Completing the BASEL phage collection to unlock hidden diversity for systematic exploration of phage-host interactions

Dorentina Humolli[1,&], Damien Piel[1,&], Enea Maffei[1,2,#], Yannik Heyer[2,#], Elia Agustoni[2,#], Aisylu Shaidullina[1,2], Luc Willi[2], Patrick Imwinkelried[2], Fabienne Estermann[2], Aline Cuénod[3], Dominik P. Buser[2], Carola Alampi[4], Mohamed Chami[4], Adrian Egli[3,5], Sebastian Hiller[2], Matthew Dunne[6], Alexander Harms[1,2]*

[1]Institute of Food, Nutrition, and Health (IFNH), ETH Zürich, Zürich, Switzerland

[2]Biozentrum, University of Basel, Basel, Switzerland

[3]Division of Clinical Bacteriology and Mycology, University Hospital Basel, Basel, Switzerland

[4]BioEM lab, Biozentrum, University of Basel, Basel, Switzerland

[5]current address: Institute for Medical Microbiology, University of Zürich, Zürich, Switzerland

[6]Micreos GmbH, Wädenswil, Switzerland

[&]These authors contributed equally.

[#]These authors contributed equally.

*Correspondence should be addressed to Prof. Alexander Harms (alexander.harms@hest.ethz.ch)

Short title: New BASEL phages requiring O-antigen or inhibited by K-12 prophages

Abstract

Research on bacteriophages, the viruses infecting bacteria, has fueled the development of modern molecular biology and inspired their therapeutic application to combat bacterial multidrug resistance. However, most work has so far focused on a few model phages which impedes direct applications of these findings in clinics and suggests that a vast potential of powerful molecular biology has remained untapped. We have therefore recently composed the BASEL collection of *Escherichia coli* phages (BActeriophage SElection for your Laboratory) which made a relevant diversity of phages infecting the *E. coli* K-12 laboratory strain accessible to the community. These phages are widely used, but their assorted diversity had remained limited by the *E. coli* K-12 host. We have therefore now genetically overcome the two major limitations of *E. coli* K-12, its lack of O-antigen glycans and the presence of resident bacterial immunity. Restoring O-antigen expression resulted in the isolation of diverse additional viral groups like *Kagunavirus*, *Nonanavirus*, *Gordonclarkvirinae*, and *Gamaleyavirus*, while eliminating all known antiviral defenses of *E. coli* K-12 additionally enabled us to isolate phages of *Wifcevirus* genus. Even though some of these viral groups appear to be common in nature, no phages from any of them had previously been isolated using *E. coli* laboratory strains and they had thus remained largely understudied. Overall, 37 new phage isolates have been added to complete the BASEL collection. These phages were deeply characterized genomically and phenotypically with regard to host receptors, sensitivity to antiviral defense systems, and host range. These results highlighted dominant roles of the O-antigen barrier for viral host recognition and of restriction-modification systems in bacterial immunity. We anticipate that the completed BASEL collection will propel research on phage-host interactions and their molecular mechanisms, deepening our understanding of viral ecology and fostering innovations in biotechnology and antimicrobial therapy.

Introduction

Bacteriophages, the viruses infecting bacteria, have attracted great interest due to their abundance and diversity in all ecosystems, their key roles as inspiration for biotechnology, and their therapeutic potential in the fight against resilient bacterial infections [1-4]. Most cultured phages are so-called tailed phages or *Caudoviricetes* that characteristically use a tail structure at the virion to inject their genome into host cells. Morphologically, tailed phages can be classified as myoviruses (contractile tail), siphoviruses (long, flexible, non-contractile tail), and podoviruses (short stubby tail). Following the "phage treaty" declared by Max Delbrück in 1944 [5], a few tailed phages infecting *Escherichia coli* have become key model systems for studying the molecular biology of phage-host interactions [6]. We therefore know comparably much about the molecular mechanisms underlying infections of *E. coli* with these seven "T phages" T1 – T7 [7, 8] and a few additional ones infecting *E. coli* or other major model organisms like *Bacillus subtilis*. While this work has resulted in many major breakthroughs since the "golden era" of molecular biology [3, 9], the limited number and diversity of these model phages is a severe limitation for systematic research that could inform the effective use of natural phage diversity for therapeutic or biotechnological applications. We and others have therefore proposed broader sets of model phages infecting commonly studied host organisms such as *E. coli* [10-12], *Klebsiella* [13], or *Pseudomonas aeruginosa* and other pseudomonads [14-16]. Our previous work had presented 68 well-characterized new *E. coli* phages as the BASEL (BActeriophage SElection for your Laboratory) collection which we have made accessible to the academic community.

The value of these BASEL phages lies in their "assorted diversity", providing access to a broad spectrum of biological diversity both within and between phage groups when exploring any aspect of phage-host interactions [6, 10]. For example, the BASEL collection has been used to study phage host recognition [17], viral dependence on host codons [18], porin functionality [19], phage replication rates [20], and bacterial immunity [21-25]. A key prerequisite for broad applicability of the BASEL phages is that they have deliberately been isolated using the *E. coli* K-12 laboratory strain as host, the most important model

organism in fundamental microbiology research [9, 26]. However, phage isolation with *E. coli* K-12 as host has also conflicted with the goal of assorted phage diversity by introducing inherent biases that prevented us from sampling a complete overview of *E. coli* phage diversity [10]. Firstly, the domesticated *E. coli* K-12 laboratory strain has lost the O-antigen glycan chains on the lipopolysaccharide (LPS) core that form a dense barrier on wild *E. coli* and shield cell surface receptors from viral recognition [27-30]. While the absence of this barrier made it easier to isolate multiple phages from different viral groups, it naturally excluded the isolation of phages which depend on the O-antigen as host receptor [31]. Furthermore, a second level of sampling bias may have been caused by known (and possibly unknown) antiviral defenses encoded by *E. coli* K-12, although we had deleted some of them already in the host strain used to isolate and study the original BASEL collection [10].

In the current study we have addressed both limitations of our previous work by expanding the BASEL collection with 37 new phages that largely depend on (restored) O-antigen expression and of which some are inhibited by resident antiviral defenses in the K-12 genome in the form of O-antigen glycosylation. Notably, while almost none of these phages belong to groups of which other representatives have previously been found to infect *E. coli* K-12, all these phages do infect *E. coli* K-12 with a few easily implemented modifications and can thus be used for systematic fundamental research as an expansion of the original BASEL collection. Similar to our previous study, we have performed a systematic phenotypic and genomic characterization of these phages with view to host receptors, bacterial immunity, and host range, and we make them accessible to the scientific community. These results extend some of our previous observations and complement them with new discoveries on previously understudied phage groups such as the *Gordonclarkvirinae* podoviruses that are abundant in nature, highly resistant to tested phage-defense systems, and have a comparably broad host range. The dependence of many newly included phages on O-antigen as host receptor highlights the dominant role of viral adsorption for *E. coli* phage host range and provides a valuable tool for studying diverse mechanisms of glycan recognition using the well-studied model strain *E. coli* K-12. Taken together, our findings provide novel insights into the ecology and

evolution of bacteriophages as well as into their arms race with each other and bacterial immunity. The now completed BASEL collection comprises 106 phages that have been isolated from diverse sources and belong to more than 30 genera (Tables S1 and S2). We anticipate that our work will serve as a reference point for future studies in fundamental phage biology and strengthen the BASEL collection as an effective and widely used tool in the scientific community.

Results

Overview: Including O-antigen-dependent phages in the BASEL collection

The cell surface of Gram-negative bacteria is dominated by LPS and proteins embedded in the outer membrane on which additional layers of protective exopolysaccharides such as a capsule can be present [32]. To recognize their host and pass through these layers, phages usually directly bind exposed glycans as an initial host receptor before moving down to the cell surface [10, 31, 33, 34]. There, the virus typically binds a second so-called terminal receptor which results in irreversible adsorption and DNA ejection [34]. The receptor-binding proteins (RBPs) that target primary and terminal receptors are typically displayed on different structures with longer, lateral tail fibers or stubby tailspikes commonly targeting primary receptors followed by central tail fibers or tail tip complexes binding terminal receptors [reviewed in reference 34].

Bacterial exopolysaccharides and the dense O-antigen layer of the LPS effectively shield the cell surface from the access of tailed phages unless the phage carries enzymatic tailspikes or glycan-targeting tail fibers to specifically bind them as part of primary receptor recognition [28, 29, 31, 33, 34] (Fig. 1A). Consequently, the variability of the ca. 180 different types of O-antigen and other surface structures greatly contribute to the observed narrow host range of most enterobacterial phages [10, 30, 35, 36]. This diversity is further increased by accessory O-antigen modification systems that can alter the polysaccharide chains, e.g., by acetylation or glycosylation [36]. However, laboratory strains of the *E. coli* K-12 lineage do not express protective exopolysaccharides such as capsules and lack even the O-antigen layer due to disruption

of the *wbbL* gene by a transposon [27] (Fig. 1A). Phages can therefore readily access the bacterial cell surface which enabled broad sampling of *E. coli* phages for the original BASEL collection [10, 28]. However, most of these phages are unable to infect *E. coli* K-12 with restored O16-type O-antigen with few exceptions that likely carry RBPs targeting this glycan such as *Hanrivervirus* JakobBernoulli (Bas07) or *Tequintavirus* IrisvonRoten (Bas32; Figs. 1A and 1B).

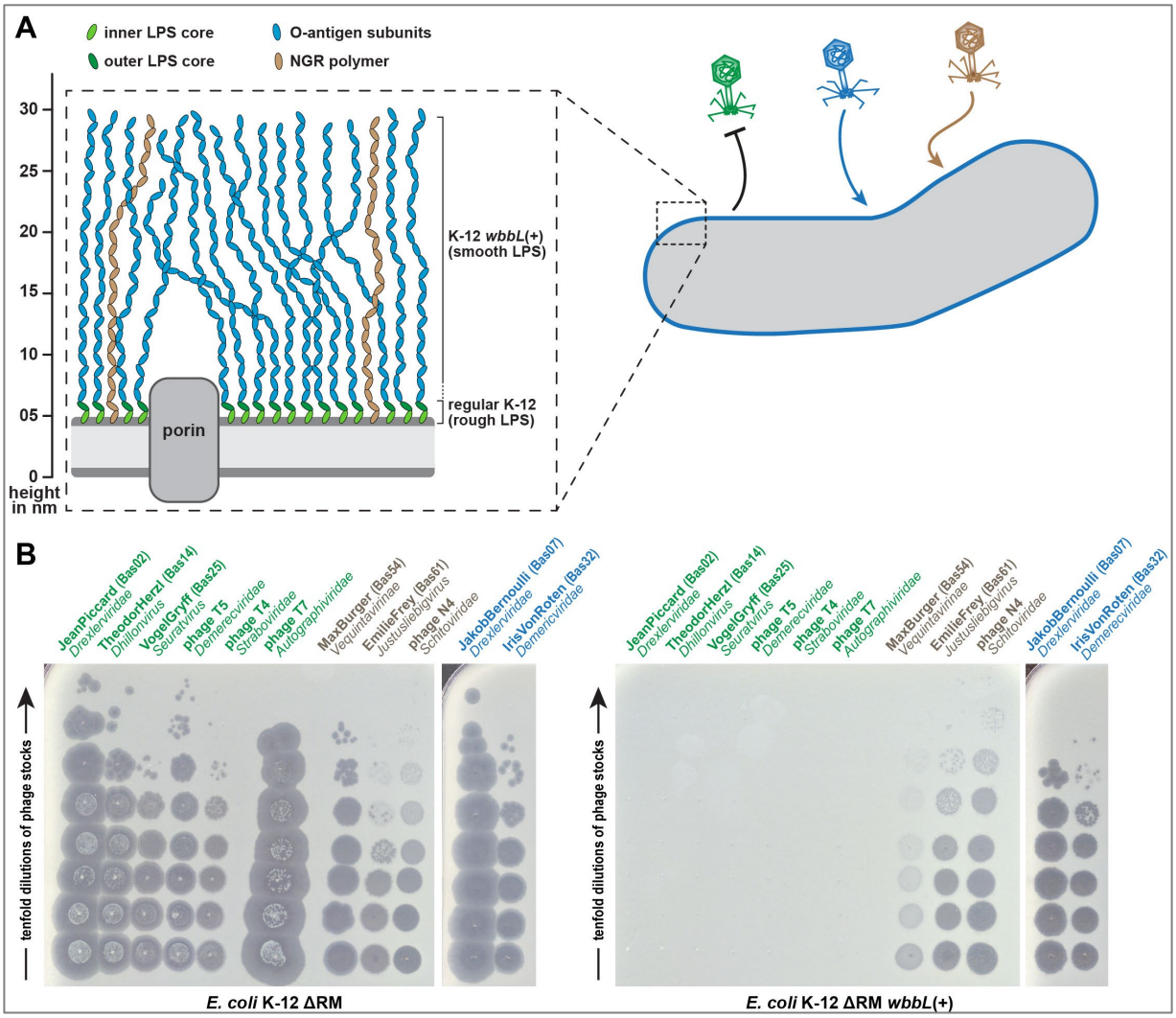


Figure 1. O-antigen glycans as barrier or host receptor for phage infections. (A) The illustration shows cell surface glycans of *E. coli* K-12 laboratory strains including rough LPS without O-antigen and smooth LPS (with genetically restored O16-type O-antigen; see details in *Materials and Methods*). The O-antigen can be a barrier to phage infection (green), used as a host receptor (blue), or be bypassed by phages that target the NGR glycans (brown). **(B)** Serial dilutions of different phages characterized previously [10] were spotted on the *E. coli* K-12 ΔRM strain without (left) or with (right) genetically restored O-antigen expression to illustrate the three different modes of interaction with these surface glycans (same color code as in (A)).

In addition, some phages like N4 or the *Vequintavirinae* can bypass the O-antigen layer by targeting the "N4 glycan receptor" (NGR), a recently described glycan that is broadly conserved among enterobacteria [10, 17] (Figs. 1A and 1B). The shielding of surface receptors by the O-antigen layer seems to be primarily due to its density because it is not more than 30nm deep while the tail fibers of model phages T4 (ca. 145nm [37]), T5 (80nm [38]), or T7 (30nm [39]) could otherwise easily reach through it to the cell surface. However, these phages are completely blocked by restored O16-type O-antigen expression [10] (Fig. 1B).

While the absence of an O-antigen barrier in the K-12 strain used for sampling the BASEL phages made it much easier to isolate many diverse phages, it also meant that phages which depend on this glycan as part of host receptor recognition could not be sampled. Using different sewage samples, we had previously shown that these are quantitatively rare though they would of course be qualitatively interesting to expand the assorted diversity of the BASEL collection [10]. We have therefore now systematically isolated and characterized diverse phages that infect *E. coli* K-12 with O16-type O-antigen but not the isogenic strain without this glycan (Figs. 2A and 2B). Fortuitously, we observed that phages depending on O16-type O-antigen as host receptor are much more common in environmental samples than in sewage (Table S1). Most, but not all of these viruses belonged to phage groups that had already been known for their dependence on surface glycans including host O-antigen such as podoviruses of *Gamaleyavirus* genus or siphoviruses of the *Kagunavirus* genus [33] (see Fig. 2B).

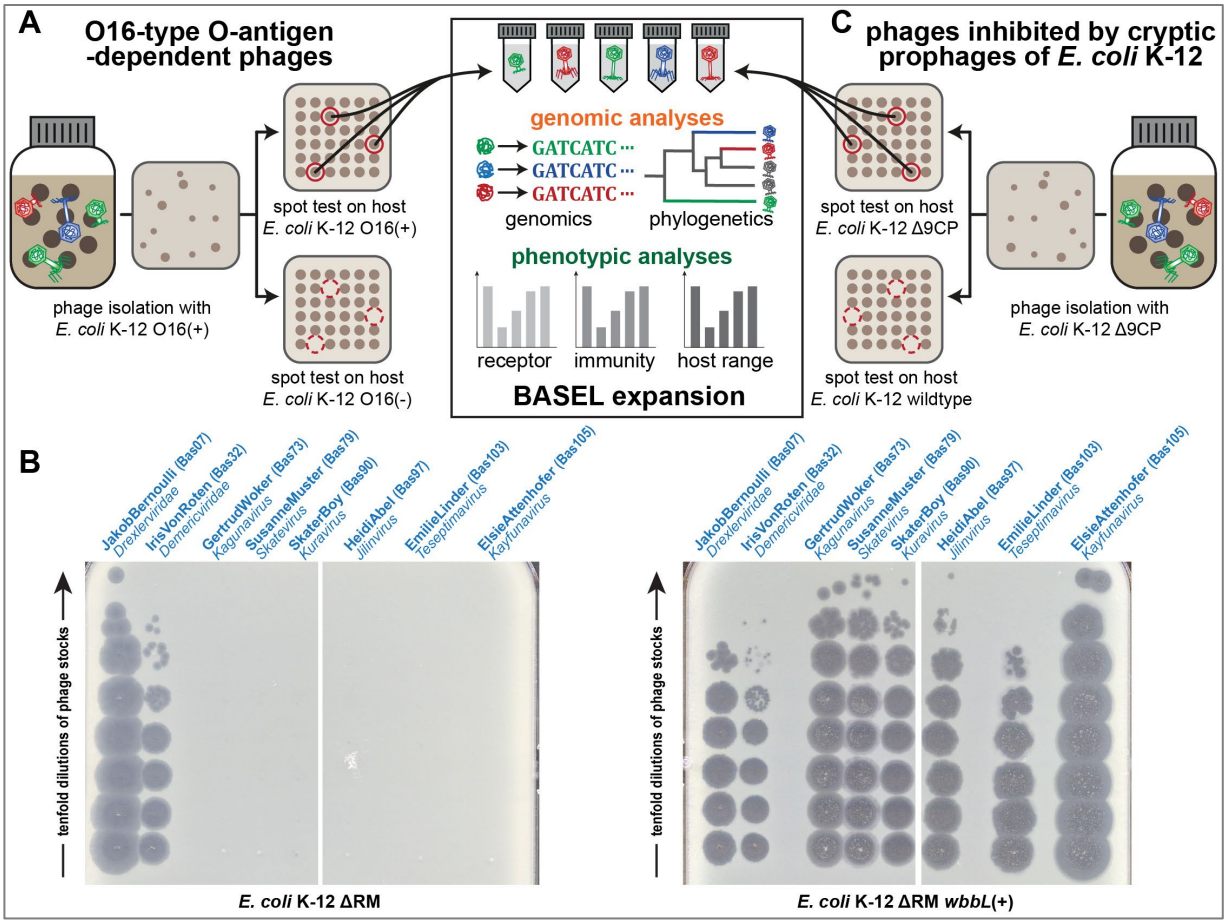


Figure 2. Expanding the BASEL collection with phages dependent on O-antigen or inhibited by K-12 cryptic prophages. (A) Phages dependent on O16-type O-antigen were isolated and characterized as shown in the illustration (analogous to our previous work [10]). **(B)** O-antigen dependence of diverse newly isolated phages characterized in this work compared to JakobBernoulli and IrisVonRoten from our previous study that can infect *E. coli* K-12 without (left) and with (right) O-antigen [10]. **(C)** Phages inhibited by the cryptic prophages in the K-12 genome were isolated analogously to those dependent on O-antigen (see (A)) with help of the *E. coli* K-12 Δ9CP strain (details outlined in *Materials and Methods*).

Overview: Including phages inhibited by resident immunity of *E. coli* K-12 in the BASEL collection

Bacteriophages encounter multiple layers of antiviral defenses outside and inside their host cells that must be overcome for successful viral replication [40]. Some of these defenses prevent phage adsorption or DNA ejection while others directly attack the invading virus such as restriction-modification (RM) systems which cleave foreign DNA based on self / non-self recognition through certain DNA modifications. Beyond direct attack, a deeper layer of defense systems protect the bacterial population

through abortive infection, i.e., the altruistic suicide of infected cells prior to successful phage replication [41]. Recent work on anti-phage defense systems has vastly expanded our knowledge about their abundance and diversity through the discovery of many new systems that target various aspects of phage biology and are countered by phages in diverse ways [40-45].

The phages of the original BASEL collection had been sampled with a derivative of the commonly used *E. coli* K-12 MG1655 laboratory strain in which we had knocked out all known RM systems, i.e., the EcoKI type I RM system and three type IV restriction systems [10]. This enabled broad sampling of *E. coli* phages including several taxonomic groups of phages that displayed considerable sensitivity to EcoKI and would otherwise have been excluded [10]. However, the literature and commonly used prediction tools such as DefenseFinder and PADLOC indicate that *E. coli* K-12 encodes several additional *bona fide* defense systems that might limit the range of phages accessible with this genomic background [46, 47] (see details at *Bacterial strains* in *Materials and Methods)*. We had not considered this to be a widely relevant factor when composing the original BASEL collection because phage defenses beyond RM systems were thought to be rather specific. However, our own results showing broad antiviral defense by the Fun/Z system and the work of others, e.g., on the Zorya system revealed that this is not necessarily the case [10, 24].

Notably, all these *bona fide* defense systems are encoded on the cryptic prophages of *E. coli* K-12 that are genomic islands descended from once active prophages, i.e., the facultatively host-integrated form of so-called temperate phages [46, 47]. This is in line with the common observation that anti-phage defenses are frequently encoded on mobile elements and genomic islands [48-50]. To overcome possible sampling biases caused by these *bona fide* phage-defense systems (and possibly others that are yet unknown), we thus performed phage isolation experiments with a variant of *E. coli* K-12 in which a previous study had removed all cryptic prophages, *E. coli* K-12 BW25113 Δ9CP (both with and without restored O-antigen) [51] (Fig. 2C). Similar to previous work in *Vibrio* that had used an analogous approach [50], these experiments indeed resulted in the isolation of an additional diversity of phages both from groups already

covered in the BASEL collection and from a genus that had never been isolated before using any *E. coli* laboratory strain (*Wifcevirus* myoviruses).

Composition of the expanded BASEL phage collection

The original BASEL phage collection had been composed of 68 new handpicked phage isolates that we characterized alongside a panel of ten well-studied model phages such as the classic T-phages [10]. These phages represent many major groups of *E. coli* phages which have been previously described in phage isolation studies with *E. coli* K-12 [52, 53]. The (mostly genus-level) groups comprising these phages showed characteristic and specific patterns of features regarding receptor usage, host range / recognition, and sensitivity or resistance to bacterial immunity in our data and those from other studies [10, 20-25]. These observations make it possible to learn about the properties of newly isolated phages merely from their genome sequences or taxonomic position, e.g., to inform a targeted selection or engineering of effective phages for individual cases of phage therapy [54-56].

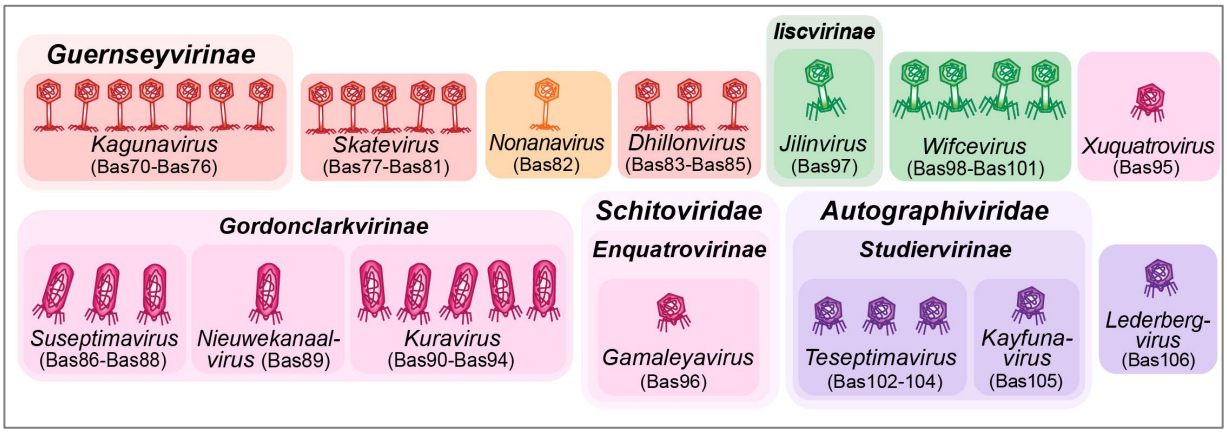


Figure 3. Overview of new phages included in the BASEL collection. The illustration shows new phage isolates included in the BASEL collection with their Bas## identifiers. Isolates are sorted by morphotype and their current ICTV classification [68]. The color code distinguishes phages based on morphotype and genome size. For siphoviruses, small phages (ca. 30-50 kb) are shown in red and medium-sized phages (ca. 50-100 kb) are shown in orange. Small myoviruses (ca. 30-100 kb) are shown in green. For podoviruses, small phages (ca. 30-50 kb) are shown in violet and medium-sized phages (ca. 50-100kb) are shown in pink.

However, a close look at published work describing phages isolated with *E. coli* hosts beyond the K-12 laboratory strain reveals important gaps in the assorted diversity of the original BASEL collection. These include, e.g., several groups of podoviruses such as the *Gordonclarkvirinae* with characteristic

prolate virions [57], diverse small siphoviruses such as the genus *Kagunavirus*, and a few myoviruses such as the genus *Wifcevirus* [35, 58, 59]. Using *E. coli* K-12 with restored O-antigen expression and / or without cryptic prophages we have now isolated and characterized more than a hundred new phages that are unable to grow on the regular *E. coli* K-12 host used for the original BASEL collection. These new isolates cover all major groups of phages (and some rarely isolated ones) that have been reported to infect *E. coli*, and we selected 37 isolates for an expansion of the BASEL collection based on their genomic and phenotypic diversity (Fig. 3 and Table S2). We systematically characterized these phages analogous to the phages in our previous study (Figs. 4A-D) and used several original BASEL phages that could overcome the O16 O-antigen barrier as controls for the phenotypic experiments. For these, no relevant differences to their phenotypes in the original paper were observed when the host strain expresses O16 O-antigen (Fig. S1B). These results suggest that the O-antigen barrier and other defenses are largely orthogonal and that our results from this study are comparable to the original publication on the BASEL collection.

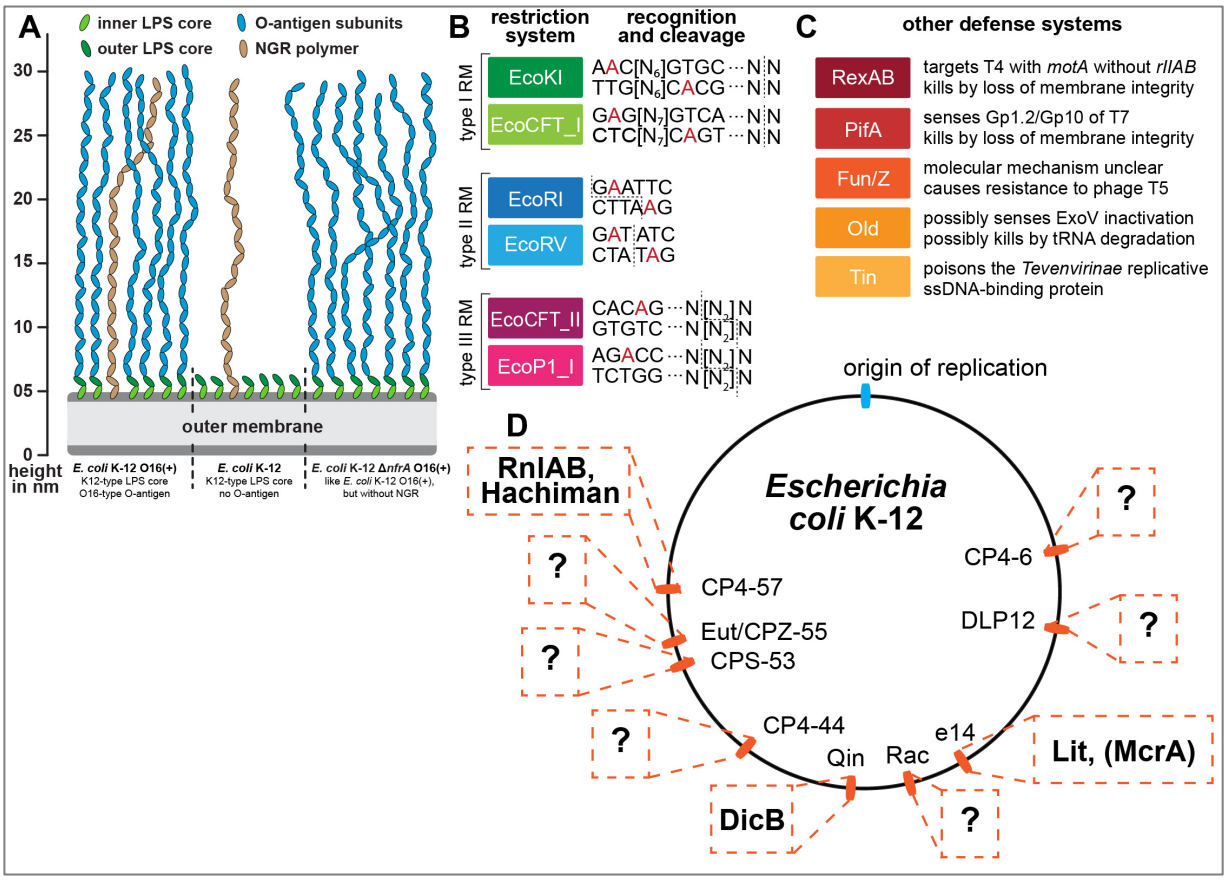


Figure 4. Overview of *E. coli* surface glycan variants and the immunity systems used in this study. **(A)** Surface glycans of different *E. coli* K-12 variants are shown schematically (compare Fig. 1A and see further details in *Materials and Methods*). None of the newly isolated phages in this work showed any dependence on the NGR glycan (see Fig. S1A) which has thus been omitted from the other illustrations. **(B, C)** Restriction-modification (B) and other defense systems (C) used to characterize phage isolates are summarized schematically (same as used in our previous work [10]) **(D)** The genome of *E. coli* K-12 contains nine cryptic prophages that encode all remaining antiviral defense systems predicted by DefenseFinder or PADLOC after deletion of all restriction systems in our previous work (including McrA; see details on these defense systems in *Materials and Methods* under *Bacterial strains*) [10, 46, 47]. Due to the known accumulation of antiviral defenses in prophages [48, 49], it seems likely that also many of the possibly still unknown defense systems of *E. coli* K-12 would be encoded in these loci (indicated by question marks). The *E. coli* K-12 Δ9CP variant of Wang et al. [51] which we use in the current study has been cured of all nine cryptic prophages.

Properties of siphoviral genus *Kagunavirus*

Small siphoviruses like the *Drexlerviridae* family (relatives of phage T1) or genus *Dhillonvirus* are frequently isolated using *E. coli* as host and typically have genomes of 40-50kb size. The diverse representatives of these groups in the original BASEL collection mostly showed considerable sensitivity to RM systems and – as expected from previous work – targeted different porins as terminal host receptor [10, 60, 61]. Only one of these phages, JakobBernoulli (Bas07), is able to infect *E. coli* K-12 with restored O-antigen expression, suggesting that its lateral tail fibers can target O16-type O-glycans [10] (see also Fig. S1A and below). Our targeted search for phages that depend on O16-type O-antigen resulted in the isolation of various small siphoviruses of genus *Kagunavirus* within the subfamily of *Guernseyvirinae* (Fig. 5A-D). These phages have small genomes of ca. 41-44kb which makes them attractive models for synthetic biology applications, e.g., in case of phage CM001 [62]. While all our isolates fully depend on restored O-antigen expression, only some of them (Bas70-73) infect a regular *E. coli* K-12 host while others (Bas74-76) could only be isolated using a K-12 host devoid of cryptic prophages (Fig. 5D). Phenotypically, both groups of *Kagunavirus* isolates behave very similarly and display the high sensitivity to RM systems and a variable sensitivity to the Old anti-phage defense system that is typical for small siphoviruses [10] (Fig. 5D). The dependence on glycans as host receptor seems to be a general property of *Guernseyvirinae* phages including, e.g., the prototypic *Kagunavirus* K1G targeting the K1 capsule of *E. coli* [15] or the *Jerseyvirus* SETP3 and several relatives targeting O-antigen of *Salmonella enterica* [63].

