

<https://doi.org/10.1038/s44298-025-00118-9>

Gp38 adhesins of *Straboviridae* phages recognize specific extracellular loops of outer membrane protein receptors



Veronika T. Lutz¹, Victor Klein- Sousa², Martin S. Bojer¹, Saar S. F. Van Overfelt¹, Sam R. Nugen³, Nicholas M. I. Taylor², Michela Gambino^{1,4} & Lone Brøndsted¹ ✉

Tailed bacteriophages (phages) depend on specific interactions with their host receptors for efficient infection and propagation. Gp38 is a unique receptor-binding protein located at the distal tail of *Straboviridae* phages characterized by a defined modular, monomeric structure. Here, we demonstrated that the similarity of Gp38 adhesins, encoded by related yet distinct *Straboviridae* phages belonging to *Tequatrovirus*, *Mosigvirus*, and *Krischvirus*, determines the recognition of receptors such as Tsx, OmpF, and OmpA. For OmpA, experimental and in silico analysis identified specific outer loops protruding from the receptor required for phage infection by interacting with Gp38. Yet, the loops involved are dependent on the adhesin variant. In-depth in silico analysis identified two groups of Gp38 adhesins differentially interacting with OmpA by expressing specific amino acids. This demonstrates that both partners' diversity affects phage binding and host range. Overall, the phylogeny of Gp38 adhesins can predict receptor binding essential for advancing phage therapy.

Phages are naturally occurring bacterial viruses that specifically infect and kill bacteria during their propagation, thus displaying antibacterial activities. Emerging antimicrobial resistance among bacterial pathogens is a major global health threat, as serious infections may no longer be treatable due to a lack of effective antibiotics¹. This includes antibiotic-resistant *Escherichia coli*, belonging to the group of critical pathogens according to the World Health Organization². A possible therapeutic alternative to antibiotics is phage therapy, used for decades in Eastern Europe³. *Straboviridae* phages, particularly the T2-, T4-, and T6-like phages⁴, infect a broad spectrum of bacterial species^{5,6}, including critical antimicrobial-resistant pathogens like *Acinetobacter baumannii*⁷, *Klebsiella pneumoniae* and *K. aerogens*⁸, and *E. coli*⁹. Due to their strictly lytic life cycle, *Straboviridae* phages are well-suited for phage therapy, and their safety has been demonstrated in vivo with no adverse effects or immune reactions^{9,10}.

Straboviridae phages are myoviruses⁵ and share many morphological similarities, including contractile tails with attached long and short tail fibers responsible for host binding. Extensive research on *E. coli* phages T4, T2, and T6 has shown that these phages utilize a two-step infection mechanism. In this process, the long-tail fibers are responsible for the initial binding to the host, while the short-tail fibers facilitate a secondary, irreversible binding¹¹. For phage T4 and similar phages, the distal part of the long tail fiber (encoded by gp37) is trimerized by chaperones (Gp38 and Gp57) and is

responsible for host binding¹². In contrast, phages T2 and T6 and their like carry an adhesin protein (encoded by gp38) at the tip of the long tail fiber (Gp37) that binds to the bacterial host receptor¹³. Overall, the molecular understanding of the specificity of phage host binding as the initial step of the infection cycle¹⁴ is essential for the target selection of suitable phages for specific therapies.

Since the Gp38 adhesins are monomeric, they differ structurally from the majority of phage receptor-binding proteins (RBPs), which are mostly trimeric. Typical of Gp38 adhesins is a modular structure consisting of a conserved N- and C-terminus flanking five conserved glycine-rich motifs (GRMs) interrupted by four hypervariable segments (HVS). The conserved GRMs have been suggested to function as recombinational hotspots¹⁵ for modular shuffling¹⁶, where modules of viral genomes can be exchanged among viruses to form new, evolutionary advantageous genes¹⁷. In contrast, the HVSs are responsible for host recognition, and minor changes such as amino acid deletions, substitution, or insertions¹⁸ can change the receptor target by the phage¹⁶. In addition to their function in modular shuffling, Drexler et al.¹⁸ postulated already in 1989 that GRMs form a compact core with the HVSs building loops sticking out to facilitate host recognition. This was later verified using crystallography, showing that Gp38 comprises three structural domains: the N-terminal attachment domain, the β -helix domain, and the C-terminal polyglycine sandwich domain. The N-terminal

¹Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg C, Denmark. ²Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ³Department of Food Science, Cornell University, Ithaca, NY, USA.

⁴Present address: Institute for Conservation, Royal Danish Academy, Copenhagen, Denmark. ✉e-mail: lobr@sund.ku.dk

attachment domain connects Gp38 to the long tail fiber Gp37 via electrostatic interactions. The β -helix domain is connected to the N-terminal attachment domain via a six amino acid linker, and the C-terminal domain is a three-layered compact polyglycine type II sandwich domain formed by the GRMs. From this last domain, the HVSs reach into the environment as four-residue β -turn loops, allowing host recognition by binding to bacterial receptors¹³.

The receptors recognized by *Straboviridae* phages encoding Gp38 adhesins are often general porins (i.e., diffusion pores forming β -barrel proteins selecting polar or nonpolar solutes¹⁹), as well as other surface structures, like lipopolysaccharides (LPS)¹⁶. For example, T2- and T6-like phages encoding Gp38 commonly bind to OmpA, OmpF, or Tsx as their receptors¹⁴. These porins often consist of 16-stranded antiparallel β -barrels with radiating β -strands forming long loops on the extracellular side and short loops at the periplasmic side²⁰. OmpF is the most abundant porin in the bacterial outer membrane. It forms a trimeric β -barrel protein²¹ with eight extracellular loops²², of which at least three (loops 5, 6, and 7) are important for coliphage K20 to bind²³. In contrast, Tsx is a less abundant β -barrel nucleoside channel with seven highly variable extracellular loops²⁴, of which loops 6 and 7 serve as binding sites for phages T6, Ox1, and H3^{24,25}. The most common phage receptor for Gp38 is OmpA¹⁶, a β -barrel protein that forms a monomer but can also exist as a homodimer^{21,26}. The phages Ox2 and K3, among others, bind to its four extracellular loops, as demonstrated by deletion mutants lacking entire extracellular loops or even just specific amino acids (25, 70, 110, and 154)^{27,28}. However, although the outer loops of these phage receptors are essential for phage binding, their molecular interactions with Gp38 adhesins have not been investigated.

To study the molecular interactions between variants of Gp38 adhesins and their bacterial receptors, we established a collection of *Straboviridae* phages encoding Gp38 and collected experimental data on Gp38 diversity, host range, and host receptors. We further analyzed the role of OmpA extracellular loop variants for phage infection by determining plaque formation on mutants expressing diverse OmpA variants and predicted binding of diverse Gp38-OmpA pairs in silico. Finally, we propose that specific amino acids of HVS interact with the receptor and that Gp38 adhesins recognizing OmpA have evolved into two groups, each recognizing a subset of OmpA variants, translating into diverse host ranges.

Results

The *Straboviridae* phage collection

To establish a diverse *Straboviridae* phage collection, we plated 12 samples containing filtrated piglet feces onto the lawns of *E. coli* MG1655. Fifty-nine plaques of diverse morphologies were picked and subjected to PCR using primers specific for the major capsid gene of *Tevenvirinae* phages belonging to the *Straboviridae* family²⁹. As a positive control, we used *Tevenvirinae* phage AV119³⁰. Twenty-five plaques, mostly small and clear, resulted in a PCR band of expected size, suggesting that they were *Tevenvirinae* phages. After purification and sequencing, whole genome BLAST analysis confirmed the isolation of 15 unique and diverse phages with genomes ranging from 162,587 bp to 170,664 bp and GC content between 35% and 41% (Table S1). The *Tequatrovirus* phage AV119³⁰ was included in our study.

A whole genome BLAST similarity search was performed against all phage genomes in the NCBI database to taxonomically assign the phages to current genera. All newly isolated phages were assigned to the family of *Straboviridae*: 11 belonging to the subfamily *Tevenvirinae*, including eight *Tequatrovirus* and three *Mosigvirus* phages. In addition, five phages belong to the *Krischvirus* genus, which is not part of the *Tevenvirinae* subfamily but still *Straboviridae*. The taxonomic assignment was verified by establishing a phylogenetic tree using VICTOR, including at least one member of all 23 genera of the *Straboviridae* family and the closest BLAST hits to the newly isolated phages (Fig. 1). The taxonomic assignment using VICTOR was consistent with the BLAST results, except for phage FL31, which, according to BLAST, belongs to the *Mosigvirus* genus. In contrast, the VICTOR analysis suggested that FL31 is closely related to both *Escherichia phage phiE142* (*Carettavirus*) and *Escherichia phage vB_EcoM_PHAPEC2*

(*Mosigvirus*). A direct genome comparison of FL31 showed a more overall identity to *Mosigvirus* PHAPEC2 than *Carettavirus* phi E142, indicating FL31 is indeed a *Mosigvirus* phage. Thus, our genomic analysis confirmed the phylogenetic assignment and demonstrated the success of the targeted isolation of *Straboviridae* phages.

Identification of *Straboviridae* phages encoding a Gp38 adhesin

Straboviridae phages bind to their host using either a Gp38 adhesin located at the tip of their long tail fibers like T2 or T6 or, like T4, by using the RBP Gp37 for binding¹⁶. To identify phages in our collection encoding *gp38* adhesin genes, all genomes were aligned to *gp38* of *Escherichia phage vB_EcoM-FiEco06* (NC_054914.1:156717-157517) and to the genome of *Escherichia phage Bp7*, known to carry the *gp38* adhesin gene (NC_019500-1). The identified *gp38* were confirmed to encode adhesins by submitting their amino acid sequence to HHpred and verifying their tertiary structure as adhesins (closest hit for all adhesin 6F45_D, receptor recognition protein, with a probability of 100%). We identified 13 unique Gp38 variants within our collection of 16 *Straboviridae* phages since phages FL23 and FL44 express identical adhesins, and we then excluded the *Mosigvirus* phages JM10 and JM17 for not encoding a *gp38* adhesin gene (Table S1). In summary, most phages (14) in our collection encode a Gp38 adhesin for specific binding to their *E. coli* host.

