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, Stigbølgen 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 0:157
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 (Kuttervirus, Aggrevirus, Limestovirus, 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 Kuttervirus TSP1s and TSP3s that were similar to specific TSP4s. A conserved motif in TSP1, TSP3 and TSP4 of Kuttervirus 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 phages. 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, tail 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 frequently 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 Viaviruses 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; Cviriniae and Agtviriniae, as well as the Taipeivirus genus [11] (Fig. 1A). The Kuttervirus is the only genus of the Cviriniae subfamily and contains phages classified as either Salmonella
or E. coli phages [12,13]. The Aglimvirinae subfamily can be divided into the Agtreivirus and Limestonevirus genera that comprise phages infecting Shigella, Enterobacter and Salmonella, or Dickeya, respectively [14,15]. Finally, the Taipeivirus genus include phages that infect 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 $\beta$-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 residues 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:21O-antigen on Salmonella, TSP2 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 orf156 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 vriC 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 (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 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/k6pynfzdch/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 µL 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 MgSO4,7H2O, 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 µM 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 µL 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 µM filters. RNase (10 µg/mL) (Thermo Fischer Scientific, Waltham, MA, USA) and DNase (20 µg/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 µg/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-isooamylalcohol (25:24:1, Ambion) and three rounds of chloroform-isooamylalcohol (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 µg/µL) (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 electrohoresis. 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 (+) vector 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 HinCII and Eco52KI 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 manufacturers description with 1 ng S117 DNA and 1 ng pET-28a (+) as templates with 0.1 μM primers. Subsequently, 1 μL 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 (OD600) 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 harvested 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 Na2HPO4, 0.5 M 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 μm filters before the His-tagged TSP was purified using HisGraviTrap™ (GE Healthcare) according to the manufacturer’s description with elution buffer (0.5 M NaCl, 20 mM Na2HPO4, 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 (Q32866) 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 μL of a bacterial overnight culture was added to 4 mL top agar and poured onto a LB agar plate. Afterwards 1.5 μg 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 where able to inhibit the infection of S117 on the respective hosts. S. Minnesota (JEO2), E. coli O157:H7 (NTCT12900) and S. Typhimurium (LT2c) was used as strains for TSP1, TSP2 and TSP3 respectively. The bacteria were grown to OD600 of 0.3 and put on ice where the four TSPs (0.5 mg/mL) or nothing were added to 100 μL 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 μL of a serial dilution (10⁻¹ to 10⁻⁶) of S117 was spotted onto the lawn. 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.

3. Results

3.1. Tsp subtypes are associated with phages genus

All classified Ackermanniviridae 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 Ackermanniviridae 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 Ackermanniviridae 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 choosen for futher analysis, 69 belonged to the Kuttervirus genus, five to the Agtevirus genus, 17 to the Limestonevirus genus and eight to the Taipeiviruses 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 Ackermanniviridae 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 Ackermanniviridae 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 ≤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 distinct 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 the 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 also encode unique TSPs, as only one or a few phages encode TSPs of the same subtypes. Furthermore, each TSP subtype were specifically associated with the phage genera. In the Ackermannviridae family express highly diverse TSPs that can be grouped into subtypes based on similarity. 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-terminus 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 TPS2 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 residue.
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 Ackermani

viridae 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 domain in all Kuttervirus TSP subtypes (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 domain in 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 domain of TSP1, TSP3 and TSP4 in

Table 1

| TSP2 subtypes representatives* | 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-24|
| 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 ScTs-2                  | 1-273 aa vs. 1-255 aa of CBA120 TSP2                                    | 6W4Q   | 99.96%      | 1.5E-27|
| TSP2-6 S12                    | 1-295 aa vs. 1-295 aa of CBA120 TSP2                                    | 6W4Q   | 99.96%      | 4.6E-27|
| TSP2-7 V11                     | 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 P46PS4                 | 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 KpSi10                | 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 KW86E43-6             | 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|

