequence similarity of the N-termini of all *Ackermannviridae* phage TSPs as well of the conserved genes flanking the TSP gene cluster provide optimal conditions for homologous recombination, allowing the entire TSP genes to be exchanged between phages thereby altering the host recognition profile. The conditions needed for exchanging whole genes or only receptor binding modules are not known, but co-infection of phages could lead to homologous recombination and thus resulting in phages acquiring new TSPs.

Host range analysis of phages are laborious when numerous species/strains must be tested, especially for phages expressing multiple RBPs and is limited by the strains available in the laboratory. Here, we determined host recognition of multiple TSPs by knowing the host recognition of only a few TSPs using our *in silico*

analysis and published data of host ranges. In addition, we purified the TSPs of *Kuttervirus* S117 and spotted these on our large *Salmonella* and *E. coli* collection, allowing us to validate our predictions as well as to identify the host for TSPs of the TSP3-1 subtype. Other studies have also used computational tools to predict phage host specificity [69–72]. For instance, the bioinformatics tool HostPhinder predicts the phage-host interactions by comparing the genomic similarities through *k*-mers by the querying phage and an established database of phages with known hosts [70]. Also, a recent paper used machine-learning to predict the host range based on the RBP sequences of phages [72]. The authors extracted RBP sequences of phages known to infect ESKAPE organisms as well as *E. coli*, *Salmonella enterica* and *Clostridium difficile* species and assign the host recognition of phages based on the sequence similarity in the RBPs. Their method allowed them to assign the host recognition of RBP down to the species level. However, Boeckaerts et al. also mentioned that their approach was not able to distinguish the RBP recognition in phages expressing multiple RBPs [72]. Our approach allowed us to identify host recognition of individual TSPs of a phage expressing multiple RBPs and predict the host range of the phage as well as the receptor recognized by each TSP. Since our approach is restricted by the species/strains available in a given lab, the host and receptors of numerous TSPs of *Ackermannviridae* are still to be identified. In summary, our work provides novel insight to the high diversity of TSPs in the *Ackermannviridae* and showed that comprehensive *in silico* analysis is an ideal tool for predicting host interaction in multiple phages belonging to the same family, when combined with biological data demonstrating host recognition. Future identification of the host recognition of individual TSPs in *Ackermannviridae* phages will allow optimization of host prediction tools and provide insight into the evolution of TSPs in the *Caudovirales* order.

CRedit authorship contribution statement

Anders Nørgaard Sørensen: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Cedric Woudstra:** Investigation, Writing – review & editing. **Martine C. Holst Sørensen:** Conceptualization, Visualization, Writing – review & editing, Funding acquisition. **Lone Brøndsted:** Conceptualization, Project administration, Supervision, Visualization, Writing – review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Danish Council for Independent Research (9041-00159B).

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.08.030>.

References

- [1] Xing S, Ma T, Zhang X, Huang Y, Mi Z, Sun Q, et al. First complete genome sequence of a virulent bacteriophage infecting the opportunistic pathogen *Serratia rubidaea*. *Arch Virol* 2017;162(7):2021–8. <https://doi.org/10.1007/s00705-017-3300-x>.
- [2] Barbirz S, Müller JJ, Uetrecht C, Clark AJ, Heinemann U, Seckler R. Crystal structure of *Escherichia coli* phage HK620 tailspike: Podoviral tailspike endoglycosidase modules are evolutionarily related. *Mol Microbiol* 2008;69:303–16. <https://doi.org/10.1111/j.1365-2958.2008.06311.x>.
- [3] Greenfield J, Shang X, Luo H, Zhou Y, Linden SB, Heselpoth RD, et al. Structure and function of bacteriophage CBA120 ORF211 (TSP2), the determinant of phage specificity towards *E. coli* O157:H7. *Sci Rep* 2020;10(1). <https://doi.org/10.1038/s41598-020-72373-0>.
- [4] Nobrega FL, Vlot M, de Jonge PA, Dreesens LL, Beaumont HJE, Lavigne R, et al. Targeting mechanisms of tailed bacteriophages. *Nat Rev Microbiol* 2018;16(12):760–73. <https://doi.org/10.1038/s41579-018-0070-8>.
- [5] Liu B, Furevi A, Perepelov A V., Guo X, Cao H, Wang Q, et al. Structure and genetics of *Escherichia coli* O antigens. *FEMS Microbiol Rev* 2020;44:655–83. <https://doi.org/10.1093/femsre/fuz028>.