Whole genome comparison of *Straboviridae* phages encoding Gp38 adhesins

To explore the overall similarity between the 14 phages encoding a Gp38 adhesin, an average nucleotide identity (ANI) comparison of whole phage genomes was performed and visualized with a heatmap (Fig. S1). Phages belonging to the same genera clustered together with high within-genus similarity. The five *Krischvirus* phages were similar (89% to 97% identity) as were the *Tequatrovirus* phages (84% to 99% identity). In contrast, the two *Mosigvirus* phages were less similar, showing 81% identity (Fig. S1). Within the *Tequatrovirus* genus, some phages were highly similar, including phages FL23 and FL44 (99.77% identity, 100% coverage) and phages FL33 and FL34 (99.95% identity, 100% coverage) according to BLAST.

Genome alignment of phages belonging to the same genus (Fig. S2) revealed a conserved genomic organization, including the position of the *gp38* adhesin gene within all three genera. The only exceptions were the *Krischvirus* phages FL37 and FL38, where the *gp38* gene was not located near the phage holin, as observed in the other phage genomes (Fig. S2a). The remaining *Krischvirus* phages (FL04, FL15, and FL20) showed high gene conservation and synteny, especially between FL15 and FL20 (Fig. S2a). Further, several distinct genes could be identified by comparison of closely related genomes. For example, compared to the other *Mosigvirus* phage, FL18, phage FL31 encodes additional genes like a dCTP pyrophosphatase, a phage-associated homing endonuclease, and a pin protease inhibitor, explaining its larger genome size (Fig. S2b). The high identity of *Tequatrovirus* phages FL23 and FL44, and FL33 and FL34 was confirmed by direct gene comparison (Fig. S3c). In contrast, among phages FL08, FL12, AV119, FL44, and FL34, a higher diversity with additional genes in the respective phage genomes could be observed (Fig. S2c). In summary, we established a collection of 14 distinct *Straboviridae* phages encoding 13 Gp38 unique adhesins belonging to three different genera: *Tequatrovirus*, *Mosigvirus*, and *Krischvirus*.

Diverse Gp38 adhesins recognize different outer membrane receptors

Gp38 adhesins are known to display a modular structure with five GRM interrupted by four HVS responsible for the host binding¹⁶. By performing an amino acid alignment of the Gp38 adhesin sequences, we identified the typical modular structure of Gp38 adhesin proteins with a conserved N-terminus and variable regions in the C-terminus. The four HVS (amino acids 122-158, 166-179, 187-222, 133-150) were flanked by five conserved GRMs (Fig. 2a). To predict the location of the identified HVS within the folded adhesins, we used the protein sequence of the adhesin of phage FL20

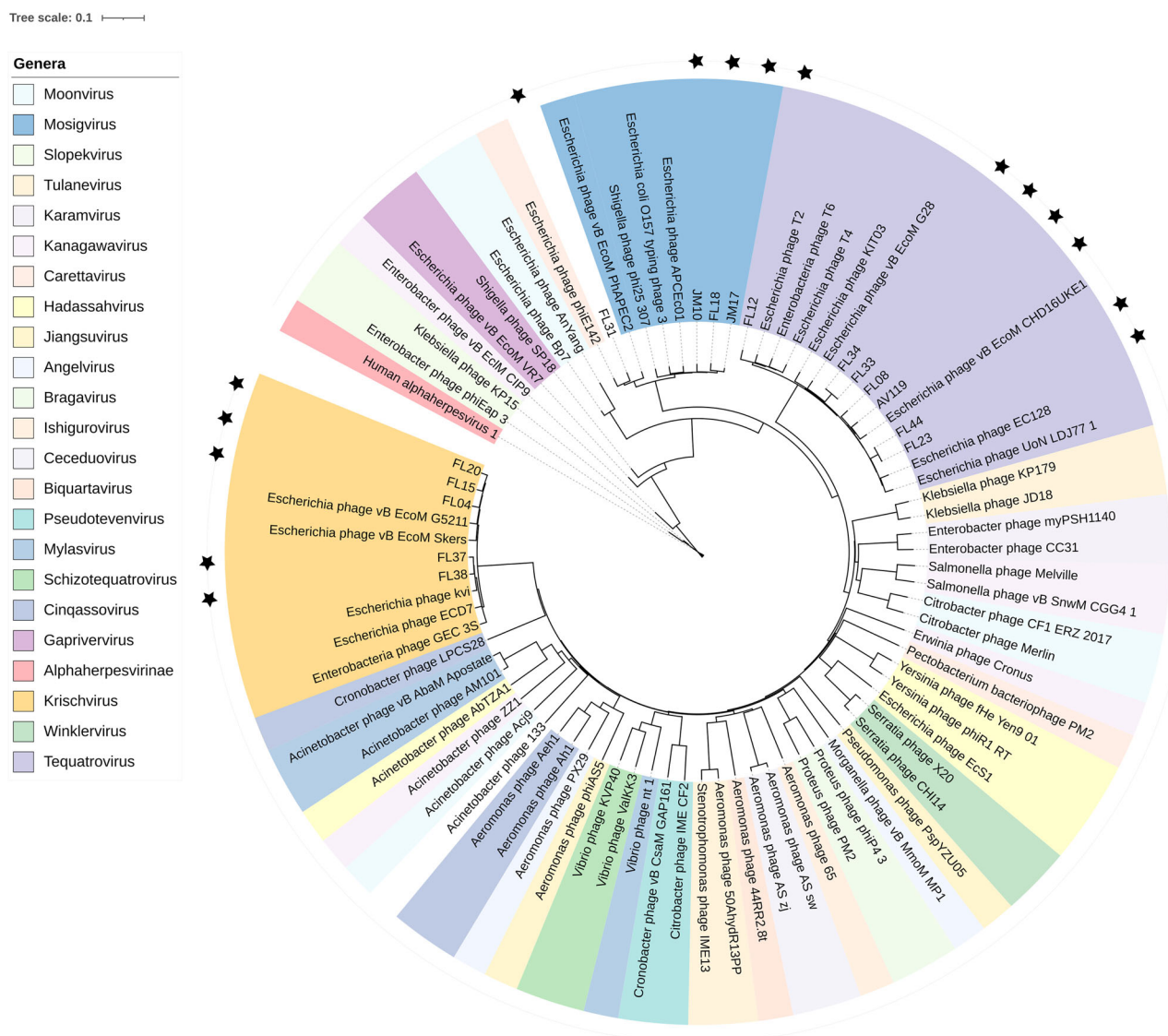


Fig. 1 | Phylogenetic assignment of Straboviridae phages. Isolated *Straboviridae* phages were phylogenetically assigned by comparing them to representatives of each *Straboviridae* genus, acquired from the ICTV master file 2022. A phylogenetic tree was constructed by whole genome comparison using VICTOR with the D4 formula

and the Human herpes virus as an outgroup and visualized with iTOL. Phages were color-coded according to the family; those from our collection are marked with a black star.

as an example for in silico protein structure prediction using AlphaFold2. The N-terminus is folded into the attachment domain and interacts with the phage particle. The β -helix domain was formed by the N- and C-terminus, and the GRM formed, together with the HVS, the outer loops at the tip of the adhesin (Fig. 2b). Taken together, the conserved GRM allowed us to identify diverse amino acid sequences of the HVS of the 13 unique Gp38.

To investigate the effect of HVS diversity on receptor recognition, we prepared deletion mutants in common phage receptors (i.e., *ompA*, *ompF*, *tsx*, *tonB*, *btuB*, *fadL*, *lamB*, *tolC*, *fhuA*, *ompW*, *fepA* and *ompC*) in *E. coli* MG1655. Serial dilutions of phage stocks were spotted on the lawns of the deletion mutants to determine the formation of single plaques. This allowed us to identify different outer membrane proteins as phage receptors (Fig. 2c). Three phages, FL31, FL33, and FL34, depend on the porins *OmpF* and *TolC* for infection. *TolC* has previously not been described as a Gp38 receptor. Still, we observed at least one log reduction in plaque formation of the *tolC* mutant, suggesting a minor role in receptor recognition. Four phages, FL04, FL15, FL23, and FL44, recognized the nucleoside-specific channel-forming protein *Tsx* as the receptor, whereas five phages (FL08, FL12, FL20, FL18, and AV119) were dependent on the outer membrane protein *OmpA* for

infection. Construction of a phylogenetic tree using the amino acid sequences of Gp38, while excluding the conserved N-terminus, showed that similar adhesins recognize the same receptors (Fig. 2c).

To determine the influence of the Gp38 diversity on phage infectivity, a host range analysis was performed using the ECOR collection, representing *E. coli* species diversity³¹. Serial dilutions of all phages were spotted on an overlay of each of the 72 ECOR strains and the propagation host MG1655. Ordering the phages according to their Gp38 phylogeny allowed the distinction of two groups using *OmpA* as the receptor. Among those, phages AV119 and FL08 infect 14 and 16 strains, respectively, and thus display a broader host range than phages FL12, FL20, and FL18, infecting only six to ten ECOR strains (Fig. 3). Thus, phages with different Gp38 variants exhibit varied host ranges, even when utilizing the same *OmpA* receptor.