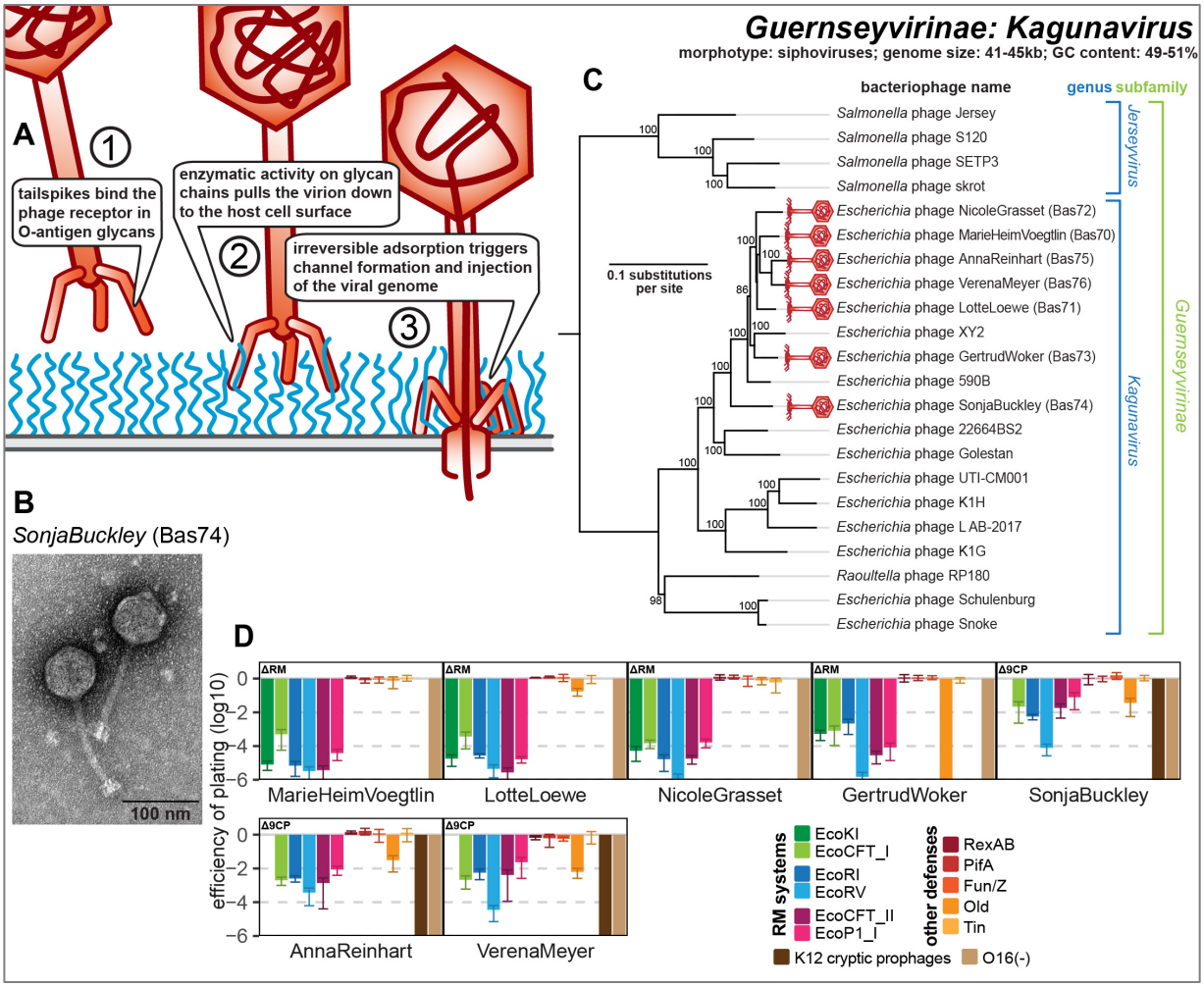


Figure 5. Overview of *Guernseyvirinae* genus *Kagunavirus* phages. (A) Schematic model of host recognition by *Kagunavirus* phages (O-antigen chains in blue). In analogy to other siphoviruses infecting *E. coli* like lambda or T5, we anticipate that these phages also use a phage-derived conduit for DNA injection across the outer membrane after the tail fibers recognizing the terminal host receptor have bent off [110, 141]. **(B)** Representative TEM micrograph of phage SonjaBuckley (Bas74). **(C)** Maximum-Likelihood phylogeny of *Kagunavirus* phages and relatives based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. Newly isolated phages of the BASEL collection are highlighted by red phage icons. The phylogeny was rooted between the genera *Kagunavirus* and *Jerseyvirus*. **(D)** The results of quantitative phenotyping experiments with *Kagunavirus* phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as efficiency of plating (EOP). Small notes of "ΔRM" or "Δ9CP" indicate on which host strain the respective phage has been characterized (see *Materials and Methods*). Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data.

Properties of other O16-dependent small siphoviruses and *Dhillonvirus*

Besides phages of the genus *Kagunavirus*, we also obtained different other small siphoviruses that depend on restored O16-type O-antigen to infect *E. coli* K-12 and new isolates of the *Dhillonvirus* genus (Fig. 6A-G). Some of the new O16-dependent isolates belong to a cluster of siphoviruses formed primarily by the genus *Skatevirus* (Bas77-81, Figs 6C and 6D). These viruses have genomes of 46-48kb which is well within the typical range for small siphoviruses. Additionally, we present phage KarinMuster (Bas82), belonging to the rarely reported *Nonanavirus* genus comprising relatives of the more commonly known *Salmonella* phage 9NA (Fig. 6E) [64]. The genome of this phage is ca. 58.5kb and thus around 10kb larger than that of the other small siphoviruses presented here, reminiscent of the *Queuovirinae* characterized in our previous work [10]. This latter group encodes a 7-deazaguanosine modification pathway in their additional genome content which provides considerable resistance to RM systems [10, 65]. However, there is no recognizable additional locus in KarinMuster and its relatives that could encode such supplemental biological functions [64], though their distant relationship to other known phages and limited knowledge about their biology make it difficult to draw firm conclusions. Very similar to small siphoviruses characterized previously and the *Kagunavirus* phages shown above, all these phages are markedly sensitive to RM systems and to the Old defense system which in the latter case results in quantifiable decrease of plaque counts only for some phages (Fig. 6G). The additional genome size of KarinMuster therefore does not seem to make this phage more robust than other small siphoviruses when facing the tested examples of bacterial immunity.

Finally, the phage isolation using a host without prophages identified additional members of the *Dhillonvirus* genus (Bas83-85) which – unlike their relatives of the original BASEL collection (Bas14-18) – fail to form plaques on regular *E. coli* K-12 (Figs. 6B, 6F, and 6G). These new isolates do not cluster separately within the *Dhillonvirus* genus but are rather split widely across this group, suggesting that no vertically inherited feature but rather highly variable aspects of phage biology are responsible for this difference (Fig. 6F). Phenotypically, the three additional phages selected for this work behave mostly

similar to the original *Dhillonvirus* phages, e.g., by showing an only moderate sensitivity to tested RM systems [10] (Fig. 6G). Like the other siphoviruses characterized for the original BASEL collection, these *Dhillonvirus* phages display the typical arrangement of two separate types of tail fibers with distinct receptor-binding proteins [10, 34] (see below and Fig. S2A). Their central tail fiber has a *bona fide* target recognition module that predicted to encode LptD specificity (Fig. S2B) [10].

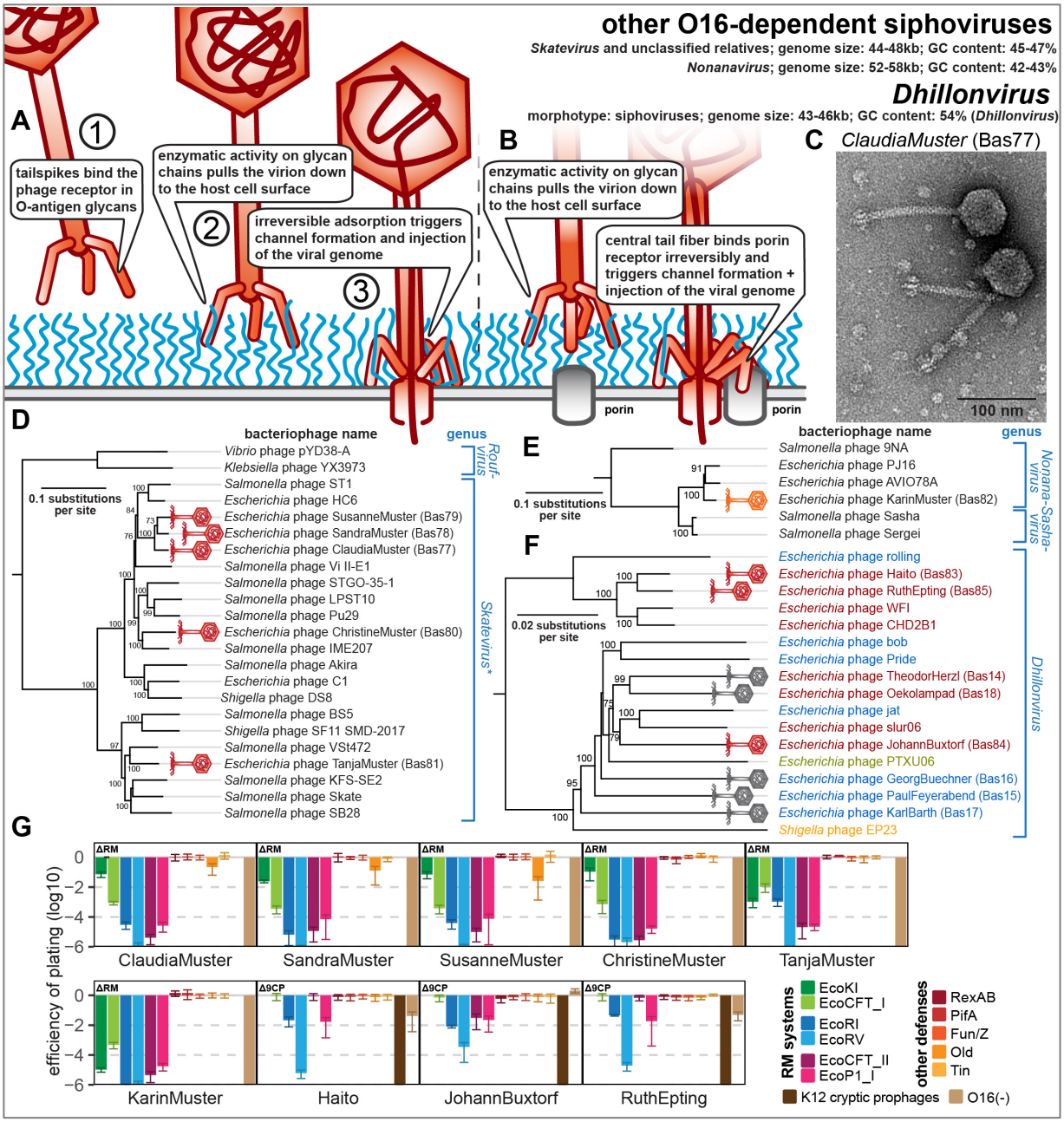


Figure 6. Overview of other O16-dependent siphoviruses of genera *Skatevirus* and *Nonanavirus* as well as new isolates of siphoviral genus *Dhillonvirus*. (A) Schematic model of host recognition by O16-dependent small siphoviruses just as shown for *Kagunavirus* phages in Fig. 5A. **(B)** Schematic model of host recognition by small siphoviruses of the *Dhillonvirus* genus that use first a lateral tail fiber to overcome the O-antigen barrier before binding a porin receptor with the central tail fiber (instead of only glycan-targeting tailspikes for the LPS-dependent small siphoviruses; see also our previous work [10]). **(C)** Representative TEM micrograph of *Skatevirus* ClaudiaMuster (Bas77). **(D)** Maximum-Likelihood phylogeny of the *Skatevirus* genus and relatives based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. The phylogeny was rooted between genus *Roufvirus* and the others. Note that most reference phages presented here as *Skatevirus* have no formal ICTV classification while others are scattered over several seemingly paraphyletic genera, prompting us to show all phages as *Skatevirus* [68]. **(E)** Maximum-Likelihood phylogeny of the *Nonanavirus* genus and relatives based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. The phylogeny was rooted between phage 9NA and the other relatives based on their much closer relationship. Note that *Sasha* and *Sergei* are formally classified as *Sashavirus* while 9NA and AVIO78A are classified as *Nonanavirus* [68] which – according to this phylogeny – would be paraphyletic. **(F)** Maximum-Likelihood phylogeny of the *Dhillonvirus* genus based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. The phylogeny was rooted as previously [10] between the group of isolates including phage WFI and the others. In (D-F), newly isolated phages of the BASEL collection are highlighted by red or orange phage icons. The color code of phage names in (F) refers to porin host receptors with blue indicating FhuA, red LptD, green BtuB, and orange LamB [10] (see also Fig. S2B). **(G)** The results of quantitative phenotyping experiments with *Skatevirus*, *Nonanavirus*, and new *Dhillonvirus* isolates regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as efficiency of plating (EOP). Small notes of "ΔRM" or "Δ9CP" indicate on which host strain the respective phage has been characterized (see *Materials and Methods*). Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data.

Properties of podoviruses of *Gordonclarkvirinae* genus *Kuravirus* and relatives

Besides small siphoviruses, our phage isolation with *E. coli* K-12 expressing its O16-type O-antigen resulted in multiple phages belonging to podoviruses in the *Gordonclarkvirinae* subfamily (Fig. 7A-D). These phages form characteristic prolate virions of the rare C3 morphology [66] (Fig. 7B) and have been frequently isolated using diverse strains of *E. coli*. However, so far no representatives infecting the *E. coli* K-12 or B laboratory strains have been obtained and the molecular biology of these phages has been only poorly studied. The genome sizes of this group differ widely from 67-69kb for one clade of the *Kuravirus* genus (at the bottom of Fig. 7C) to 76-77kb for other *Kuravirus* phages and genera *Suseptimavirus* as well as *Nieuwekanaalvirus* (see details for all in Table S2). Our phylogeny largely reproduces the previously identified subgroups of *Gordonclarkvirinae* which have in the meantime been recognized as different genera in the ICTV taxonomy [67, 68] (Fig. 7C). The isolates infecting *E. coli* K-

12 are widely scattered over the different genera of *Gordonclarkvirinae*, suggesting that no vertically inherited factor plays major roles in the ability of these phages to infect that host (Fig. 7C).

All our isolates completely depend on the restored expression of O16-type O-antigen for infection, suggesting that the O-antigen and possibly other surface glycans are the conserved host receptor of these phages (Fig. 7D). This is in line with previous work reporting that spontaneous *E. coli* mutants resistant to *Kuravirus* phage ES17 had always lost O-antigen expression [69]. In one of the few dedicated studies on *Gordonclarkvirinae*, the Plevka group reported the virion structure of *Kuravirus* SU10 which – in the absence of decades-long research available for standard model phages – was critical for the functional understanding and annotation of structural genes in *Kuravirus* genomes [57]. Unlike O-antigen-dependent siphoviruses (see above) and other well-studied podoviruses such as T7 [70, 71], the *Gordonclarkvirinae* have two separate *bona fide* RBPs, identifiable as long tail fibers and short tail fibers, that likely recognize different host receptors. These two sets of tail fibers appear to be used successively analogous to classical T-even myoviruses where the long tail fibers mediate initial host recognition by targeting porins and O-antigen while the short tail fibers subsequently bind conserved motifs in the LPS core as the terminal receptor [10, 37, 57, 72-74] (see also below). The functionality of long and short tail fibers of T-even phages is linked in a way that host recognition by long tail fibers is required for the deployment of short tail fibers and, consequently, viral genome injection [10, 73, 74]. Similarly, direct observation of *Kuravirus* SU10 infection using cryogenic electron tomography shows how the lateral tail fibers initially target an exposed receptor – likely O-antigen glycans – before the short tail fibers target the cell surface and contribute to the formation of a nozzle which extends the virion tail for genome delivery [57] (see Fig. 7A).

The *Gordonclarkvirinae* are remarkably resistant to RM systems despite the lack of known pathways of covalent genome modification or restriction inhibitors (Fig. 7D and Table S2). However, they encode a *bona fide* dCTP deaminase and a nucleotidyltransferase with their DNA replication genes (Fig. S3), suggesting that these phages might feature a covalent DNA modification – most simply deoxyuracil instead of thymidine like *Bacillus* phage PBS1 – which may protect their genome from

restriction [75]. The only detectable sensitivity to tested anti-phage defense systems was observed for Fun/Z which targets phages of the *Kuravirus* genus (Fig. 7D).

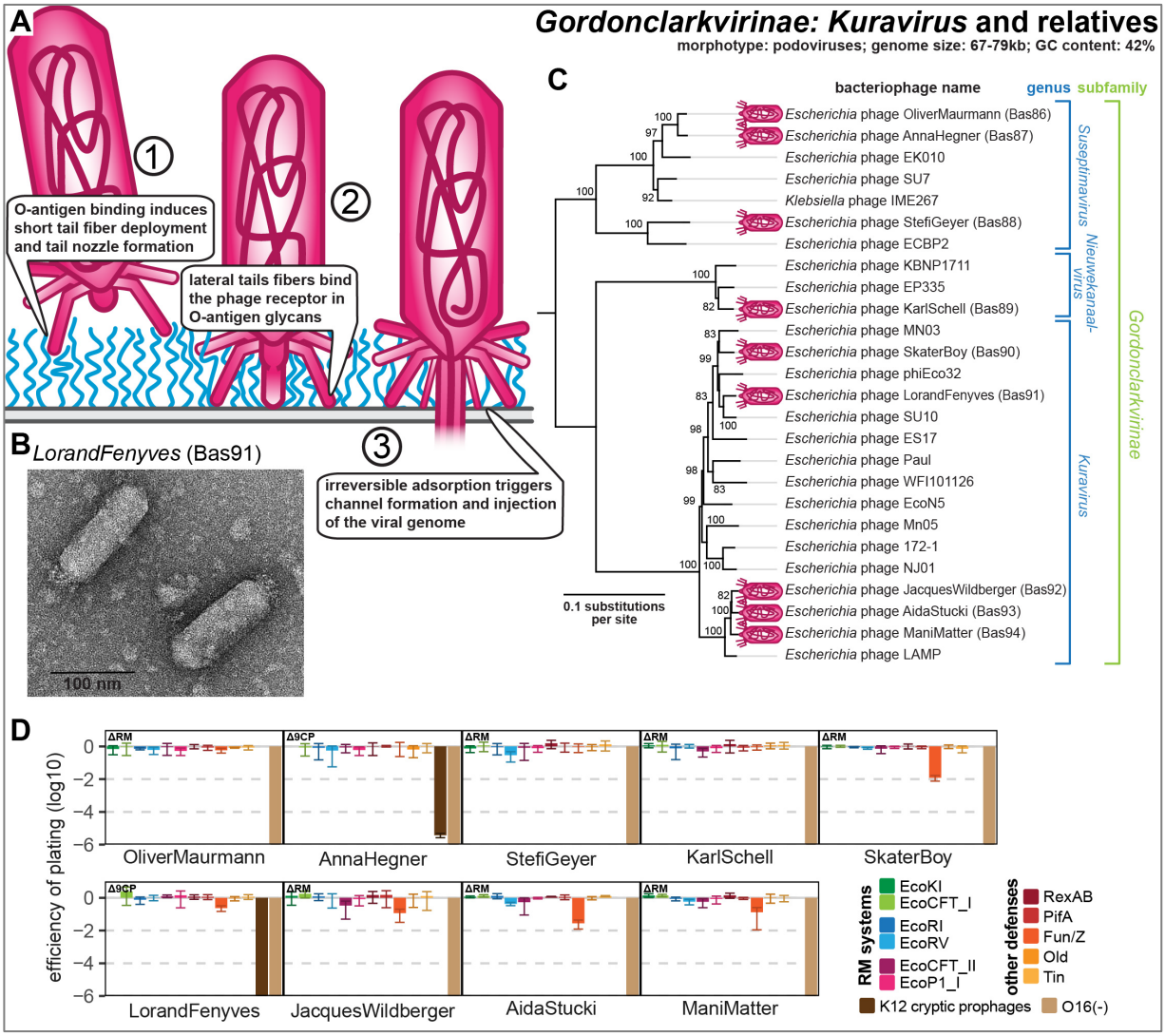


Figure 7. Overview of *Gordonclarkvirinae* phages. (A) Schematic model of host recognition by *Gordonclarkvirinae* phages as an interpretation of different molecular studies [57, 117]. **(B)** Representative TEM micrograph of phage LorandFenyves (Bas91). **(C)** Maximum-Likelihood phylogeny of the *Gordonclarkvirinae* subfamily based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. Phages of the BASEL collection are highlighted by little violet phage icons. The phylogeny was rooted between the genus *Suseptimavirus* and the genera *Nieuwekanaalvirus* + *Kuravirus* based on closer relationship of the latter two. **(D)** The results of quantitative phenotyping experiments with *Gordonclarkvirinae* phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as efficiency of plating (EOP). Small notes of "ΔRM" or "Δ9CP" indicate on which host strain the respective phage has been characterized (see *Materials and Methods*). Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data.

Properties of other groups of podoviruses and myoviruses

Xuquatrovirus relative HelliStehle (Bas95)

Besides small siphoviruses and *Gordonclarkvirinae* podoviruses, all other phage groups isolated for this work were individually much rarer and only one or a few isolates of each could be obtained. This includes two podoviruses and one myovirus that are each the only isolates of their respective groups (Figs. 8A-G). One of these is a podovirus related to the *Xuquatrovirus* genus for which there are only two closely related representatives, the original PTXU04 of the study by Korf et al. [58] and phage Pondi of Sidi Mabrouk et al. [76] (Fig. 8D). These phages have an intermediate genome size for a podovirus of around 60kb which is ca. 20kb larger than *Autographiviridae* (T7 relatives; see below) but ca. 15-20kb smaller than *Gordonclarkvirinae* (see above). Our new isolate HelliStehle (Bas95) is clearly related to the *Xuqutatovirus* phages with a largely colinear genome of similar size (Fig. S4) but is phylogenetically distinct (Fig. 8D). The biology of these phages has largely remained enigmatic besides a genomic and proteomic analysis of PTXU04 when it had originally been described, which was crucial for a functional annotation of at least the structural genes due to low sequence similarity to well-characterized phages [58]. Consequently, a recent transcriptome-wide CRISPR interference study determined the essential genes of PTXU04 but could in the absence of additional phenotypic data not use this information for deeper insights as has been possible, e.g., with model phages T4 and T5 [77]. Our experiments show that phage HelliStehle is completely dependent on restored O-antigen expression, indicating that it is using this glycan as an essential receptor (Fig. 8G). Furthermore, this phage is very sensitive to RM systems, suggesting that – unlike other larger podoviruses such as N4 or *Gordonclarkvirinae* – it lacks potent mechanisms to protect its genomic DNA (Fig. 8G).

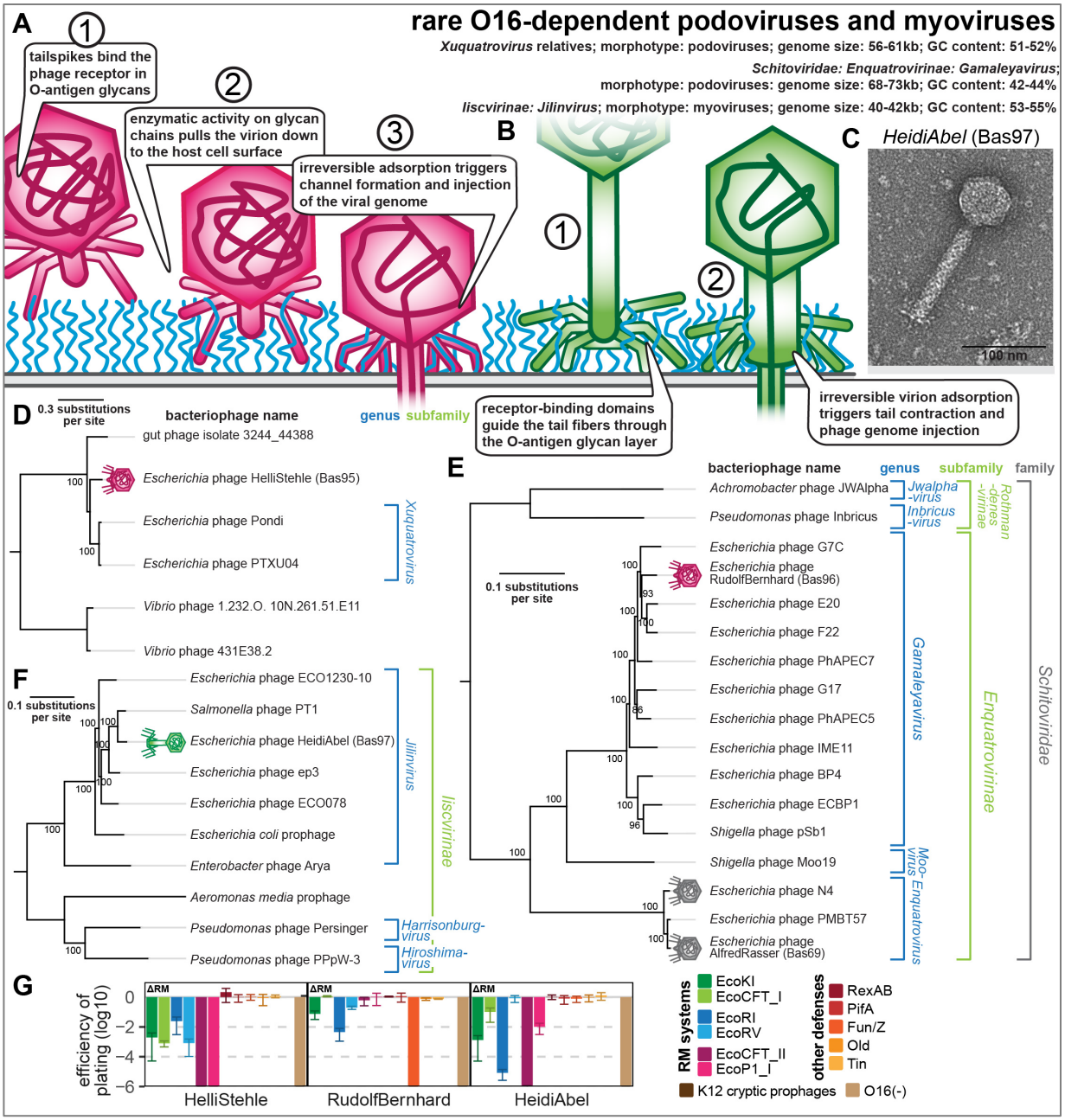


Figure 8. Overview of rare O-antigen-dependent phage groups (*Xuquatrovirus* relatives, *Gamaleyavirus*, *Jilinvirus*). (A) Schematic model of host recognition by podoviruses with enzymatic tailspikes targeting O-antigen (e.g., *Gamaleyavirus*) based on previous work on phage T7 and P22 [71, 101, 203]. **(B)** Schematic model of host recognition by myoviruses targeting O-antigen with a tail fiber (e.g., *Jilinvirus*). Note that *Jilinvirus* HeidiAbel encodes two sets of tail fibers (see below) whose relationship is unknown, so only a simple schematic with one set is shown. **(C)** Representative TEM micrograph of *Jilinvirus* phage HeidiAbel (Bas97). This phage forms myovirus virions with similar appearance to other Jilinvirus phages and relatives that have been studied previously [79, 80, 82]. **(D)** Maximum-Likelihood phylogeny of the genus *Xuquatrovirus* and relatives based on a curated concatenated alignment of major capsid protein, portal protein, and terminase large subunit protein sequences with bootstrap support of branches shown if > 70/100. The phylogeny was rooted between

two marine *Vibrio* phages and the others based on a large phylogenetic distance between these groups. **(E)** Maximum-Likelihood phylogeny of the *Enquatrovirinae* subfamily of *Schitoviridae* and relatives based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. The phylogeny was rooted between subfamilies *Rothmandenesvirinae* and subfamily *Enquatrovirinae*. **(F)** Maximum-Likelihood phylogeny of the *Iiscvirinae* subfamily based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. The phylogeny was rooted between the genus *Jilinvirus* and the other genera. The exemplary prophages of *E. coli* and *Aeromonas media* were identified in strains USECESBL382 (NCBI GenBank AATBGZ010000031) and T0.1-19 (NCBI GenBank CP038441), respectively, using PHASTEST [204]. In (D-F), phages of the BASEL collection are highlighted by pink and green (new) or grey (old) phage icons. **(G)** The results of quantitative phenotyping experiments with the three newly described phage isolates regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as efficiency of plating (EOP). Small notes of "ΔRM" or "Δ9CP" indicate on which host strain the respective phage has been characterized (see *Materials and Methods*). Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data.

***Gamaleyavirus* RudolfBernhard (Bas96)**

Podoviruses of the *Gamaleyavirus* genus within the *Enquatrovirinae* subfamily are close relatives of the genus *Enquatrovirus* including N4 and AlfredRasser which we have characterized previously [10]. They typically target O-antigen glycans as essential host receptor [33, 78], and our isolate RudolfBernhard (Bas96) is also completely dependent on restored O16 O-antigen expression to infect *E. coli* K-12 (Fig. 8G). This isolate is closely related to previously studied phage G7C (targeting O22-type O-antigen; Fig. 8E) but encodes different tailspikes that likely mediate O-antigen recognition (see in the dedicated section below) [78]. Consequently, RudolfBernhard does not depend on the NGR glycan unlike N4 and AlfredRasser which target this enigmatic polysaccharide as their essential host receptor (Figs. 8G and S1B) [10, 17]. Like its relatives N4 and AlfredRasser, RudolfBernhard shows considerable resistance to RM systems despite not encoding any known genome protection mechanism, but unlike those phages it does have a detectable sensitivity to some systems (Fig. 8G and Table S2). It is also sensitive to the Fun/Z defense system which targets most phages with a genome length above ca. 60kb [10]), but the molecular basis of this antiviral defense system has remained unknown.

***Jilinvirus* HeidiAbel (Bas97)**

The genus *Jilinvirus* of the *Iiscvirinae* subfamily of myoviruses comprises only a few isolates that have not been well characterized but that all infect enterobacteria (Figs. 8B and 8F). Related genera infect

other hosts such as pseudomonads (Fig. 8F). These phages have very small genomes for phages of the myovirus morphotype (41-44 kb) and their lifestyle has remained enigmatic. On one hand, previous work has never observed any temperate behavior of *Jilinvirus* phages or their relatives [79-82], and also our isolate HeidiAbel (Bas97) forms regular clear plaques with no signs of lysogen growth (Fig. S5A). On the other hand, some *Jilinvirus* phages including HeidiAbel feature seemingly intact genes encoding an integrase and a cI-like repressor – indicative of a temperate lifestyle - which are deteriorated or absent in other representatives as noted previously (Figs. S5B and S5C) [81, 82]. In addition, diverse bacterial genomes contain *bona fide* prophages that would be classified in this viral group (see examples in Fig. 8F) [79]. It thus appears that these phages may have a facultatively temperate lifestyle that is lost in different representatives, possibly as an adaptation to mostly lytic growth. We therefore also interpret HeidiAbel as a virulent phage, though it would be interesting to see future work exploring if or when the phage makes use of its seemingly intact lysogeny control module. Notably, there also appear to be temperate members of the well-studied *Autographiviridae* family of virulent podoviruses which may even be their ancestral state [83], suggesting that such a transition between temperate and virulent lifestyles would not be unprecedented.

Our data show that HeidiAbel is very sensitive to RM systems except EcoRV for which its genome does not contain a recognition site (Fig. 8G and Table S2). This sensitivity is shared with all temperate phages that we have tested in our previous work, but it is also very common among virulent phages with small genomes that encode no dedicated protection mechanisms (see, e.g., the small siphoviruses presented in this study or our previous work [10]). The infectivity of HeidiAbel depends totally on the presence of restored O16-type O-antigen on *E. coli* K-12, suggesting that its tail fibers target this glycan as essential host receptor (Fig. 8G).

Properties of myoviruses in genus *Wifcevirus*

Phages of the genus *Wifcevirus* within *Caudoviricetes* have been described comparably rarely with currently less than 20 sequenced isolates in NCBI GenBank that all infect *E. coli* or *Shigella* (Fig. 9A-E).

They have genomes of ca. 66-72kb and were described as a genus when two new isolates, WFC and WFH, had been linked to several phages that had been sequenced earlier [58]. Interestingly, distant relatives of these phages in the *Pbunavirus* genus are among the *P. aeruginosa* phages most used for personalized therapy [84, 85]. Based on clear homology, the cryogenic electron microscopy structure of virions of a *Pbunavirus* has greatly helped us to annotate the structural genes of *Wifcevirus* phages [86]. No *Wifcevirus* infecting *E. coli* laboratory models K-12 or B has ever been reported, limiting our knowledge about the biology of these phages and their interaction with the host. However, one previous study on two *Wifcevirus* isolates reported remarkably fast replication with only 3-5 minutes between adsorption and release of progeny [87]. This is much faster than even mutants of phage T7 – another phage considered to be fast – which had been evolved for rapid replication but did not get faster than seven minutes in comparable assays [88]. This property of *Wifcevirus* phages could be of great interest for biotechnology or clinical applications, but the underlying molecular mechanisms have not been explored.

Our four isolates (Bas98-101) belong to both major subgroups of *Wifcevirus* phages (Fig. 9B) and all depend totally on restored O16-type O-antigen as host receptor (Fig. 9E), suggesting that their tail fibers target this glycan as host receptor (see also the dedicated section below). This is reminiscent of their *Pbunavirus* relatives which are known to target the LPS of *P. aeruginosa* as host receptor [86, 89, 90]. All our isolates could only be obtained using *E. coli* K-12 without cryptic prophages as host and their plaque formation is severely impaired in their presence (Fig. 9D), though a quantitative reduction of plaque counts was only observed for VerenaGrendelmeier (Bas101; Fig. 9E). All isolates are not sensitive to any tested bacterial defense system including RM systems despite the lack of any known genome protection mechanism (Fig. 9E and Tabe S2). One notable exception is the EcoP1_I type III RM system to which they are very sensitive even though they have similar numbers of recognition sites for this system as for the EcoCFT_II type III RM system to which they are highly resistant (Fig. 9E and Table S2).