- [6] Casjens SR, Thuman-Commike PA. Evolution of mosaicity related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology* 2011;411(2):393–415. <https://doi.org/10.1016/j.virol.2010.12.046>.
- [7] Schwarzer D, Buettner FFR, Browning C, Nazarov S, Rabsch W, Bethe A, et al. A Multivalent Adsorption Apparatus Explains the Broad Host Range of Phage phi92: a Comprehensive Genomic and Structural Analysis. *J Virol* 2012;86(19):10384–98. <https://doi.org/10.1128/JVI.00801-12>.
- [8] Gebhart D, Williams SR, Scholl D. Bacteriophage SP6 encodes a second tailspike protein that recognizes *Salmonella enterica* serogroups C2 and C3. *Virology* 2017;507:263–6. <https://doi.org/10.1016/j.virol.2017.02.025>.
- [9] Latka A, Leiman PG, Drulis-Kawa Z, Briers Y. Modeling the architecture of depolymerase-containing receptor binding proteins in *Klebsiella* phages. *Front Microbiol* 2019;10. <https://doi.org/10.3389/fmicb.2019.02649.03389/fmicb.2019.02649.s00110.3389/fmicb.2019.02649.s003>.
- [10] Adriaenssens EM, Ackermann H-W, Anany H, Blasdel B, Connerton IF, Goulding D, et al. A suggested new bacteriophage genus: ‘Viunlikevirus’. *Arch Virol* 2012;157(10):2035–46. <https://doi.org/10.1007/s00705-012-1360-5>.
- [11] Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res* 2018;46:D708–17. <https://doi.org/10.1093/nar/gkx932>.
- [12] Kutter EM, Skutt-Kakaria K, Blasdel B, El-Shibiny A, Castano A, Bryan D, et al. Characterization of a *Vil*-like phage specific to *Escherichia coli* O157:H7. *Virol J* 2011;8:1–14. <https://doi.org/10.1186/1743-422X-8-430>.
- [13] Casjens SR, Jacobs-Sera D, Hatfull GF, Hendrix RW. Genome sequence of *Salmonella enterica* phage Det7. *Genome Announc* 2015;3:2046. <https://doi.org/10.1128/genomeA.00279-15>.
- [14] Adriaenssens EM, van Vaerenbergh J, Vandenheuvel D, Dunon V, Ceyssens PJ, de Proft M, et al. T4-related bacteriophage LIMEstone isolates for the control of soft rot on potato caused by ‘*Dickeya solani*’. *PLoS One* 2012;7. <https://doi.org/10.1371/journal.pone.0033227>.
- [15] Anany H, Lingohr EJ, Villegas A, Ackermann H-W, She Y-M, Griffiths MW, et al. A *Shigella boydii* bacteriophage which resembles *Salmonella* phage *Vil*. *Virol J* 2011;8(1). <https://doi.org/10.1186/1743-422X-8-242>.
- [16] Min L, Lessor L, O’Leary C, Bonasera R, Gill J, Liu M. Complete Genome Sequence of *Klebsiella pneumoniae* Myophage Menlow. *Microbiol Resour Announc* 2019;8:1–2. <https://doi.org/10.1128/mra.01338-19>.
- [17] Hsu CR, Lin TL, Pan YJ, Hsieh PF, Wang JT. Isolation of a Bacteriophage Specific for a New Capsular Type of *Klebsiella pneumoniae* and Characterization of Its Polysaccharide Depolymerase. *PLoS One* 2013;8. <https://doi.org/10.1371/journal.pone.0070092>.
- [18] Fan C, Tie D, Sun Y, Jiang J, Huang H, Gong Y, et al. Characterization and Genomic Analysis of *Escherichia coli* O157:H7 Bacteriophage FEC14, a New Member of Genus *Kuttervirus*. *Curr Microbiol* 2021;78(1):159–66. <https://doi.org/10.1007/s00284-020-02283-x>.
- [19] Shahrabak SS, Khodabandehlou Z, Shahverdi AR, Skurnik M, Ackermann HW, Varjosalo M, et al. Isolation, characterization and complete genome sequence of Phaxi: A phage of *Escherichia coli* O157: H7. *Microbiol (United Kingdom)* 2013;159:1629–38. <https://doi.org/10.1099/mic.0.063776-0>.