Gp38 adhesin variants recognize different *OmpA* outer loop combinations

Sequencing was conducted to investigate the *OmpA* diversity among the ECOR strains and confirmed high-level conservation interrupted by hypervariable regions corresponding to the extracellular loops. The

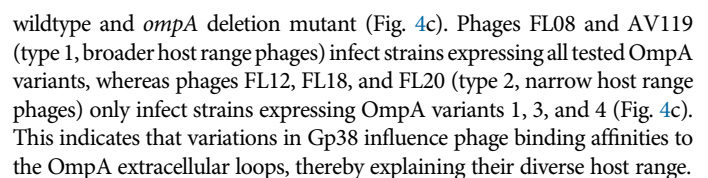


Fig. 2 | Identification of hypervariable regions in Gp38 adhesin and their receptors. An alignment of identified Gp38 adhesin proteins in our collection to the reference adhesin protein of *Escherichia phage vB_EcoM-fFiEco06*, NC_054914.1:156717-157517 was performed in CLC, showing the typical modularity of Gp38 in our phage collection, with a conserved N- and C-terminus (gray and green respectively), which are flanking the alternating conserved GRMs (blue) and diverse HVS (red) (a). The position of the different protein modules is indicated in an AlphaFold prediction of phage FL20, in which the N- and C-terminus (gray and green) are building the attachment domain and beta-helix domain, and the GRM and HVS (blue and red) are building the polyglycine rich sandwich domain, important for binding (b). The variability of *gp38* adhesin genes was analyzed by

constructing a phylogenetic tree excluding their conserved N-terminus. Receptor recognition of the isolated phages carrying Gp38 adhesin was analyzed by spotting phage dilutions on MG1655 deletion mutants lacking common phage receptors, i.e., *ompA*, *ompF*, *tsx*, *tonB*, *btuB*, *fadL*, *lamB*, *tolC*, *ompW*, and *ompC*. Numbers display the log of the EOP (efficiency of plaquing) to show infection reduction. No infection or reduced infection (defined as more than three log reduction) is highlighted according to the deleted gene. For FL31, only a one log reduction could be observed for *ompF* and *tolC* deletion mutants, which is a higher reduction than for any other deletion mutant but lower than the threshold. Individual Gp38 variants were ordered according to a phylogenetic tree built using the Gp38 sequences lacking the conserved N-terminus (c).

As OmpA contains variable sequences of the extracellular loops, we set out to investigate how combinations of different OmpA loop variants may influence phage binding. We constructed a phylogenetic tree of the OmpA sequences from the ECOR collection and identified their combinations of loop variants (Fig. 4d). We then noted OmpA variants allowing phage infection using the host range data and infection capabilities of OmpA mutant strains. Interestingly, we observed that the combination of loop variants affected phage infection by expectedly affecting phage binding. Since type 1 phages FL08 and AV119 infect strains carrying all combinations of OmpA loops of the ECOR collection and MG1655 expressing any of the eight OmpA variants tested, we concluded that FL08 and AV119 may bind to more than one and possibly different combinations of OmpA receptor outer loops. On the contrary, type 2 phages FL12, FL18, and FL20 show a much narrower host range and can only infect strains expressing OmpA variants 1, 3, and 4. The only combination of outer loops that allow this infection pattern are loops 3a and 3b in combination with loop 2a (Fig. 4d). Unexpectedly, phage FL20 can also infect ECOR37, and FL18 can infect ECOR33, but infectivity is very low (Fig. 3), which may suggest inefficient binding.

In summary, we concluded that Gp38 of phages FL08 and AV119 binds to several OmpA outer loop variants, enabling a broader host range, whereas Gp38 of phages FL12, FL18, and FL20 binds only to two outer loop variants, likely 3a and 3b, possibly stabilized by loop 2a. Thus, with this analysis, we propose the molecular interaction behind these phages' broad and narrow host ranges, respectively.

AlphaFold provides reliable binding predictions of Gp38-OmpA interactions

To further understand the influence of the variations of OmpA loops on Gp38 binding, their interactions were investigated in silico. Initially, we assessed whether AlphaFold2-multimer models could predict adhesin-receptor-binding. We performed predictions of the five diverse Gp38 adhesins expressed by AV119, FL8, FL12, FL18, and FL20 with the eight OmpA variants. OmpX and Phage Mu gpU served as negative controls. This generated 1,535 binding models that included at least 25 models for each pair. Only models with the correct directionality that show the Gp38's HVS interacting with the outer regions of membrane receptors are biologically feasible. Models with the opposite directionality, showing the attachment domain interacting with the outer region of membrane receptors, are biologically incorrect, as this domain is attached to the long tail fibres in the matured virion. Therefore, the principle axis of inertia between adhesin-receptor pairs was determined. Here, a theoretical angle threshold for true binding partners could be identified since the proportion of binding models with incorrect versus acceptable directionality at $\alpha = 90^\circ$ between the two binding partners is expected to be 25% with random placement, decreasing sigmoidally with lower angles (Fig. 5a). In our dataset, however, only 21% of all generated models displayed the acceptable directionality at $\alpha = 90^\circ$ (Fig. 5a). Analysis of each pair revealed a wide range of proportions of acceptable directionality, with a non-random distribution with a notable peak at 0.0 (p value $\ll 0.001$, Mann-Whitney U test). Overall, the distribution of adhesin-receptor pair directionality was significantly different

compared to the distribution of random pairs (p value = 0.007, Mann-Whitney U test) (Fig. 5b).

Using the directionality criteria, we predicted Gp38-OmpA binding (Fig. 5c) and compared it to experimental data of phage infectivity of MG1655 expressing various OmpA variants (Fig. 4c). When predicting the binding of all 40 Gp38-OmpA pairs, 28 were predicted as non-binding, with 23 experimentally confirmed as negative. In addition, 12 pairs were predicted to bind, with 11 confirmed as positives (Fig. 5d). Thus, using directionality resulted in 85% overall accuracy, including a true positive rate of 92% and a true negative rate of 82%. Changing $\alpha < 10^\circ$ resulted in maximum precision (100%) but fewer correct results (20%). For larger angles, precision decreased, but fewer pairs were identified as binding (Fig. S3c–e), indicating a trade-off between precision and identification. In addition, other metrics, such as interface predicted aligned error, pDockQ, and interface pLDDT, were unsuitable due to their inherently lower values in disordered loop regions (Fig. S3a, b). In summary, we can conclude that analysis of adhesin-receptor directionality provides a highly accurate in silico indication of possible binding.

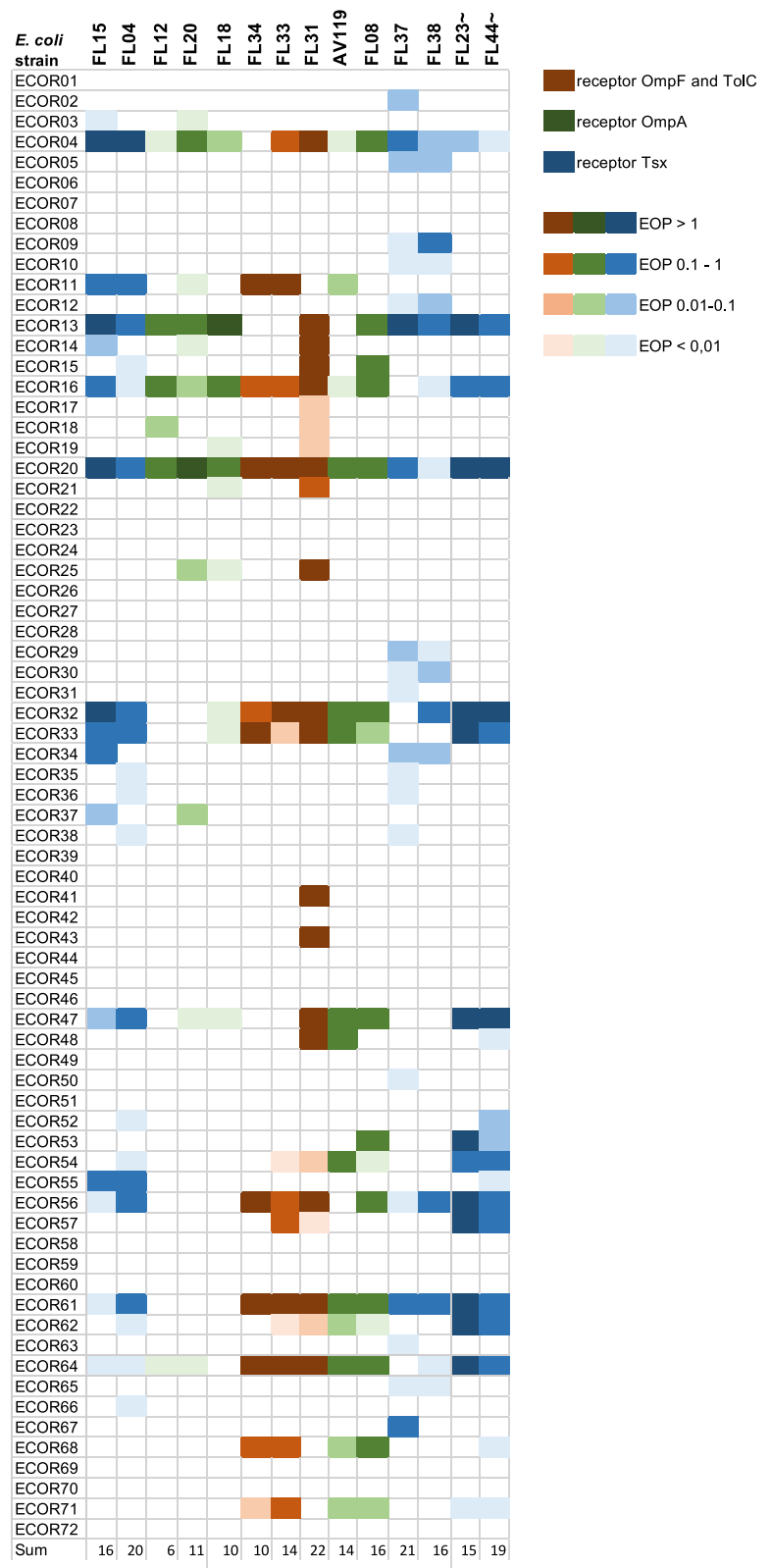
In silico predictions of amino acid interactions between Gp38-OmpA

To predict amino acid interactions involved in Gp38-OmpA binding, we selected true positive AlphaFold multimer prediction and used molecular dynamics simulations without solvents between chains to observe potential non-covalent protein-protein interactions. After excluding models exhibiting zero contacts, the remaining 115 models confirmed that most interactions occurred near the four OmpA loops. All loops displayed significant hydrogen bonds, with most hydrogen bonds and hydrophobic interactions in loop 4. Loop 2 had few hydrophobic interactions and fewer salt bridges, especially compared to loop 3, which had the largest number of these bonds. Stacking interactions primarily occurred in loops 1, 2, and 3, suggesting they may also play a role in stabilizing specific binding (Fig. 6a), as supported by our experimental data.