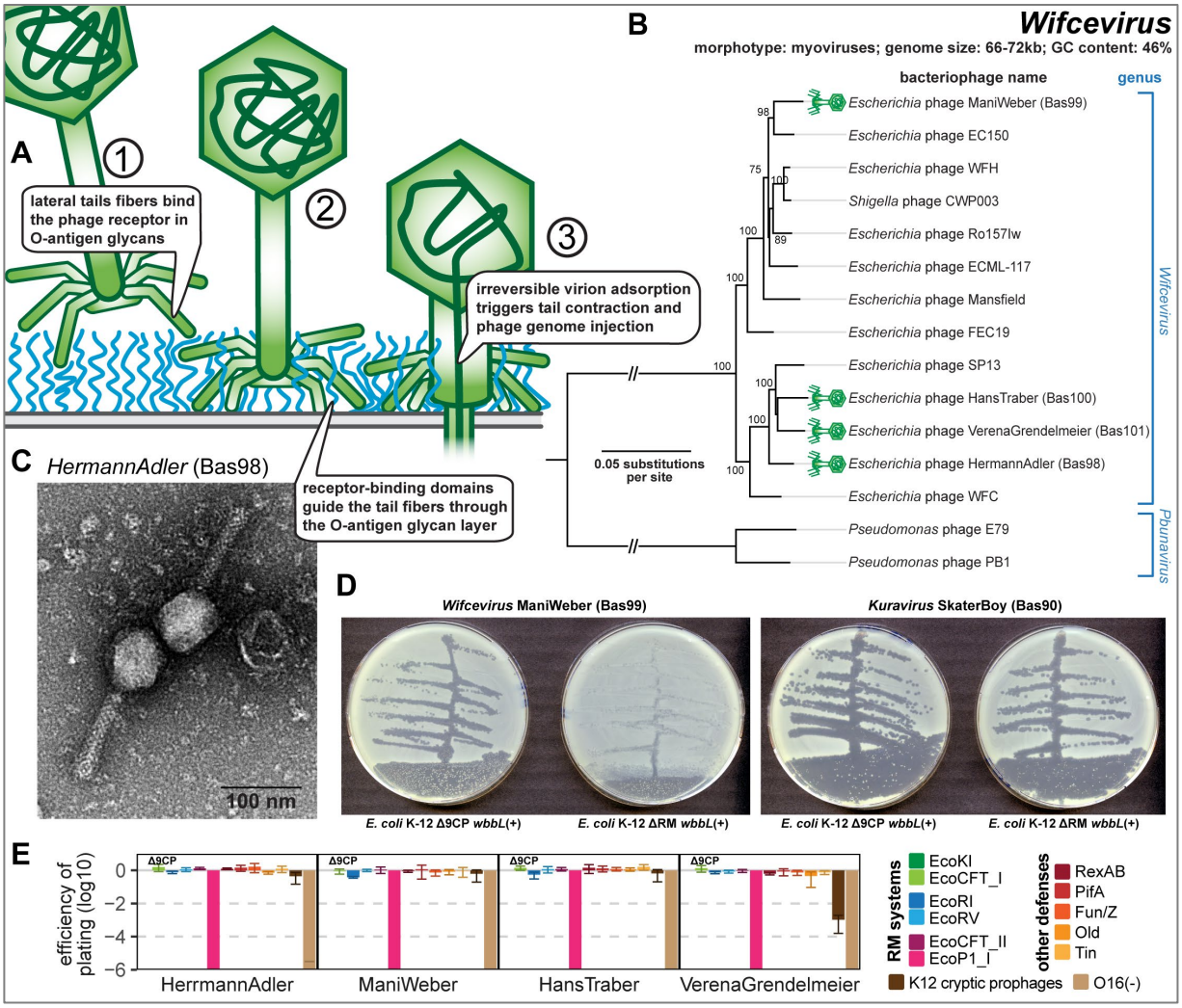


Figure 9. Overview of *Wifcevirus* phages. (A) Schematic model of host recognition by *Wifcevirus* myoviruses targeting host O-antigen (blue) using tail fibers. Note that these phages carry two sets of tail fibers (see below) whose relationship is unclear, so only a simple schematic with one set is shown. **(B)** Maximum-Likelihood phylogeny of the *Wifcevirus* genus and relatives based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. The phylogeny was midpoint-rooted between the distantly related *Wifcevirus* and *Pbunavirus* genera. Phages of the BASEL collection are highlighted by little green phage icons. **(C)** Representative TEM micrograph of phage HermannAdler (Bas98) with a virion morphology very similar to other representatives that have been visualized previously [58]. **(D)** *Wifcevirus* isolate ManiWeber (Bas99) was streaked for single plaques on *E. coli* K-12 strains with and without cryptic prophages. *Kuravirus* SkaterBoy (Bas90; see Fig. 7) is shown as a control without sensitivity to the presence of cryptic prophages. **(E)** The results of quantitative phenotyping experiments with *Wifcevirus* phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as efficiency of plating (EOP). Small notes of "ΔRM" or "Δ9CP" indicate on which host strain the respective phage has been characterized (see *Materials and Methods*). Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data.

Properties of *Autographiviridae: Studiervirinae* genera *Teseptimavirus* and *Kayfunavirus*

Podoviruses of *Autographiviridae* subfamily *Studiervirinae* are well-known from classical representatives T3 (*Teetrevirus* genus) and T7 (*Teseptimavirus* genus) and have been frequently isolated against diverse enterobacterial hosts. In our previous work we had characterized T3 and T7 together with several new isolates from different genera of this subfamily [10]. As also suggested by the literature, we found that all these phages depended on the LPS core of *E. coli* K-12 as host receptor when infecting this host though they were invariably unable to infect this host when O-antigen expression was restored [10, 91]. These phages exhibited considerable flexibility in their glycan recognition in a way that truncations of the LPS core reduced infectivity but that usually no individual truncation fully abolished it [10, 91]. Given the ubiquity of O-antigen expression among wild *E. coli* strains and its barrier function, it seems likely that these phages have evolved to target these or other surface glycans as host receptor and more accidentally recognize also the LPS core of *E. coli* K-12. Previous work has already well established that diverse representatives of *Autographiviridae* target specific types of O-antigen as host receptor which determines the host range of these phages [29, 33, 92-94].

In the current study, we indeed isolated several different *Teseptimavirus* phages that depend on restored O16-type O-antigen expression of *E. coli* K-12 and included three of these in the expanded BASEL collection (Bas102-104; Fig. 10). Furthermore, we isolated ElsieAttenhofer (Bas105), an O16-dependent phage of the distantly related *Kayfunavirus* genus that also contains various representatives known to target host glycans such as phage K1F targeting the K1 capsule [95] or CLB_P1 targeting the O104-type O-antigen of *E. coli* strain 55989 [96] (Fig. 10B). Our four new isolates encode a single tailspike or tail fiber in their genomes analogous to other *Autographiviridae* phages and likely use it to target O16-type O-antigen (see in the dedicated section below). Like their relatives, they have genomes of ca. 40kb size and are sensitive particularly to type II and type III RM systems if they have a considerable number of recognition sites (Fig. 10E and Table S2) [10].

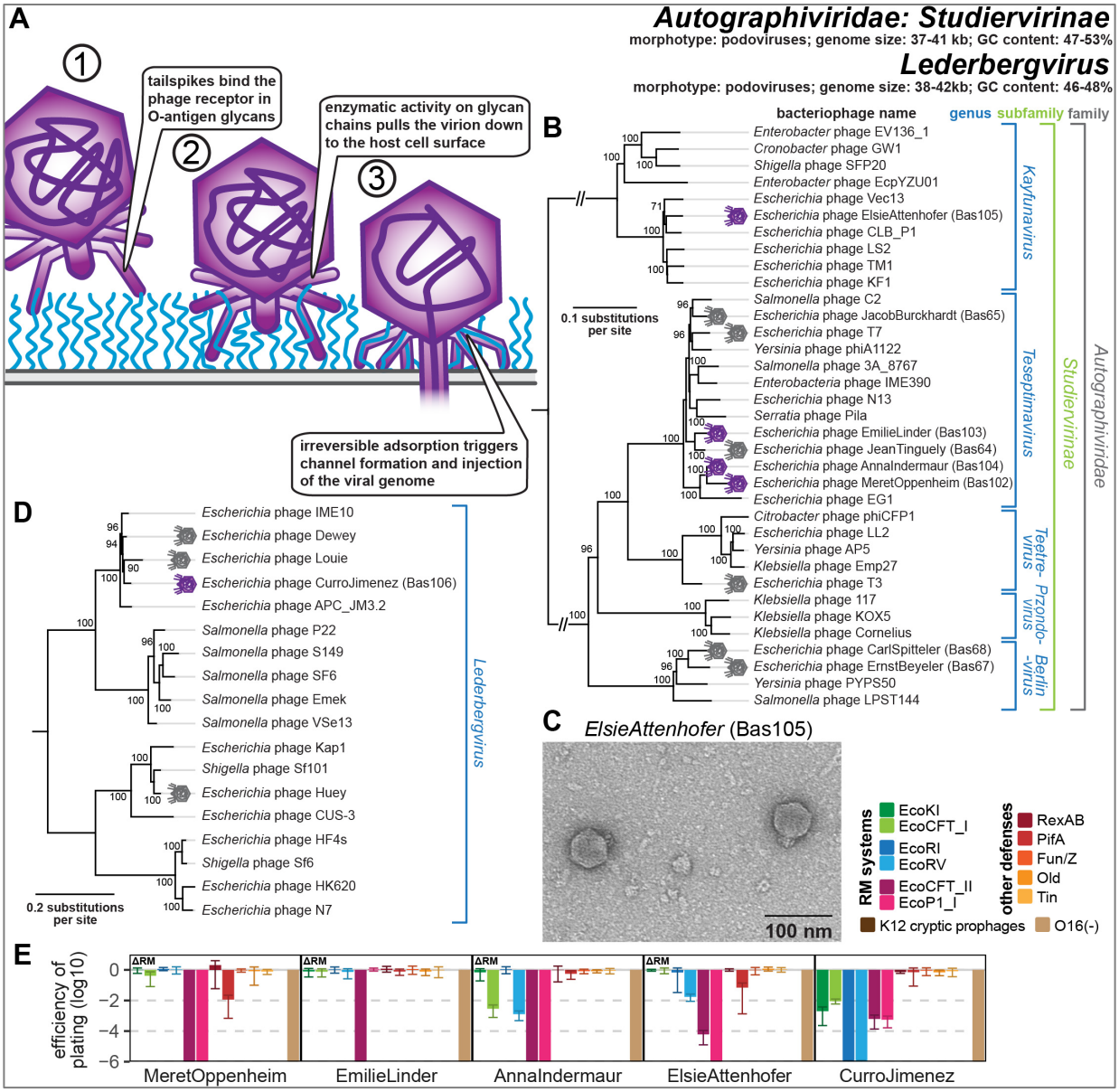


Figure 10. Overview of *Autographiviridae*: *Studiervirinae* and *Lederbergvirus* phages. (A) Schematic model of host recognition by podoviruses with enzymatic tailspikes targeting O-antigen (e.g., *Kayfunavirus* and *Lederbergvirus*) based on previous work on phage T7 and P22 [71, 101, 203]. **(B)** Maximum-Likelihood phylogeny of the *Studiervirinae* subfamily of *Autographiviridae* based on a curated whole-genome alignment with bootstrap support of branches shown if > 70/100. The phylogeny was midpoint-rooted between distantly related genus *Kayfunavirus* and the other genera. Phages of the BASEL collection are highlighted by violet (new) or grey (old) phage icons. **(C)** Representative TEM micrograph of *Kayfunavirus* ElsieAttenhofer (Bas105). **(D)** Maximum-Likelihood phylogeny of the *Lederbergvirus* genus based on an alignment of their structural genes with bootstrap support of branches shown if >70/100. This phylogeny was midpoint-rooted between two major clades of this genus. Phage CurroJimenez (Bas106) is highlighted by a violet phage icon while phages Huey, Dewey, and Louie are highlighted by grey phage icons. **(E)** The results of quantitative phenotyping experiments with *Autographiviridae* phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as efficiency of plating (EOP). Small notes of "ΔRM" or "Δ9CP" indicate on which host

strain the respective phage has been characterized (see *Materials and Methods*). Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data.

Beyond restriction, studies on model phages T3 and T7 as well as our previous work suggested that *Autographiviridae* phages are specifically targeted by the PifA defense protein dependent on the allelic variant of the *gp1.2* dGTPase inhibitor gene [10, 97, 98]. Of the three new *Teseptimavirus* phages, only MeretOppenheim (Bas102) encodes a *gp1.2* allele that is similar those of sensitive phages T7 and JacobBurckhardt (Bas65), and it is indeed significantly inhibited by PifA (Figs. 10E and S6A). However, also the *Kayfunavirus* ElsieAttenhofer (Bas105) is sensitive to PifA (Fig. 10E), but it carries a distinct variant of the *gp1.2* gene (Fig. S6A). These results suggest that *Autographiviridae* sensitivity to PifA is more complex than two allelic variants of one locus and / or that additional factors such as the major capsid protein may play a role (as indicated by previous work [98]).

Properties of podoviruses in genus *Lederbergvirus*

Phages in the genus *Lederbergvirus* are among the most well-known phages that use tailspikes to target enterobacterial surface glycans and particularly the O-antigen as host receptor (Fig. 10A) [33, 99, 100]. These phages are podoviruses with genomes of around 40kb size and a temperate lifestyle typically governed by lambda-like lysogeny control elements comprising a cI-like repressor, cII, and Cro [99]. The prototypic *Lederbergvirus* is phage P22 that has become a key model for viral host interaction. This phage targets the O:4 O-antigen of its *Salmonella* Typhimurium host with its tailspike before a tail needle penetrates the outer membrane and eventually a trans-envelope channel for DNA injection is formed (Fig. 10A) [101, 102]. Similarly, well-known *Lederbergvirus* phages like *E. coli* phage HK620 and *Shigella* phage Sf6 bind specific O-antigen variants of their respective hosts while others target different glycans like the K1 capsule in case of *E. coli* phage CUS-3 [103, 104]. Its potency in the transduction of genes between *Salmonella* Typhimurium strains has made P22 a crucial tool for research on this organism [105], and the propensity of this phage for lateral transduction due to the initiation of viral replication and packaging before excision from the host genome can result in extremely high rates of gene transfer [106].

Beyond these classical models, *Lederbergvirus* phages infecting enterobacterial hosts are known to be common and diverse [107]. It was thus unsurprising that we isolated three different phages of this genus – Huey, Dewey, and Louie – which showed homogeneous and highly turbid plaques indicative of lysogen formation and self-immunity (Fig. S6B; NCBI GenBank accessions PQ850600, PQ850609, and PQ850616) [108]. However, we also isolated one representative called CurroJimenez (Bas106) that did not show temperate behavior when infecting *E. coli* K-12 with restored O-antigen and was included in the BASEL collection that focuses on virulent phages (Fig. S6B). Since it does not have obvious defects in its lysogeny control genes, we speculate that *E. coli* K-12 may lack the specific *attB* site which would be targeted by the integrase of this phage. The phylogeny of *Lederbergvirus* phages shows different distinct clades among which CurroJimenez, Dewey, and Louie are closely related in a sister group of P22 itself while Huey is found in a more distant clade closer to CUS-3 and also HK620 (Fig. 10D). However, the high genomic mosaicism of this viral genus makes it difficult to draw conclusions about the relationships of individual phages beyond the specific genes that have been used to calculate the phylogeny [33, 109]. Phage CurroJimenez totally depends on the expression of restored O16-type O-antigen for the infection of *E. coli* K-12 which identifies this glycan as the host receptor targeted by its tailspike (Fig. 10E). The phage is also highly sensitive to restriction-modification systems which is not unexpected given the absence of any predictable anti-restriction mechanisms (Fig. 10E and Table S2) seems to be common among temperate phages [10].

Tailspikes and tail fibers to overcome the O16-type O-antigen barrier

Since the O-antigen barrier is ubiquitous among enterobacteria, it must be overcome or bypassed by any phage infecting these hosts (Fig. 1). For classical model phages the O-antigen has been seen as a facultative first receptor for initial host recognition followed by the binding to a second, terminal receptor for irreversible adsorption with a distinct receptor-binding protein [29, 34]. As an example, phage T5 efficiently infects the *E. coli* K-12 model strain – not expressing any O-antigen – by directly targeting the FhuA porin as terminal receptor with its central tail fiber [110]. While its lateral tail fibers are not required

to infect a host without O-antigen, increased expression of the polymannose O8-type O-antigen that they target results in concomitantly faster adsorption of T5 to these hosts [111]. Conversely, increased expression of O8-type O-antigen results in decreasing adsorption of a T5 mutant specifically lacking these lateral tail fibers [111]. Restoration of the O16-type O-antigen of *E. coli* K-12 (which is not bound by the lateral tail fibers of T5) results in complete abrogation of T5 infections (Fig. 1B and our previous work [10]). Similar results have been obtained for other phages of different morphotypes that express two distinct RBPs for successive recognition of O-antigen and a terminal cell surface receptor ([recently reviewed in reference 29]), and the concept also applies to analogous recognition of other initial receptors like capsules [112]. However, previous work had already indicated that some phages – primarily podoviruses like classical models T7 and N4 – can also merely depend on a single receptor such as O-antigen or other glycans for host recognition and irreversible adsorption [17, 29, 31, 70]. Similar to these, the O16-dependent phages characterized in this study typically feature a single RBP to target this glycan, seemingly for both initial host recognition and subsequent irreversible adsorption (Figs. 11 and 12).

Our new isolates of small siphoviruses in genera *Kagunavirus*, *Skatevirus*, and others encode tailspikes for recognition and enzymatic processing of the O16-type O-antigen that are encoded directly downstream of a J-like central tail fiber protein (Fig. S2). Notably, this is the same locus where the *bona fide* RBPs described in our previous work are encoded [10] as well as the capsule-targeting tailspikes of small siphoviruses targeting *Klebsiella* spp. (Fig. S2) [113]. We therefore conclude that these different RBPs are used analogously for terminal receptor recognition. Interestingly, the tailspikes of these distinct siphoviruses as well as *Kayfunavirus* ElsieAttenhofer (Bas105), *Gamaleyavirus* RudolfBernhard (Bas96), and *Lederbergvirus* CurroJimenez (Bas106) are all homologous and feature an enzymatic pectin lyase-like receptor-binding domain (InterPro accession IPR011050 [114]) connected to variable anchor domains (Figs. 11A and 12A). A phylogeny including distant relatives known to target the O:4 O-antigen of *Salmonella* Typhimurium such as P22 shows that the tailspikes of our O16-targeting phages form a distinct, though diverse, group alongside tailspikes from a few phages isolated in previous studies (Fig. 11A) [64].

This close relationship between tailspikes from very different phages targeting O16-type O-antigen highlights the frequent horizontal exchange of loci encoding these RBPs [33, 115].

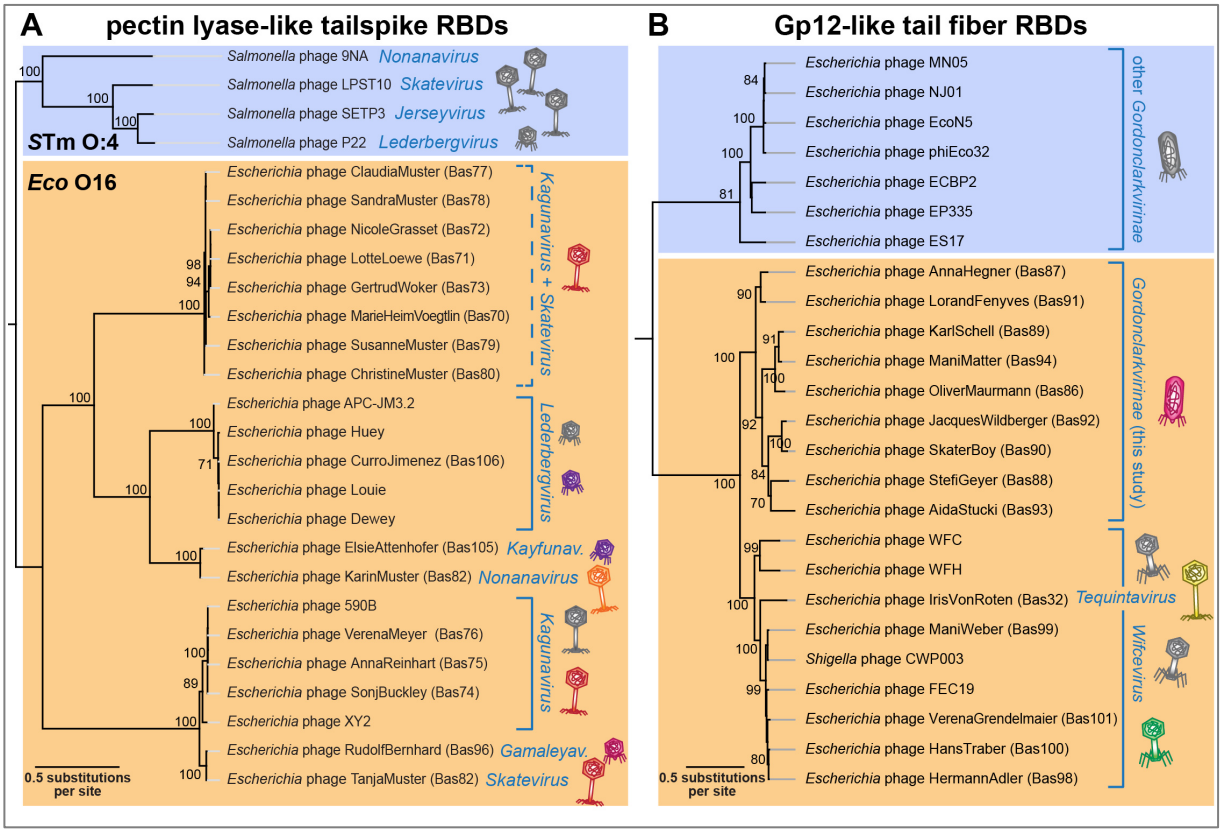


Figure 11. Phylogenetic analyses of tailspikes and tail fibers. (A) Maximum-Likelihood phylogeny of pectin lyase-like receptor-binding domains (InterPro accession IPR011050 [114]) of tailspikes encoded by our new BASEL phages and relevant controls. The phylogeny was midpoint-rooted between previously characterized tailspikes targeting the O:4 O-antigen of *Salmonella* Typhimurium (blue) and those of our new isolates targeting *E. coli* O16-type O-antigen (orange; including a few additional previously published phages with very similar tailspikes). Bootstrap support of branches is shown if > 70/100. Phages of the BASEL collection are highlighted by colorful phage icons. **(B)** Maximum-Likelihood phylogeny of Gp12-like receptor-binding domains (InterPro accession SSF88874 [114]) of tail fibers encoded by our new BASEL phages and relevant controls. The phylogeny was midpoint-rooted between a clade of distant *Gordonclarkvirinae* tail fibers at the top (blue) and those of *Wifcevirus*, *Tequintavirus*, and the *Gordonclarkvirinae* presented in this study (orange). Note that a clade with *Wifcevirus* receptor-binding domains (bottom) includes diverse phages from the database as well as *Tequintavirus* IrisVonRoten (Bas32 [10]) scattered between the *Wifcevirus* isolates characterized in the current study.

Based on prominent examples of glycan-targeting tailspikes, we had initially expected that phages depending on O-antigen as host receptor would mostly use enzymatic tailspikes to penetrate the glycan barrier [33, 78, 94]. However, many of our O16-dependent phage isolates had tail fibers instead. While

these varied in length and the types of receptor-binding domains (RBDs), many contained an RBD homologous to that of Gp12 short tail fibers of T-even phages targeting the *E. coli* LPS core [116] (InterPro accession SSF88874 [114]). These Gp12-like RBDs are located in the distal parts of lateral tail fibers across various phages, including myoviruses of the *Wifcevirus* genus, siphovirus IrisVonRoten (Bas32), and podoviruses of the *Gordonclarkvirinae* subfamily (Fig. 11B). However, such a broad distribution of closely related RBDs is not unexpected given the frequent horizontal transfer of tail fiber components [115].

Interestingly, a phylogeny shows that the Gp12-like RBDs of *Wifcevirus* isolates targeting O16-type O-antigen are scattered between homologs from diverse other *Wifcevirus* phages, suggesting that they either mediate very broad host recognition or that these other *Wifcevirus* isolates could also target this specific glycan (Fig. 11B). Conversely Gp12-like RBDs of *Gordonclarkvirinae* targeting O16-type O-antigen are only very distantly related to homologous domains of other phages from this group (Fig. 11B). For one of these others, EP335, it had been shown previously that its lateral tail fiber directly binds O157-type O-antigen, suggesting that this may be a conserved feature of these tail fibers with Gp12-like RBDs [117]. Other phages from the expanded BASEL collection encode different types of tail fibers with other RBDs that all likely target this glycan, highlighting the versatility of these non-enzymatic RBPs (see predicted structures in Figs. 12B, 12C, and S7).

Phages with several receptor-binding proteins or dual receptor-binding domains

Notably, some of our new isolates that depend on O16-tpe O-antigen of *E. coli* K-12 encode two and not only one *bona fide* RBP. Of these, *Jilinvirus* HeidiAbel (Bas97) encodes separate long and short tail fibers that are likely used successively during host recognition analogous to the long and short tail fibers of much larger T-even myoviruses [10, 73, 74] (Fig. S7). A very different situation is apparent for *Gamaleyavirus* RudolfBernhard (Bas96) which encodes two tailspikes of which one carries a pectin lyase-like domain closely related to others from our collection (Figs. 11A and 12A) and the other one seems unrelated. Such a setup of two tailspikes is common for *Gamaleyavirus* phages and other *Schitoviridae* (see also Fig. S8) [33, 78]. It seems likely that the two tailspikes are used for alternative target recognition

similar to *Autographiviridae* phage SP6 which carries two separate tailspikes that it can use to target different types of O-antigen to infect different strains of *Salmonella* [94].

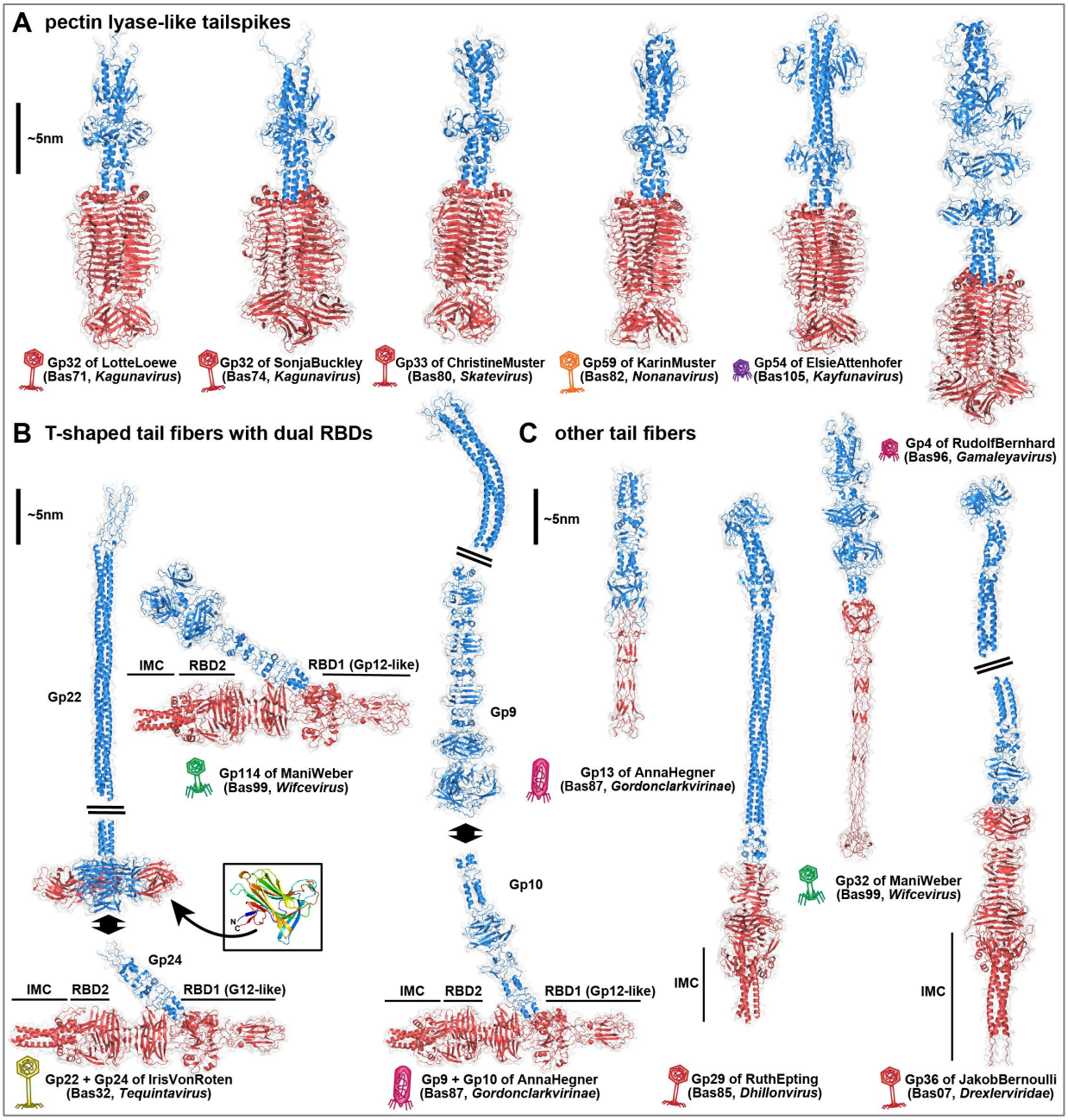


Figure 12. Predicted structures of representative tailspikes and tail fibers. (A) AlphaFold-Multimer [198] predicted structures of homotrimeric tailspikes from selected BASEL phages (see also Fig. 11A). **(B)** Predicted structures of the representative T-shaped tail fiber complexes of IrisVonRoten (Bas32, *Tequintavirus*), AnnaHegner (Bas87, *Gordonclarkvirinae*), and ManiWeber (Bas99, *Wifcevirus*) which all feature the dual Gp12-like RBD1 and T5-like RBD2 (predicted with the C-terminal intramolecular chaperone domain (IMC) attached; see also Figs. 11B and S10). While ManiWeber features an N-terminal anchor domain for baseplate attachment, the T-shaped tail fibers of IrisVonRoten and

AnnaHegner attach to proximal fibers (Gp22 and Gp9, respectively) which connect the complete fiber structure to the distal baseplate complex. The proximal fiber protein Gp22 of IrisVonRoten contains an additional galectin-like domain (inset box; colored N- (blue) to C- (red) terminus) that protrudes from the C-terminus of the fiber and could contribute to host receptor interaction as an additional RBD. While the overall setup of the T-shaped tail fibers is highly conserved, the RBDs are highly variable and seem to be frequently transferred horizontally (Figs. 11B, S9C, and S10D). **(C)** Predicted structures of other tail fibers with single RBDs (more shown in Fig. S7). The short tail fiber (Gp13) of AnnaHegner (Bas87, *Gordonclarkvirinae*) and the second tail fiber (Gp32) of ManiWeber (Bas99, *Wifcevirus*) have an elongated tip similar to the long tail fiber of phage T4 including conserved HxH motifs for ion coordination [118]. Tail fibers of JakobBernoulli (Bas07, *Drexlerviridae*; Gp36) and RuthEpting (Bas85, *Dhillonvirus*; Gp29) both contain elongated N-terminal coiled coils connected to receptor-binding T5-like fiber tips with C-terminal IMC domains. In (A)-(C), N-terminal anchor domains and fiber shafts are colored blue while the RBDs are colored red. A ~5 nm vertical scale bar is provided for reference. In all panels, viral morphotypes are annotated by phage icons analogous to Figs. 11A and 11B.

While the presentation of two independent RBPs is thus not uncommon for podoviruses, the *Gordonclarkvirinae* are the only phages of this morphotype in the BASEL collection that encode separate long and short tail fibers which are likely used successively and not alternatively [57]. The short tail fibers of these phages are highly conserved and appear to be vertically inherited (Figs. S9A and 12C) [57]. Similarly, the *Wifcevirus* phages also feature two different tail fibers of which one – predicted to have an elongated RBD tip similar to the T4 Gp37 long tail fiber [37] (Fig. 12C) – is highly conserved across this genus (Fig. S9B). We therefore suggest that these conserved RBPs have a more mechanical role during the infection process rather than recognizing divergent host receptors. Conversely, both the *Gordonclarkvirinae* and *Wifcevirus* phages feature a second tail fiber with a predicted distinctive T-shaped structure presenting dual RBDs, the Gp12-like RBD1 (see above) and a second, T5-like RBD2 that could not be identified from primary sequence alone (Figs. 12B and S10A-C). Both the Gp12-like RBD1 and the T5-like RBD2 differ widely between homologs of the T-shaped tail fibers (Fig. S9C) and have a phylogenetic distribution that is scattered across viral groups and morphotypes (Figs. S10D and 11B). These observations are in line with the recognition of highly variable host receptors like O-antigen, suggesting that these T-shaped tail fibers may generally target these glycans and possibly others like capsules.