- [20] Plattner M, Shneider MM, Arbatsky NP, Shashkov AS, Chizhov AO, Nazarov S, et al. Structure and Function of the Branched Receptor-Binding Complex of Bacteriophage CBA120. *J Mol Biol* 2019;431(19):3718–39. <https://doi.org/10.1016/j.jmb.2019.07.022>.
- [21] Walter M, Fiedler C, Grassl R, Biebl M, Rachel R, Hermo-Parrado XL, et al. Structure of the Receptor-Binding Protein of Bacteriophage Det7: a Podoviral Tail Spike in a Myovirus. *J Virol* 2008;82(5):2265–73. <https://doi.org/10.1128/JVI.01641-07>.
- [22] Steinbacher S, Baxa U, Miller S, Weintraub A, Seckler R, Huber R. Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. *Proc Natl Acad Sci U S A* 1996;93(20):10584–8. <https://doi.org/10.1073/pnas.93.20.10584>.
- [23] Chen C, Bales P, Greenfield J, Heselpoth RD, Nelson DC, Herzberg O, et al. Crystal structure of ORF210 from *E. coli* O157:H1 phage CBA120 (TSP1), a putative tailspike protein. *PLoS ONE* 2014;9(3):e93156. <https://doi.org/10.1371/journal.pone.0093156>. <https://doi.org/10.1371/journal.pone.0093156.g00110.1371/journal.pone.0093156.g00210.1371/journal.pone.0093156.g00310.1371/journal.pone.0093156.g00410.1371/journal.pone.0093156.g00510.1371/journal.pone.0093156.g00610.1371/journal.pone.0093156.g00710.1371/journal.pone.0093156.t00110.1371/journal.pone.0093156.t002>.
- [24] Greenfield J, Shang X, Luo H, Zhou Y, Heselpoth RD, Nelson DC, et al. Structure and tailspike glycosidase machinery of ORF212 from *E. coli* O157:H7 phage

- CBA120 (TSP3). *Sci Rep* 2019;9:1–11. <https://doi.org/10.1038/s41598-019-43748-9>.
- [25] Taylor NMI, Prokhorov NS, Guerrero-Ferreira RC, Shneider MM, Browning C, Goldie KN, et al. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* 2016;533(7603):346–52. <https://doi.org/10.1038/nature17971>.
- [26] Fong K, Tremblay DM, Delaquis P, Goodridge L, Levesque RC, Moineau S, et al. Diversity and Host Specificity Revealed by Biological Characterization and Whole Genome Sequencing of Bacteriophages Infecting *Salmonella enterica*. *Viruses* 2019;11. <https://doi.org/10.3390/v11090854>.
- [27] Yan T, Liang Lu, Yin P, Zhou Y, Mahdy Sharoba A, Lu Q, et al. Application of a novel phage LPSEYT for biological control of *Salmonella* in foods. *Microorganisms* 2020;8(3):400. <https://doi.org/10.3390/microorganisms8030400>.
- [28] Bai J, Jeon B, Ryu S. Effective inhibition of *Salmonella* Typhimurium in fresh produce by a phage cocktail targeting multiple host receptors. *Food Microbiol* 2019;77:52–60. <https://doi.org/10.1016/j.fm.2018.08.011>.
- [29] Newase S, Kapadnis BP, Shashidhar R. Isolation and Genome Sequence Characterization of Bacteriophage vB_SalM_PM10, a Cba120virus, Concurrently Infecting *Salmonella enterica* Serovars Typhimurium, Typhi, and Enteritidis. *Curr Microbiol* 2019;76(1):86–94. <https://doi.org/10.1007/s00284-018-1588-8>.
- [30] Park M, Lee J-H, Shin H, Kim M, Choi J, Kang D-H, et al. Characterization and comparative genomic analysis of a novel bacteriophage, SFP10, simultaneously inhibiting both *Salmonella enterica* and *Escherichia coli* O157:H7. *Appl Environ Microbiol* 2012;78(1):58–69. <https://doi.org/10.1128/AEM.06231-11>.
- [31] Kabanova AP, Shneider MM, Korzhenkov AA, Bugaeva EN, Miroshnikov KK, Zdorovenko EL, et al. Host specificity of the dickeya bacteriophage PP35 is directed by a tail spike interaction with bacterial o-antigen, enabling the infection of alternative non-pathogenic bacterial host. *Front Microbiol* 2019;9. <https://doi.org/10.3389/fmicb.2018.03288>. <https://doi.org/10.3389/fmicb.2018.03288.s001>.