Focussing on Gp38 amino acid residues involved in OmpA binding, a multiple sequence alignment of the five Gp38 variants revealed that key interaction residues of type 1 (AV119 and FL08) were located near unmatched regions compared to type 2 (FL12, FL18, FL20). For instance, residue R165 played a major role in interacting through a salt bridge with all variants except OmpA variant 2. In variants 3 and 8, this residue interacted primarily with asparagine in loop 3 but could also interact with loop 4 in fewer cases. In contrast, in variants 4 to 7, the interaction occurred mostly with another asparagine in loop 4. Other neighboring residues of R165 (V166, H167, S168, G169, A170) also displayed interactions with OmpA loops. The residue D217 of Gp38 type 1, present in another unmatched region, was important in interacting with Y84 in loop 2 of variants 1, 2, 5, and 6. This interaction was also found in variant 7, where D217 primarily interacted with T137 in loop 3 (Figs. 6b, c and S3). We can thus conclude that residues R165 and its neighboring residues, as well as D217, enable adhesins expressed by phages AV119 and FL08 to interact with a variety of amino acid residues displayed in loops 2, 3, and 4 of OmpA.

To pinpoint differences in both Gp38 variants, we compared their interactions with OmpA variants 1, 3, and 4. For both types, loop 3 was used

Fig. 3 | Phage infection of the *E. coli* reference (ECOR) collection. Dilutions of all phages were spotted on the 72 *E. coli* reference strain collection (ECOR) strains and MG1655 as their isolation host. EOP was calculated by dividing each strain's plaque count by the count on MG1655. Colors indicate the receptors the phages bind: brown for OmpF and TonB, green for OmpA, and blue for Tsx. Shades indicate infection efficiency, ranging from an EOP of more than one (dark) to an EOP of <0.01 (light). Phages marked with ~ have an identical Gp38 protein but a different genome and infection patterns.



as a secondary interaction point, where an arginine interacted with N137. They also targeted similar residues in OmpA variant 3. However, for type 1, R262 and R181 interacted with E89 in loop 2, and an asparagine (N217) also stabilized loop 2 (T84). In contrast, type 2 utilized loop 1 as a hotspot, where R172 of Gp38 interacted with N41 of variant 1, and two alanines (A189 and A253) interacted with F44. We can thus conclude that type 1 interacted

mostly with loop 3 of variant 4, whereas type 2 docked on loops 1 and 2 (Fig. 6d, e). This analysis enabled us to pinpoint the specific amino acid residues involved in Gp38-OmpA binding. It confirmed that Gp38 type 1 binds to a larger variety of residues and loop combinations than Gp38 type 2, consistent with our OmpA variant mutant analysis, leading to a larger host range of phages expressing Gp38 variants of type 1.

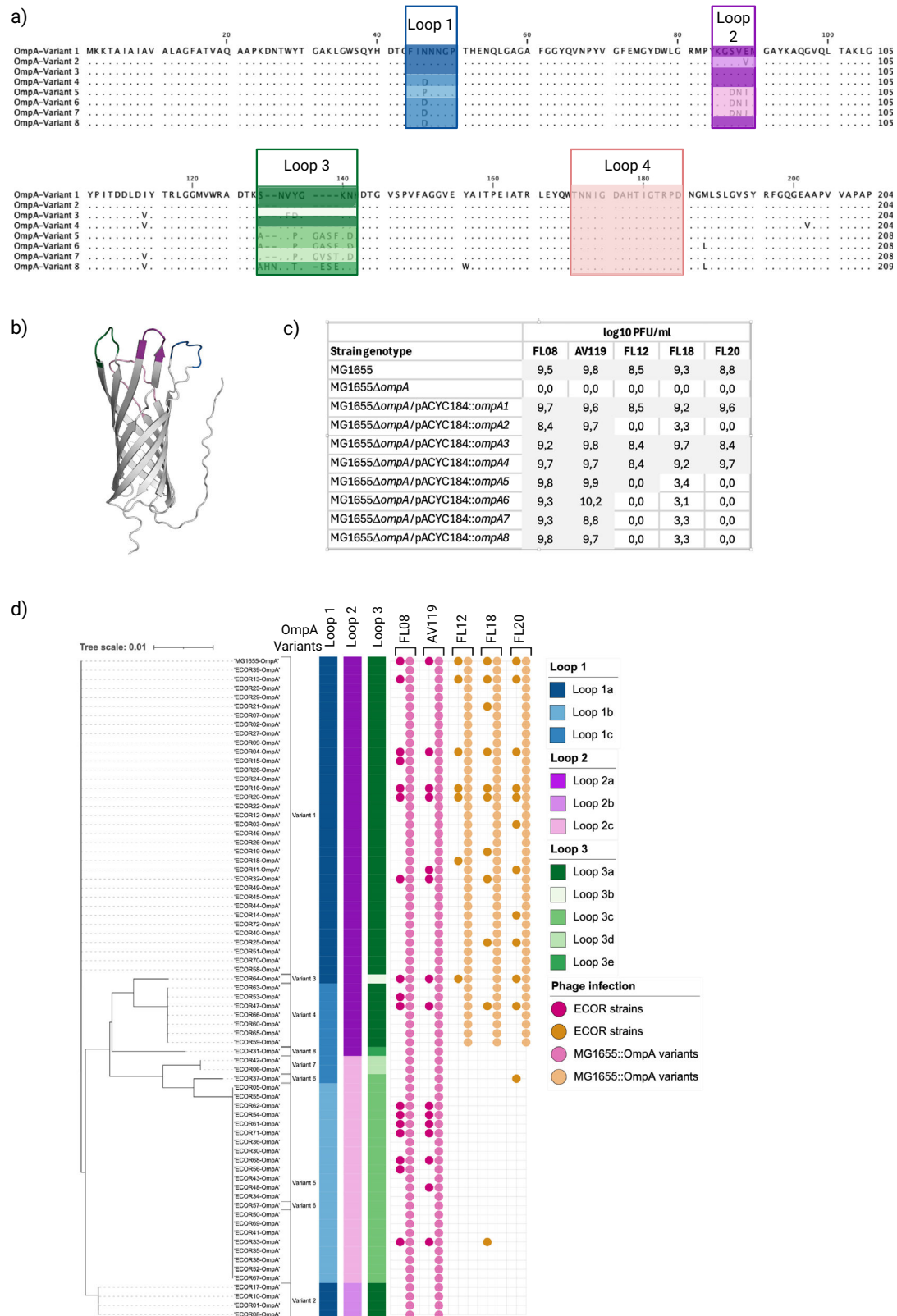


Fig. 4 | Importance of variation of outer loops of receptor OmpA for Gp38 infection. Representative sequences of the eight identified OmpA receptor variants were aligned using CLC. Three outer loops were identified as variable regions (a). These identified extracellular loops were marked in an OmpA AlphaFold model in their respective colors (b). MG1655 mutant strains were constructed, expressing the eight different OmpA variants. Phages were spotted on these strains, MG1655 wildtype as a positive control and MG1655 with an OmpA deletion as a negative control. The resulting plaque formation is calculated as log10 PFU/ml (c). A

phylogenetic tree using the OmpA amino acid sequences of all ECOR strains was constructed using CLC. Strains were colored according to the expression of the different OmpA variants. Outer loop variants of individual ECOR strains were marked in blue (loop 1), purple (loop 2), and green (loop 3). Infectivity, determined by host range and infection of MG1655 mutants expressing the respective OmpA variant, are indicated for phages FL08 and AV119 in dark and light magenta circles, respectively. A similar analysis was conducted for phages FL12, FL18, and FL20; the results are marked by dark and light orange, respectively (d).