This remarkable T-shaped tail fiber structure is predicted to have a very similar arrangement in *Gordonclarkvirinae* (exemplified by Gp10 of AnnaHegner, Bas87), *Wifcevirus* (exemplified by Gp114 of ManiWeber, Bas99) and also the lateral tail fibers of some *Tequintavirus* phages like IrisVonRoten (Gp24,

Bas32; see Fig. 12B). Using Gp24 of IrisvonRoten as a representative, the DALI server identified structural similarities and shared structural elements between the Gp12-like RBD1 with X-ray crystal structures of the distal knob domain (D10) of the T4 long tail fiber Gp37 [118] (Z-score 8.8, RMSD 3.2, 100 residues) and – as expected from primary sequence – with the Gp12 short tail fiber distal tip [118] (Z-score 6.2, RMSD 4.1, 127 residues; Fig. S10A-C). Likewise, the T5-like RBD2 of Gp24 of IrisVonRoten also shares structural similarity with the distal binding tip and intramolecular chaperone (IMC) domain of the phage T5 lateral tail fiber [119](Z-score 6.3, RMSD 4.8, 179 residues; see Fig. S10A-C). Finally, while the T-shaped RBPs of IrisVonRoten (Gp24) and AnnaHegner (Gp10) connect to a long proximal fiber (Gp23 and Gp9, respectively), the T-shaped RBP of ManiWeber (Gp114) instead features a prominent attachment domain at its N-terminus that likely facilitates direct baseplate attachment (Fig. 12B). This diversity highlights the modular versatility of bacteriophage RBPs for host recognition and the initiation of viral infections.

Host range across pathogenic *E. coli* and *Salmonella* Typhimurium

The host range of bacteriophages is a crucial property both with regard to their ecology and for applications in clinics and biotechnology [120-122]. It can be defined in different ways, but researchers generally distinguish a phage's lytic host range – estimated by the ability to lyse a host strain in a spot test – and its plating host range, i.e., the ability to form plaques [120, 121]. While plaque formation requires the phage to bind the host cell and to overcome or bypass all layers of bacterial immunity at least partially, the lytic host range is primarily informative about viral host recognition due to "lysis from without" at high phage-to-host ratios even in the absence of viral replication [123]. In our previous work we had tested the host range of BASEL phages and classical model phages against commonly used model strains of *E. coli* and *S.* Typhimurium [10]. For the current study we had anticipated that the diverse O-antigen variants of these strains would largely exclude the mostly O16-dependent phages presented here. This would make it impossible to estimate their interaction with bacterial immunity. We therefore included three additional

E. coli strains encoding O16-type O-antigen that belong to sequence types (ST)-131, -144, and -8281 (see section *Bacterial strains and strain construction* in *Materials and Methods*).

Our experiments to determine the lysis host range indeed revealed that most O16-dependent phages lysed all or some of the O16-expressing strains but barely any of the strains with rough LPS or another type of O-antigen (Fig. 13A). This is in line with the expected dominant role of O-antigen for the host recognition of these phages, especially since most of them do not encode additional receptor-binding proteins besides the O16-targeting tailspikes or tail fibers (Figs. 11 and 12) [29-31, 35]. However, there were stark differences between the different groups of phages that we tested. All small siphoviruses (genera *Kagunavirus*, *Skatevirus*, and relatives; Bas70-82) and many podoviruses (Bas95/96 Bas102-106) had a very narrow host range beyond strains encoding O16-type O-antigen and lysed almost none of the strains without it (Fig. 13A). These phages have either an O16-targeting enzymatic tailspike or simple tail fibers (Figs. 11, 12, and S7), suggesting that their ability to target additional host receptors is limited. Conversely, isolates of *Gordonclarkvirinae* and *Wifcevirus* groups showed a very broad host recognition and could in some cases even lyse strains with rough LPS (Fig. 13A). This indicates that the T-shaped tail fibers of these phages are more versatile and might be able to bind different glycan structures with their two RBDs (Figs. 11, 12, S7, S9B-C, and S10). Previous work had indeed already observed a comparably broad recognition of some isolates of large myoviruses (e.g., *Tevenvirinae*, *Vequintavirinae* and relatives, etc.) or large siphoviruses (e.g., *Markadamsvirinae*) that commonly carry complex long tail fibers [10, 35, 59, 124, 125]. Our data do not show unambiguously which of the two RBDs of the T-shaped tail fibers – or both – may be critical for O-antigen recognition and bacterial infection in our experiments. However, the phylogeny of the T5-like RBD2 does not correlate with any host recognition data (compare Figs. S10D and 13A). Conversely, the phylogeny of the Gp12-like RBD1 shows different clusters for *Gordonclarkvirinae* with broadest host recognition (Bas87 and Bas91) and others for those with narrower host range (Bas86/89/94 and Bas88/89/82/83; compare Figs. 11B and 13A). This may suggest a bigger role for the Gp12-like RBD in host recognition of *Gordonclarkvirinae* across our set of host strains.

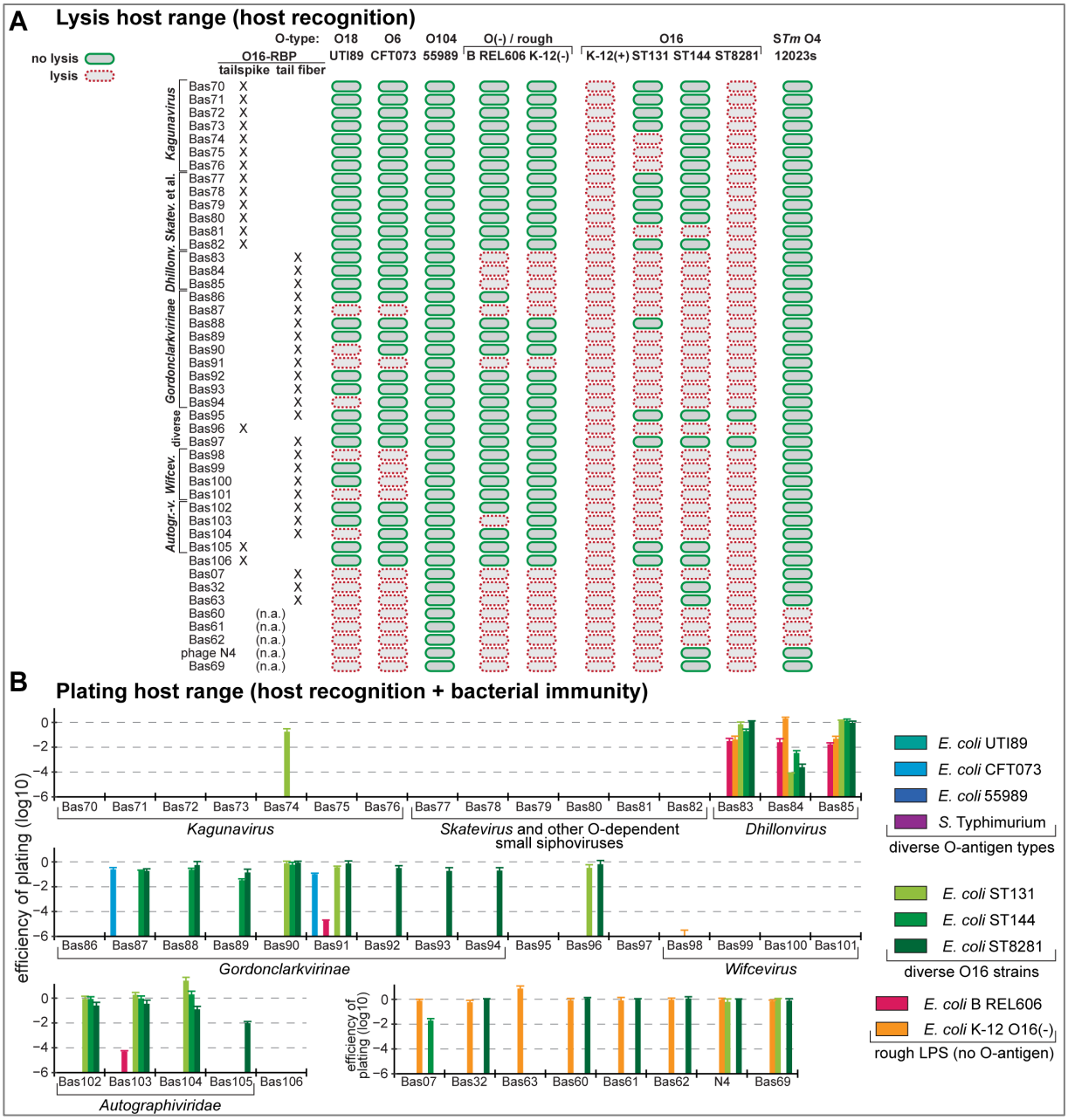


Figure 13. Host range of new phage isolates in the BASEL collection. (A) Lysis host range of all new BASEL phages and controls (same as in Fig. S1B) as determined by qualitative spot test assays with high-titer phage stocks (>10^9 pfu/ml if possible; see *Materials and Methods*). The O-antigen type of different *E. coli* and *Salmonella* Typhimurium strains and possibly rough LPS (as O(-)) are highlighted at the top while phages with O16-targeting tail spikes or tail fibers are distinguished on the left side. Lysis of a host strain is shown by a red cell lysis icon while the inability to lyse a strain is shown as a green icon in the shape of a cell with full integrity. **(B)** Plating host range of all new BASEL phages and controls (like in (A)) as determined by quantitative efficiency-of-plating (EOP) assays using serial dilutions of phage stocks. Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data. In (A) and (B), K-12 strains have been selected differently for phages inhibited by its cryptic prophages (Δ9CP) and for those who are not (ΔRM).

The siphoviruses with O16-targeting lateral tail fibers and additional central fibers to bind a porin as terminal receptor – new *Dhillonvirus* isolates Bas83-85 as well as Bas07 (JakobBernoulli) and Bas32 (IrisVonRoten) – lysed all rough strains but differed in their ability to infect others (Fig. 13A). This confirms the barrier function of mismatched O-antigen types for these phages and the expected facultative nature of O-antigen binding when the terminal receptor on the cell surface is accessible [10, 28]. As shown already previously [10, 12, 17], phages that can bypass O-antigen by targeting the NGR glycan (Bas60-62 and N4 as well as Bas69) have a very broad host recognition with the *Justusliebigvirus* phages Bas60-62 featuring the largest lysis host range which even includes *Salmonella* Typhimurium (Fig. 13A). This is likely due to the array of different tailspikes and tail fibers expressed by these phages which enables them to overcome diverse glycan barriers [126]. Conversely, we also observed that many apparently O16-dependent phages failed to lyse one or more of the strains with this O-antigen type that we tested, especially when the phages use enzymatic tailspikes for host recognition (Fig. 13A). Since these phages do not encode any other receptor-binding proteins, we find it unlikely that bacterial variation in other potential receptors such as the LPS core may be responsible for this behavior. Instead, it seems likely to us that strain-specific O-antigen modifications – which are abundant and diverse ([recently reviewed in reference 36]) – could prevent the binding of some highly specific tailspikes to certain variants of O16-type O-antigen.

The plating host range of all phages was usually much narrower than their lysis host range (Figs. 13A and 13B), which shows the impact of bacterial immunity in the diverse strains that we used. This effect was particularly strong for all phages with considerable sensitivity to RM systems such as small siphoviruses Bas70-82 (Fig. 13B; compare Figs. 5D and 6G). Conversely, phages with pronounced resistance to RM systems of the *Gordonclarkvirinae* (Bas86-93) and *Gamaleyavirus* (Bas96) groups could form plaques on many of the strains that they can lyse, indicating that this abundant and broadly targeting form of defense systems has a major impact on phage host range (Fig. 13B; compare Figs. 7D and 8G) [40, 46, 47]. This is in line with previous work showing a broad plating host range for individual isolates of *Gordonclarkvirinae* [117, 121, 127] and a narrower one for individual *Wifcevirus* phages or small

siphoviruses like *Kagunavirus* isolates [35, 59]. Intriguingly, *Autographiviridae* phages (Bas102-105) and in particular the new *Dhillonvirus* isolates (Bas83-85) formed plaques on nearly all strains that they could lyse despite their only moderate resistance to RM systems (Figs. 13A and 13B; compare Figs. 6G and 10D). It would be interesting to study the underlying molecular mechanisms that apparently enable these phages to overcome the native RM systems of these strains while the *Wifcevirus* phages – displaying considerable resistance to restriction in the K-12 host (Fig. 9E) – fail to form plaques on any of the many strains which they can lyse (Figs. 13A and 13B).

Cryptic prophages inhibit some BASEL phages via O-antigen glycosylation

One of the most curious biological results of this study has been the observation that phages belonging to very different viral groups – Bas74-76 (*Kagunavirus*), Bas83-85 (*Dhillonvirus*), Bas87 + Bas91 (*Gordonclarkvirinae*), and Bas98-Bas101 (*Wifcevirus*) – are inhibited by the presence of the cryptic prophages of *E. coli* K-12 (see Figs. 5D, 6G, 7D, and 9E). Based on the diverse *bona fide* anti-phage defense systems encoded in these genomic islands (Fig. 4D), we had initially assumed that one or more of these would target the phages and impair their replication. The phylogenies of these phages show that sensitivity to inhibition by the cryptic prophages does not correlate with certain clades of their respective viral groups (Figs. 5C, 6F, 7C, and 9B). This observation suggested that sensitivity to this inhibition must be due to variable genes of the susceptible phage groups that could, e.g., be sensed / targeted by bacterial immunity or may serve as anti-defense factors as shown in other cases [40, 45, 128]. However, we also realized that the loci encoding the receptor-binding proteins of the phages sensitive to inhibition by the cryptic prophages were distinct from their insensitive relatives. This can be seen, e.g., for the *Kagunavirus* tailspikes in Fig. 11A where those of the isolates inhibited by cryptic prophages (Bas74-76) cluster in a completely different clade than the insensitive ones (Bas70-73). Similarly, the only features specifically shared by *Gordonclarkvirinae* AnnaHegner (Bas87) and LorandFenyves (Bas91) which are sensitive to the cryptic prophages is their variant of their Gp12-like tail fiber RBD which is different from those of all other isolates

(Fig. 11B). We thus suspected that viral adsorption and not classical anti-phage defense systems could be responsible at least for some of the observed differences in sensitivity to the presence of cryptic prophages.

Interestingly, the cryptic prophage CPS-53 (also known as KplE1) of *E. coli* K-12 encodes a GtrABS O-antigen glycosylation system that adds the glucose side chain to the O16-type O-antigen of this strain for which its O-antigen biosynthesis locus itself lacks a dedicated glycosyltransferase (Fig. 14A) [129, 130]. Consistently, this side chain could not be found when the O-antigen structure of another *E. coli* strain with O16-type O-antigen was solved [131]. Given that O-antigen modifications can impair the adsorption of bacteriophages [36], we tested the role of the CPS-53 GtrABS system for the host range of our phage isolates that were inhibited by the K-12 cryptic prophages. Remarkably, genetic inactivation of this system resulted in susceptibility of *E. coli* K-12 to all these phages, suggesting that O-antigen glycosylation and no other aspect of bacterial immunity had inhibited their growth on *E. coli* K-12 with cryptic prophages (Fig. 14B). These results were reminiscent of previous work showing that homologous tailspikes of phages EP75 and HK620 target differently glucosylated variants of *E. coli* O18 O-antigen, possibly due to differences in glycan repeat recognition [117]. Unlike the other phages that depend on O-antigen as host receptor, the *Dhillonvirus* isolates Bas83-85 were only affected by the GtrABS system in presence of restored O-antigen expression while they infected *E. coli* K-12 with and without cryptic prophages equally well in the absence of this glycan barrier (Figs. 14B and S11). The *Dhillonvirus* phages have two RBPs and do not need recognition of the O-antigen as primary receptor to target their terminal porin receptor, but they depend on O-antigen recognition – which is apparently impaired by GtrABS glycosylation – to cross this layer when it is present [10, 100].

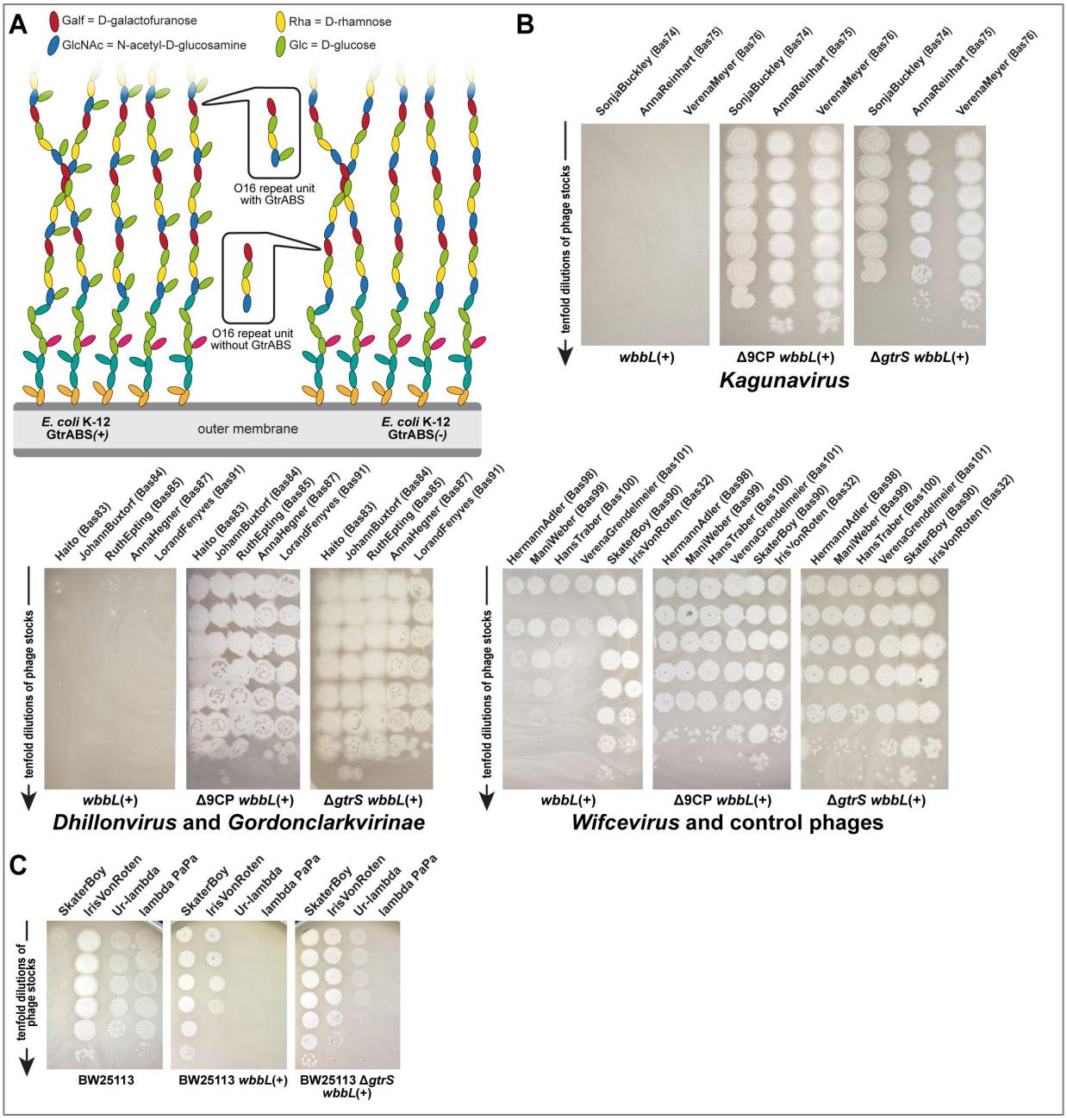


Figure 14. GtrABS-mediated O-antigen glycosylation inhibits phage infection. (A) O16-type O-antigen of *E. coli* K-12 with (left) or without (right) the GtrABS system of CPS-53 differs in the presence or absence of the glucose side chain, respectively [129, 130]. **(B)** Serial dilutions of different phages were spotted on variants of *E. coli* K-12 BW25113 with restored O-antigen expression as well as deletion of the GtrS glucosyltransferase (Δ*gtrS*) or the nine cryptic prophages including CPS-53 (Δ9CP). Three *Dhillonvirus* phages (Bas83-85), two *Gordonclarkvirinae* (Bas87 + Bas91), and the *Wifcevirus* phages (Bas98-101) are inhibited by the K-12 cryptic prophages if the GtrABS system in CPS-53 is intact (see Fig. S11 for the same experiment without O-antigen expression). *Gordonclarkvirinae* phage SkaterBoy (Bas90) and *Tequintavirus* IrisVonRoten (Bas32) were used as GtrABS-insensitive controls that do or do not depend on O16-type O-antigen, respectively. A small growth defect of the three *Kagunavirus* phages in the *gtrS* mutant compared to the Δ9CP strain (top right) is indicative of a slight inhibition of

these phages by other factor(s) encoded in the cryptic prophages **(C)** Serial dilutions of lambda PaPa and Ur-lambda were spotted on variants of *E. coli* K-12 BW25113 either in its wildtype form (left) or with restored O-antigen expression and either active (middle) or inactivated (right) GtrABS system. Phages SkaterBoy (Bas90, *Gordonclarkvirinae*) and IrisVonRoten (Bas32, *Tequintavirus*) were added as controls that do or do not depend on the O16-type O-antigen and are both insensitive to GtrABS.

These observations with the *Dhillonvirus* phages were reminiscent of the conundrum that phage lambda – initially found as a prophage in *E. coli* K-12 – was unable to infect this strain after restoration of its O-antigen [10]. No plaque formation in the presence of this glycan barrier could be seen both with the commonly used laboratory variant lambda PaPa or the original Ur-lambda (Fig. 14C) [10, 132]. Unlike the domesticated lambda PaPa, Ur-lambda does not only feature a central tail fiber targeting the LamB porin as terminal receptor but also side tail fibers suggested to bind OmpC as primary receptor [132]. Our results raised the question how lambda would initially have lysogenized the original *E. coli* K-12 strain that expressed O16-type O-antigen [27]. Interestingly, we find that Ur-lambda with its side tail fibers – but not lambda PaPa without them – can overcome the O16-type O-antigen of *E. coli* K-12 in the absence of GtrABS-mediated glycosylation (Fig. 14C). These results suggest that the side tail fibers of Ur-lambda can bind unmodified but not glucosylated O16-type O-antigen as host receptor similar to the recognition of O-antigen as primary receptor by other siphoviruses like JakobBernoulli (Bas07) or IrisVonRoten (Bas32; see Fig. 1) [10]. Phage lambda might thus have lysogenized *E. coli* K-12 either before acquisition of CPS-53 or while the GtrABS system was transiently not expressed since related systems, e.g., in *Salmonella*, are controlled by phase variation [133].

Discussion

Phages depending on O-antigen glycans are diverse and abundant in nature

While it is curious that we find O16-targeting phages to be much more abundant in environmental samples compared to sewage treatment plant inflow (Table S1), previous work has also observed such differences in other contexts. As an example, phages targeting the conjugative pili of different

enterobacterial plasmids were much more common in wastewater compared to river water [134]. While the underlying ecological phenomena remain elusive, it seems certain that wastewater does not merely contain a large abundance of *E. coli* phages compared to most environmental samples but also has its own biases.

One interesting observation for many O16-dependent phages characterized in this work is that they seem to use these glycans as terminal host receptors while their genomes lack any recognizable strategy to inactivate this receptor during infections. Conversely, virulent phages targeting porins characteristically encode proteins such as Cor of T1 or Llp of T5 that disable their porin receptors to prevent adsorption of viral offspring to the debris of the lysed virocell [135, 136]. Similar mechanisms are also common among temperate phages including those targeting O-antigen glycans which they can chemically modify, e.g., by glycosylation or acetylation ([recently reviewed in reference 36]), for "superinfection exclusion" of related viruses [14, 137]. These examples show that such exclusion mechanisms would in principle also be possible for virulent phages, but it seems likely that the abundance of O-antigen molecules on the cell surface (e.g., compared to porins) makes a complete modification during lytic infections futile. Instead, recent work suggested an alternative strategy of phages that transiently block their offspring's host recognition by "caps" made of chaperones on receptor-binding proteins instead of disabling the host receptors [138]. We expect that this strategy and similar concepts may be more widespread among glycan-targeting phages, and we look forward to future research uncovering their molecular mechanisms and evolutionary context.

Traditionally, siphoviruses infecting *E. coli* were considered to exclusively target porins as terminal host receptors like well-studied model phages T1 or T5 [60, 61]. However, more recent studies had already described siphoviruses infecting *E. coli* or other enterobacteria that target flagella, conjugative pili, or surface glycans as host receptor [33, 93, 139, 140]. Using *E. coli* K-12 with restored O16-type O-antigen expression as host, we indeed isolated diverse small siphoviruses from genera *Kagunavirus*, *Skatevirus*, and others which use this surface glycan as essential host receptor. It seems likely that the abundance and diversity of these groups had previously been overlooked due to the traditional focus on *E. coli* laboratory strains with rough LPS (see Fig. 1A). Our results strongly suggest that O-antigen is the only and terminal

host receptor of these isolates because they only carry a single receptor-binding protein – in all cases enzymatic tailspikes (Figs. 11A and 12A) – which is encoded directly next to the J-like central tail fiber hub (Fig. S2A). This is the same locus where other small siphoviruses infecting *E. coli* encode *bona fide* receptor-binding proteins as well as superinfection exclusion factors targeting different porins (Fig. S2B [10, 135]), and host recognition by these central tail fibers is known to trigger irreversible adsorption and genome injection [110, 141]. We look forward to future studies characterizing the mechanistic link of O-antigen recognition and genome injection in more detail. One specific aspect deserving further investigation is the way how these processes are linked on the molecular level compared to phages with two distinct tail fibers carrying receptor-binding domains. Electron microscopy of small siphoviruses depending on O-antigen clearly shows three distal tail appendages that are likely the tailspikes presented on the J-like fibers (see Figs. 5B and 6C). Conversely, siphoviruses targeting porins generally have only one central tail fiber targeting this terminal receptor besides three lateral tail fibers targeting O-antigen or other initial receptors [100]. This indicates that the single RBPs of siphoviruses targeting only O-antigen may be presented on the virion like lateral tail fibers while being functionally equivalent to central tail fibers for irreversible adsorption and triggering DNA genome ejection.

Our study highlighted once more that – besides a few models like P22, G7C, and T4 [37, 72, 78, 142] – we lack a molecular understanding how many phages use their diverse tail fibers and tailspikes to recognize host receptors and initiate infections. While we had initially expected to isolate many phages with tailspikes targeting O16-type O-antigen that would move the virion to the cell surface by enzymatic activity, many instead carried tail fibers lacking enzymatic properties (Figs. 12B-C and S7). Beyond a mere binding to their LPS receptor (which has been well studied for several examples [37, 70, 119]), we speculate that successive transient interactions of tail fibers with O-antigen receptors may enable the fibers to “glide” through this barrier towards the cell surface (see Figs. 7A and 9A). Of particular interest are the novel T-shaped tail fibers with separate Gp12-like and T5-like RBDs (Figs. 12B and S10) that – to the best of our knowledge – are the first example of a tail fiber protein with two distinct RBDs in the same polypeptide.

Conversely, branched structures combining distinct RBPs have been well described [143] and the long tail fiber tip Gp37 of phage T4 features different patches in the same RBD to target distinct receptors [37].

Phage host range, glycan receptors, and bacterial immunity

Understanding the molecular mechanisms underlying the host range of different bacteriophages is important not only for ecological insights but also to enable effective applications in clinics [35, 54-56, 122]. While diverse aspects of phage-host interactions like bacterial metabolism and antiviral immunity all influence phage host range, the access to and recognition of viral host receptors generally play major roles [144, 145]. As expected from previous work [10, 28, 30, 35], we observed a dominant role of the O-antigen barrier for phage host range in a way that O-antigen types mismatched to the phages' receptor-binding proteins abolished host recognition while phages with no other host receptor also fully depended on its presence. Interestingly, our data show that phages with tail fibers tend to have a broader host recognition than phages with tailspikes among the tested viruses (Fig. 13A). Understanding these differences on the molecular level by studying distinct functionalities of these different types of RBPs more systematically would not only expand our knowledge of phage-host interactions but could also help applications in clinics and biotechnology, e.g., by phage host range engineering [146, 147]. The expanded BASEL collection with diverse phages that target the same receptor – O16-type O-antigen – could be a useful tool to investigate these interactions.

The dual role of O-antigen as a barrier and as a host receptor is reminiscent, e.g., of the well-studied capsules of *Klebsiella* spp. which form a robust barrier to phage infection but also serve as a dominant factor for phage host range by specific recognition of capsule types by certain variants of RBPs [113, 148, 149]. The specificity and dominance of these interactions even enable their use for quite reliable phage host range predictions across *Klebsiella* strains [150]. Conversely, for other hosts like *P. aeruginosa* the availability of long retractable type IV pili reduces the importance of glycans as host receptors, and for many Gram-positives like *Staphylococcus aureus* at least the host recognition is usually broader due to a lower inter-strain diversity of bacterial surface structures [14, 151-153].

For *E. coli*, the picture is certainly more nuanced because, e.g., most of the O16-targeting phages that can lyse our different tested strains with this O-antigen type fail to form plaques on them (compare Figs. 13A and 13B). While this shows a clear influence of bacterial immunity, a recent study that experimentally tested the host range of many *E. coli* phages across many strains did not observe a simple correlation between phage host range and the number of predicted antiviral defense systems [35], unlike what had been observed for *P. aeruginosa* and *Vibrio* spp. [16, 154, 155]. In agreement with recent work by David, Plantady, et al. on a different strain of *E. coli* [145], our results primarily suggest a link between the ability to form plaques on diverse strains and viral resistance to RM systems as the most common and most broadly targeting form of bacterial immunity (see above) [46, 47, 156]. However, the relative importance of different antiviral defense systems for observed phage host range would certainly be a promising field for future study beyond mere predictions, e.g., by contrasting the phage susceptibility of wildtype hosts with mutants specifically lacking one or more defense systems. Our results indicate that – after removal of restriction systems in our previous work [10] – the *E. coli* K-12 genomic background may not encode anti-phage defense systems in narrow sense that play a major role for phage host range. Even when all predictable defense systems were absent in the Δ9CP strain lacking all cryptic prophages (Fig. 4D; see *Bacterial strains* in *Materials and Methods*), the new viruses that could be isolated were merely susceptible to an O-antigen glycosylation system encoded in one of these genomic islands (Fig. 14).

Expanding the BASEL collection to cover all major groups of *E. coli* phages

The value of the BASEL collection as a tool for molecular biology research lies in its assorted diversity across and within phage groups and on their ability to infect the model organism *E. coli* K-12 [6]. In this study we therefore aimed at enhancing the assorted diversity to cover all known major groups of *E. coli* phages without a major compromise on domesticated *E. coli* K-12 as host strain (Fig. 15). After restoring its O-antigen and inactivating all predicted antiviral defenses – of which only O-antigen glycosylation played a detectable role –we indeed readily isolated representatives of diverse previously inaccessible viral groups most of which strictly depend on O-antigen as host receptor. These include, e.g.,

diverse groups of small siphoviruses like the *Kagunavirus* genus, myoviruses of the *Wifcevirus* genus, and podoviruses of *Gordonclarkvirinae* or genus *Kayfunavirus.* Since phages from most of these diverse groups had never been isolated with any *E. coli* lab strain, there were only scarce – if any – previous studies on their molecular biology. This limitation restricted our analysis of the genomic and phenotypic data for some of them largely to data from the current study, while in our previous work on more well-studied groups of phages we could draw heavily from the generations of previous research in the field [10]. To improve the genome annotation of our new phage isolates, we therefore created a new tool – GAPS (Gene Annotation Pipeline based on sequence and structure Similarity; see *Materials and Methods*) to infer functional information from remote homologies and structure predictions.