* 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

| TSP4 subtypes representatives | 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 |
| 82-197 aa vs. 67-185 aa in T4 Gp9 | 12SE | 99.5% | 5.4E-12 | XD3 |
| 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 |
| 82-176 aa vs. 67-164 aa in T4 Gp9 | 12SE | 99.41% | 3.9E-12 | XD3 |
| TSP4-5 Sh19 82-329 aa vs. 49-383 aa in T4 Gp10 | 5XH2 | 98.77% | 2.2E-6 | XD1-XD2 |
| 82-176 aa vs. 67-164 aa in T4 Gp9 | 12SE | 99.45% | 3E-12 | XD3 |
| TSP4-6 LPST94 82-329 aa vs. 49-383 aa in T4 Gp10 | 5XH2 | 99.15% | 1.8E-6 | XD1-XD2 |
| 82-209 aa vs. 67-198 aa in T4 Gp9 | 12SE | 99.47% | 6.7E-12 | XD3 |
| 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 |
| 82-197 aa vs. 67-185 aa in T4 Gp9 | 12SE | 99.46% | 1.2E-11 | XD3 |
| TSP4-9 Det7 82-329 aa vs. 49-383 aa in T4 Gp10 | 5XH2 | 98.75% | 3.7E-6 | XD1-XD2 |
| 82-197 aa vs. 67-185 aa in T4 Gp9 | 12SE | 99.47% | 7.7E-12 | XD3 |
| 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 |
| 82-197 aa vs. 67-185 aa in T4 Gp9 | 12SE | 99.48% | 6.4E-12 | XD3 |
| 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 | 5XH2 | 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% | 6.6E-16 | |
| TSP4-20 UMP2146 84-332 aa vs. 49-383 aa in T4 Gp10 | 5XH2 | 98.52% | 2.3E-5 | XD1-XD2 |
| 83-178 aa vs. 66-164 aa in T4 Gp9 | 12SE | 99.32% | 2.5E-11 | XD3 |
| 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 | |

*Subtypes expressed by phages belonging to Kuttervirus (Blue), Agrevirus (yellow), Limestonevirus (grey), Taipeivirus (red).
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 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:21-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 hingeing 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:157-antigen on Shiga toxin (Stx)-producing *Escherichia coli* (STEC) strains, suggesting that all 32 phages expressing TSP2-1 recognize the O:157-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:3-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]. TSP3 of phage Det7 binds and degrades the O:4-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)-j-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 taipeiviruss 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...
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, TSP3-1 | 9 | GG32, Kage, S117, SPF10, STP07, Agejoakim, S8, 4HA11 and CBA120 [20] |
| E. coli O:157 | TSP1-1, TSP4-7 | 32 | Rabagast, Moki, Heyday, SeHz-1, S118, S115, Se-U, Pa-sanjiao, S8, Se-D, Se-B, BSP101, Kage, SPP10, STP07, Se_EM1, GG32, 4HA11, Se-E, FEC14, S117, EP75, S3, Phad., FSL_SP-063, FSL_SP-029, STM13-1, ECM1-4, Agejoakim 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, Moottan, STM1-13-1, Matapan, Maynard, FSL-SP-029, FSL-SP-063, S3, Sal157sw, Bering, ISTP3, Se-H, BSP101, Se-I, Se-J, Se-N, Se-S, SP1, Dinky, SenASZ3, Sh19, Moki, S117, Pa-sanjiao, S115, S118, Se-U, SPF10, Agejoakim, Kage, Rabagast, 4HA11, SeHz-1, Se-E, Heyday, S8, GG32, STP07, Se-F, Se-G, SenALZ1, SeTs-2, Se-B, Se-D, Perispoe, S35, SeE-1, Se_A01, Se_EM3, Se_EM1, and Se_EM4 |
| E. coli O:77 | TSP3-1, TSP3-4 | 1 | CBA120 [20] |
| E. coli O:78 | TSP3-1 | 1 | CBA120 [20] |
| S. Typhimurium O:4 | TSP3-2 | 9 | SJ3, EP75, LPSTP4, Mutine, PM10, Chennai, SenM2 and barely and Det7 [21] |
| S. Anatum O:3 | TSP2-10 | 3 | SenM-2 and barely and Det7 [42] |
| E. coli O:18A | TSP1-3 | 1 | SenM-2 and barely and Det7 [42] |
| Dickeya Solani and Lelliottia F154 | TSP1-26 | 17 | PhaDTP23.1, Coodle, X04, PhaD3, Kamild, Limestone1, JA15, PhaDTP10.3, RC-2014, Ds16CZ, Ds9CZ, Ds5CZ, Ds3CZ, Ds23CZ, Ds2CZ and Ds25CZ and PTP5 [31] |
| Klebsiella KN2 capsular | TSP2-18 | 1 | 0507-KN2-1 [17] |

Fig. 4. TSP3-1 was not able to produce a translucent zone on an LPS mutant (ArfbP). Phage S117 and TSP3-1 were spotted on Salmonella Typhimurium (LT2C) mutant strains lacking known phage receptors; O-Ag (fhuA), flagella (flikG), 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.

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 [23]. 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 enzymatic functional (Fig. 54). 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 → 4), 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-
ing zones on to strains expressing both serogroups, similar to phages P22, KB1 and P22 and 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* [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 (10^5 PFU/mL) of their respective bacteria like *E. coli* bacteria like *Acinetobacter baumannii* and *Klebsiella pneumoniae* (ECOR), and to predict putative hosts for 52 phages encoding TSPs in the *Salmonella Typhimurium* family [63].