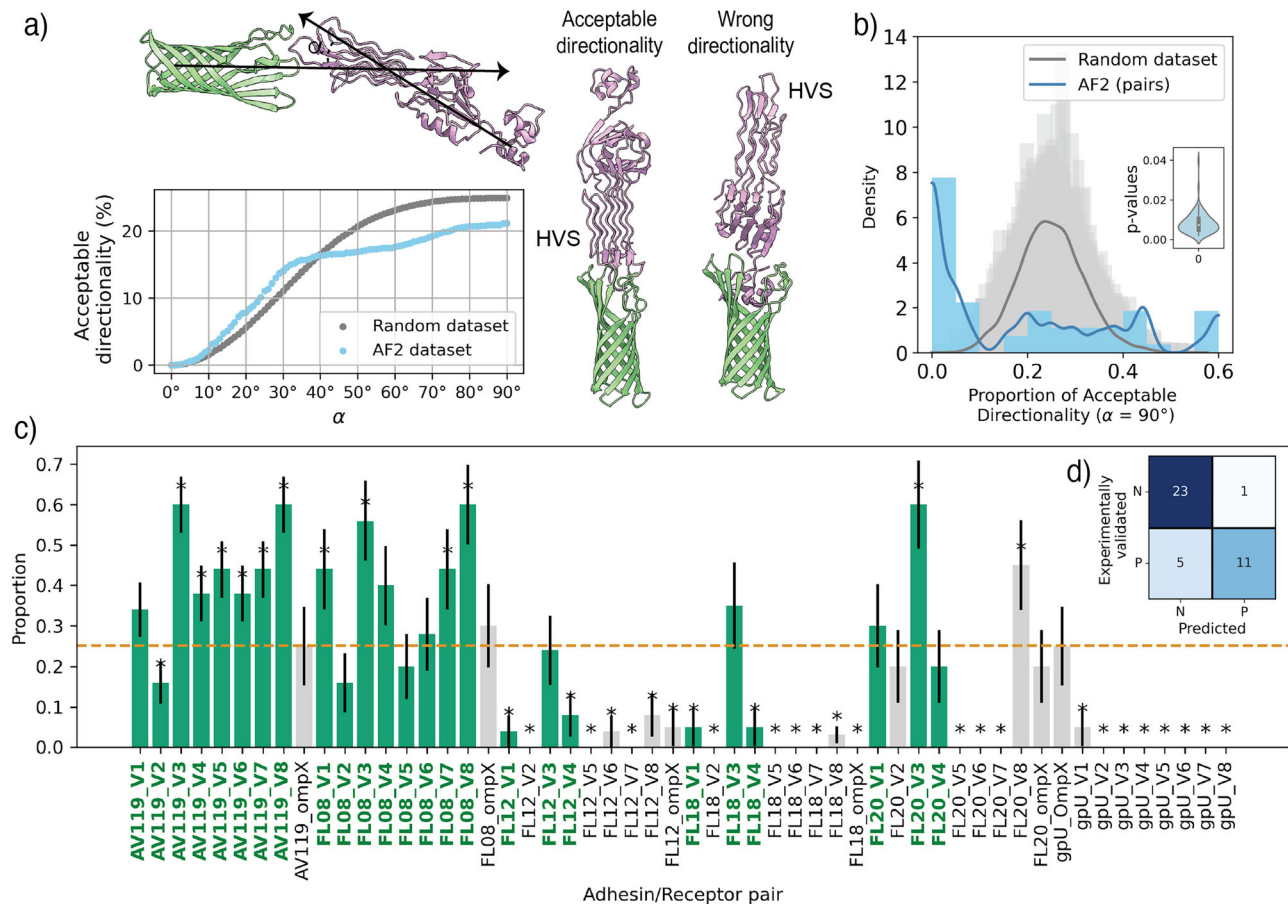


Fig. 5 | AlphaFold2 prediction of Gp38-OmpA binding. To compare MG1655 OmpA variant infection data with all adhesin-receptor predictions using AlphaFold2-multimer, the directionality of receptor/adhesin pairs was determined by defining the principal axis of inertia for each subunit, serving as a proxy for potential binding pairs. A correct interaction model must exhibit acceptable directionality. The percentage of predicted models with acceptable directionality deviates from random placement across different angle (α) thresholds (a). Proportion values for each adhesin/receptor pair were calculated, as well as the proportion for a thousand

generated random datasets, each containing 54 samples with random vector orientations. The distribution of the AlphaFold2 pair models was significantly different from all random distributions (Mann-Whitney U test), with a mean p value of 0.007 (b). We tested (two-sided t test) if each individual proportion was different from 0.25. * Represents a p value < 0.05 ; green represents the true binding partners (confirmed by experiments in Fig. 4c). The dashed line represents the expected proportion for a randomly generated dataset at $\alpha = 90^\circ$ (0.25) (c). A confusion matrix shows the reliability of using such criteria to select true positives (d).

Gp38 variant analysis across different genera allows receptor protein prediction

To explore the distribution of Gp38 adhesin diversity across various genera and determine whether our receptor identification and analysis can be used to predict receptor recognition, we constructed a phylogenetic tree, including Gp38 proteins expressed by *Tequatrovirus*, *Krischvirus*, and *Mosigvirus* phages. According to the literature^{5,16,30,34–37}, selected Gp38 proteins were associated with their respective receptors. This analysis confirmed that closely related Gp38 proteins recognize the same bacterial receptor. We also observed that most Gp38 adhesins are predicted to recognize Tsx, followed by OmpA and OmpF. Only a small group of Gp38 adhesins bind to OmpC (Fig. 7a). Remarkably, while most Gp38 adhesins recognizing similar receptors were phylogenetically similar, the two groups of Gp38 adhesins recognizing OmpA (type 1 and type 2) were far apart (Fig. 7a). Indeed, type 1 Gp38 seemed phylogenetically closer to adhesins recognizing Tsx than to type 2 Gp38 proteins, which could be confirmed by directly comparing their amino acid sequences (Fig. S4). Furthermore, the type 1 Gp38 adhesins were expressed by more phages than the type 2 Gp38. Interestingly, analyzing their protein sequence alignment revealed that the amino acid sequences are conserved within the two different types. The previously identified variable regions between type 1 and type 2 (Fig. 6b) were confirmed for all Gp38 adhesins binding to OmpA (Fig. 7b). This indicates that other phages expressing type 1 Gp38 adhesins may also have a

broader host range than those encoding type 2 Gp38 adhesins. Thus, determining Gp38 diversity reveals bacterial receptor recognition and may further predict the host range of OmpA-dependent phages.

Discussion

The rising number of antibiotic-resistant bacteria seriously threatens global health, turning once harmless infections into life-threatening conditions¹. An alternative to antibiotic treatment is phage therapy, where phages, often combined in phage cocktails, are used to kill bacteria³⁸. Yet, an effective method for selecting appropriate phages is essential for successful therapeutic outcomes. As receptor recognition and binding are the initial steps of the life cycle of phages¹⁴, understanding molecular determinants of receptor recognition is essential for predicting phage hosts. Thus, a common strategy in recent approaches is utilizing RBPs for such predictions^{39,40}. However, these models depend on detailed knowledge of interactions between phage RBPs and their bacterial receptors. Here, we have investigated the molecular mechanism of host recognition by *Straboviridae* phages mediated by Gp38 adhesins carrying HVS. Using a targeted isolation approach, we successfully isolated *Straboviridae* phages belonging to the genera of *Mosigvirus*, *Tequatrovirus*, and *Krischvirus* and identified phages encoding Gp38 adhesins. By identifying the receptors of all Gp38 adhesins and phylogenetically comparing them to other Gp38, we could extrapolate our findings and propose receptors of related yet uncharacterized phages of

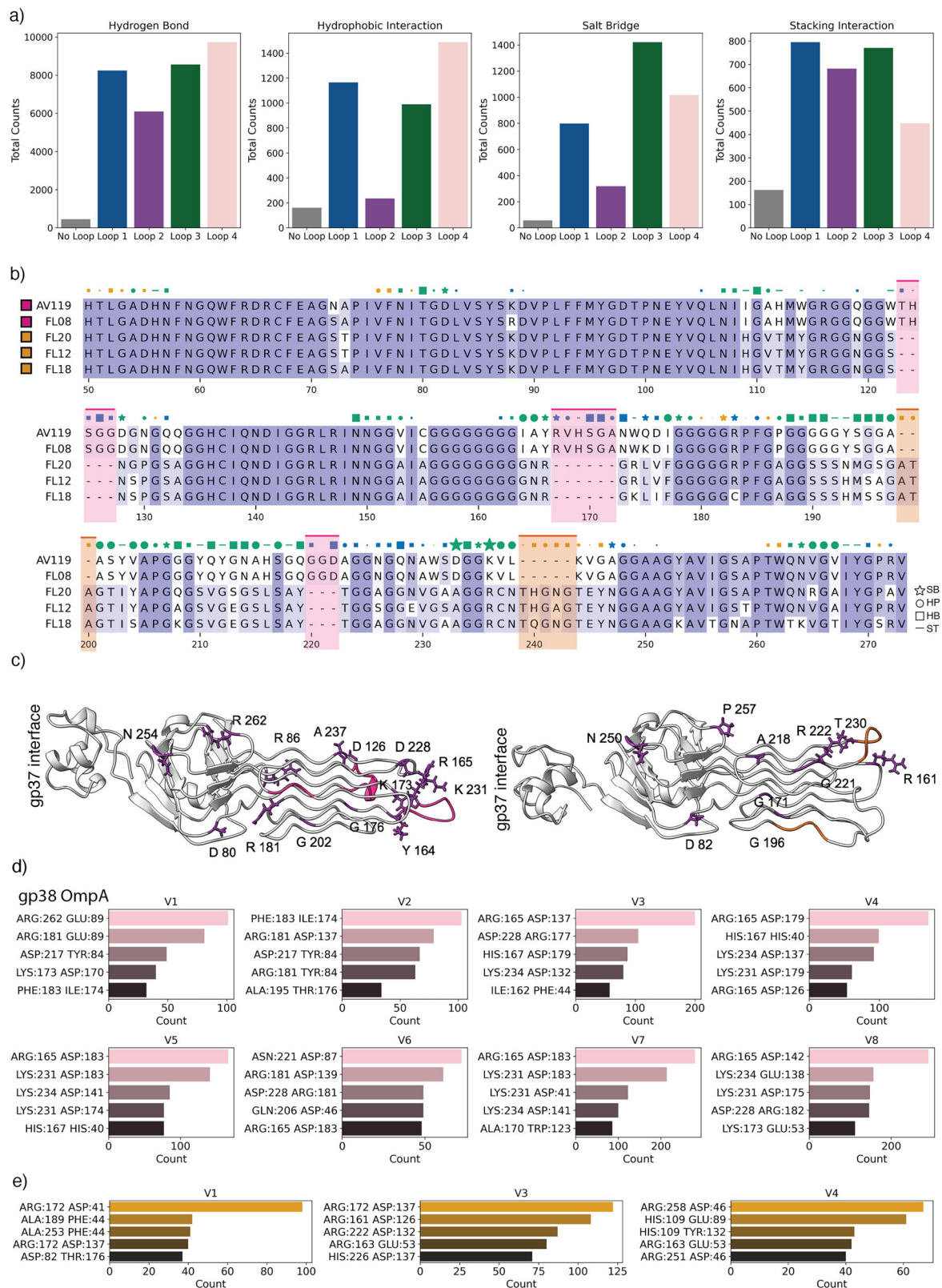
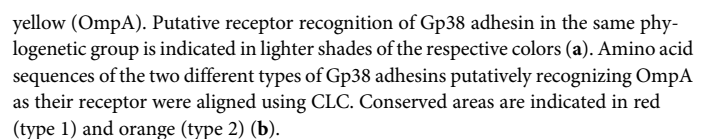