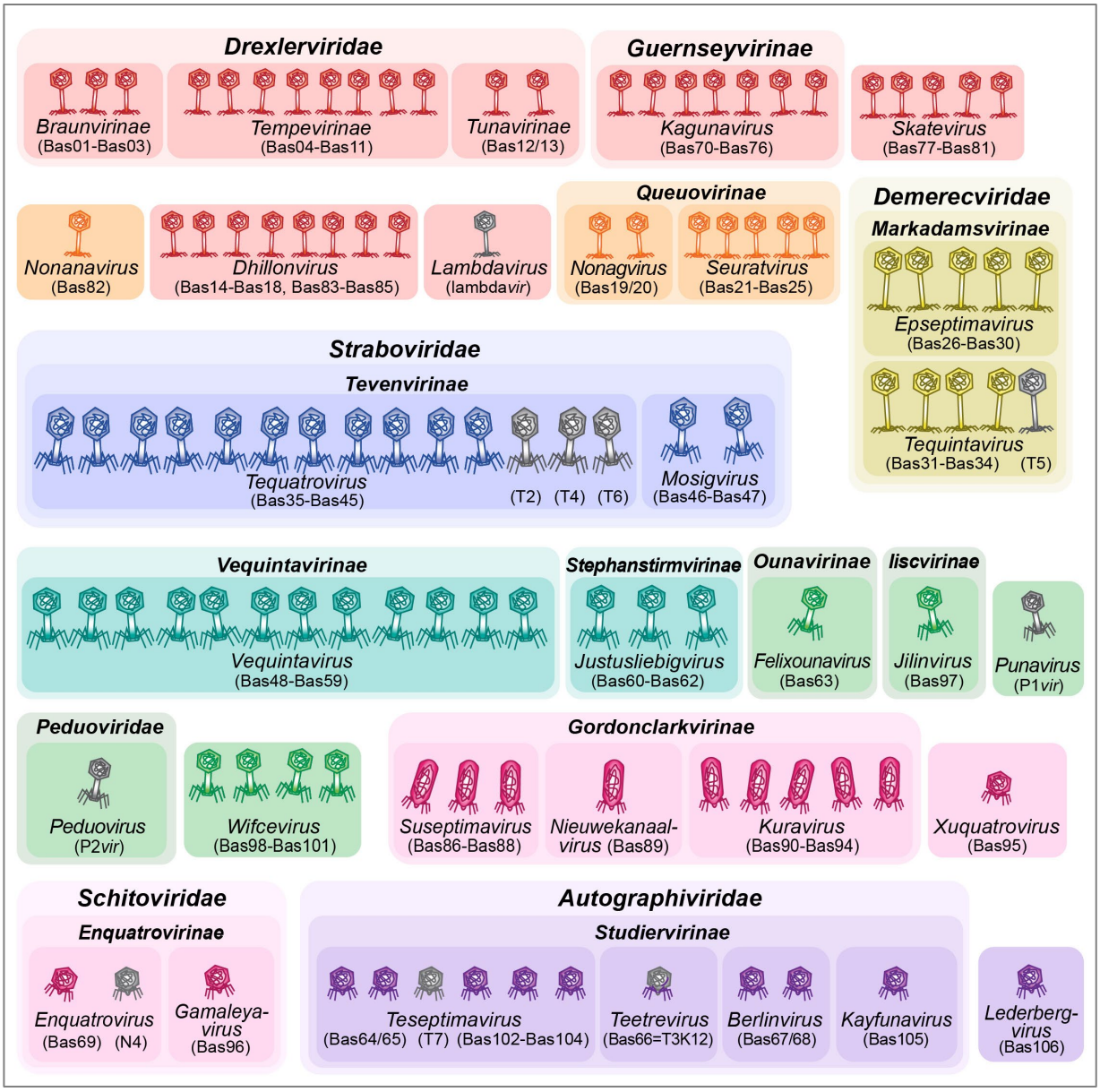


Figure 15. Overview of the complete BASEL collection. The illustration shows all isolates of the complete BASEL collection with their Bas## identifiers. Isolates are sorted by morphotype and their current ICTV classification [68]. The color code distinguishes phages based on morphotype and genome size. For siphoviruses, small phages (ca. 30 50 kb) are shown in red, medium-sized phages (ca. 50-100 kb) are shown in orange, and large phages (>100 kb) are shown in yellow. For myoviruses, small phages (ca. 30-100 kb) are shown in green, medium-sized phages (ca. 100-150kb) are shown in cyan, and large phages (>100kb) are shown in blue. For podoviruses, small phages (ca. 30-50 kb) are shown in violet and medium-sized phages (ca. 50-100kb) are shown in pink.

We anticipate that the completed BASEL collection will support various applications from fundamental research on viral ecology and evolution over the work on new antiviral defense systems to studies on phage therapy. With an expanded biological diversity in still manageable numbers, these phages

can all be grown on a regular *E. coli* K-12 by using a strain with inactivated GtrABS system in two variants, once with and once without restored O-antigen expression. Given the increasing interest in phage therapy but also its notable incidence of treatment failure [85, 157, 158], it is clear that not only fundamental research but also clinical applications could profit from a better understanding of links between phage genomes and taxonomy on one side and relevant biology / phenotypes + clinical properties on the other side [35, 54-56]. This is becoming increasingly evident as more and more engineered phages for therapeutic applications are developed [159]. For these applications, phages are mostly used as a vehicle to deliver proteins from non-viral origin such as CRISPR-Cas nucleases or antibacterial toxins [62, 159-162]. This limitation highlights our lack of knowledge about the diverse molecular biology of phage-host interactions beyond a few traditional models. Unraveling the molecular mechanisms underlying new aspects of phage-host interactions – e.g., with tools like the BASEL collection or modern biochemistry and structural biology – could thus open up the untapped potential of engineered phages for applications in clinics or biotechnology.

Materials and Methods

Preparation of culture media and solutions

Lysogeny Broth (LB) was prepared by dissolving 10 g/l tryptone, 5 g/l yeast extract, and 10 g/l sodium chloride in Milli-Q H_2O and then sterilized by autoclaving. LB agar plates were prepared by supplementing LB medium with agar at 1.5% w/v before autoclaving. Phosphate-buffered saline (PBS) was prepared as a solution containing 8 g/l NaCl, 0.2 g/l KCl, 1.44 g/l $NA_2HPO_4 \cdot 2H_2O$, and 0.24 g/l KH_2PO_4 with the pH adjusted to 7.4 using 10 M NaOH and sterilized by autoclaving. SM buffer was prepared as 0.1 M NaCl, 10 mM $MgSO_4$, and 0.05 M Tris (pH 7.5) and sterilized by autoclaving.

Bacterial handling and culturing

Escherichia coli and *Salmonella enterica* subsp. *enterica* serovar Typhimurium strains were routinely cultured in LB medium at 37°C in glass tubes, plastic tubes, or Erlenmeyer flasks with agitation at 170 rpm. Selection for genetic modifications or plasmid maintenance was performed with ampicillin at 50 µg/ml, kanamycin at 25 µg/ml, and zeocin at 50 µg/ml.

Bacteriophage handling and culturing

Bacteriophages were generally cultured using the double-agar overlay method [163] with improvements of classical techniques described by Kauffman and Polz [164]. Top agar was prepared as LB agar with only 0.5% w/v agar supplemented with 20 mM $MgSO_4$ and 5 mM $CaCl_2$. Top agar plates were incubated at 37°C and plaques were counted as soon as they were identified by visual inspection. However, the plates were always incubated for at least 24 hours to record also slow-growing plaques.

High-titer stocks of bacteriophages were generated using plate-grown viruses. Briefly, phages were spread with a loop throughout the top agar before solidification to achieve confluent lysis after incubation for 16-24 hours at 37°C. Subsequently, top agars were scraped off from the bottom agar into each 15 ml of SM buffer and centrifuged at 8’000 g for 10 minutes. Supernatants were used as high-titer stocks and stored in the dark at 4°C. Phage Ur-lambda was obtained from the supernatant of an exponentially growing culture

of the *E. coli* K-12 EMG2 strain (CGSC #4401) – carrying an Ur-lambda prophage – after two hours of treatment with 0.5 µg/ml of mitomycin C to trigger prophage induction [132, 165].

Phages of the original BASEL collection from our previous work [10] were generally grown on *E. coli* K-12 MG1655 ΔRM. New isolates from this study were propagated on *E. coli* K-12 MG1655 ΔRM with restored O-antigen (carrying pAS001 or the *wbbL*(+) genotype) except those with sensitivity to GtrABS-mediated O-antigen glycosylation. These phages were cultured on *E. coli* K-12 BW25113 Δ9CP or *E. coli* K-12 BW25113 Δ*gtrS* (both with restored O-antigen, i.e., carrying pAS001 or the *wbbL*(+) genotype). All bacteriophages used in this study are listed in Table S2.

Bacterial strains and strain construction

All bacterial strains used in this work are listed in Table S3, all oligonucleotide primers in Table S4, and all plasmids in Table S5.

Variants of *Escherichia coli* K-12 MG1655 and BW25113

The primary host of our experimentation with *E. coli* K-12 was the ΔRM mutant of the MG1655 laboratory strain which we had reported in our previous work [10]. Briefly, we had engineered this strain to be a more permissive host for bacteriophages by knocking out all known restriction systems of *E. coli* K-12, i.e., the EcoKI type I RM system (encoded by *hsdRMS*) and the McrA, Mrr, and McrBC type IV restriction systems. For phage isolation and phenotyping, the strain carried a variant of the F-plasmid in which the *pifA* immunity gene had been deleted [10]. Phenotyping with plasmid-encoded antiviral defense systems was done with a strain carrying pBR322_ΔP*tet* as an empty vector control [10, 166] (see Table S3). Whenever indicated, we used the *E. coli* K-12 ΔRM variant with a repaired chromosomal *wbbL* gene that has restored O16-type O-antigen expression from our previous work [10] (see Table S3). To rapidly restore O-antigen expression in diverse strains (e.g., porin mutants of the Keio single-gene knockout collection [167]) we created plasmid pAS001 that constitutively expresses *wbbL* (see below in section *Plasmid construction*). Fortuitously, the plasmid-encoded *wbbL* does not fully restore the O-antigen barrier

(because most phages inhibited by it still show infectivity; compare Figs. 1 and S11A) but enough of it to enable robust infections by all O16-dependent phages (compare Figs. 2 and S11B).

To identify possible limitations of phage infectivity beyond host recognition, we focused on possible antiviral defenses in the K-12 genomic background beyond the restriction systems knocked out in *E. coli* K-12 MG1655 ΔRM. One of these is the enigmatic enterobacterial type I-E CRISPR-Cas system known to be not functional in phage defense without dedicated genetic engineering [168, 169]. In addition, *E. coli* K-12 strains encode Lit (reported to cleave bacterial EF-Tu when sensing the major capsid protein of some *Tevenvirinae* [170]), RnlAB (toxin-antitoxin system that aborts T4 infections in absence of the viral *dmd* "master key" antitoxin [171]), DicB (inhibitor of the bacterial ManYZ inner membrane conduit of lambda phage DNA translocation [172]), and a Hachiman type I system (targeting DNA replication of diverse phages; also known as AbpBA [44, 47, 173]). With the exception of the seemingly inactive type I-E CRISPR-Cas system, all these *bona fide* antiviral defenses are encoded in the cryptic prophages of *E. coli* K-12. We therefore used the *E. coli* K-12 BW25113 Δ9CP strain devoid of all cryptic prophages as an additional host for phage isolation and characterization [51]. This strain is derived from the BW25113 laboratory wildtype of *E. coli* K-12 that has a few minor differences compared to the MG1655 laboratory wildtype [174, 175] (see also Fig. 4D). Most important for our work, *E. coli* K-12 BW25113 carries active variants of the three type IV restriction systems knocked out in our MG1655 ΔRM strain as well as the *hsdRMS* locus encoding the EcoKI type I RM system. Unfortunately, due to the high sensitivity of *E. coli* K-12 BW25113 Δ9CP to stress and possibly also as a consequence of its already extensive genomic engineering, our attempted knockouts of these systems consistently failed. However, the EcoKI locus of BW25113 is restriction-deficient and we are not aware of any wildtype phages that would be sensitive to the three type IV restriction systems of *E. coli* K-12, suggesting that the presence of these loci might have only a negligible effect. Analogous to our work with *E. coli* K-12 MG1655 ΔRM [10], we also generated a variant of *E. coli* K-12 BW25113 Δ9CP with chromosomally restored *wbbL(+)* to isolate and study phages that would depend on this glycan (see Table S3). The different variants of *E. coli* K-12 BW25113 Δ9CP

were used for phage isolation and characterization just like the ΔRM variants with plasmid-encoded antiviral defenses and a *pifA*-deficient F-plasmid (see Table S3).

Later during our study, we found that only the GtrABS O-antigen modification system of CPS-53 was responsible for the observed inhibition of phage growth by the cryptic prophages of *E. coli* K-12 (Fig. 14). We therefore additionally used an *E. coli* K-12 BW25113 Δ*gtrS* mutant as well as a derivative of this knockout with chromosomally repaired *wbbL* – restoring O-antigen expression – for different experiments shown in this manuscript (see Table S3). Briefly, a clean deletion of *gtrS* in the *E. coli* K-12 BW25113 strain was generated by lambda red recombineering in two steps analogous to the knockouts in our previous work [10]. In a first step, *gtrS* was replaced with a double-selectable cassette encoding kanamycin resistance and sucrose sensitivity (amplified from pUA139_cat-sacB_v3 using prDP0079/80). Subsequently, the double-selectable cassette was removed by recombineering with annealed 80 nt oligonucleotides spanning the desired deletion with 40 nt on each side (prDP0081/82). Oligonucleotide primers prDP0083/84 were used to probe the success of each step of recombineering by overspanning colony PCR over the *gtrS* gene (all listed in Table S4). Finally, the *wbbL(+)* gene was restored by precisely removing the IS5 element disrupting this gene as described in our previous work [10].

Other enterobacterial strains used for bacteriophage phenotyping

To determine the host range of our new phage isolates and relevant controls, we initially used the same strains as in our previous work (see there for details [10]). Briefly, besides the *E. coli* K-12 variants with and without restored O-antigen (see above; phylogroup A, ST10, O16(+/-):H48, K-12 LPS core) these were the rough *E. coli* B REL606 laboratory strain (phylogroup A, O7(-):K5:H(-), truncated R1 LPS core), uropathogenic *E. coli* model strains UTI89 (phylogroup B2, ST95, O18:K1:H7, R1 LPS core) and CFT073 (phylogroup B2, ST73, O6:K2:H1, R1 LPS core), and enteroaggregative *E. coli* model strain 55989 (phylogroup B1, ST678, O104:H4, R3 LPS core). Similarly, we again used the common *Salmonella enterica* subsp. *enterica* serovar Typhimurium laboratory strains 12023s (also known as ATCC 14028) and SL1344 that share the *Salmonella* Typhimurium O4-type O-antigen. Because the results with these strains

were indistinguishable, only those with *S.* Typhimurium 12023s are included in this manuscript. Beyond these strains already included for host range analyses in our previous work, we included three additional *E. coli* strains with O16-type O-antigen (i.e., the same as *E. coli* K-12) belonging to ST131 (Egli lab collection isolate 720834-18; phylogroup B2, O16:H5), ST144 (Egli lab collection isolate 714478-19; phylogroup B2, O16:H6), and ST8281 (Egli lab collection isolate 711043-19; phylogroup B2, O16:H5) [176].

Plasmids and plasmid construction

For the phenotyping of phage resistance or sensitivity to diverse bacterial immunity systems, we used plasmids encoding different antiviral defenses constructed in our previous work (see Figs. 4B and 4C as well as Table S5). Plasmid pAS001 encoding *wbbL* was cloned by inserting the *wbbL* gene downstream of a constitutively expressed chloramphenicol resistance gene on a plasmid with SC101 origin of replication (see details of plasmid construction in Table S6).

Bacteriophage isolation

Basic procedure

Phages isolates presented in this study have been isolated from diverse samples between March of 2020 and July of 2022. Initial experiments aimed at isolating phages that would depend on O16-type O-antigen used *E. coli* K-12 MG1655 ΔRM *wbbL*(+) as host and subsequent plating on its ancestor without O-antigen to identify O16-dependent isolates. Later experiments aimed at isolating phages inhibited by presence of the K-12 cryptic prophages used *E. coli* K-12 BW25113 Δ9CP transformed with *wbbL*-expressing pAS001 as host to obtain phages irrespective of whether they depend on O-antigen or not (compare Figs. 1, 2, and S12). These isolates were later plated on *E. coli* K-12 BW25113 transformed with pAS001 to screen for impaired growth in the presence of cryptic prophages and on *E. coli* K-12 BW25113 Δ9CP without pAS001 to screen for dependence on O-antigen expression. The isolation and screening procedure is summarized in Figs. 2A and 2B.

We generally isolated phages by direct plating without enrichment step to avoid biases in favor of fast-growing isolates. The procedure was performed similar to the approach in our previous work [10]. Aqueous samples were directly used in batches of 50 ml while solid samples were first agitated overnight in PBS for suspension and release of viral particles. Subsequently, we centrifuged all samples at 8'000 g for 10 minutes to separate viral particles – which remain in the supernatant – from larger unorganic particles or bacterial cells. Viral particles were then precipitated by supplementation with $ZnCl_2$ at a final concentration of 40 mM and an incubation without agitation for one hour at 37°C. Subsequently, precipitates were pelleted by centrifugation at 8'000 g for 10 minutes and pellets were resuspended carefully in each 0.5-1 ml of SM buffer depending on how much volume was required for a given pellet. For plating of phage isolates, the suspensions were mixed with 500 µl of bacterial overnight culture and, after some minutes at room temperature, with 10 ml of top agar. In case of sewage plant inflow, serial dilutions of the resuspended precipitate were used due to the usually very high amount of phages. Top agars with samples and bacterial inoculum were then poured onto a square LB agar plate (ca. 12 cm x 12 cm). After solidification, we incubated the plates at 37°C for up to 24 hours.

Isolation of clonal bacteriophage isolates

After incubation, phages grew as plaques – i.e., spots of clearance – scattered across the top agar that was otherwise opaque from dense bacterial growth. To screen for phages that depended on restored O16-type O-antigen expression and / or showed impaired growth in presence of K-12 cryptic prophages, ca. 150 plaques per plate were separately picked with sterile toothpicks and copied onto freshly inoculated soft agars of hosts without O-antigen expression and / or cryptic prophages (see *Bacterial strains* above). When O-antigen dependency and / or inhibition by cryptic prophages were observed, the respective isolates were propagated at least three times via single plaques on top agars of their isolation host.

Composition of the completed BASEL collection

After phage isolation and initial screening, we sequenced the genomes of all ca. 100 isolates that depended on O16-type O-antigen or were inhibited by the K-12 cryptic prophages (see below). Based on

phylogenetic analyses, a set of ca. 40 phages was then chosen based primarily on largest genomic distances and characterized phenotypically alongside 8 phages from our previous work that grew well on *E. coli* K-12 with restored O-antigen expression (see below). We then chose 37 new phage isolates for inclusion in the completed BASEL collection by excluding a few more closely related isolates that behaved very similarly (see list in Table S2).

Efficiency of Plating assays

We quantified the infectivity of a phage on a host strain as the efficiency of plating (EOP) [177], i.e., as the ratio of its plaque formation on that host over the plaque formation on susceptible reference strains. As references we used *E. coli* K-12 MG1655 ΔRM *wbbL*(+) or *E. coli* K-12 BW251139 Δ9CP *wbbL*(+), depending on the isolates' sensitivity to presence of cryptic prophages, since all tested phages grew very well with restored O16-type O-antigen expression. All variants of *E. coli* K-12 used for phenotyping carried an F-plasmid without *pifA* defense gene (unless that was to be tested) and pBR322_ΔP*tet* as empty vector control or plasmids expressing different antiviral defenses (see *Bacterial strains* above and Table S3 for a list). Note that the phenotyping of phages sensitive to the presence of cryptic prophages enabled us to use the isogenic host with prophages as a test condition for efficiency of plating but prevented us from testing phage sensitivity to the EcoKI type I RM system because the *E. coli* K-12 BW251139 Δ9CP strain has active EcoKI methylation (see *Bacterial strains* above and the genotype in Table S3). It is therefore highlighted at all graphs whether ΔRM (EcoKI tested, cryptic prophage sensitivity not tested) or Δ9CP (EcoKI not tested, cryptic prophage sensitivity tested) have been used as hosts. To test for the effect of surface glycans on phage infectivity, we used host variants expressing no O-antigen but – unlike in our previous work [10] – no mutants with deeper truncations in the LPS core because almost all phages completely depended on O-antigen expression. The results with an *E. coli* K-12 mutant lacking NGR expression due to the inactivation of *nfrA* are merely shown in Fig. S1A since evidently no new phage isolate used this glycan for host recognition.

For these quantitative experiments, we prepared top agars on regular LB agar plates by overlaying them with a soft agar composed of LB agar containing only 0.5% agar and additionally 20 mM $MgSO_4$ as well as 5 mM $CaCl_2$ supplemented with a suitable bacterial inoculum. We used 3 ml of this soft agar with 100 µl of bacterial overnight culture as inoculum for regular round Petri dishes (ca. 9.4 cm diameter) and 10 ml of top agar supplemented with 200 µl of bacterial overnight culture for larger square Petri dishes (ca. 12 cm x 12 cm). EOP experiments were performed with serial dilutions of phage stocks grown on the respective reference hosts prepared in sterile phosphate-buffered saline (PBS). For each experiment, 2.5 µl of all serial dilutions were spotted on the top agar plates and dried into the top agar before incubation at 37°C for at least 24 hours. We then recorded plaque formation repeatedly throughout this time to accurately quantify the infectivity of fast- and slow-growing phages. Finally, we determined the EOP of each phage on each host as the ratio of plaques obtained on the experimental host over the number of plaques obtained on the respective reference strain [177]. Plaque counts and the calculation of EOP of all replicates of all experiments reported in this study (Figs. 5D, 6G, 7D, 8G, 9E, 10E, 13B, and S1A/B) as well as their summary statistics are compiled in S1 Data.

In cases when visual inspection could not unambiguously record plaque formation, we determined the EOP to be below the detection limit of 10^{-6} even if significant lysis from without was apparent (i.e., lysis zones caused by bacterial cell death without significant phage replication simply from the death of infected cells [123]). In cases where a phage / host pair was on the edge between strong lysis from without and very poor plaque formation, we cautiously recorded the result as an EOP below detection limit.

Qualitative top agar assays

We also determined the lysis host range of all tested phages on diverse *E. coli* strains (including K-12 ΔRM or Δ9CP with and without restored O-antigen) as well as *Salmonella* Typhimurium, i.e., the range of host strains on which a certain phage caused lysis zones irrespective of whether these were accompanied by plaque formation or not (Fig.13A). These experiments were performed as qualitative top agar assays as described in our previous work [10]. Briefly, top agar plates of each host strain were prepared

as described for efficiency-of-plating assays. After solidification, we spotted each 2.5 µl of undiluted high-titer stocks of all tested phage isolates (typically >10^9 pfu/ml) onto the top agar plates. When the spots had dried in, we incubated the plates at 37°C for at least 24 hours before visual inspection. The appearance of a lysis zone in more than half of all replicates for a phage-host pair meant that the host was counted as part of the lysis host range for this phage. In all such cases we subsequently quantified phage infectivity in efficiency-of-plating assays (see above).

Bacteriophage genome sequencing and assembly

Genomic DNA of phage isolates was prepared from culture lysates using the Norgen Biotek Phage DNA Isolation Kit and sequenced at SeqCenter (https://www.seqcenter.com/) or SeqCoast Genomics (https://seqcoast.com/) using Illumina technology. The short reads (151bp) were trimmed and assembled using the Geneious Assembler implemented in Geneious Prime 2024.0.2 with a coverage of typically 50-100x (see Table S2 for each phage). Circular contigs (indicating complete assembly of a viral chromosome due to the fusion of characteristic repeats at the genome ends [178]) were typically obtained easily using the "Low Sensitivity / Fast" setting with ca. 100'000 reads as input for the assembly. If necessary, ambiguities or incomplete assemblies were resolved by PCR amplification the high-fidelity polymerase Phusion (NEB) followed by Sanger Sequencing. Sequences were linearized at the 5' end either to the first nucleotide of the small or large terminase subunit gene or to the first position of the operon containing the small or large terminase subunit gene.

Initial bacteriophage genome annotation

The genomes of new phage isolates were annotated successively in several steps. To start, we used Pharokka to generate an automated, preliminary annotation of the genes in all phage genomes which included tRNA prediction using tRNAscan-SE [179, 180]. Subsequently, the annotations were refined manually based on identified orthologs in whole-genome alignments to different related phages generated using MAFFT v7.490 [181] implemented in Geneious Prime 2024.0.2. In parallel, we used the InterPro

101.0 protein domain signature database to maximize the functional information accessible from the primary sequence of phage proteins [114].

Curation of genome annotations using remote homologies and structure prediction

For the curation of phage genome annotations, we developed a computational pipeline called GAPS (Gene Annotation Pipeline based on sequence and structure Similarity). Briefly, this pipeline employs a combination of sequence and structure analyses, database searches, interpretation, annotation and visualization routines to infer functional information for a given set of genes, of any organism, in an automated way. For the current study, phage gene sequences were fed to GAPS as features in a .gb file and translated to protein. For each protein sequence: (i) the protein and domain families were determined by searching the Pfam, NCBI_CD, CATH_S40 and PHROGS databases with HHblits [182-187]; (ii) identical sequences were found by searching the PDB, SwissProt and RefSeq databases [188-190]; (iii) similar sequences were found by searching the UniRef30 and PDB70 databases with HHblits [186-189]; (iv) the protein structures were predicted with AlphaFold2 [191]; (v) similar structures were found by searching the PDB, AlphaFold-Proteome and AlphaFold-UniProt50 databases with Foldseek [188, 191, 192]. The results were parsed and integrated programmatically for presentation in an interactive report file that summarizes the available information for all genes of the input genome in one combined overview page. Detailed results of all individual genes are presented in sub-pages. An exemplary report for *Jilinvirus* HeidiAbel (Bas97) is included in this publication as Data S2 and the added value of GAPS for the functional annotation of its genes is highlighted in Fig. S13. GAPS is written in Python 3.8 and is open-source. The code used to run GAPS for generating the data presented in this manuscript is available on GitHub (https://github.com/hiller-lab/gaps) and on Zenodo (https://www.doi.org/10.5281/zenodo.14277981) in the form of a Jupyter notebook.

Bacteriophage naming and taxonomy

Phage isolates were classified according to applicable rules of the International Committee on the Taxonomy of Viruses (ICTV) [68, 193] largely as described in our previous work BASEL [10] (see Table

S2). Briefly, phages were first roughly assigned to viral family and, if possible, genus based on whole-genome blastn searches against the non-redundant nucleotide collection database [https://blast.ncbi.nlm.nih.gov/Blast.cgi; see also reference 194]. Subsequently, we unraveled the evolutionary relationships between our new isolates and reference sequences from NCBI GenBank across the diversity of relevant viral groups by calculating phylogenies based on either whole-genome alignments or alignments of conserved marker genes (see details in the dedicated section below). We then compared the clusters of viral sequences observed in these phylogenies to phage taxonomy as established by the ICTV and the relevant literature [68, 195]. For classification on the species level, whole-genome alignments of new phage isolates with close relatives were generated as described below and the nucleotide sequence identity was determined as the fraction of aligned genome sequence multiplied by the nucleotide identity of aligned segments. Phage isolates were classified as the same species as previously described relatives if their genomes showed >95% nucleotide sequence identity [196].

We named most phages in honor of scientists or historically significant persons linked to the city of Basel, Switzerland, where most of the phages had been isolated. To overcome the inherent gender imbalance due to historical biases, we included diverse female personalities from across the whole of Switzerland. Phages of *Skatevirus* and *Nonanavirus* genera were given names composed of common female first names in Switzerland in 1971 (when Swiss women obtained voting rights in federal elections) and "Muster" as surname (German word for paradigm or example).

Sequence alignments and phylogenetic analyses

Sequence alignments of bacteriophage genomes, specific loci, or certain genes / proteins were generated using MAFFT v7.490 implemented in Geneious Prime 2024.0.2 [181]. Occasionally, we supplemented poor or missing annotations in these sequences using the ORF finder tool of Geneious Prime 2024.0.2 guided by orthologous loci. Alignments were typically constructed using default settings of MAFFT with the fast FFT-NS-2 algorithm and 200PAM/k=2 or BLOSUM62 scoring matrices for

nucleotide and amino acid sequences, respectively. Subsequently, we curated the alignments manually and masked non-homologous parts for exclusion during phylogenetic analyses.

Maximum-Likelihood phylogenies were calculated using PhyML 3.3.20180621 implemented in Geneious Prime 2024.0.2 from sequence alignments of orthologous genes / proteins or orthologous sequence stretches of different bacteriophage genomes [197]. We used the HYK85 and LG substitution models for nucleotide and amino acid sequences, respectively, and always calculated 100 bootstraps. Phylogenetic relationships between phage genomes were inferred from whole-genome alignments (whenever possible) or from concatenated sequence alignments of conserved core genes on nucleotide or amino acid level as indicated for each phylogeny. Relevant reference sequences and controls were selected to represent the diversity of taxonomic units or protein families and typically identified by BLAST searches against the non-redundant nucleotide collection database [https://blast.ncbi.nlm.nih.gov/Blast.cgi; see also reference 194]. When available, well-studied model phages such as P22 and their sequences were always included to provide intuitive reference points for data interpretation. The tree files and underlying sequence alignments in NEXUS format are included in this manuscript as Data S3 file.

Structure predictions and analyses of receptor-binding proteins

Structure predictions were performed using AlphaFold-Multimer on the Cosmic^2 platform or on the AlphaFold3 webserver [198-200]. All predictions presented high internal confidence scores which are presented in Table S7. Illustrations were generated using the PyMOL Molecular Graphics System (Version 2.4.1, Schrodinger LLC). The Dali server was used to identify structural homologs in the Protein Data Bank (PDB) [201].

Morphological analyses by transmission electron microscopy

We analyzed the virion morphology of representative isolates from the different taxonomic groups covered in this manuscript by negative-stain transmission electron microscopy [202]. Briefly, we applied 5 µl drops of high-titer phage lysates to 400 mesh carbon-coated grids that had been rendered hydrophilic using a glow-discharger at low vacuum conditions. Subsequently, virions were stained on 5 µl drops of 2%

(w/v) uranyl acetate and examined using an FEI Tecnai G2 Spirit transmission electron microscope (FEI Company) at 80-kV accelerating voltage, a Hitachi HT7700 transmission electron microscope (Hitachi) at 100 kV, or an FEI Talos F200C transmission electron microscope (FEI Company) at 200 kV. Images were recorded with a side-mounted Olympus Veleta CCD camera 4k (Tecnai G2 Spirit), an AMT XR81B Peltier cooled CCD camera (8 M pixel; Hitachi HT7700), or a Thermo Fisher Scientific Ceta 16M CCD camera (Talos F200C) using EMSIS RADIUS software at a nominal magnification of typically 150'000x.

Quantification and statistical analysis

Data sets were analyzed quantitatively by calculating mean and standard deviation of at least three independent biological replicates per experiment. Detailed information about replicates and statistical analyses is provided in the figure legends. In addition, we determined the probability for all efficiency-of-plating values that they are significantly different from the *E. coli* K-12 vector control using Student's T-test (two-sample unequal variance / heteroscedastic) with a result considered significant if $p<0.01$ (included in Data S1).

Acknowledgments

The authors are grateful to Dr. Julie Sollier and Valerie Oriet for valuable input and critical reading of the manuscript. We thank high school students participating in the Biozentrum Summer Science Academy 2022 for assistance with the isolation of multiple bacteriophages described in this study (Table S2). Carmen Saez Garcia Wanzenried is acknowledged for her invaluable help with replicating the BASEL phages and for naming phage CurroJimenez (Bas106). The authors are indebted to the ScopeM facility at ETH Zürich (Switzerland) and the team at the BioEM lab of Biozentrum, University of Basel (Switzerland) for their support with transmission electron microscopy. We further thank ARA Basel, ARA Canius (Lenzerheide), and ARA Triengen for samples of sewage plant inflow. The authors are grateful to David Schwegler for pioneering work on the annotation of phage genomes using remote homologies and

structure prediction with GAPS. Dr. Toomas Mets (University of Tartu, Estonia) is acknowledged for the initial discovery that partial restoration of O16-type O-antigen expression with pAS001 can still inhibit the growth of some phages. The authors are grateful to Dr. Jiemin Du, Dr. Samuel Kilcher, and Prof. Dr. Martin Loessner (ETH Zürich, Switzerland) for helpful discussions about the impact of bacterial O-antigen on phage host range. Dr. Artem Isaev is acknowledged for a helpful comment on anti-phage defense systems in cryptic prophages. We are grateful to the Coli Genetic Stock Center (CGSC) for the *E. coli* K-12 EMG2 strain and to the German Collection of Microorganisms and Cell Cultures (DSMZ) for phage lambda PaPa. This work was supported by the Swiss National Science Foundation (SNSF) Ambizione Fellowship PZ00P3_180085 (to A.H.), SNSF Starting Grant TMSGI3_211369 (to A.H.), and SNSF National Centre of Competence in Research (NCCR) AntiResist (grant number 180541).

References

1. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. Nat Rev Microbiol. 2020;18(3):125-38. Epub 2020/02/06. doi: 10.1038/s41579-019-0311-5. PubMed PMID: 32015529.

2. Cobian Güemes AG, Youle M, Cantu VA, Felts B, Nulton J, Rohwer F. Viruses as Winners in the Game of Life. Annu Rev Virol. 2016;3(1):197-214. Epub 2016/10/16. doi: 10.1146/annurev-virology-100114-054952. PubMed PMID: 27741409.