- [32] Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005;33(Web Server):W244–8. <https://doi.org/10.1093/nar/gki408>.
- [33] Gency YE, Gambino M, Prüssing TF, Brøndsted L. The genera of bacteriophages and their receptors are the major determinants of host range. *Environ Microbiol* 2019;21(6):2095–111. <https://doi.org/10.1111/emi.2019.21.issue-6>. <https://doi.org/10.1111/emi.2019.21.issue-6>.
- [34] García-Nafria J, Watson JF, Greger IH. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci Rep* 2016;6:1–12. <https://doi.org/10.1038/srep27459>.
- [35] Casjens SR, Grose JH. Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages. *Virology* 2016;496:255–76. <https://doi.org/10.1016/j.virol.2016.05.022>.
- [36] Leiman PG, Molineux IJ. Evolution of a new enzyme activity from the same motif fold. *Mol Microbiol* 2008;69:287–90. <https://doi.org/10.1111/j.1365-2958.2008.06241.x>.
- [37] Akter M, Brown N, Clorie M, Yeasmin M, Tareq TM, Baddam R, et al. Prevalence of Shigella boydii in Bangladesh: Isolation and Characterization of a Rare Phage MK-13 That Can Robustly Identify Shigellosis Caused by Shigella boydii Type 1. *Front Microbiol* 2019;10. <https://doi.org/10.3389/fmicb.2019.02461>. <https://doi.org/10.3389/fmicb.2019.02461.s001>. <https://doi.org/10.3389/fmicb.2019.02461.s003>. <https://doi.org/10.3389/fmicb.2019.02461.s004>. <https://doi.org/10.3389/fmicb.2019.02461.s005>.
- [38] Chae SJ, Kwon T, Lee S, Kang YH, Chung GT, Kim DW, et al. Genome sequence of bacteriophage GG32, which can infect both *Salmonella enterica* serovar Typhimurium and *Escherichia coli* O157: H7. *Genome Announc* 2016;4:2015–6. <https://doi.org/10.1128/genomeA.00802-16>.
- [39] van Mierlo J, Hagens S, Witte S, Klamert S, van de Straat L, Fieseler L, et al. Complete Genome Sequences of *Escherichia coli* Phages vB_EcoM-EP75 and vB_EcoP-EP335. *Microbiol Resour Annu* 2019;8(16). <https://doi.org/10.1128/MRA.00078-19>.
- [40] C.E.C. Kaz I, Soil V.E. Berezin Complete Genome Sequence of *Escherichia coli* Phage vB_EcoM Sa157lw, Isolated from Surface Water Collected in 2019 Salinas, California 6 7
- [41] Carter CD, Parks A, Abuladze T, Li M, Woolston J, Magnone J, et al. Bacteriophage cocktail significantly reduces *Escherichia coli* O157: H7 contamination of lettuce and beef, but does not protect against recontamination. *Bacteriophage* 2012;2(3):178–85. <https://doi.org/10.4161/bact.22825>.
- [42] Broecker NK, Roske Y, Valleriani A, Stephan MS, Andres D, Koetz J, et al. Time-resolved DNA release from an O-antigen-specific *Salmonella* bacteriophage with a contractile tail. *J Biol Chem* 2019;294(31):11751–61. <https://doi.org/10.1074/jbc.RA119.008133>.
- [43] Kosznik-Kwaśnicka K, Ciemińska K, Grabski M, Grabowski Ł, Górnjak M, Jurczak-Kurek A, et al. Characteristics of a series of three bacteriophages infecting *salmonella enterica* strains. *Int J Mol Sci* 2020;21(17):6152. <https://doi.org/10.3390/ijms21176152>.
- [44] Hyeon SH, Lim WK, Shin HJ. Novel surface plasmon resonance biosensor that uses full-length Det7 phage tail protein for rapid and selective detection of *Salmonella enterica* serovar Typhimurium. *Biotechnol Appl Biochem* 2021;68(1):5–12. <https://doi.org/10.1002/bab.v68.1>. <https://doi.org/10.1002/bab.1883>.
- [45] Gutierrez J, Xie Y, Gill JJ, Liu M, Dennehy JJ. Complete Genome Sequence of *Salmonella enterica* Serovar Typhimurium Myophage Mutine. *Microbiol Resour Annu* 2019;8(19). <https://doi.org/10.1128/MRA.00401-19>.