Fig. 6 | Interaction analysis of adhesin-receptor-binding. The total count of non-covalent interactions for 154 true positive models was calculated using molecular dynamics (MD) simulations, run for 10,000 steps with frames captured every 1000 steps. The interactions predominantly occurred in the vicinity of the OmpA loops. **a** Sequence alignment of the five Gp38 adhesins highlights residues with the most interactions. The alignment is colored based on sequence conservation. Legend: (☆) indicates Salt Bridges, (○) indicates Hydrophobic interactions, (●) indicates Hydrogen Bonds, and (-) indicates Stacking interactions. Blue marks

interactions were found only in type 1, yellow marks interactions were found only in type 2, and green marks shared interactions between different types. Magenta highlights unique peptide sequences in type 1, and orange highlights unique peptide sequences in type 2. **b** Side view of models of FL08 and FL20. Residues prone to salt bridge formation are colored in purple. Magenta regions indicate unique sequences of type 1, while orange regions indicate unique sequences of type 2. **c** The top 5 interaction pairs of type 1 (**d**) and type 2 (**e**) with the OmpA variants are shown. The left residue belongs to Gp38, and the right residue belongs to OmpA.



these genera. Interestingly, phages recognizing OmpA as a receptor showed two different host range patterns dependent on their Gp38 variant. Thus, identifying the Gp38 adhesin variant of *Straboviridae* phages may allow the prediction of receptors and phage host range and further improve prediction tools.

It was previously hypothesized that minor amino acid changes may enable phages encoding Gp38 to switch between recognizing OmpA and Tsx¹⁶. Here, we showed that variation in the HVS of Gp38 influences the host range of the phages even if they recognize the same receptor, OmpA. Moreover, Gp38s of type 2 are phylogenetically further apart from type 1 than Gp38 adhesins recognizing Tsx. This suggests that type 2 Gp38 adhesins have more recently adapted from binding to Tsx to recognizing OmpA. This may also explain the narrow host range of phages expressing type 2 Gp38 since these adhesins may not fully have adapted to OmpA as their receptor, causing a less specific binding than type 1 Gp38 adhesins. In addition, our *in silico* analysis revealed that amino acids of type 1 Gp38 adhesins can potentially interact with a larger variety of amino acids in different OmpA variants than type 2 Gp38 adhesins. Interestingly, amino acids involved in this broader host range are conserved among similar Gp38 proteins, which may allow host range prediction of newly isolated phages based on their Gp38 variant. Previous studies utilizing mutant analysis showed that a minor amino acid change in the RBP can cause a change in receptor recognition^{16,41}. Our analysis expands this understanding and shows that different amino acids in the Gp38 adhesin confer binding to different receptors and variants of the same receptor.

This *in-depth* *in silico* amino acid interaction analysis was made possible using AlphaFold2 multimer models in which directionality was used as a measure for adhesin-receptor interaction. Here, AlphaFold2 models showed a directionality bias, accurately placing most true interaction pairs with the correct interface. This likely reflects strong coevolutionary signals in the variable Gp38 sandwich domain and OmpA outer loops, enabling binding predictions that other metrics (pDockQ, local PAE, or pLDDT) failed to capture. While these findings underscore the need for improved general metrics to assess local interactions in AlphaFold2 models with disordered regions, the directionality criterion was validated for the Gp38-OmpA system. This approach could potentially be generalized to other RBP-receptor systems for future RBP-receptor recognition prediction and analysis. For example, it was experimentally shown for other phages, like T5, that RBPs and receptors can interact via loop interaction of both proteins, involving a variety of amino acids⁴².

It is known that common phage receptors, including OmpA, OmpF, and OmpC, are diverse across *E. coli* populations³³. However, the variations in extracellular loops providing a diversity related to phage recognition is an overlooked factor influencing host recognition. Since hypervariable extracellular loop variants in OmpA could also be identified in pathogenic species like *Mahheimia haemolytica*, *Mannheimia glucosida* and *Pasterella trehalosi*⁴³, we hypothesize that our analysis could help predict Gp38 adhesin binding to these variants and may not be limited to *E. coli* strains. One drawback of phage therapy is the high specificity of phages, which often requires the isolation of new phages for each patient. Overall, phage binding analysis at the amino acid level will allow for a more targeted and efficient approach in selecting phages for therapy.

Materials and methods

Specific isolation of *Tevenvirinae* phages using PCR

Pig feces filtrates were diluted 1:100 in SM buffer (200 mM NaCl₂, 10 mM MgSO₄, 50 mM Tris-HCl, pH 7.5) and screened for *T-even* phages by running a PCR protocol using primers targeting the conserved major head protein (CCC TGC TGT TCC AGA TCG ANA ARG ARG C and CTG CCT GGC GTA CTG GTC DAT RWA NAC) previously developed²⁹. Phage AV119³⁰ was used as a positive control for identifying *T-even* phages by PCR. Filtrates that showed a positive PCR result by amplifying a 240 bp DNA fragment were plated with *E. coli* K12 strain MG1655. Briefly, 100 µl of 10⁰–10² dilutions of pig feces filtrates were mixed with 100 µl of the five-hour

culture of MG1655 grown in LB broth (BD 240230) and 3.5 ml 0.6% LB agar and then poured on fresh 1.2% LB agar (BD 240110) plates. The resulting plaques were picked three times and verified using PCR with the primers.

Phage propagation

Phages were propagated by mixing 100 µl of SM buffer containing the picked plaques with 100 µl of a five-hour culture of MG1655 and 3.5 ml 0.6% LB agar. The mix was poured on a fresh 1.2% LB agar plate and incubated overnight at 37 °C. 5 ml SM buffer was added and incubated overnight at 4 °C and 150 rpm. SM buffer was collected from the plates and sterile-filtered using a 0.2 µm filter (Sarstedt).

Extraction of phage DNA

Phages were concentrated before phage DNA extraction. Phage lysate was mixed with a polyethylene glycol (PEG) and NaCl solution to reach a final concentration of 10% PEG and 0.5 M NaCl. The resulting phage solution was incubated at 4 °C overnight and centrifuged for 1.5 hours at 4 °C and 15,000 × g. The supernatant was discarded, and pelleted phages were resuspended in 1:10 of the original volume of SM buffer. Phage DNA extraction was adapted from ref. 44. In short, phages were incubated with a final concentration of 10 µg/ml RNase and 20 µg/ml DNase (Thermo-Scientific) at 37 °C for one hour. EDTA (20 mM final concentration, pH 8) was added, and phage capsids were digested with a final concentration of 50 µg/ml proteinase K (Invitrogen) at 56 °C for 2 hours. One volume of phenol was added, mixed with the digested phages, and centrifuged at 11,000 × g for 15 minutes. The aqueous phase was extracted, and one volume of phenol-chloroform-isoamyl alcohol (25:24:19) was added, followed by centrifugation and aqueous phase extraction as before. This aqueous phase was mixed with one volume of chloroform-isoamyl alcohol (24:1) and centrifuged. This step was repeated once. 0.1 volume of sodium acetate (3 M, pH 5.5), to a final concentration of 0.05 µg/l, and 2.5 volumes of ice-cold absolute ethanol were added. DNA samples were incubated for six days at −20 °C. Samples were centrifuged (20 minutes, 4 °C, at 20,000 × g) and washed three times with 70% ice-cold ethanol. The pellet was dried at 37 °C and resuspended in nuclease-free water. Phage DNA was further purified using the Genomic DNA Clean and Concentrator kit (Zymo, D4011) according to the kit's protocol.

Phage DNA sequencing and genome analysis

Phage DNA was prepared for sequencing using the Illumina Nexterra Flex library preparation kit, and Illumina MiSeq sequencing was performed using 2 × 250 bp paired end approach. Reads are accessible under <https://data.mendeley.com/datasets/gwx5zpwgs5/1>. Reads were assembled to contigs using the CLC genomics workbench software by first trimming the reads and performing a *de novo* assembly with default parameters. The resulting contig was annotated using PATRIC⁴⁵. Here, the Genome Annotation function “Bacteriophages” and the closest BLAST hit for the taxonomy name were used. Whole genome comparison of the bacteriophage genomes was performed using a heatmap by completing a whole genome alignment first, followed by ANI comparison in CLC Main Workbench. Gene synteny of phages within one genus was visualized using Easyfig⁴⁶. Phage taxonomy was determined by running a BLAST analysis and confirmed using VICTOR (Virus classification and tree building online resource)⁴⁷. For the VICTOR analysis, 85 genomes, including two representatives of each genus in the *Straboviridae* family, and the closest BLAST hits to the isolated phage genomes were analyzed using the D4 formula. VICTOR results were visualized using iTOL⁴⁸. Gp38 sequences were identified by performing a whole genome alignment using CLC Main Workbench with a previously identified gp38 adhesin sequence (*Escherichia* phage vB_EcoM-ffEco06, NC_054914.1:156717-157517) and a published phage genome carrying the gp38 gene (*Enterobacteria* phage Bp7, NC_019500-1) to confirm the location of the gp38 gene. The resulting sequences and their correct folding were confirmed by submitting the whole Gp38 protein sequence to HHpred^{49,50} and to AlphaFold^{51,52}.