3. Ofir G, Sorek R. Contemporary Phage Biology: From Classic Models to New Insights. Cell. 2018;172(6):1260-70. doi: 10.1016/j.cell.2017.10.045. PubMed PMID: 29522746.

4. Gordillo Altamirano FL, Barr JJ. Phage Therapy in the Postantibiotic Era. Clin Microbiol Rev. 2019;32(2). Epub 2019/01/18. doi: 10.1128/CMR.00066-18. PubMed PMID: 30651225; PubMed Central PMCID: PMCPMC6431132.

5. Kolter R. Small Things Considered [Internet]. Schaechter M, Kolter R, editors. https://schaechter.asmblog.org/schaechter/2022/03/the-phage-treaty.html2022. [cited 2022]. Available from: https://schaechter.asmblog.org/schaechter/2022/03/the-phage-treaty.html.

6. Shaidullina A, Harms A. Toothpicks, logic, and next-generation sequencing: systematic investigation of bacteriophage-host interactions. Curr Opin Microbiol. 2022;70:102225. Epub 20221031. doi: 10.1016/j.mib.2022.102225. PubMed PMID: 36327691.

7. Demerec M, Fano U. Bacteriophage-Resistant Mutants in *Escherichia Coli*. Genetics. 1945;30(2):119-36. Epub 1945/03/01. PubMed PMID: 17247150; PubMed Central PMCID: PMCPMC1209279.

8. Abedon ST. The murky origin of Snow White and her T-even dwarfs. Genetics. 2000;155(2):481-6. Epub 2000/06/03. PubMed PMID: 10835374; PubMed Central PMCID: PMCPMC1461100.

9. Ruiz N, Silhavy TJ. How *Escherichia coli* Became the Flagship Bacterium of Molecular Biology. J Bacteriol. 2022;204(9):e0023022. Epub 20220802. doi: 10.1128/jb.00230-22. PubMed PMID: 35916528; PubMed Central PMCID: PMCPMC9487582.

10. Maffei E, Shaidullina A, Burkolter M, Heyer Y, Estermann F, Druelle V, et al. Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. PLoS Biol. 2021;19(11):e3001424. Epub 2021/11/17. doi: 10.1371/journal.pbio.3001424. PubMed PMID: 34784345; PubMed Central PMCID: PMCPMC8594841.

11. Kelly A, Went SC, Mariano G, Shaw LP, Picton DM, Duffner SJ, et al. Diverse Durham collection phages demonstrate complex BREX defense responses. Appl Environ Microbiol. 2023;89(9):e0062323. Epub 20230905. doi: 10.1128/aem.00623-23. PubMed PMID: 37668405; PubMed Central PMCID: PMCPMC10537673.

12. Nicolas M, Trotereau A, Culot A, Moodley A, Atterbury R, Wagemans J, et al. Isolation and Characterization of a Novel Phage Collection against Avian-Pathogenic *Escherichia coli*. Microbiol Spectr. 2023;11(3):e0429622. Epub 20230504. doi: 10.1128/spectrum.04296-22. PubMed PMID: 37140373; PubMed Central PMCID: PMCPMC10269720.

13. Rothschild-Rodriguez D, Lambon KS, Latka A, Costa AR, Mantzouratou A, King C, et al. KlebPhaCol: A community-driven resource for *Klebsiella* research identified a novel gut phage order associated with the human gut. bioRxiv. 2024:2024.12.02.626457. doi: 10.1101/2024.12.02.626457.

14. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, et al. Prophages mediate defense against phage infection through diverse mechanisms. ISME J. 2016;10(12):2854-66. Epub 2016/06/04. doi: 10.1038/ismej.2016.79. PubMed PMID: 27258950; PubMed Central PMCID: PMCPMC5148200.

15. Brauer A, Rosendahl S, Kangsep A, Lewanczyk AC, Rikberg R, Horak R, et al. Isolation and characterization of a phage collection against *Pseudomonas putida*. Environ Microbiol. 2024;26(6):e16671. doi: 10.1111/1462-2920.16671. PubMed PMID: 38863081.

16. Costa AR, van den Berg DF, Esser JQ, Muralidharan A, van den Bossche H, Bonilla BE, et al. Accumulation of defense systems in phage-resistant strains of *Pseudomonas aeruginosa*. Sci Adv. 2024;10(8):eadj0341. Epub 20240223. doi: 10.1126/sciadv.adj0341. PubMed PMID: 38394193; PubMed Central PMCID: PMCPMC10889362.

17. Sellner B, Prakapaite R, van Berkum M, Heinemann M, Harms A, Jenal U. A New Sugar for an Old Phage: a c-di-GMP-Dependent Polysaccharide Pathway Sensitizes *Escherichia coli* for Bacteriophage Infection. mBio. 2021;12(6):e0324621. Epub 2021/12/15. doi: 10.1128/mbio.03246-21. PubMed PMID: 34903045; PubMed Central PMCID: PMCPMC8669472.

18. Nyerges A, Vinke S, Flynn R, Owen SV, Rand EA, Budnik B, et al. A swapped genetic code prevents viral infections and gene transfer. Nature. 2023;615(7953):720-7. Epub 20230315. doi: 10.1038/s41586-023-05824-z. PubMed PMID: 36922599; PubMed Central PMCID: PMCPMC10151025.

19. Hantke K, Friz S. The TonB-dependent uptake of pyrroloquinoline-quinone (PQQ) and secretion of gluconate by *Escherichia coli* K-12. Mol Microbiol. 2022;118(4):417-25. Epub 20220914. doi: 10.1111/mmi.14975. PubMed PMID: 36054785.

20. Mulla Y, Müller J, Trimcev D, Bollenbach T. Extreme diversity of phage amplification rates and phage-antibiotic interactions revealed by PHORCE. bioRxiv. 2024:2024.06.07.597930. doi: 10.1101/2024.06.07.597930.

21. Zhang T, Tamman H, Coppieters 't Wallant K, Kurata T, LeRoux M, Srikant S, et al. Direct activation of a bacterial innate immune system by a viral capsid protein. Nature. 2022;612(7938):132-40. Epub 20221116. doi: 10.1038/s41586-022-05444-z. PubMed PMID: 36385533; PubMed Central PMCID: PMCPMC9712102.

22. Zhong A, Jiang X, Hickman AB, Klier K, Teodoro GIC, Dyda F, et al. Toxic antiphage defense proteins inhibited by intragenic antitoxin proteins. Proc Natl Acad Sci U S A. 2023;120(31):e2307382120. Epub 20230724. doi: 10.1073/pnas.2307382120. PubMed PMID: 37487082; PubMed Central PMCID: PMCPMC10400941.

23. Skutel M, Andriianov A, Zavialova M, Kirsanova M, Shodunke O, Zorin E, et al. T5-like phage BF23 evades host-mediated DNA restriction and methylation. microLife. 2023;4. doi: 10.1093/femsml/uqad044.

24. Hu H, Popp PF, Hughes TCD, Roa-Eguiara A, Rutbeek NR, Martin FJO, et al. Structure and mechanism of the Zorya anti-phage defense system. Nature. 2024. Epub 20241211. doi: 10.1038/s41586-024-08493-8. PubMed PMID: 39662505.

25. Vizzarro G, Lemopoulos A, Adams DW, Blokesch M. *Vibrio cholerae* pathogenicity island 2 encodes two distinct types of restriction systems. J Bacteriol. 2024:e0014524. Epub 20240812. doi: 10.1128/jb.00145-24. PubMed PMID: 39133004.

26. Blount ZD. The unexhausted potential of *E. coli*. Elife. 2015;4. doi: 10.7554/eLife.05826. PubMed PMID: 25807083; PubMed Central PMCID: PMCPMC4373459.

27. Liu D, Reeves PR. *Escherichia* coli K12 regains its O antigen. Microbiology. 1994;140 (Pt 1):49-57. doi: 10.1099/13500872-140-1-49. PubMed PMID: 7512872.

28. van der Ley P, de Graaff P, Tommassen J. Shielding of *Escherichia coli* outer membrane proteins as receptors for bacteriophages and colicins by O-antigenic chains of lipopolysaccharide. J Bacteriol. 1986;168(1):449-51. doi: 10.1128/jb.168.1.449-451.1986. PubMed PMID: 2428812; PubMed Central PMCID: PMCPMC213476.

29. Letarov AV. Bacterial Virus Forcing of Bacterial O-Antigen Shields: Lessons from Coliphages. Int J Mol Sci. 2023;24(24). Epub 20231212. doi: 10.3390/ijms242417390. PubMed PMID: 38139217; PubMed Central PMCID: PMCPMC10743462.

30. Berryhill BA, Burke KB, Fontaine J, Brink CE, Harvill MG, Goldberg DA, et al. Enteric Populations of *Escherichia coli* are Likely to be Resistant to Phages Due to O Antigen Expression. bioRxiv. 2024:2023.11.08.566299. doi: 10.1101/2023.11.08.566299.

31. Broeker NK, Barbirz S. Not a barrier but a key: How bacteriophages exploit host's O-antigen as an essential receptor to initiate infection. Mol Microbiol. 2017;105(3):353-7. Epub 2017/06/16. doi: 10.1111/mmi.13729. PubMed PMID: 28618013.

32. Henderson JC, Zimmerman SM, Crofts AA, Boll JM, Kuhns LG, Herrera CM, et al. The Power of Asymmetry: Architecture and Assembly of the Gram-Negative Outer Membrane Lipid Bilayer. Annu Rev Microbiol. 2016;70:255-78. Epub 2016/07/01. doi: 10.1146/annurev-micro-102215-095308. PubMed PMID: 27359214.

33. Pas C, Latka A, Fieseler L, Briers Y. Phage tailspike modularity and horizontal gene transfer reveals specificity towards *E. coli* O-antigen serogroups. Virol J. 2023;20(1):174. Epub 20230807. doi: 10.1186/s12985-023-02138-4. PubMed PMID: 37550759; PubMed Central PMCID: PMCPMC10408124.

34. Iglesias SM, Li F, Briani F, Cingolani G. Viral Genome Delivery Across Bacterial Cell Surfaces. Annu Rev Microbiol. 2024. Epub 20240709. doi: 10.1146/annurev-micro-041222-124727. PubMed PMID: 38986128.

35. Gaborieau B, Vaysset H, Tesson F, Charachon I, Dib N, Bernier J, et al. Prediction of strain level phage-host interactions across the *Escherichia* genus using only genomic information. Nat Microbiol. 2024;9(11):2847-61. Epub 20241031. doi: 10.1038/s41564-024-01832-5. PubMed PMID: 39482383.

36. Whitfield C, Williams DM, Kelly SD. Lipopolysaccharide O-antigens-bacterial glycans made to measure. J Biol Chem. 2020;295(31):10593-609. Epub 20200518. doi: 10.1074/jbc.REV120.009402. PubMed PMID: 32424042; PubMed Central PMCID: PMCPMC7397119.

37. Islam MZ, Fokine A, Mahalingam M, Zhang Z, Garcia-Doval C, van Raaij MJ, et al. Molecular anatomy of the receptor binding module of a bacteriophage long tail fiber. PLoS Pathog. 2019;15(12):e1008193. Epub 20191219. doi: 10.1371/journal.ppat.1008193. PubMed PMID: 31856258; PubMed Central PMCID: PMCPMC6957217.

38. Zivanovic Y, Confalonieri F, Ponchon L, Lurz R, Chami M, Flayhan A, et al. Insights into bacteriophage T5 structure from analysis of its morphogenesis genes and protein components. J Virol. 2014;88(2):1162-74. Epub 20131106. doi: 10.1128/JVI.02262-13. PubMed PMID: 24198424; PubMed Central PMCID: PMCPMC3911642.

39. Steven AC, Trus BL, Maizel JV, Unser M, Parry DA, Wall JS, et al. Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. J Mol Biol. 1988;200(2):351-65. doi: 10.1016/0022-2836(88)90246-x. PubMed PMID: 3259634.

40. Georjon H, Bernheim A. The highly diverse antiphage defence systems of bacteria. Nat Rev Microbiol. 2023;21(10):686-700. Epub 20230717. doi: 10.1038/s41579-023-00934-x. PubMed PMID: 37460672.

41. Lopatina A, Tal N, Sorek R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annu Rev Virol. 2020;7(1):371-84. Epub 2020/06/20. doi: 10.1146/annurev-virology-011620-040628. PubMed PMID: 32559405.

42. Yirmiya E, Leavitt A, Lu A, Ragucci AE, Avraham C, Osterman I, et al. Phages overcome bacterial immunity via diverse anti-defence proteins. Nature. 2023. Epub 20231122. doi: 10.1038/s41586-023-06869-w. PubMed PMID: 37992756.

43. Samson JE, Magadan AH, Sabri M, Moineau S. Revenge of the phages: defeating bacterial defences. Nat Rev Microbiol. 2013;11(10):675-87. Epub 2013/08/28. doi: 10.1038/nrmicro3096. PubMed PMID: 23979432.

44. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. Science. 2018;359(6379). Epub 2018/01/27. doi: 10.1126/science.aar4120. PubMed PMID: 29371424; PubMed Central PMCID: PMCPMC6387622.

45. Mayo-Munoz D, Pinilla-Redondo R, Camara-Wilpert S, Birkholz N, Fineran PC. Inhibitors of bacterial immune systems: discovery, mechanisms and applications. Nat Rev Genet. 2024;25(4):237-54. Epub 20240130. doi: 10.1038/s41576-023-00676-9. PubMed PMID: 38291236.

46. Tesson F, Herve A, Mordret E, Touchon M, d'Humieres C, Cury J, et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun. 2022;13(1):2561. Epub 2022/05/11. doi: 10.1038/s41467-022-30269-9. PubMed PMID: 35538097; PubMed Central PMCID: PMCPMC9090908.

47. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, et al. Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. Nucleic Acids Res. 2021;49(19):10868-78. Epub 2021/10/05. doi: 10.1093/nar/gkab883. PubMed PMID: 34606606; PubMed Central PMCID: PMCPMC8565338.

48. Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT. A functional selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome. Nat Microbiol. 2022;7(10):1568-79. Epub 20220919. doi: 10.1038/s41564-022-01219-4. PubMed PMID: 36123438; PubMed Central PMCID: PMCPMC9519451.

49. Beamud B, Benz F, Bikard D. Going viral: The role of mobile genetic elements in bacterial immunity. Cell Host Microbe. 2024;32(6):804-19. doi: 10.1016/j.chom.2024.05.017. PubMed PMID: 38870898.

50. Lopatina A, Ferenc M, Bartlau N, Wolfram M, Steensen K, Muscia S, et al. Interpretable machine learning reveals a diverse arsenal of anti-defenses in bacterial viruses. bioRxiv. 2024:2024.06.14.598830. doi: 10.1101/2024.06.14.598830.

51. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, et al. Cryptic prophages help bacteria cope with adverse environments. Nat Commun. 2010;1:147. doi: 10.1038/ncomms1146. PubMed PMID: 21266997; PubMed Central PMCID: PMCPMC3105296.

52. Olsen NS, Forero-Junco L, Kot W, Hansen LH. Exploring the Remarkable Diversity of Culturable *Escherichia coli* Phages in the Danish Wastewater Environment. Viruses. 2020;12(9). Epub 2020/09/10. doi: 10.3390/v12090986. PubMed PMID: 32899836; PubMed Central PMCID: PMCPMC7552041.

53. Michniewski S, Redgwell T, Grigonyte A, Rihtman B, Aguilo-Ferretjans M, Christie-Oleza J, et al. Riding the wave of genomics to investigate aquatic coliphage diversity and activity. Environ Microbiol. 2019;21(6):2112-28. Epub 2019/03/19. doi: 10.1111/1462-2920.14590. PubMed PMID: 30884081; PubMed Central PMCID: PMCPMC6563131.

54. Mutalik VK, Arkin AP. A Phage Foundry Framework to Systematically Develop Viral Countermeasures to Combat Antibiotic-Resistant Bacterial Pathogens. iScience. 2022;25(4):104121. Epub 2022/04/12. doi: 10.1016/j.isci.2022.104121. PubMed PMID: 35402883; PubMed Central PMCID: PMCPMC8983348.

55. Venturini C, Petrovic Fabijan A, Fajardo Lubian A, Barbirz S, Iredell J. Biological foundations of successful bacteriophage therapy. EMBO Mol Med. 2022;n/a(n/a):e12435. Epub 2022/05/28. doi: 10.15252/emmm.202012435. PubMed PMID: 35620963.

56. Lood C, Boeckaerts D, Stock M, De Baets B, Lavigne R, van Noort V, et al. Digital phagograms: predicting phage infectivity through a multilayer machine learning approach. Curr Opin Virol. 2022;52:174-81. Epub 20211221. doi: 10.1016/j.coviro.2021.12.004. PubMed PMID: 34952265.

57. Siborova M, Fuzik T, Prochazkova M, Novacek J, Benesik M, Nilsson AS, et al. Tail proteins of phage SU10 reorganize into the nozzle for genome delivery. Nat Commun. 2022;13(1):5622. Epub 20220924. doi: 10.1038/s41467-022-33305-w. PubMed PMID: 36153309; PubMed Central PMCID: PMCPMC9509320.

58. Korf IHE, Meier-Kolthoff JP, Adriaenssens EM, Kropinski AM, Nimtz M, Rohde M, et al. Still Something to Discover: Novel Insights into *Escherichia coli* Phage Diversity and Taxonomy. Viruses. 2019;11(5). doi: 10.3390/v11050454. PubMed PMID: 31109012; PubMed Central PMCID: PMCPMC6563267.

59. Vitt AR, Sørensen AN, Bojer MS, Bortolaia V, Sørensen MCH, Brøndsted L. Diverse bacteriophages for biocontrol of ESBL- and AmpC-β-lactamase-producing *E. coli*. iScience. 2024;27(2):108826. doi: https://doi.org/10.1016/j.isci.2024.108826.

60. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiol Lett. 2016;363(4). Epub 2016/01/13. doi: 10.1093/femsle/fnw002. PubMed PMID: 26755501.

61. Hantke K. Compilation of *Escherichia coli* K-12 outer membrane phage receptors - their function and some historical remarks. FEMS Microbiol Lett. 2020;367(2). Epub 2020/02/06. doi: 10.1093/femsle/fnaa013. PubMed PMID: 32009155.

62. Du J, Meile S, Baggenstos J, Jaggi T, Piffaretti P, Hunold L, et al. Enhancing bacteriophage therapeutics through in situ production and release of heterologous antimicrobial effectors. Nat Commun. 2023;14(1):4337. Epub 20230720. doi: 10.1038/s41467-023-39612-0. PubMed PMID: 37474516; PubMed Central PMCID: PMCPMC10359290.

63. Mohammed M, Casjens SR, Millard AD, Harrison C, Gannon L, Chattaway MA. Genomic analysis of Anderson typing phages of *Salmonella* Typhimrium [sic]: towards understanding the basis of bacteria-phage interaction. Sci Rep. 2023;13(1):10484. Epub 20230628. doi: 10.1038/s41598-023-37307-6. PubMed PMID: 37380724; PubMed Central PMCID: PMCPMC10307801.

64. Zeng C, Gilcrease EB, Hendrix RW, Xie Y, Jalfon MJ, Gill JJ, et al. DNA Packaging and Genomics of the *Salmonella* 9NA-Like Phages. J Virol. 2019;93(22). Epub 20191029. doi: 10.1128/JVI.00848-19. PubMed PMID: 31462565; PubMed Central PMCID: PMCPMC6819911.

65. Hutinet G, Kot W, Cui L, Hillebrand R, Balamkundu S, Gnanakalai S, et al. 7-Deazaguanine modifications protect phage DNA from host restriction systems. Nat Commun. 2019;10(1):5442. Epub 2019/12/01. doi: 10.1038/s41467-019-13384-y. PubMed PMID: 31784519; PubMed Central PMCID: PMCPMC6884629.

66. Khan Mirzaei M, Eriksson H, Kasuga K, Haggard-Ljungquist E, Nilsson AS. Genomic, proteomic, morphological, and phylogenetic analyses of vB_EcoP_SU10, a podoviridae phage with C3 morphology. PLoS One. 2014;9(12):e116294. Epub 20141231. doi: 10.1371/journal.pone.0116294. PubMed PMID: 25551446; PubMed Central PMCID: PMCPMC4281155.

67. Batinovic S, Fujii Y, Nittami T. Expansion of *Kuravirus*-like Phage Sequences within the Past Decade, including *Escherichia* Phage YF01 from Japan, Prompt the Creation of Three New Genera. Viruses. 2023;15(2). Epub 20230211. doi: 10.3390/v15020506. PubMed PMID: 36851720; PubMed Central PMCID: PMCPMC9965538.

68. Zerbini FM, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, et al. Changes to virus taxonomy and the ICTV Statutes ratified by the International Committee on Taxonomy of Viruses (2023). Arch Virol. 2023;168(7):175. Epub 20230610. doi: 10.1007/s00705-023-05797-4. PubMed PMID: 37296227; PubMed Central PMCID: PMCPMC10861154.

69. Zulk JJ, Clark JR, Ottinger S, Ballard MB, Mejia ME, Mercado-Evans V, et al. Phage Resistance Accompanies Reduced Fitness of Uropathogenic *Escherichia coli* in the Urinary Environment. mSphere. 2022;7(4):e0034522. Epub 20220803. doi: 10.1128/msphere.00345-22. PubMed PMID: 35920561; PubMed Central PMCID: PMCPMC9429881.

70. Gonzalez-Garcia VA, Pulido-Cid M, Garcia-Doval C, Bocanegra R, van Raaij MJ, Martin-Benito J, et al. Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor. J Biol Chem. 2015;290(16):10038-44. Epub 2015/02/24. doi: 10.1074/jbc.M114.614222. PubMed PMID: 25697363; PubMed Central PMCID: PMCPMC4400320.

71. Hu B, Margolin W, Molineux IJ, Liu J. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science. 2013;339(6119):576-9. Epub 2013/01/12. doi: 10.1126/science.1231887. PubMed PMID: 23306440; PubMed Central PMCID: PMCPMC3873743.

72. Trojet SN, Caumont-Sarcos A, Perrody E, Comeau AM, Krisch HM. The gp38 adhesins of the T4 superfamily: a complex modular determinant of the phage's host specificity. Genome Biol Evol.

2011;3:674-86. Epub 2011/07/13. doi: 10.1093/gbe/evr059. PubMed PMID: 21746838; PubMed Central PMCID: PMCPMC3157838.

73. Washizaki A, Yonesaki T, Otsuka Y. Characterization of the interactions between *Escherichia coli* receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. Microbiologyopen. 2016;5(6):1003-15. Epub 2016/06/09. doi: 10.1002/mbo3.384. PubMed PMID: 27273222; PubMed Central PMCID: PMCPMC5221442.

74. Efimov AD, Golomidova AK, Kulikov EE, Belalov IS, Ivanov PA, Letarov AV. RB49-like Bacteriophages Recognize O Antigens as One of the Alternative Primary Receptors. Int J Mol Sci. 2022;23(19). Epub 20220926. doi: 10.3390/ijms231911329. PubMed PMID: 36232640; PubMed Central PMCID: PMCPMC9569957.

75. Weigele P, Raleigh EA. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. Chem Rev. 2016;116(20):12655-87. Epub 2016/06/21. doi: 10.1021/acs.chemrev.6b00114. PubMed PMID: 27319741.

76. Sidi Mabrouk A, Ongenae V, Claessen D, Brenzinger S, Briegel A. A Flexible and Efficient Microfluidics Platform for the Characterization and Isolation of Novel Bacteriophages. Appl Environ Microbiol. 2023;89(1):e0159622. Epub 20230105. doi: 10.1128/aem.01596-22. PubMed PMID: 36602353; PubMed Central PMCID: PMCPMC9888219.

77. Adler BA, Al-Shimary MJ, Patel JR, Armbruster E, Colognori D, Charles EJ, et al. Genome-wide Characterization of Diverse Bacteriophages Enabled by RNA-Binding CRISPRi. bioRxiv. 2023:2023.09.18.558157. doi: 10.1101/2023.09.18.558157.

78. Prokhorov NS, Riccio C, Zdorovenko EL, Shneider MM, Browning C, Knirel YA, et al. Function of bacteriophage G7C esterase tailspike in host cell adsorption. Mol Microbiol. 2017;105(3):385-98. Epub 2017/05/18. doi: 10.1111/mmi.13710. PubMed PMID: 28513100.

79. Lv M, Wang S, Yan G, Sun C, Feng X, Gu J, et al. Genome sequencing and analysis of an *Escherichia coli* phage vB_EcoM-ep3 with a novel lysin, Lysep3. Virus Genes. 2015;50(3):487-97. Epub 20150405. doi: 10.1007/s11262-015-1195-8. PubMed PMID: 25842152.

80. Santos TM, Bicalho RC. Complete genome sequence of vB_EcoM_ECO1230-10: a coliphage with therapeutic potential for bovine metritis. Vet Microbiol. 2011;148(2-4):267-75. Epub 20100908. doi: 10.1016/j.vetmic.2010.08.034. PubMed PMID: 20875932.

81. Tikhe CV, Gissendanner CR, Husseneder C. Whole-Genome Sequence of the Novel *Enterobacter* Bacteriophage Arya with an Integrase Pseudogene, Isolated from the Gut of the Formosan Subterranean Termite. Genome Announc. 2018;6(1). Epub 20180104. doi: 10.1128/genomeA.00838-17. PubMed PMID: 29301894; PubMed Central PMCID: PMCPMC5754474.

82. Kawato Y, Yasuike M, Nakamura Y, Shigenobu Y, Fujiwara A, Sano M, et al. Complete genome sequence analysis of two *Pseudomonas plecoglossicida* phages, potential therapeutic agents. Appl Environ Microbiol. 2015;81(3):874-81. Epub 20141121. doi: 10.1128/AEM.03038-14. PubMed PMID: 25416766; PubMed Central PMCID: PMCPMC4292507.

83. Boeckman J, Korn A, Yao G, Ravindran A, Gonzalez C, Gill J. Sheep in wolves' clothing: Temperate T7-like bacteriophages and the origins of the *Autographiviridae*. Virology. 2022;568:86-100. Epub 20220131. doi: 10.1016/j.virol.2022.01.013. PubMed PMID: 35149347.

84. Onallah H, Hazan R, Nir-Paz R, group Ps, Brownstein MJ, Fackler JR, et al. Refractory *Pseudomonas aeruginosa* infections treated with phage PASA16: A compassionate use case series. Med. 2023;4(9):600-11 e4. Epub 20230809. doi: 10.1016/j.medj.2023.07.002. PubMed PMID: 37562400.

85. Pirnay JP, Djebara S, Steurs G, Griselain J, Cochez C, De Soir S, et al. Personalized bacteriophage therapy outcomes for 100 consecutive cases: a multicentre, multinational, retrospective observational study. Nat Microbiol. 2024;9(6):1434-53. Epub 20240604. doi: 10.1038/s41564-024-01705-x. PubMed PMID: 38834776; PubMed Central PMCID: PMCPMC11153159.

86. Li F, Hou CD, Lokareddy RK, Yang R, Forti F, Briani F, et al. High-resolution cryo-EM structure of the *Pseudomonas* bacteriophage E217. Nat Commun. 2023;14(1):4052. Epub 20230708. doi: 10.1038/s41467-023-39756-z. PubMed PMID: 37422479; PubMed Central PMCID: PMCPMC10329688.

87. Necel A, Bloch S, Nejman-Falenczyk B, Grabski M, Topka G, Dydecka A, et al. Characterization of a bacteriophage, vB_Eco4M-7, that effectively infects many *Escherichia coli* O157 strains. Sci Rep. 2020;10(1):3743. Epub 20200228. doi: 10.1038/s41598-020-60568-4. PubMed PMID: 32111934; PubMed Central PMCID: PMCPMC7048862.

88. Heineman RH, Bull JJ. Testing optimality with experimental evolution: lysis time in a bacteriophage. Evolution. 2007;61(7):1695-709. doi: 10.1111/j.1558-5646.2007.00132.x. PubMed PMID: 17598749; PubMed Central PMCID: PMCPMC1974807.

89. Maffei E, Woischnig A-K, Burkolter MR, Heyer Y, Humolli D, Thürkauf N, et al. Phage Paride can kill dormant, antibiotic-tolerant cells of Pseudomonas aeruginosa by direct lytic replication. Nature Communications. 2024;15(1):175. doi: 10.1038/s41467-023-44157-3.

90. Kropinski AM, Chan L, Jarrell K, Milazzo FH. The nature of *Pseudomonas aeruginosa* strain PAO bacteriophage receptors. Can J Microbiol. 1977;23(6):653-8. doi: 10.1139/m77-098. PubMed PMID: 406024.

91. Mutalik VK, Adler BA, Rishi HS, Piya D, Zhong C, Koskella B, et al. High-throughput mapping of the phage resistance landscape in *E. coli*. PLoS Biol. 2020;18(10):e3000877. Epub 2020/10/14. doi: 10.1371/journal.pbio.3000877. PubMed PMID: 33048924; PubMed Central PMCID: PMCPMC7553319 following competing interests: VKM, AMD, and APA consult for and hold equity in Felix Biotechnology, Inc.

92. Kiljunen S, Vilen H, Pajunen M, Savilahti H, Skurnik M. Nonessential genes of phage phiYeO3-12 include genes involved in adaptation to growth on *Yersinia enterocolitica* serotype O:3. J Bacteriol. 2005;187(4):1405-14. Epub 2005/02/03. doi: 10.1128/JB.187.4.1405-1414.2005. PubMed PMID: 15687205; PubMed Central PMCID: PMCPMC545605.

93. Adler BA, Kazakov AE, Zhong C, Liu H, Kutter E, Lui LM, et al. The genetic basis of phage susceptibility, cross-resistance and host-range in *Salmonella*. Microbiology (Reading). 2021;167(12). Epub 2021/12/16. doi: 10.1099/mic.0.001126. PubMed PMID: 34910616; PubMed Central PMCID: PMCPMC8744999.

94. Tu J, Park T, Morado DR, Hughes KT, Molineux IJ, Liu J. Dual host specificity of phage SP6 is facilitated by tailspike rotation. Virology. 2017;507:206-15. Epub 20170426. doi: 10.1016/j.virol.2017.04.017. PubMed PMID: 28456019; PubMed Central PMCID: PMCPMC5503682.

95. Scholl D, Merril C. The genome of bacteriophage K1F, a T7-like phage that has acquired the ability to replicate on K1 strains of *Escherichia coli*. J Bacteriol. 2005;187(24):8499-503. Epub 2005/12/03. doi:

10.1128/JB.187.24.8499-8503.2005. PubMed PMID: 16321955; PubMed Central PMCID: PMCPMC1317022.

96. Lourenco M, Chaffringeon L, Lamy-Besnier Q, Titecat M, Pedron T, Sismeiro O, et al. The gut environment regulates bacterial gene expression which modulates susceptibility to bacteriophage infection. Cell Host Microbe. 2022;30(4):556-69 e5. Epub 2022/04/15. doi: 10.1016/j.chom.2022.03.014. PubMed PMID: 35421351.

97. Molineux IJ, Spence JL. Virus-plasmid interactions: mutants of bacteriophage T3 that abortively infect plasmid F-containing (F+) strains of *Escherichia coli*. Proc Natl Acad Sci U S A. 1984;81(5):1465-9. Epub 1984/03/01. doi: 10.1073/pnas.81.5.1465. PubMed PMID: 6324192; PubMed Central PMCID: PMCPMC344857.

98. Schmitt CK, Molineux IJ. Expression of gene 1.2 and gene 10 of bacteriophage T7 is lethal to F plasmid-containing *Escherichia coli*. J Bacteriol. 1991;173(4):1536-43. Epub 1991/02/01. doi: 10.1128/jb.173.4.1536-1543.1991. PubMed PMID: 1995595; PubMed Central PMCID: PMCPMC207293.

99. Kropinski AM. P22-like viruses. In: Tidona CA, Darai G, Büchen-Osmond C, editors. The Springer Index of Viruses. Berlin, Heidelberg: Springer Berlin Heidelberg; 2001. p. 804-9.

100. Nobrega FL, Vlot M, de Jonge PA, Dreesens LL, Beaumont HJE, Lavigne R, et al. Targeting mechanisms of tailed bacteriophages. Nat Rev Microbiol. 2018;16(12):760-73. Epub 2018/08/15. doi: 10.1038/s41579-018-0070-8. PubMed PMID: 30104690.

101. Wang C, Tu J, Liu J, Molineux IJ. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. Nat Microbiol. 2019;4(6):1049-56. Epub 20190318. doi: 10.1038/s41564-019-0403-z. PubMed PMID: 30886360; PubMed Central PMCID: PMCPMC6533119.

102. Steinbacher S, Baxa U, Miller S, Weintraub A, Seckler R, Huber R. Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. Proc Natl Acad Sci U S A. 1996;93(20):10584-8. doi: 10.1073/pnas.93.20.10584. PubMed PMID: 8855221; PubMed Central PMCID: PMCPMC38196.

103. Barbirz S, Muller JJ, Uetrecht C, Clark AJ, Heinemann U, Seckler R. Crystal structure of *Escherichia coli* phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. Mol Microbiol. 2008;69(2):303-16. doi: 10.1111/j.1365-2958.2008.06311.x. PubMed PMID: 18547389.