- [46] Witte S, Zinsli LV, Gonzalez-Serrano R, Matter C, Loessner MJ, van Mierlo JT, et al. Structural and functional characterization of the receptor binding proteins of *Escherichia coli* O157 phages EP75 and EP335. *Comput Struct Biotechnol J* 2021;19:3416–26. <https://doi.org/10.1016/j.csbi.2021.06.001>.
- [47] Thanh NC, Nagayoshi Y, Fujino Y, Iiyama K, Furuya N, Hiromasa Y, et al. Characterization and Genome Structure of Virulent Phage EspM4VN to Control *Enterobacter* sp. M4 Isolated From Plant Soft Rot. *Front Microbiol* 2020;11. <https://doi.org/10.3389/fmicb.2020.00885>. <https://doi.org/10.3389/fmicb.2020.00885.s001>. <https://doi.org/10.3389/fmicb.2020.00885.s002>. <https://doi.org/10.3389/fmicb.2020.00885.s003>. <https://doi.org/10.3389/fmicb.2020.00885.s004>. <https://doi.org/10.3389/fmicb.2020.00885.s005>.
- [48] Day A, Ahn J, Salmond GPC. Jumbo bacteriophages are represented within an increasing diversity of environmental viruses infecting the emerging phytopathogen. *Dickeya solani*. *Front Microbiol* 2018;9:1–15. <https://doi.org/10.3389/fmicb.2018.02169>.
- [49] Czajkowski R, Ozymko Z, Lojkowska E. Isolation and characterization of novel soilborne lytic bacteriophages infecting *Dickeya* spp. biovar 3 ('D. solani'). *Plant Pathol* 2014;63(4):758–72. <https://doi.org/10.1111/ppa.12157>.
- [50] Day A, Ahn J, Fang X, Salmond GPC. Environmental bacteriophages of the emerging enterobacterial phytopathogen, *dickeya solani*, show genomic conservation and capacity for horizontal gene transfer between their bacterial hosts. *Front Microbiol* 2017;8:1–9. <https://doi.org/10.3389/fmicb.2017.01654>.
- [51] Czajkowski R, Ozymko Z, Zwirowski S, Lojkowska E. Complete genome sequence of a broad-host-range lytic *Dickeya* spp. bacteriophage φD5. *Arch Virol* 2014;159(11):3153–5. <https://doi.org/10.1007/s00705-014-2170-8>.
- [52] Czajkowski R, Ozymko Z, De Jager V, Siwinska J, Smolarska A, Ossowicki A, et al. Genomic, proteomic and morphological characterization of two novel broad host lytic bacteriophages PdbIPD10.3 and PdbIPD23.1 infecting pectinolytic *Pectobacterium* spp. and *Dickeya* spp. *PLoS One* 2015;10. <https://doi.org/10.1371/journal.pone.0119812>.
- [53] Carstens A, Djurhuus A, Kot W, Jacobs-Sera D, Hatfull G, Hansen L. Unlocking the potential of 46 new bacteriophages for biocontrol of *Dickeya Solani*. *Viruses* 2018;10(11):621. <https://doi.org/10.3390/v10110621>.
- [54] Knecht LE, Veljkovic M, Fieseler L. Diversity and Function of Phage Encoded Depolymerases. *Front Microbiol* 2020;10:1–16. <https://doi.org/10.3389/fmicb.2019.02949>.
- [55] Oliveira H, Costa AR, Konstantinides N, Ferreira A, Akturk E, Sillankorva S, et al. Ability of phages to infect *Acinetobacter calcoaceticus-Acinetobacter baumannii* complex species through acquisition of different peptidase lyase depolymerase domains. *Environ Microbiol* 2017;19:5060–77. <https://doi.org/10.1111/1462-2920.13970>.
- [56] Oliveira H, Pinto G, Mendes B, Dias O, Hendrix H, Akturk E, et al. Tailspike with EPS-depolymerase activity, encoded by a phage belonging to a new genus, makes *Providencia stuartii* susceptible to serum-mediated killing. *Appl Environ Microbiol* 2020. <https://doi.org/10.1128/aem.00073-20>.
- [57] Curd H, Liu D, Reeves PR. Relationships among the O-antigen gene clusters of *Salmonella enterica* groups B, D1, D2, and D3. *J Bacteriol* 1998;180(4):1002–7. <https://doi.org/10.1128/JB.180.4.1002-1007.1998>.
- [58] Wollin R, Eriksson U, Lindberg AA. *Salmonella* Bacteriophage Glycanases: Endorhamnosidase Activity of Bacteriophages P27, 9NA, and KB1. *J Virol* 1981;38(3):1025–33. <https://doi.org/10.1128/jvi.38.3.1025-1033.1981>.