Bacterial DNA sequencing and ECOR OmpA variant identification

The genomic DNA from the ECOR strains was extracted using the DNeasy blood and tissue kit (Qiagen), following the manufacturer's instructions, and normalized to 10 ng/μl. Next, the sequencing library was prepared according to the manufacturer's instructions using the rapid barcoding kit, SQK-RBK114.24 (Oxford Nanopore Technologies), starting with 100 ng DNA per sample and pooling 24 barcoded samples in an equimolar ratio. The final library was sequenced on a MinION R10.4.1 flow cell (ONT) for 48 to 72 h. Basecalling and demultiplexing were performed with Dorado v0.6.1 using the super-high accuracy DNA model. The resulting reads were mapped to the eight known OmpA variants³³ using CLC workbench, and its extracellular loops were identified compared to the literature³².

Phage host range analyses

A two-step approach was used to determine the phage's host range. First, all undiluted phage stocks were spotted on bacterial lawns of each of the *E. coli* strains from the ECOR collection, using square LB agar plates containing an overlay of 330 μl five-hour culture of the respective strains, mixed with 11 ml of 0.6% LB agar. In the next step, all the phage-bacteria combinations that showed a lysis spot were further investigated by spotting 10 μl of 10¹–10⁶ dilutions of the phage stocks on lawns of respective ECOR strains by preparing plates with 100 μl of a five-hour culture of the ECOR strains and 3.5 ml 50% Brain Heart Infusion (BHI) agar (Oxoid CM1138). Phages that showed a lysis spot for 10^{−2} and 10^{−3} dilutions but no single plaques were further analyzed in a whole plate assay. Here, 100 μl of phage 10^{−1} to 10^{−6} dilutions were mixed with 100 μl of a five-hour culture of the respective strain and 0.5% BHI agar. Host range analysis was repeated once for positive strains. Plaques were counted, and the efficiency of plaquing (EOP) was calculated using the plaque counts calculated on MG1655 as a reference.

Construction of bacterial receptor deletion mutants

Receptor mutants were constructed according to³⁰. Briefly, outer membrane protein genes *ompA*, *ompC*, *ompF*, *btuB*, *fadL*, *tolC*, *fhuA*, *tsx*, *tonB*, *ompW*, *lamB* and *fepA* were deleted by exchanging the genes with a deletion cassette using the Quick and Easy *E. coli* Gene deletion Kit (GeneBridges). PCR was performed using primers amplifying a Kanamycin gene with overhangs corresponding to the gene to be deleted in MG1655 (Table S2). Electrocompetent MG1655 containing the plasmid pRedET for the expression of the λ red integrase was freshly prepared. MG1655-pRedET was grown together in 100 μg/ml Ampicillin at 30 °C to an OD₆₀₀ of 0.3, the λ red integrase was induced by adding arabinose to a final concentration of 0.35% and then further grown for one hour at 37 °C. Cells were pelleted (5000 × g, 4 °C, 5 minutes) and washed four times with ice-cold sterile H₂O with constantly reducing volume, leading to an end volume of 300 μl electrocompetent cells. DNA was transformed into the electrocompetent MG1655 cells expressing the λ red integrase by electroporating 2 μl of PCR product with 50 μl competent cells (2.1 mV, 5 sec). Cells were revived in 1200 μl prewarmed LB broth and incubated at 37 °C and 650 rpm for 3 hours followed by plating on LB agar plates containing Kanamycin (50 μg/ml). Positive clones were identified by PCR and verified with Sanger sequencing (Eurofins).

Phage infection of receptor mutants

100 μl of a five-hour culture of MG1655 deletion mutants (MG1655Δ*ompA*, MG1655Δ*ompC*, MG1655Δ*ompF*, MG1655Δ*btuB*, MG1655Δ*fadL*, MG1655Δ*tolC*, MG1655Δ*fhuA*, MG1655Δ*tsx*, MG1655Δ*tonB*, MG1655Δ*ompW*, MG1655Δ*lamB*, and MG1655Δ*fepA*) grown in LB as mixed with 3.5 ml preheated 0.6% LB agar and poured on LB agar plates. Plates were dried for 30 minutes, and 10 μl of phage dilutions (10^{−1}–10^{−6}) were spotted. EOP was calculated using the PFU of the respective wild-type strains as a reference.

Cloning of OmpA receptor variants

To express different OmpA variants, we PCR amplified the corresponding ORF, including its promoter region, using conserved primers (Table S4), the

Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific), and chromosomal DNA from respective variant-encoding strains as template. Amplicons were cloned into pACYC184 at *Hind*III and *Sal*I restriction sites, transformed into the *E. coli* cloning host IM08B, and selected on 25 mg/l chloramphenicol. Successful clones were identified by subsequent selection for loss of tetracycline resistance. The resulting plasmid constructs were transformed into MG1655Δ*ompA* to determine the phage efficiency of plating.

AlphaFold2-multimer prediction and binding analysis

AlphaFold2-multimer (v2.3.1) was used to generate dimeric models of adhesins with the antiparallel β-barrel domain of different OmpA proteins. For each adhesin/receptor pair, 25 to 60 models were generated using an Nvidia Quadro RTX 8000 graphics card. The interface pLDDT and pDockQ scores were calculated as described in the referenced study⁵³. For each generated model, the principal axis of inertia for both the adhesin and receptor subunits was determined by calculating the eigenvectors of the covariance matrix based on the recentered coordinates relative to their geometric centers. The geometric center was calculated as the mean of the atomic coordinates within each chain. The angle between the principal axes of the adhesin and receptor was then computed to assess the biological relevance of the models. Additionally, the distance between the most distal points of both subunits was measured to define the directionality. The proportion of models with acceptable directionality for each adhesin/receptor pair was determined by calculating the ratio of models with potential binding interactions (given an angle α) to the total number of generated models for that pair.

To evaluate the statistical significance of the observed proportions, we generated random datasets for comparison. We conducted 100 iterations, each generating 40 random subsets. For each subset, a random number of vectors (*n*) between 20 and 60 was generated. The angle between each vector pair was calculated, and a binary outcome for similarity was determined based on a random value: if the value was less than 0.25, the pair was marked as True (similar), otherwise as False (dissimilar). The proportions of True outcomes with angles less than 90° were calculated and stored, forming a distribution of random data proportions. A Mann–Whitney *U* test was then performed to compare the random data proportions with the observed data proportions, assessing if the random data had significantly lower proportions of True similarity outcomes than the observed data. The resulting *p*-values were collected for further analysis.

A two-tailed *t* test was employed, in which adhesin-receptor pairs with a significantly higher proportion (*p*) than 25% were classified as binding, those with a significantly lower proportion as no binding, and others as not classified. To determine if the proportion values for specific pairs differed from 0.25, we first estimated the standard error using the formula $std = \sqrt{\frac{p(1-p)}{n}}$, where *n* is the total number of models for each pair. We then conducted a *t* test to compare the observed proportions against the expected value from a similarly generated random dataset. The statistical analyses were performed according to Virtanen et al.⁵⁴.

Molecular dynamics and non-covalent interactions estimation

The selected true positive models were amber-relaxed using the Amber14 force field. The system was set up in OpenMM⁵⁵ with periodic boundary conditions and solvated with the appropriate parameters. A Langevin integrator was employed at a temperature of 298 K with a time step of 2 femtoseconds. The number of clashes and PDB manipulation was calculated using ChimeraX-1.6.1(60). Models with fewer than 10 clashes were selected for MD simulation. The PDB files were centered, and the system was prepared using the Amber14 force field, with periodic boundary conditions applied at a box size of 20.0 nm. The system was then energy minimized and subjected to molecular dynamics simulations for 10,000 steps at 310 K using a Langevin integrator. The simulation utilized GPU acceleration and was run on an Nvidia A100. The non-covalent interactions were calculated using the GetContacts dynamics script (<https://getcontacts.github.io/>) with default parameters.

Phylogenetic analysis of Gp38 adhesins

Bacteriophage genomes of the genera *Tequatrovirus*, *Krischvirus*, and *Mosigvirus* were downloaded from NCBI using CLC. Gp38 adhesins were identified by performing a whole genome alignment of all phage genomes to the Gp38 adhesin of phage FL04. Additionally, Gp38 adhesins of phages isolated in this study were blasted against the NCBI database to enable the identification of Gp38 adhesins outside of the *Tequatrovirus*, *Krischvirus*, and *Mosigvirus* genera. The alignment of the identified Gp38 protein sequences, excluding the conserved N-terminus, was used to construct a phylogenetic tree, using the Neighbor Joining algorithm. Previously identified receptors from this study and the literature^{5,16,30,34–37} were marked in the phylogenetic tree using iTOL⁴⁸.

Data availability

Sequence data raw reads are accessible under <https://data.mendeley.com/datasets/gwx5zpwgs5/1>.