### Table 4

| Genus         | Species or serovar | O-antigen | Genotype                        | Infected or translucent zone/Tested strains |
|---------------|--------------------|-----------|--------------------------------|---------------------------------------------|
| *Salmonella*  | Derby              | O:4       |                                 | 2/17                                      |
| *Salmonella*  | Typhimurium        | O:4       |                                 | 6/8                                       |
| *Salmonella*  | 4.12i:--           | O:4       |                                 | 2/4                                       |
| *Salmonella*  | 4.5.12i:--         | O:4       |                                 | 5/6                                       |
| *Salmonella*  | Brandenburg        | O:4       |                                 | 0/1                                       |
| *Salmonella*  | Bradford           | O:4       |                                 | 0/1                                       |
| *Salmonella*  | Dublin             | O:9       |                                 | 4/4                                       |
| *Salmonella*  | Enteritidis        | O:9       |                                 | 3/3                                       |
| *Salmonella*  | Infantis           | O:7       |                                 | 0/1                                       |
| *Salmonella*  | Selenberg          | O:3       |                                 | 0/1                                       |
| *Salmonella*  | Adelaide           | O:35      |                                 | 0/1                                       |
| *Salmonella*  | Wesleco            | O:42      |                                 | 0/1                                       |
| *Salmonella*  | Montevideo         | O:54      |                                 | 0/1                                       |
| *Salmonella*  | Tanger             | O:13      |                                 | 0/1                                       |
| *Salmonella*  | Cerro              | O:18      |                                 | 0/1                                       |
| *Salmonella*  | Basel              | O:58      |                                 | 0/1                                       |
| *Salmonella*  | Anatum             | O:3,10    |                                 | 0/1                                       |
| *Salmonella*  | Eilbeck            | O:61      |                                 | 0/1                                       |
| *Salmonella*  | Worthington        | O:13      |                                 | 0/1                                       |
| *Salmonella*  | Ondersteapoort     | O:6       |                                 | 0/1                                       |
| *Salmonella*  | Deversoir          | O:45      |                                 | 0/1                                       |
| *Salmonella*  | Telaviv            | O:28      |                                 | 0/1                                       |
| *Salmonella*  | Munster            | O:3       |                                 | 0/1                                       |
| *Salmonella*  | Aberdeen           | O:11      |                                 | 0/1                                       |
| *Salmonella*  | Inverness          | O:38      |                                 | 0/1                                       |
| *Salmonella*  | Bergen             | O:47      |                                 | 0/1                                       |
| *Salmonella*  | Ruiru              | O:21      |                                 | 1/1                                       |
| *Salmonella*  | Minnesota          | O:21      |                                 | 2/2                                       |
| *Escherichia* | coli               | O:157     |                                 | 2/3                                       |
| *Escherichia* | coli (K12)         |           |                                 | 1/1                                       |
| *Escherichia* | coli (ECOR)        |           |                                 | 1/1                                       |
| *Shigella*    | sonnei             |           |                                 | 1/1                                       |
| *Acinetobacter* | baumannii         |           |                                 | 0/1                                       |
| *Klebsiella*  | spp.               |           |                                 | 0/1                                       |
| *Enterobacter* | aerogenes          |           |                                 | 0/1                                       |
| *Cronobacter* | sokazkii           |           |                                 | 0/1                                       |
| *Citrobacter* | freundii           |           |                                 | 0/1                                       |

*) Highly reduced infection (EOP of 10^-4^).

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*, *Agtrevirus*, *Limenovirus* 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].
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 suggests 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 S6 belong 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 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 altering 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 kuttervirus Det7, suggesting an exchange between even distantly related phages [21]. Thus, Ackermannviridae phages may alter their host recognition by exchanging the receptor binding module between TSPs thereby altering their host recognition. We could lead to 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 may lead to homologous recombination and thus resulting in phages acquiring new TSPs.

Host range analysis of phages is 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 approach.
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 rubidea. Arch Virol 2017;162(7):2021–8. https://doi.org/10.1007/s00705-017-3390-4.

[2] Barbiris S, Müller JJ, Uetrecht C, Clark AJ, Heinemann U, Seckler R. Crystal structure of Escherichia coli phage HK620 tailspike: Podoviral tailspike endoglucosidase 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, Heselkamp RD, et al. Structure and function of bacteriophage CBA120 ORF211 (TSF2), the determinant of phage specificity towards E. coli O157:H7. Sci Rep 2020;10(1). https://doi.org/10.1038/s41598-020-67765-3.

[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;16:66–73. https://doi.org/10.1038/s41579-018-00705-01611-1365-2958.2008.06311.x.