- [59] Kintz E, Davies MR, Hammarlöf DL, Canals R, Hinton JCD, van der Woude MW. A BTP1 prophage gene present in invasive non-typhoidal *Salmonella* determines composition and length of the O-antigen of the lipopolysaccharide. *Mol Microbiol* 2015;96(2):263–75. <https://doi.org/10.1111/mmi.2015.96.issue-2>. <https://doi.org/10.1111/mmi.2015.96.issue-2>.
- [60] Duc HM, Son HM, Yi HPS, Sato J, Ngan PH, Masuda Y, et al. Isolation, characterization and application of a polyvalent phage capable of controlling *Salmonella* and *Escherichia coli* O157:H7 in different food matrices. *Food Res Int* 2020;131:108977. <https://doi.org/10.1016/j.foodres.2020.108977>.
- [61] Luna AJ, Wood TL, Chamakura KR, Kutty Everett GF. Complete genome of *Salmonella enterica* serovar Enteritidis myophage Marshall. *Genome Announc* 2013;1:3–4. <https://doi.org/10.1128/genomeA.00867-13>.
- [62] Chamblee J, Zeng C, O'Leary CJ, Gill JJ, Liu M, Dennehy JJ. Complete Genome Sequence of *Salmonella enterica* Serovar Enteritidis Myophage Mooltan. *Microbiol Resour Annu* 2019;8(17). <https://doi.org/10.1128/MRA.00187-19>.
- [63] Liu B, Knirel YA, Feng Lu, Perepelov AV, Senchenkova SN, Reeves PR, et al. Structural diversity in *Salmonella* O antigens and its genetic basis. *FEMS Microbiol Rev* 2014;38(1):56–89. <https://doi.org/10.1111/1574-6976.12034>.
- [64] Pan Y-J, Lin T-L, Chen C-T, Chen Y-Y, Hsieh P-F, Hsu C-R, et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp. *Sci Rep* 2015;5(1). <https://doi.org/10.1038/srep15573>.
- [65] Prokhorov NS, Riccio C, Zdorovenko EL, Shneider MM, Browning C, Knirel YA, et al. Function of bacteriophage G7C esterase tailspike in host cell adsorption. *Mol Microbiol* 2017;105(3):385–98. <https://doi.org/10.1111/mmi.2017.105.issue-3>. <https://doi.org/10.1111/mmi.13710>.
- [66] Tu J, Park T, Morado DR, Hughes KT, Molineux IJ, Liu J. Dual host specificity of phage SP6 is facilitated by tailspike rotation. *Virology* 2017;507:206–15. <https://doi.org/10.1016/j.virol.2017.04.017>.

- [67] Latka A, Maciejewska B, Majkowska-Skrobek G, Briens Y, Drulis-Kawa Z. Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. *Appl Microbiol Biotechnol* 2017;101(8):3103–19. <https://doi.org/10.1007/s00253-017-8224-6>.
- [68] Stummeyer K, Schwarzer D, Claus H, Vogel U, Gerardy-Schahn R, Muhlenhoff M. Evolution of bacteriophages infecting encapsulated bacteria: Lessons from *Escherichia coli* K1-specific phages. *Mol Microbiol* 2006;60(5):1123–35. <https://doi.org/10.1111/mmi.2006.60.issue-510.1111/j.1365-2958.2006.05173.x>.
- [69] Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free d2* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res* 2017;45:39–53. <https://doi.org/10.1093/nar/gkw1002>.
- [70] Villarroel J, Kleinheinz K, Jurtz V, Zschach H, Lund O, Nielsen M, et al. HostPhinder: A phage host prediction tool. *Viruses* 2016;8(5):116. <https://doi.org/10.3390/v8050116>.
- [71] Leite DMC, Brochet X, Resch G, Que Y-A, Neves A, Peña-Reyes C. Computational prediction of inter-species relationships through omics data analysis and machine learning. *BMC Bioinf* 2018;19(S14). <https://doi.org/10.1186/s12859-018-2388-7>.
- [72] Boeckaerts D, Stock M, Criel B, Gerstmans H, De Baets B, Briens Y. Predicting bacteriophage hosts based on sequences of annotated receptor - binding proteins. *Sci Rep* 2021;11(1). <https://doi.org/10.1038/s41598-021-81063-4>.