104. King MR, Vimr RP, Steenbergen SM, Spanjaard L, Plunkett G, 3rd, Blattner FR, et al. *Escherichia coli* K1-specific bacteriophage CUS-3 distribution and function in phase-variable capsular polysialic acid O acetylation. J Bacteriol. 2007;189(17):6447-56. Epub 20070629. doi: 10.1128/JB.00657-07. PubMed PMID: 17601779; PubMed Central PMCID: PMCPMC1951898.

105. Sternberg NL, Maurer R. Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella typhimurium*. Methods Enzymol. 1991;204:18-43. doi: 10.1016/0076-6879(91)04004-8. PubMed PMID: 1943777.

106. Fillol-Salom A, Bacigalupe R, Humphrey S, Chiang YN, Chen J, Penades JR. Lateral transduction is inherent to the life cycle of the archetypical *Salmonella* phage P22. Nat Commun. 2021;12(1):6510. Epub 20211108. doi: 10.1038/s41467-021-26520-4. PubMed PMID: 34751192; PubMed Central PMCID: PMCPMC8575938.

107. Casjens SR, Grose JH. Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages. Virology. 2016;496:255-76. Epub 20160630. doi: 10.1016/j.virol.2016.05.022. PubMed PMID: 27372181; PubMed Central PMCID: PMCPMC4969182.

108. Valencia-Toxqui G, Ramsey J. How to introduce a new bacteriophage on the block: a short guide to phage classification. J Virol. 2024;98(10):e0182123. Epub 20240912. doi: 10.1128/jvi.01821-23. PubMed PMID: 39264154; PubMed Central PMCID: PMCPMC11494874.

109. Casjens SR, Thuman-Commike PA. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. Virology. 2011;411(2):393-415. Epub 20110218. doi: 10.1016/j.virol.2010.12.046. PubMed PMID: 21310457.

110. Linares R, Arnaud CA, Effantin G, Darnault C, Epalle NH, Boeri Erba E, et al. Structural basis of bacteriophage T5 infection trigger and *E. coli* cell wall perforation. Sci Adv. 2023;9(12):eade9674. Epub 20230324. doi: 10.1126/sciadv.ade9674. PubMed PMID: 36961893; PubMed Central PMCID: PMCPMC10038345.

111. Heller K, Braun V. Polymannose O-antigens of *Escherichia coli*, the binding sites for the reversible adsorption of bacteriophage T5+ via the L-shaped tail fibers. J Virol. 1982;41(1):222-7. Epub 1982/01/01. doi: 10.1128/JVI.41.1.222-227.1982. PubMed PMID: 7045389; PubMed Central PMCID: PMCPMC256742.

112. Gong Q, Wang X, Huang H, Sun Y, Qian X, Xue F, et al. Novel Host Recognition Mechanism of the K1 Capsule-Specific Phage of *Escherichia coli*: Capsular Polysaccharide as the First Receptor and Lipopolysaccharide as the Secondary Receptor. J Virol. 2021;95(18):e0092021. Epub 20210825. doi: 10.1128/JVI.00920-21. PubMed PMID: 34191582; PubMed Central PMCID: PMCPMC8387036.

113. Beamud B, Garcia-Gonzalez N, Gomez-Ortega M, Gonzalez-Candelas F, Domingo-Calap P, Sanjuan R. Genetic determinants of host tropism in *Klebsiella* phages. Cell Rep. 2023;42(2):112048. Epub 20230206. doi: 10.1016/j.celrep.2023.112048. PubMed PMID: 36753420; PubMed Central PMCID: PMCPMC9989827.

114. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2023;51(D1):D418-D27. doi: 10.1093/nar/gkac993. PubMed PMID: 36350672; PubMed Central PMCID: PMCPMC9825450.

115. Smug BJ, Szczepaniak K, Rocha EPC, Dunin-Horkawicz S, Mostowy RJ. Ongoing shuffling of protein fragments diversifies core viral functions linked to interactions with bacterial hosts. Nat Commun. 2023;14(1):7460. Epub 20231128. doi: 10.1038/s41467-023-43236-9. PubMed PMID: 38016962; PubMed Central PMCID: PMCPMC10684548.

116. Hu B, Margolin W, Molineux IJ, Liu J. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. Proc Natl Acad Sci U S A. 2015;112(35):E4919-28. Epub 2015/08/19. doi: 10.1073/pnas.1501064112. PubMed PMID: 26283379; PubMed Central PMCID: PMCPMC4568249.

117. Witte S, Zinsli LV, Gonzalez-Serrano R, Matter CI, Loessner MJ, van Mierlo JT, et al. Structural and functional characterization of the receptor binding proteins of *Escherichia coli* O157 phages EP75 and EP335. Comput Struct Biotechnol J. 2021;19:3416-26. Epub 20210604. doi: 10.1016/j.csbj.2021.06.001. PubMed PMID: 34194667; PubMed Central PMCID: PMCPMC8217332.

118. Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, et al. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proc Natl Acad Sci U S A. 2010;107(47):20287-92.

Epub 20101101. doi: 10.1073/pnas.1011218107. PubMed PMID: 21041684; PubMed Central PMCID: PMCPMC2996694.

119. Garcia-Doval C, Caston JR, Luque D, Granell M, Otero JM, Llamas-Saiz AL, et al. Structure of the Receptor-Binding Carboxy-Terminal Domain of the Bacteriophage T5 L-Shaped Tail Fibre with and without Its Intra-Molecular Chaperone. Viruses. 2015;7(12):6424-40. Epub 20151208. doi: 10.3390/v7122946. PubMed PMID: 26670244; PubMed Central PMCID: PMCPMC4690869.

120. Hyman P. Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. Pharmaceuticals (Basel). 2019;12(1). Epub 2019/03/14. doi: 10.3390/ph12010035. PubMed PMID: 30862020; PubMed Central PMCID: PMCPMC6469166.

121. Khan Mirzaei M, Nilsson AS. Isolation of phages for phage therapy: a comparison of spot tests and efficiency of plating analyses for determination of host range and efficacy. PLoS One. 2015;10(3):e0118557. doi: 10.1371/journal.pone.0118557. PubMed PMID: 25761060; PubMed Central PMCID: PMCPMC4356574.

122. Holtappels D, Alfenas-Zerbini P, Koskella B. Drivers and consequences of bacteriophage host range. FEMS Microbiol Rev. 2023;47(4). doi: 10.1093/femsre/fuad038. PubMed PMID: 37422441.

123. Abedon ST. Lysis from without. Bacteriophage. 2011;1(1):46-9. doi: 10.4161/bact.1.1.13980. PubMed PMID: 21687534; PubMed Central PMCID: PMCPMC3109453.

124. Meile S, Du J, Staubli S, Grossmann S, Koliwer-Brandl H, Piffaretti P, et al. Engineered reporter phages for detection of *Escherichia coli*, *Enterococcus*, and *Klebsiella* in urine. Nat Commun. 2023;14(1):4336. Epub 20230720. doi: 10.1038/s41467-023-39863-x. PubMed PMID: 37474554; PubMed Central PMCID: PMCPMC10359277.

125. Loose M, Saez Moreno D, Mutti M, Hitzenhammer E, Visram Z, Dippel D, et al. Natural Bred ε2-Phages Have an Improved Host Range and Virulence against Uropathogenic *Escherichia coli* over Their Ancestor Phages. Antibiotics (Basel). 2021;10(11). Epub 20211101. doi: 10.3390/antibiotics10111337. PubMed PMID: 34827275; PubMed Central PMCID: PMCPMC8614997.

126. Schwarzer D, Buettner FF, Browning C, Nazarov S, Rabsch W, Bethe A, et al. A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. J Virol. 2012;86(19):10384-98. Epub 2012/07/13. doi: 10.1128/JVI.00801-12. PubMed PMID: 22787233; PubMed Central PMCID: PMCPMC3457257.

127. Manohar P, Tamhankar AJ, Lundborg CS, Ramesh N. Isolation, characterization and in vivo efficacy of *Escherichia* phage myPSH1131. PLoS One. 2018;13(10):e0206278. Epub 20181024. doi: 10.1371/journal.pone.0206278. PubMed PMID: 30356310; PubMed Central PMCID: PMCPMC6200275.

128. Tesson F, Huiting E, Wei L, Ren J, Johnson M, Planel R, et al. Exploring the diversity of anti-defense systems across prokaryotes, phages and mobile genetic elements. Nucleic Acids Res. 2025;53(1). doi: 10.1093/nar/gkae1171. PubMed PMID: 39657785; PubMed Central PMCID: PMCPMC11724313.

129. Stevenson G, Neal B, Liu D, Hobbs M, Packer NH, Batley M, et al. Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. J Bacteriol. 1994;176(13):4144-56. doi: 10.1128/jb.176.13.4144-4156.1994. PubMed PMID: 7517391; PubMed Central PMCID: PMCPMC205614.

130. Mann E, Ovchinnikova OG, King JD, Whitfield C. Bacteriophage-mediated Glucosylation Can Modify Lipopolysaccharide O-Antigens Synthesized by an ATP-binding Cassette (ABC) Transporter-

dependent Assembly Mechanism. J Biol Chem. 2015;290(42):25561-70. Epub 20150901. doi: 10.1074/jbc.M115.660803. PubMed PMID: 26330553; PubMed Central PMCID: PMCPMC4646201.

131. Jann B, Shashkov AS, Kochanowski H, Jann K. Structure of the O16 polysaccharide from *Escherichia coli* O16:K1: an NMR investigation. Carbohydr Res. 1994;264(2):305-11. doi: 10.1016/s0008-6215(05)80014-x. PubMed PMID: 7528642.

132. Hendrix RW, Duda RL. Bacteriophage lambda PaPa: not the mother of all lambda phages. Science. 1992;258(5085):1145-8. doi: 10.1126/science.1439823. PubMed PMID: 1439823.

133. Broadbent SE, Davies MR, van der Woude MW. Phase variation controls expression of *Salmonella* lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. Mol Microbiol. 2010;77(2):337-53. Epub 20100512. doi: 10.1111/j.1365-2958.2010.07203.x. PubMed PMID: 20487280; PubMed Central PMCID: PMCPMC2909390.

134. He Z, Parra B, Nesme J, Smets BF, Dechesne A. Quantification and fate of plasmid-specific bacteriophages in wastewater: Beyond the F-coliphages. Water Res. 2022;227:119320. Epub 20221104. doi: 10.1016/j.watres.2022.119320. PubMed PMID: 36395568.

135. Wietzorrek A, Schwarz H, Herrmann C, Braun V. The genome of the novel phage Rtp, with a rosette-like tail tip, is homologous to the genome of phage T1. J Bacteriol. 2006;188(4):1419-36. Epub 2006/02/03. doi: 10.1128/JB.188.4.1419-1436.2006. PubMed PMID: 16452425; PubMed Central PMCID: PMCPMC1367250.

136. Decker K, Krauel V, Meesmann A, Heller KJ. Lytic conversion of *Escherichia coli* by bacteriophage T5: blocking of the FhuA receptor protein by a lipoprotein expressed early during infection. Mol Microbiol. 1994;12(2):321-32. doi: 10.1111/j.1365-2958.1994.tb01020.x. PubMed PMID: 8057856.

137. Silveira CB, Luque A, Rohwer F. The landscape of lysogeny across microbial community density, diversity and energetics. Environ Microbiol. 2021;23(8):4098-111. Epub 2021/06/15. doi: 10.1111/1462-2920.15640. PubMed PMID: 34121301.

138. Gonzalez-Serrano R, Rosselli R, Roda-Garcia JJ, Martin-Cuadrado AB, Rodriguez-Valera F, Dunne M. Distantly related *Alteromonas* bacteriophages share tail fibers exhibiting properties of transient chaperone caps. Nat Commun. 2023;14(1):6517. Epub 20231016. doi: 10.1038/s41467-023-42114-8. PubMed PMID: 37845226; PubMed Central PMCID: PMCPMC10579305.

139. Quinones-Olvera N, Owen SV, McCully LM, Marin MG, Rand EA, Fan AC, et al. Diverse and abundant phages exploit conjugative plasmids. Nat Commun. 2024;15(1):3197. Epub 20240412. doi: 10.1038/s41467-024-47416-z. PubMed PMID: 38609370; PubMed Central PMCID: PMCPMC11015023.

140. Gambino M, Sorensen MCH. Flagellotropic phages: common yet diverse host interaction strategies. Curr Opin Microbiol. 2024;78:102451. Epub 20240306. doi: 10.1016/j.mib.2024.102451. PubMed PMID: 38452595.

141. Ge X, Wang J. Structural mechanism of bacteriophage lambda tail's interaction with the bacterial receptor. Nat Commun. 2024;15(1):4185. Epub 20240517. doi: 10.1038/s41467-024-48686-3. PubMed PMID: 38760367; PubMed Central PMCID: PMCPMC11101478.

142. Iglesias SM, Lokareddy RK, Yang R, Li F, Yeggoni DP, David Hou CF, et al. Molecular Architecture of *Salmonella Typhimurium* [sic]Virus P22 Genome Ejection Machinery. J Mol Biol. 2023;435(24):168365.

Epub 20231110. doi: 10.1016/j.jmb.2023.168365. PubMed PMID: 37952769; PubMed Central PMCID: PMCPMC10842050.

143. Ouyang R, Ongenae V, Muok A, Claessen D, Briegel A. Phage fibers and spikes: a nanoscale Swiss army knife for host infection. Curr Opin Microbiol. 2024;77:102429. Epub 20240126. doi: 10.1016/j.mib.2024.102429. PubMed PMID: 38277900.

144. Gordillo Altamirano FL, Barr JJ. Unlocking the next generation of phage therapy: the key is in the receptors. Curr Opin Biotechnol. 2020;68:115-23. Epub 2020/11/18. doi: 10.1016/j.copbio.2020.10.002. PubMed PMID: 33202354.

145. David E, Plantady C, Poissonnier S, Le Boulch J, Gutierrez A, Chevallereau A. Systematic functional assessment of antiphage systems in their native host. bioRxiv. 2024:2024.12.20.629700. doi: 10.1101/2024.12.20.629700.

146. Dunne M, Prokhorov NS, Loessner MJ, Leiman PG. Reprogramming bacteriophage host range: design principles and strategies for engineering receptor binding proteins. Curr Opin Biotechnol. 2021;68:272-81. Epub 20210318. doi: 10.1016/j.copbio.2021.02.006. PubMed PMID: 33744824; PubMed Central PMCID: PMCPMC10163921.

147. Klein-Sousa V, Roa-Eguiara A, Kielkopf CS, Sofos N, Taylor NMI. Towards a complete phage tail fiber structure atlas. bioRxiv. 2024:2024.10.28.620165. doi: 10.1101/2024.10.28.620165.

148. Haudiquet M, Le Bris J, Nucci A, Bonnin RA, Domingo-Calap P, Rocha EPC, et al. Capsules and their traits shape phage susceptibility and plasmid conjugation efficiency. Nat Commun. 2024;15(1):2032. Epub 20240306. doi: 10.1038/s41467-024-46147-5. PubMed PMID: 38448399; PubMed Central PMCID: PMCPMC10918111.

149. Ferriol-Gonzalez C, Concha-Eloko R, Bernabeu-Gimeno M, Fernandez-Cuenca F, Canada-Garcia JE, Garcia-Cobos S, et al. Targeted phage hunting to specific *Klebsiella pneumoniae* clinical isolates is an efficient antibiotic resistance and infection control strategy. Microbiol Spectr. 2024:e0025424. Epub 20240828. doi: 10.1128/spectrum.00254-24. PubMed PMID: 39194291.

150. Boeckaerts D, Stock M, Ferriol-Gonzalez C, Oteo-Iglesias J, Sanjuan R, Domingo-Calap P, et al. Prediction of *Klebsiella* phage-host specificity at the strain level. Nat Commun. 2024;15(1):4355. Epub 20240522. doi: 10.1038/s41467-024-48675-6. PubMed PMID: 38778023; PubMed Central PMCID: PMCPMC11111740.

151. Wright RCT, Friman VP, Smith MCM, Brockhurst MA. Cross-resistance is modular in bacteria-phage interactions. PLoS Biol. 2018;16(10):e2006057. Epub 2018/10/04. doi: 10.1371/journal.pbio.2006057. PubMed PMID: 30281587; PubMed Central PMCID: PMCPMC6188897.

152. Göller PC, Elsener T, Lorge D, Radulovic N, Bernardi V, Naumann A, et al. Multi-species host range of staphylococcal phages isolated from wastewater. Nat Commun. 2021;12(1):6965. Epub 2021/12/01. doi: 10.1038/s41467-021-27037-6. PubMed PMID: 34845206; PubMed Central PMCID: PMCPMC8629997.

153. Walsh SK, Imrie RM, Matuszewska M, Paterson GK, Weinert LA, Hadfield JD, et al. The host phylogeny determines viral infectivity and replication across *Staphylococcus* host species. PLoS Pathog. 2023;19(6):e1011433. Epub 20230608. doi: 10.1371/journal.ppat.1011433. PubMed PMID: 37289828; PubMed Central PMCID: PMCPMC10284401.

154. Piel D, Bruto M, Labreuche Y, Blanquart F, Goudenege D, Barcia-Cruz R, et al. Phage-host coevolution in natural populations. Nat Microbiol. 2022;7(7):1075-86. Epub 2022/06/28. doi: 10.1038/s41564-022-01157-1. PubMed PMID: 35760840.

155. Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, et al. Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. Science. 2021;374(6566):488-92. Epub 20211021. doi: 10.1126/science.abb1083. PubMed PMID: 34672730.

156. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 2015;43(Database issue):D298-9. Epub 2014/11/08. doi: 10.1093/nar/gku1046. PubMed PMID: 25378308; PubMed Central PMCID: PMCPMC4383893.

157. Kortright KE, Chan BK, Koff JL, Turner PE. Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. Cell Host Microbe. 2019;25(2):219-32. Epub 2019/02/15. doi: 10.1016/j.chom.2019.01.014. PubMed PMID: 30763536.

158. Aslam S, Lampley E, Wooten D, Karris M, Benson C, Strathdee S, et al. Lessons Learned From the First 10 Consecutive Cases of Intravenous Bacteriophage Therapy to Treat Multidrug-Resistant Bacterial Infections at a Single Center in the United States. Open Forum Infect Dis. 2020;7(9):ofaa389. Epub 2020/10/03. doi: 10.1093/ofid/ofaa389. PubMed PMID: 33005701; PubMed Central PMCID: PMCPMC7519779.

159. Meile S, Du J, Dunne M, Kilcher S, Loessner MJ. Engineering therapeutic phages for enhanced antibacterial efficacy. Curr Opin Virol. 2022;52:182-91. Epub 20211221. doi: 10.1016/j.coviro.2021.12.003. PubMed PMID: 34952266.

160. Gencay YE, Jasinskyte D, Robert C, Semsey S, Martinez V, Petersen AO, et al. Engineered phage with antibacterial CRISPR-Cas selectively reduce *E. coli* burden in mice. Nat Biotechnol. 2024;42(2):265-74. Epub 20230504. doi: 10.1038/s41587-023-01759-y. PubMed PMID: 37142704; PubMed Central PMCID: PMCPMC10869271.

161. Brödel AK, Charpenay LH, Galtier M, Fuche FJ, Terrasse R, Poquet C, et al. In situ targeted base editing of bacteria in the mouse gut. Nature. 2024. Epub 20240710. doi: 10.1038/s41586-024-07681-w. PubMed PMID: 38987595.

162. Kim P, Sanchez AM, Penke TJR, Tuson HH, Kime JC, McKee RW, et al. Safety, pharmacokinetics, and pharmacodynamics of LBP-EC01, a CRISPR-Cas3-enhanced bacteriophage cocktail, in uncomplicated urinary tract infections due to *Escherichia coli* (ELIMINATE): the randomised, open-label, first part of a two-part phase 2 trial. Lancet Infect Dis. 2024. Epub 20240809. doi: 10.1016/S1473-3099(24)00424-9. PubMed PMID: 39134085.

163. Kropinski AM, Mazzocco A, Waddell TE, Lingohr E, Johnson RP. Enumeration of bacteriophages by double agar overlay plaque assay. Methods Mol Biol. 2009;501:69-76. doi: 10.1007/978-1-60327-164-6_7. PubMed PMID: 19066811.

164. Kauffman KM, Polz MF. Streamlining standard bacteriophage methods for higher throughput. MethodsX. 2018;5:159-72. Epub 2019/01/10. doi: 10.1016/j.mex.2018.01.007. PubMed PMID: 30622914; PubMed Central PMCID: PMCPMC6318102.

165. Browning DF, Hobman JL, Busby SJW. Laboratory strains of *Escherichia coli* K-12: things are seldom what they seem. Microb Genom. 2023;9(2). doi: 10.1099/mgen.0.000922. PubMed PMID: 36745549; PubMed Central PMCID: PMCPMC9997739.

166. Pleska M, Qian L, Okura R, Bergmiller T, Wakamoto Y, Kussell E, et al. Bacterial Autoimmunity Due to a Restriction-Modification System. Curr Biol. 2016;26(3):404-9. Epub 2016/01/26. doi: 10.1016/j.cub.2015.12.041. PubMed PMID: 26804559.

167. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006 0008. doi: 10.1038/msb4100050. PubMed PMID: 16738554; PubMed Central PMCID: PMC1681482.

168. Xue C, Sashital DG. Mechanisms of Type I-E and I-F CRISPR-Cas Systems in *Enterobacteriaceae*. EcoSal Plus. 2019;8(2). doi: 10.1128/ecosalplus.ESP-0008-2018. PubMed PMID: 30724156; PubMed Central PMCID: PMCPMC6368399.

169. Strotskaya A, Savitskaya E, Metlitskaya A, Morozova N, Datsenko KA, Semenova E, et al. The action of *Escherichia coli* CRISPR-Cas system on lytic bacteriophages with different lifestyles and development strategies. Nucleic Acids Res. 2017;45(4):1946-57. Epub 2017/01/29. doi: 10.1093/nar/gkx042. PubMed PMID: 28130424; PubMed Central PMCID: PMCPMC5389539.

170. Yu YT, Snyder L. Translation elongation factor Tu cleaved by a phage-exclusion system. Proc Natl Acad Sci U S A. 1994;91(2):802-6. Epub 1994/01/18. doi: 10.1073/pnas.91.2.802. PubMed PMID: 8290603; PubMed Central PMCID: PMCPMC43037.

171. Koga M, Otsuka Y, Lemire S, Yonesaki T. *Escherichia coli rnlA* and *rnlB* compose a novel toxin-antitoxin system. Genetics. 2011;187(1):123-30. Epub 2010/10/29. doi: 10.1534/genetics.110.121798. PubMed PMID: 20980243; PubMed Central PMCID: PMCPmc3018318.

172. Ragunathan PT, Vanderpool CK. Cryptic-Prophage-Encoded Small Protein DicB Protects *Escherichia coli* from Phage Infection by Inhibiting Inner Membrane Receptor Proteins. J Bacteriol. 2019;201(23). Epub 2019/09/19. doi: 10.1128/JB.00475-19. PubMed PMID: 31527115; PubMed Central PMCID: PMCPMC6832061.

173. Sasaki T, Takita S, Fujishiro T, Shintani Y, Nojiri S, Yasui R, et al. Phage single-stranded DNA-binding protein or host DNA damage triggers the activation of the AbpAB phage defense system. mSphere. 2023:e0037223. Epub 20231026. doi: 10.1128/msphere.00372-23. PubMed PMID: 37882551.

174. Grenier F, Matteau D, Baby V, Rodrigue S. Complete Genome Sequence of *Escherichia coli* BW25113. Genome Announc. 2014;2(5). Epub 20141016. doi: 10.1128/genomeA.01038-14. PubMed PMID: 25323716; PubMed Central PMCID: PMCPMC4200154.

175. Bachmann BJ. Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol Rev. 1972;36(4):525-57. doi: 10.1128/br.36.4.525-557.1972. PubMed PMID: 4568763; PubMed Central PMCID: PMCPMC408331.

176. Cuenod A, Agnetti J, Seth-Smith HMB, Roloff T, Walchli D, Shcherbakov D, et al. Bacterial genome-wide association study substantiates *papGII* of *Escherichia coli* as a major risk factor for urosepsis. Genome Med. 2023;15(1):89. Epub 20231030. doi: 10.1186/s13073-023-01243-x. PubMed PMID: 37904175; PubMed Central PMCID: PMCPMC10614358.

177. Abedon ST, Katsaounis TI. Basic Phage Mathematics. Methods Mol Biol. 2018;1681:3-30. Epub 2017/11/15. doi: 10.1007/978-1-4939-7343-9_1. PubMed PMID: 29134583.

178. Garneau JR, Depardieu F, Fortier LC, Bikard D, Monot M. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci

Rep. 2017;7(1):8292. Epub 2017/08/16. doi: 10.1038/s41598-017-07910-5. PubMed PMID: 28811656; PubMed Central PMCID: PMCPMC5557969.

179. Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald PJ, Vreugde S. Pharokka: a fast scalable bacteriophage annotation tool. Bioinformatics. 2023;39(1). doi: 10.1093/bioinformatics/btac776. PubMed PMID: 36453861; PubMed Central PMCID: PMCPMC9805569.

180. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol Biol. 2019;1962:1-14. Epub 2019/04/26. doi: 10.1007/978-1-4939-9173-0_1. PubMed PMID: 31020551; PubMed Central PMCID: PMCPMC6768409.

181. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. Epub 2013/01/19. doi: 10.1093/molbev/mst010. PubMed PMID: 23329690; PubMed Central PMCID: PMCPMC3603318.

182. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-D9. doi: 10.1093/nar/gkaa913. PubMed PMID: 33125078; PubMed Central PMCID: PMCPMC7779014.

183. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, et al. The conserved domain database in 2023. Nucleic Acids Res. 2023;51(D1):D384-D8. doi: 10.1093/nar/gkac1096. PubMed PMID: 36477806; PubMed Central PMCID: PMCPMC9825596.

184. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, et al. CATH: increased structural coverage of functional space. Nucleic Acids Res. 2021;49(D1):D266-D73. doi: 10.1093/nar/gkaa1079. PubMed PMID: 33237325; PubMed Central PMCID: PMCPMC7778904.

185. Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Perez Bucio RE, Mom R, et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. NAR Genom Bioinform. 2021;3(3):lqab067. Epub 20210805. doi: 10.1093/nargab/lqab067. PubMed PMID: 34377978; PubMed Central PMCID: PMCPMC8341000.

186. Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics. 2019;20(1):473. Epub 20190914. doi: 10.1186/s12859-019-3019-7. PubMed PMID: 31521110; PubMed Central PMCID: PMCPMC6744700.

187. Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. Nat Commun. 2018;9(1):2542. Epub 20180629. doi: 10.1038/s41467-018-04964-5. PubMed PMID: 29959318; PubMed Central PMCID: PMCPMC6026198.

188. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42. doi: 10.1093/nar/28.1.235. PubMed PMID: 10592235; PubMed Central PMCID: PMCPMC102472.

189. UniProt C. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523-D31. doi: 10.1093/nar/gkac1052. PubMed PMID: 36408920; PubMed Central PMCID: PMCPMC9825514.

190. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids

Res. 2016;44(D1):D733-45. Epub 20151108. doi: 10.1093/nar/gkv1189. PubMed PMID: 26553804; PubMed Central PMCID: PMCPMC4702849.

191. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. Epub 2021/07/16. doi: 10.1038/s41586-021-03819-2. PubMed PMID: 34265844; PubMed Central PMCID: PMCPMC8371605.

192. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nat Biotechnol. 2024;42(2):243-6. Epub 20230508. doi: 10.1038/s41587-023-01773-0. PubMed PMID: 37156916; PubMed Central PMCID: PMCPMC10869269.

193. Adriaenssens E, Brister JR. How to Name and Classify Your Phage: An Informal Guide. Viruses. 2017;9(4). doi: 10.3390/v9040070. PubMed PMID: 28368359; PubMed Central PMCID: PMCPMC5408676.

194. Tolstoy I, Kropinski AM, Brister JR. Bacteriophage Taxonomy: An Evolving Discipline. Methods Mol Biol. 2018;1693:57-71. Epub 2017/11/10. doi: 10.1007/978-1-4939-7395-8_6. PubMed PMID: 29119432.

195. Grose JH, Casjens SR. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family *Enterobacteriaceae*. Virology. 2014;468-470:421-43. Epub 2014/09/23. doi: 10.1016/j.virol.2014.08.024. PubMed PMID: 25240328; PubMed Central PMCID: PMCPMC4301999.

196. Turner D, Adriaenssens EM, Lehman SM, Moraru C, Kropinski AM. Bacteriophage Taxonomy: A Continually Evolving Discipline. Methods Mol Biol. 2024;2734:27-45. doi: 10.1007/978-1-0716-3523-0_3. PubMed PMID: 38066361.

197. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307-21. Epub 2010/06/09. doi: 10.1093/sysbio/syq010. PubMed PMID: 20525638.

198. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2022:2021.10.04.463034. doi: 10.1101/2021.10.04.463034.

199. Cianfrocco MA, Wong-Barnum M, Youn C, Wagner R, Leschziner A. COSMIC2: A Science Gateway for Cryo-Electron Microscopy Structure Determination. Practice and Experience in Advanced Research Computing 2017: Sustainability, Success and Impact; New Orleans, LA, USA: Association for Computing Machinery; 2017. p. Article 22.

200. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493-500. Epub 20240508. doi: 10.1038/s41586-024-07487-w. PubMed PMID: 38718835; PubMed Central PMCID: PMCPMC11168924.

201. Holm L. Dali server: structural unification of protein families. Nucleic Acids Res. 2022;50(W1):W210-W5. doi: 10.1093/nar/gkac387. PubMed PMID: 35610055; PubMed Central PMCID: PMCPMC9252788.

202. Aziz RK, Ackermann HW, Petty NK, Kropinski AM. Essential Steps in Characterizing Bacteriophages: Biology, Taxonomy, and Genome Analysis. Methods Mol Biol. 2018;1681:197-215. Epub 2017/11/15. doi: 10.1007/978-1-4939-7343-9_15. PubMed PMID: 29134597.

203. Swanson NA, Lokareddy RK, Li F, Hou CD, Leptihn S, Pavlenok M, et al. Cryo-EM structure of the periplasmic tunnel of T7 DNA-ejectosome at 2.7 A resolution. Mol Cell. 2021;81(15):3145-59 e7. Epub 20210701. doi: 10.1016/j.molcel.2021.06.001. PubMed PMID: 34214465; PubMed Central PMCID: PMCPMC8349896.

204. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res. 2023;51(W1):W443-W50. doi: 10.1093/nar/gkad382. PubMed PMID: 37194694; PubMed Central PMCID: PMCPMC10320120.

205. Pick K, Ju T, Willing BP, Raivio TL. Isolation and Characterization of a Novel Temperate *Escherichia coli* Bacteriophage, Kapi1, Which Modifies the O-Antigen and Contributes to the Competitiveness of Its Host during Colonization of the Murine Gastrointestinal Tract. mBio. 2022;13(1):e0208521. Epub 20220125. doi: 10.1128/mbio.02085-21. PubMed PMID: 35073745; PubMed Central PMCID: PMCPMC8787464.

206. Bertani B, Ruiz N. Function and Biogenesis of Lipopolysaccharides. EcoSal Plus. 2018;8(1). Epub 2018/08/02. doi: 10.1128/ecosalplus.ESP-0001-2018. PubMed PMID: 30066669; PubMed Central PMCID: PMCPMC6091223.

207. North OI, Sakai K, Yamashita E, Nakagawa A, Iwazaki T, Buttner CR, et al. Phage tail fibre assembly proteins employ a modular structure to drive the correct folding of diverse fibres. Nat Microbiol. 2019;4(10):1645-53. Epub 20190617. doi: 10.1038/s41564-019-0477-7. PubMed PMID: 31209305.

Data reporting

All data generated or analyzed during this study are included in this published article. The raw data of every replicate experiment underlying the plots shown in Figs. 5D, 6G, 7D, 8G, 9E, 10E, 11A/B, and S1A/B, all calculations underlying the summary data in these figures, and additional analyses are compiled in S1 Data. Tree files of all phylogenies shown in this work and the underlying sequence alignments in NEXUS format are available in Data S3 file. The code used to run GAPS for the results presented in this work is available on Github (https://github.com/hiller-lab/gaps) and on Zenodo (https://www.doi.org/10.5281/zenodo.14277981).

Accession numbers

The annotated genome sequences of all 37 newly isolated phages as well as *Lederbergvirus* isolates Dewey, Huey, and Louie (Figs. 10D and S6B) are available in NCBI BioProject PRJNA1207239 with GenBank accession numbers for the BASEL phages listed in Table S2 and PQ850600, PQ850609, and PQ850616 for Huey, Dewey, and Louie (to be published elsewhere).

Supporting Information

Table S1. Isolation of O16-dependent and K12-prophage-inhibited phages from sewage and environmental samples

Table S2. List of all phages used in this study.

This table presents the taxonomic classification, isolation / source, host receptors, and genomic features of all phages characterized in this study. The original BASEL phages from our previous work (Bas01-Bas69; [10]) and classical reference phages are included for completeness. Predictions of anti-

defense proteins encoded by all BASEL phages (using AntiDefenseFinder [128]) are included on a second sheet for a better overview.