Received: 10 January 2025; Accepted: 7 April 2025;

Published online: 25 April 2025

References

- Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature* **529**, 336–343 (2016).
- World Health, O. WHO bacterial priority pathogens list, 2024: bacterial pathogens of public health importance, to guide research, development and strategies to prevent and control antimicrobial resistance. xii, 56 p. (World Health Organization, 2024).
- Hatfull, G. F., Dedrick, R. M. & Schooley, R. T. Phage therapy for antibiotic-resistant bacterial infections. *Annu Rev. Med* **73**, 197–211 (2022).
- Turner, D. et al. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Arch. Virol.* **168**, 74, (2023).
- Salem, M., Pajunen, M. I., Jun, J. W. & Skurnik, M. T4-like bacteriophages isolated from pig stools infect yersinia pseudotuberculosis and yersinia pestis using LPS and OmpF as receptors. *Viruses* **13**, <https://doi.org/10.3390/v13020296> (2021).
- Li, M. et al. Characterization of the novel T4-like Salmonella enterica bacteriophage STP4-a and its endolysin. *Arch. Virol.* **161**, 377–384 (2016).
- Jin, J. et al. Genome organisation of the Acinetobacter lytic phage ZZ1 and comparison with other T4-like Acinetobacter phages. *BMC Genomics* **15**, 793 (2014).
- Groover, K. E. et al. Complete genome sequence of Klebsiella aerogenes myophage metamorpho. *Microbiol. Resour. Announc.* **10**, e01420-20 (2021).
- Denou, E. et al. T4 phages against Escherichia coli diarrhea: potential and problems. *Virology* **388**, 21–30 (2009).
- Sarker, S. A. et al. Oral T4-like phage cocktail application to healthy adult volunteers from Bangladesh. *Virology* **434**, 222–232 (2012).
- Riede, I. Receptor specificity of the short tail fibres (gp12) of T-even type Escherichia coli phages. *Mol. Gen. Genet.* **206**, 110–115 (1987).
- Islam, M. Z. et al. Molecular anatomy of the receptor binding module of a bacteriophage long tail fiber. *PLoS Pathog.* **15**, e1008193 (2019).
- Dunne, M. et al. Salmonella phage S16 tail fiber adhesin features a rare polyglycine rich domain for host recognition. *Structure* **26**, 1573–1582.e1574 (2018).
- Bertozzi Silva, J., Storms, Z. & Sauvageau, D. Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett* **363**, fnw002 (2016).
- Tétart, F., Desplats, C. & Krisch, H. M. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *J. Mol. Biol.* **282**, 543–556 (1998).
- Trojet, S. N., Caumont-Sarcos, A., Perrody, E., Comeau, A. M. & Krisch, H. M. The gp38 adhesins of the T4 superfamily: a complex modular determinant of the phage's host specificity. *Genome Biol. Evol.* **3**, 674–686 (2011).
- Botstein, D. A theory of modular evolution for bacteriophages. *Ann. N. Y Acad. Sci.* **354**, 484–490 (1980).
- Drexler, K., Riede, I., Montag, D., Eschbach, M. L. & Henning, U. Receptor specificity of the Escherichia coli T-even type phage Ox2. Mutational alterations in host range mutants. *J. Mol. Biol.* **207**, 797–803 (1989).
- Schulz, G. E. Porins: general to specific, native to engineered passive pores. *Curr. Opin. Struct. Biol.* **6**, 485–490 (1996).
- Schulz, G. E. Bacterial porins: structure and function. *Curr. Opin. Cell Biol.* **5**, 701–707 (1993).
- Stenberg, F. et al. Protein complexes of the Escherichia coli cell envelope. *J. Biol. Chem.* **280**, 34409–34419 (2005).
- Zhu, Y. et al. Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting. *Mol. Cell Proteom.* **8**, 1999–2010 (2009).
- Traurig, M. & Misra, R. Identification of bacteriophage K20 binding regions of OmpF and lipopolysaccharide in Escherichia coli K-12. *FEMS Microbiol Lett.* **181**, 101–108 (1999).
- Nieweg, A. & Bremer, E. The nucleoside-specific Tsx channel from the outer membrane of Salmonella typhimurium, Klebsiella pneumoniae and Enterobacter aerogenes: functional characterization and DNA sequence analysis of the tsx genes. *Microbiology* **143**(Pt 2), 603–615 (1997).
- Schneider, H., Fsihi, H., Kottwitz, B., Mygind, B. & Bremer, E. Identification of a segment of the Escherichia coli Tsx protein that functions as a bacteriophage receptor area. *J. Bacteriol.* **175**, 2809–2817 (1993).
- Samsudin, F., Ortiz-Suarez, M. L., Piggot, T. J., Bond, P. J. & Khalid, S. OmpA: a flexible clamp for bacterial cell wall attachment. *Structure* **24**, 2227–2235 (2016).
- Morona, R., Klose, M. & Henning, U. Escherichia coli K-12 outer membrane protein (OmpA) as a bacteriophage receptor: analysis of mutant genes expressing altered proteins. *J. Bacteriol.* **159**, 570–578 (1984).
- Koebnik, R. Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein OmpA from Escherichia coli. *J. Bacteriol.* **181**, 3688–3694 (1999).
- Born, Y. et al. A major-capsid-protein-based multiplex PCR assay for rapid identification of selected virulent bacteriophage types. *Arch. Virol.* **164**, 819–830 (2019).
- Vitt, A. R. et al. Diverse bacteriophages for biocontrol of ESB- and AmpC-β-lactamase-producing E. coli. *iScience* **27**, 108826. <https://doi.org/10.1016/j.isci.2024.108826> (2024).
- Ochman, H. & Selander, R. K. Standard reference strains of Escherichia coli from natural populations. *J. Bacteriol.* **157**, 690–693 (1984).
- Liao, C. et al. Allelic variation of Escherichia coli outer membrane protein A: impact on cell surface properties, stress tolerance and allele distribution. *PLoS One* **17**, e0276046 (2022).
- Chen, X. et al. Mosaic evolution of beta-barrel-porin-encoding genes in Escherichia coli. *Appl. Environ. Microbiol.* **88**, e0006022 (2022).
- Maffei, E. et al. Systematic exploration of Escherichia coli phage-host interactions with the BASEL phage collection. *PLoS Biol.* **19**, e3001424 (2021).
- Goodridge, L., Gallaccio, A. & Griffiths, M. W. Morphological, host range, and genetic characterization of two coliphages. *Appl. Environ. Microbiol.* **69**, 5364–5371 (2003).
- Roush, C., Chan, B. K., Turner, P. E. & Burmeister, A. R. Assembly and annotation of the complete genome sequence of T4-Like bacteriophage 132. *Microbiol. Resour. Announc.* **10**, e0064921 (2021).
- An, W. et al. Assembly and annotation of Escherichia coli bacteriophage U115. *Microbiol. Resour. Announc.* **11**, e0094921 (2022).

38. Luong, T., Salabarria, A. C. & Roach, D. R. Phage therapy in the resistance era: where do we stand and where are we going?. *Clin. Ther.* **42**, 1659–1680 (2020).
39. Boeckaerts, D. et al. Predicting bacteriophage hosts based on sequences of annotated receptor-binding proteins. *Sci. Rep.* **11**, 1467 (2021).
40. Boeckaerts, D. et al. Prediction of *Klebsiella* phage-host specificity at the strain level. *Nat. Commun.* **15**, 4355 (2024).
41. Taslem Mourosi, J. et al. Understanding bacteriophage tail fiber interaction with host surface receptor: the key “blueprint” for reprogramming phage host range. *Int. J. Mol. Sci.* **23**, 12146 (2022).
42. Degroux, S., Effantin, G., Linares, R., Schoehn, G. & Breyton, C. Deciphering bacteriophage T5 host recognition mechanism and infection trigger. *J. Virol.* **97**, e0158422 (2023).
43. Davies, R. L. & Lee, I. Sequence diversity and molecular evolution of the heat-modifiable outer membrane protein gene (*ompA*) of *Mannheimia*(*Pasteurella*) *haemolytica*, *Mannheimia glucosida*, and *Pasteurella trehalosi*. *J. Bacteriol.* **186**, 5741–5752 (2004).
44. Gambino, M. et al. Phage S144, A New Polyvalent Phage Infecting *Salmonella* spp. and *Cronobacter sakazakii*. *Int J Mol Sci* **21**, <https://doi.org/10.3390/ijms21155196> (2020).
45. Wattam, A. R. et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* **45**, D535–d542 (2017).
46. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
47. Meier-Kolthoff, J. P. & Göker, M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* **33**, 3396–3404 (2017).
48. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–w296 (2021).
49. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
50. Gabler, F. et al. Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinforma.* **72**, e108. <https://doi.org/10.1002/cpbi.108> (2020).
51. Varadi, M. et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–d444 (2022).
52. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
53. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).
54. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
55. Eastman, P. et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).

Acknowledgements

This work was funded by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (1R01EB027895). The Novo Nordisk Foundation Center for Protein Research is supported

financially by the Novo Nordisk Foundation (NNF14CC0001). N.M.I.T. acknowledges support from NNF Hallas-Møller Emerging Investigator grant (NNF17OC0031006), NNF Hallas-Møller Ascending Investigator grant (NNF23OC0081528), and NNF Project grant (NNF21OC0071948). N.M.I.T. is a member of the Integrative Structural Biology Cluster (ISBUC) at the University of Copenhagen. V.K.S acknowledges the Novo Nordisk Foundation Copenhagen PhD Program for grant NNF0069780.

Author contributions

V.T.L.: conceptualization, methodology, validation, formal analysis, investigation, visualization, writing—original draft, writing—review & editing. V.K.-S.: conceptualization, formal analysis, investigation, visualization, writing—original draft. M.S.B.: Investigation. S.S.F.V.O.: investigation. S.R.N.: project administration, funding acquisition. N.M.I.T.: conceptualization, writing -review & editing, funding acquisition. M.G.: conceptualization, supervision, methodology, funding acquisition. writing—review and editing. L.B.: conceptualization, supervision, project administration, funding acquisition, writing—review & editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44298-025-00118-9>.

Correspondence and requests for materials should be addressed to Lone Brøndsted.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025