[5] Liu B, Furevi A, Pererepol A V, Guo X, Cao H, Wang Q, et al. Structure and genetics of escherichia coli O157. FEMS Microbiol Rev 2020;44:655–83. https://doi.org/10.1016/j.femsre.2020.04.007.

[6] Casjens SR, Thuman C, Det7 PA. Evolution of mosaically 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, Buerstner FFR, Browning C, Nazarov S, Raibach W, Behe A, et al. A Multivalent Adsortion Apparatus Explains the Broad Host Range of Phage phi29: A Comprehensive Genomic and Structural Analysis. J Virol 2012;86 (10):3848–98. https://doi.org/10.1128/jvi.00101-12.

[8] Gebhart D, Williams SR, Scholl D. Bacteriophage SP6 encodes a second tailspike protein that recognizes Salmonella enterica serogroups C2 and C. J Virol 2017;507:263–6. https://doi.org/10.1128/jvi.01720-17.

[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;10:2044. https://doi.org/10.3389/fmicb.2019.02044.

[10] Adriaenssens EM, Ackermann H-W, Anany H, Blasdel B, Connerford JT, Goulding D, et al. A suggested new bacteriophage genus: ‘Vibroniaviridae’. Arch Virol 2012;157(10):2035–46. https://doi.org/10.1007/s00705-012-1365-5.

[11] Lefkowitz RJ, 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/gky1032.

[12] Kutter EM, Skurt-Kakaria K, Blasdel B, El-Shibiny A, Castano A, Bryan D, et al. Characterization of a Vib-like phage specific to Escherichia coli O157:H7. Virol J 2011;8:1–14. https://doi.org/10.1186/1743-422X-8-420.

[13] Casjens SR, Jacobs-Sera D, Hatfull GF, Hendrix RW. Genome sequence of Salmonella enterica enterica phage Det7. Genome Announc 2015;3:2046. https://doi.org/10.1128/genomeA.00279-15.

[14] Adriaenssens EM, van Varenbergh J, Vandenheuvel D, Dunon V, Ceyssens PJ, de Prof M, et al. T4-related bacteriophage LIMEestone isolates for the control of soft rot on potato caused by ‘Dickeyea solani’. PLoS One 2012;7:2. 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 VII. Virol J 2011;8(1):1. https://doi.org/10.1186/1743-422X-8-242.

[16] Miles AJ, Lessor L, O’Toole GA, Beresny R, Gill J. Complete Genome Sequence of Klebsiella pneumoniae Mucoidy Mesophilic, Microb Rev Announc 2019:9:1–2. https://doi.org/10.1128/mcr.00138-19.

[17] Hsu CR, Lin TL, Pan YJ, Hsieh PF, Wang JT. Isolation of a Bacteriophage Specific for the Capsular Type K of Klebsiella pneumoniae and Characterization of Its Polysaccharide Depolymerase. PLoS One 2013:8. https://doi.org/10.1371/journal.pone.0070902.

[18] Fan C, Tie D, Sun Y, Jiang J, Huang H, Gong Y, et al. Characterization and Genomic Analysis of the Bacteriophage φE14 of Escherichia coli O157:H7, a New Member of Genus Kuttervirus. Curr Microbiol 2021;78(1):159–66. https://doi.org/10.1007/s00284-020-02283-x.

[19] Shahrbabak SS, Khodabandeillou Z, Shahrverdi AR, Skurik M, Ackermann HW, Varjoumal M, et al. Isolation, characterization and complete genome sequence of Phako: a phage of Escherichia coli O157: H7. Microbiol (United Kingdom) 2013;159:1629–38. https://doi.org/10.1099/mic.0.057770-0.

[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] Walse M, Forder C, Gravel R, Biebel M, Rachel R, Hervo-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, Weintraub A, Jarschke A, Huber R, Huber C, et al. 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] Chang C, Bales P, Greaves C, Heselkamp RD, Nelson DC, Herzberg O, et al. Crystal structure of ORF210 from E. coli O157:H7 phage CBA120 (TSF1): a putative tailspike protein. PLoS ONE 2014;9(3):e93156. https://doi.org/10.1371/journal.pone.0093156.

[24] Greenfield J, Bang X, Luo H, Zhou Y, Heselkamp RD, Nelson DC, et al. Structure and tailspike glycosidase machinery of ORF212 from E. coli O157:H7 phage
[67] Latka A, Maciejewska B, Majkowska-Skrobek G, Briers 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-5.10.1111/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(5):14. https://doi.org/10.1186/s12859-018-2388-7.

[72] Boeckaerts D, Stock M, Criel B, Gerstmans H, De Baets B, Briers Y. Predicting bacteriophage hosts based on sequences of annotated receptor - binding proteins. Sci Rep 2021;11(1). https://doi.org/10.1038/s41598-021-81061-4.