Table S3. List of all bacterial strains used in this study.

This table presents all bacterial strains that have been used in this work. The abbreviations in the selection column indicate the drug and its concentration that were used. Amp = ampicillin, Cam = chloramphenicol, Kan = kanamycin, Zeo = zeocin; 25 / 50 / 100 refer to 25 µg/ml, 50 µg/ml, and 100 = 100 µg/ml, respectively.

Table S4. List of all oligonucleotide primers used in this study.

Table S5. List of all plasmids used in this study.

Abbreviations in the selection column highlight the drug and its concentration that were used to select for plasmid maintenance. Amp = ampicillin, Cam = chloramphenicol, Kan = kanamycin; 25 / 50 / 100 refer to 25 µg/ml, 50 µg/ml, and 100 = 100 µg/ml, respectively.

Table S6. Construction of all plasmids generated in this study.

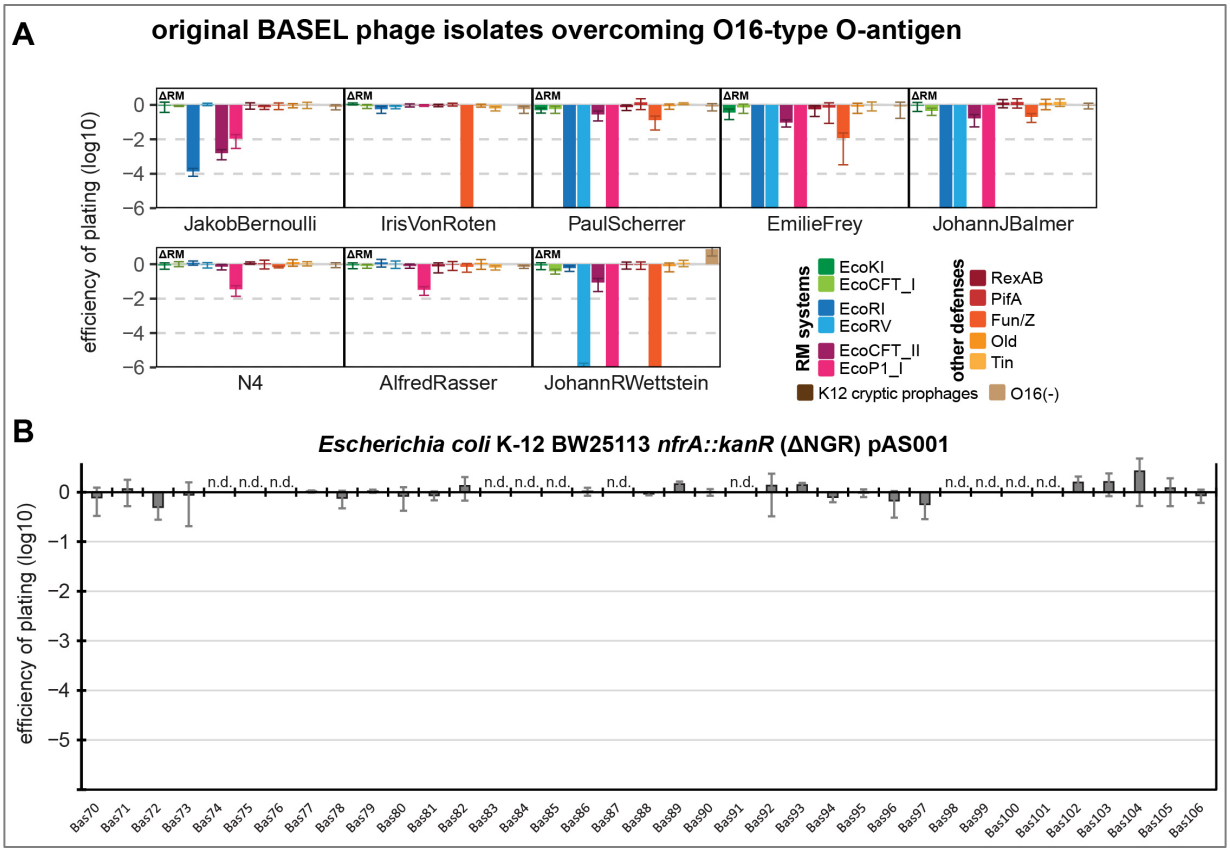
Table S7. Confidence scores of AlphaFold-generated models.

Confidence scores of AlphaFold-generated models. Confidence per residue is calculated as a predicted Local Distance Difference Test score (0-100), with an average of all residues within the models provided below. A pLDDT ≥ 90 have very high model confidence, residues with 90 > pLDDT ≥ 70 are classified as confident, while residues with 70 > pLDDT > 50 have low confidence, and residues with pLDDT <50 correspond to very low confidence. Interface pTM scores ("iptm+ptm") are a measure of predicted protein multimer accuracy by AlphaFold-Multimer using the with confidence scored 0 to 1 for the complete model. All structures were predicted using the COSMIC2 platform, except ManiWeber Gp32 (*) solved using the AlphaFold 3 Server.

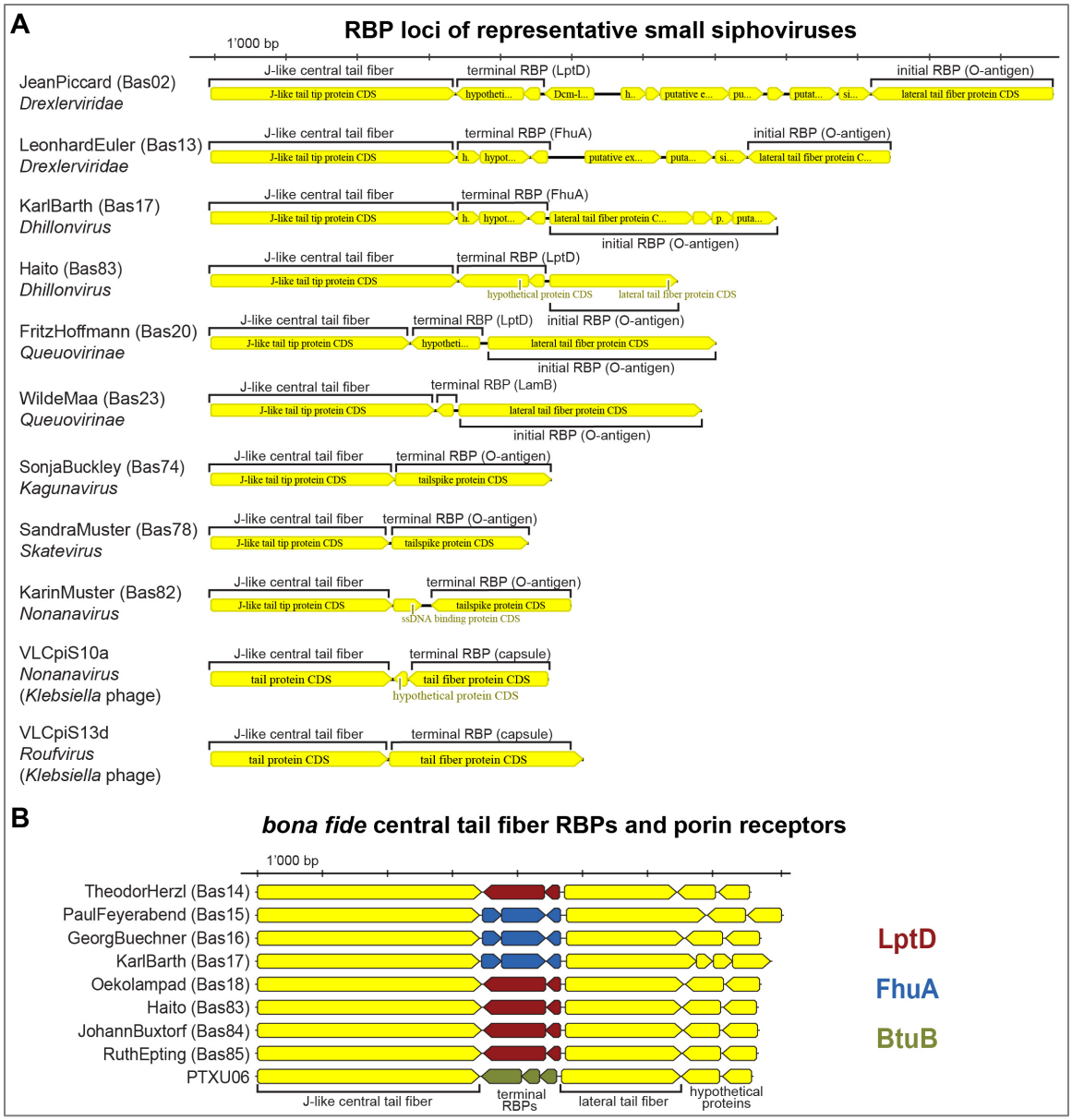
S1 Data. Raw data and calculations of all efficiency-of-plating (EOP) assays and qualitative top agar experiments.

S2 Data. GAPS report for the HeidiAbel genome.

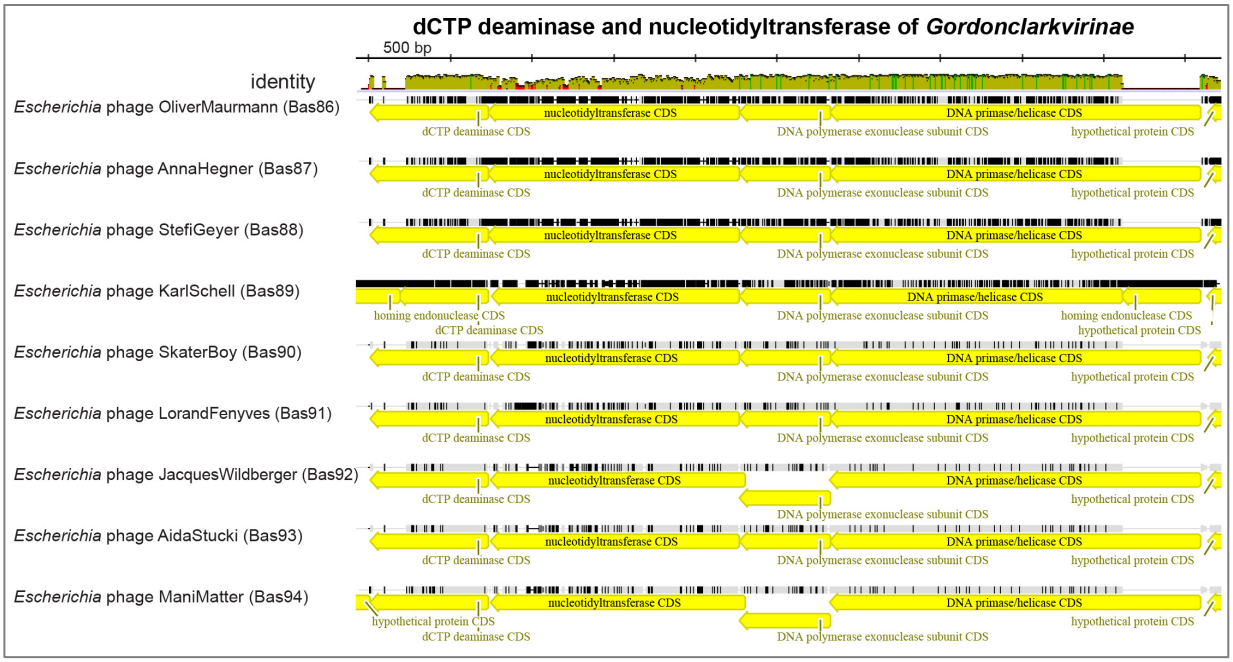
S3 Data. Tree files with underlying sequence alignments of all phylogenies.



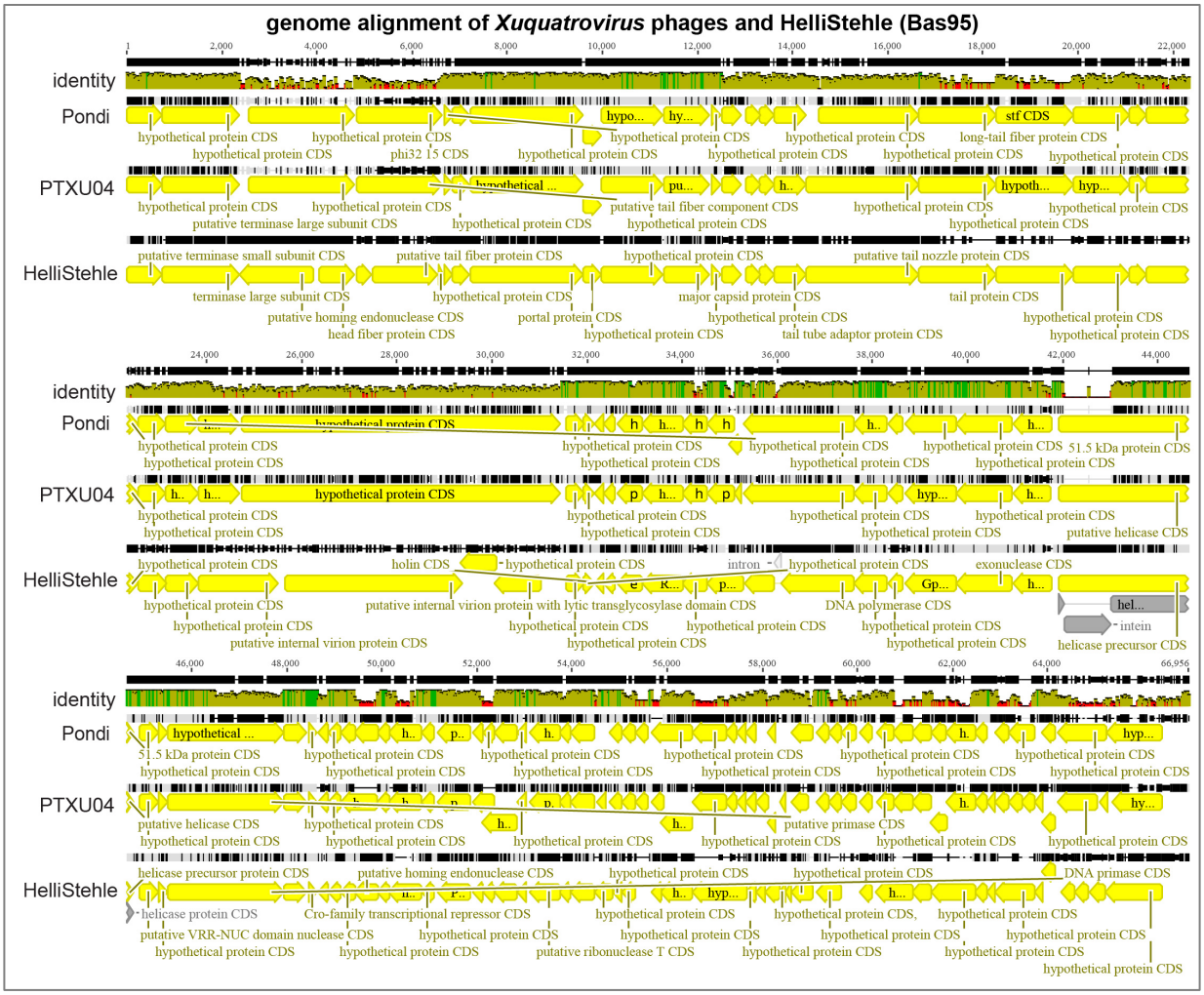
S1 Figure. Supplemental data for Fig. 4. (A) Phages that had shown significant growth on *E. coli* K-12 with restored O-antigen in our previous work [10] were now phenotypically characterized in the presence of this glycan. The results of these quantitative phenotyping experiments regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as efficiency of plating (EOP). The small note of "ΔRM" indicates that this experiment has been performed using *E. coli* K-12 ΔRM *wbbL*(+) as host (i.e., with restored O-antigen expression). Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data. **(B)** The results of quantitative phenotyping experiments with all newly presented BASEL phages infecting an *nfrA* knockout (specifically lacking the NGR glycan [17]) are presented as efficiency of plating (EOP). This strain had been transformed with plasmid pAS001 to restore O16-type O-antigen expression (see *Materials and Methods*). Data points and error bars represent average and standard deviation of at least three independent experiments. Raw data and calculations are available in S1 Data. It is evident that no newly included BASEL phage depends partially or completely on the NGR glycan for infection of *E. coli* K-12. Phages sensitive to the presence of cryptic prophages in the K-12 genomic background could not be tested. However, since these phages all carry tailspikes or tail fibers targeting O16-type O-antigen as receptor-binding proteins (Figs. 11 and 12) it seems highly unlikely that they would be affected by the absence of NGR.



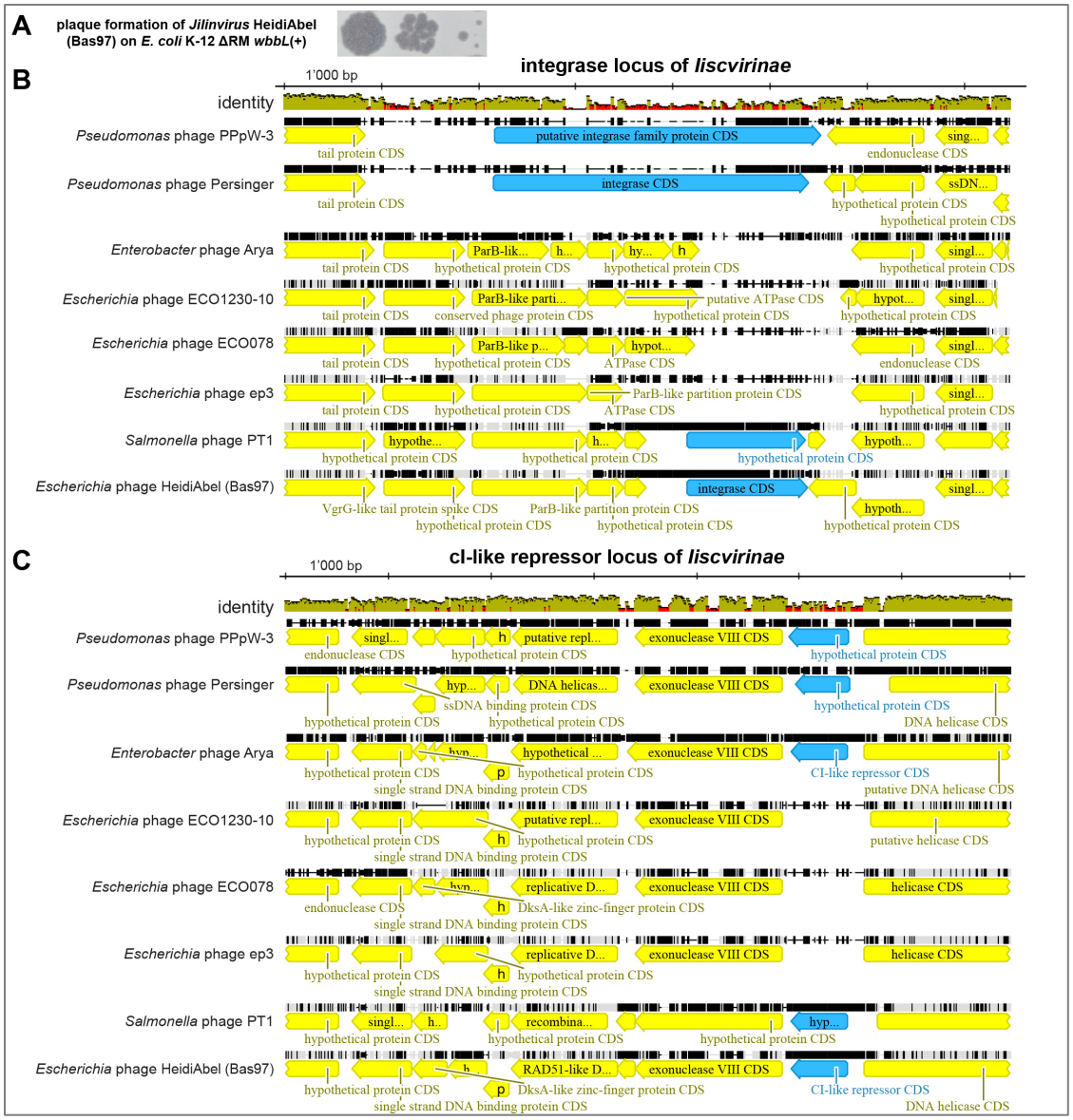
S2 Figure. Supplemental data for Fig. 6. (A) The illustration compares the tail fiber and tailspike loci of diverse small siphoviruses from the original BASEL collection and the current study and porin-targeting receptor-binding proteins (RBPs) are annotated based on the *bona fide* central tail fiber caps identified previously [10]. It is apparent that the single RBPs of the three exemplary O16-dependent phages SonjaBuckley (*Kagunavirus*, Bas74), SandraMuster (*Skatevirus*, Bas78), and KarinMuster (*Nonanavirus*, Bas82) are encoded at the same locus as the porin-targeting RBPs for terminal receptor recognition in the other genomes. For comparison, we also included two small siphoviruses targeting *Klebsiella* capsules that had been classified in previous work as belonging to *Nonanavirus* (see Fig. 6E) and *Roufvirus* (see Fig. 6D) [49]. Despite the annotation as "tail fiber", both viruses encode tailspikes directly downstream of the orthologs of the J-like central tail fiber proteins [49]. **(B)** Comparison of the *bona fide* central tail fiber RBP loci of all *Dhillonvirus* isolates in the BASEL collection as well as PTXU06 for comparison [10, 58]. The association of different variants of this locus with porin specificity is color-coded as in our previous work – all new isolates from the current study are predicted to target LptD (red) as terminal receptor [10].



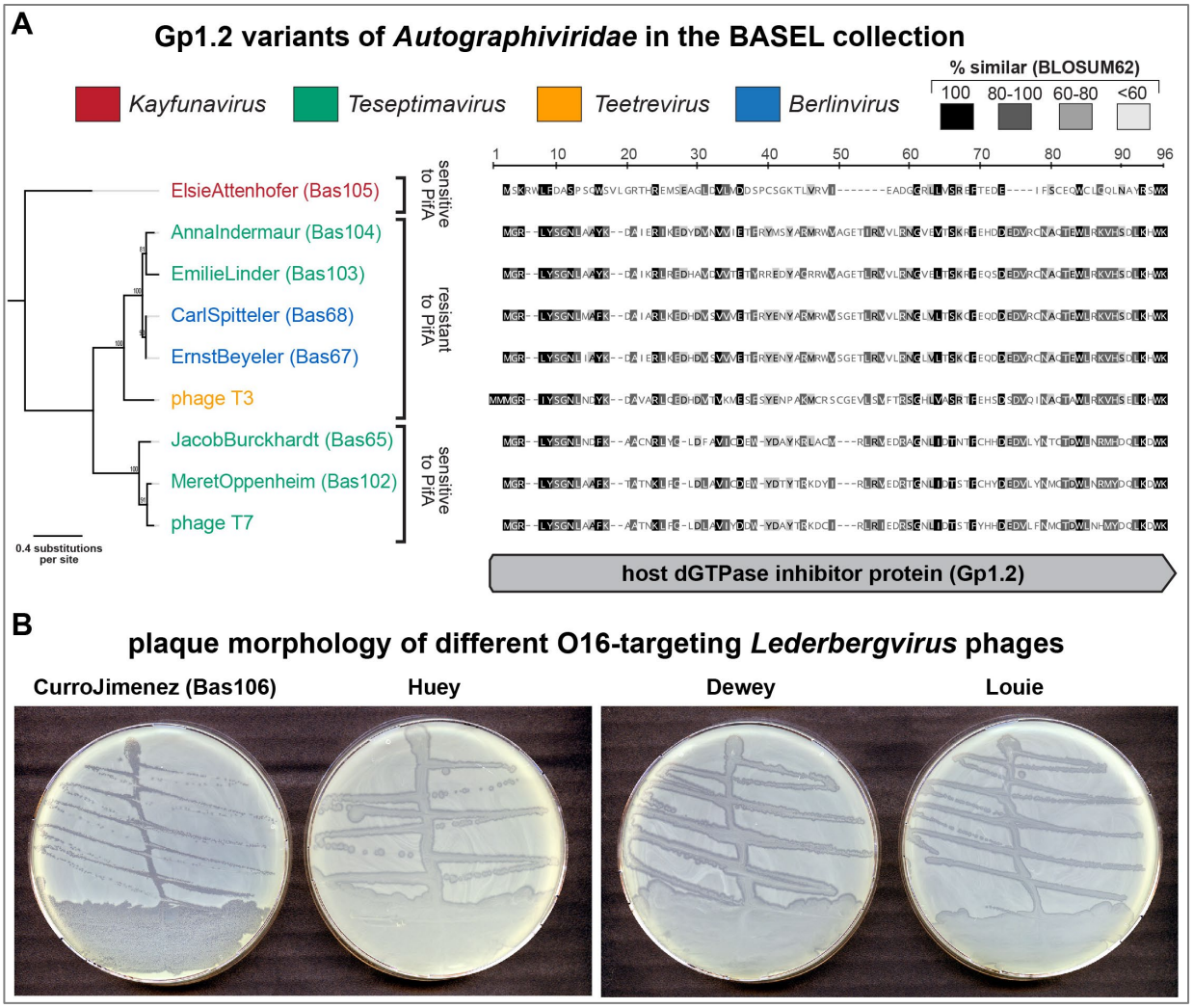
S3 Figure. Supplemental data for Fig. 7. The conserved locus encoding dCTP deaminase, nucleotidyltransferase, and DNA replication genes of all included *Gordonclarkvirinae* isolates is shown in a sequence alignment (see *Materials and Mehtods*). Colors in the sequence identity graph above the alignment indicate the sequence identity at each individual position with green representing 100% identity, greenish brown 30%-99% identity, and red <30% identity. Besides homing endonuclease insertions, the locus is fully syntenic across these isolates (and many other *Gordonclarkvirinae*).



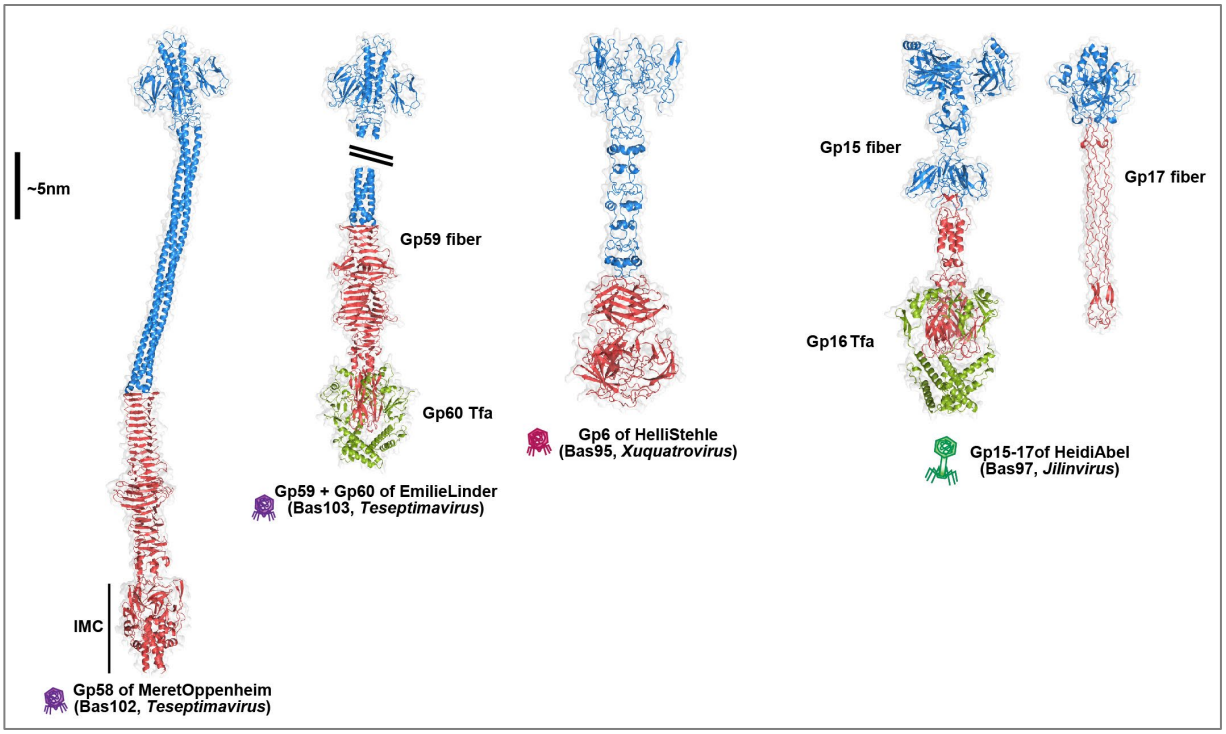
S4 Figure. Supplemental data for Fig. 8. The genomes of our new isolate HelliStehle (Bas95) and *Xuquatrovirus* phages Pondi (GenBank accession OP136151.1) and PTXU04 (NCBI GenBank accession NC_048193.1) are shown in a sequence alignment with the putative small terminase subunit as starting point (see *Materials and Methods*). Colors in the sequence identity graph above the alignment indicate the sequence identity at each individual position with green representing 100% identity, greenish brown 30%-99% identity, and red <30% identity. The pairwise sequence identity of prototypic phage PTXU04 to fellow *Xuquatrovirus* Pondi is 83.8% and to HelliStehle (Bas95) merely 58.3%.



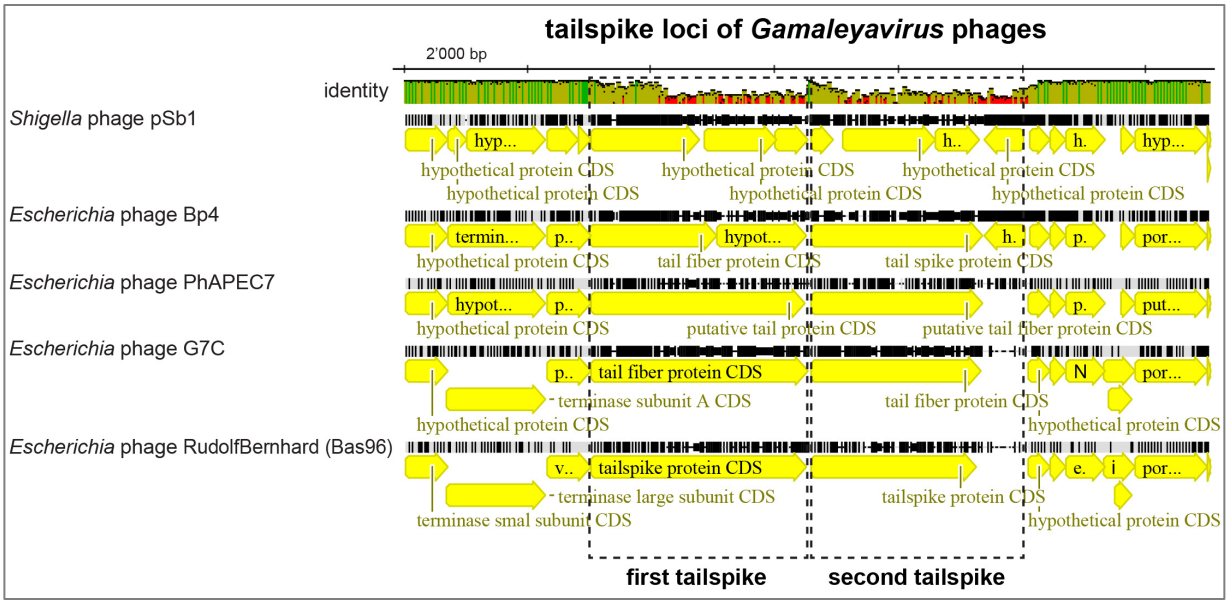
S5 Figure. Supplemental data for Fig. 9. (A) Serial dilutions of *Jilinvirus* HeidiAbel (from left to right) were spotted on a soft agar of *E. coli* K-12 ΔRM *wbbL*(+) and grown overnight at 37°C before imaging. The lysis zones and plaques show no classical signs of lysogen formation such as turbidity or marked growth of (self-immune) lysogens in the middle [108] – compare, e.g., the growth of temperate *Lederbergvirus* phages Huey, Dewey, and Louie in Fig. S6B or of phage Kap1 on a host enabling lysogeny and a host without its preferred integration site in Fig. 3 of Pick et al. [205]. Instead, upon prolonged incubation merely an abundance of dotted individual colonies is found that we typically observe for phages requiring the LPS O-antigen as host receptor which is a huge mutational target [206]. **(B, C)** The illustrations show sequence alignments of the integrase locus of *Iiscvirinae* phages (B) and – encoded slightly downstream – their cI-like repressor locus (C; see *Materials and Methods*). Both genes of interest are highlighted in blue. Colors in the sequence identity graphs above the alignments indicate the sequence identity at each individual position with green representing 100% identity, greenish brown 30%-99% identity, and red <30% identity. The alignments show that only the two *Pseudomonas* phages (top) and *Salmonella* phage PT1 as well as our isolate HeidiAbel encode seemingly intact integrase and cI-like repressor genes. These observations are further discussed in the main text.



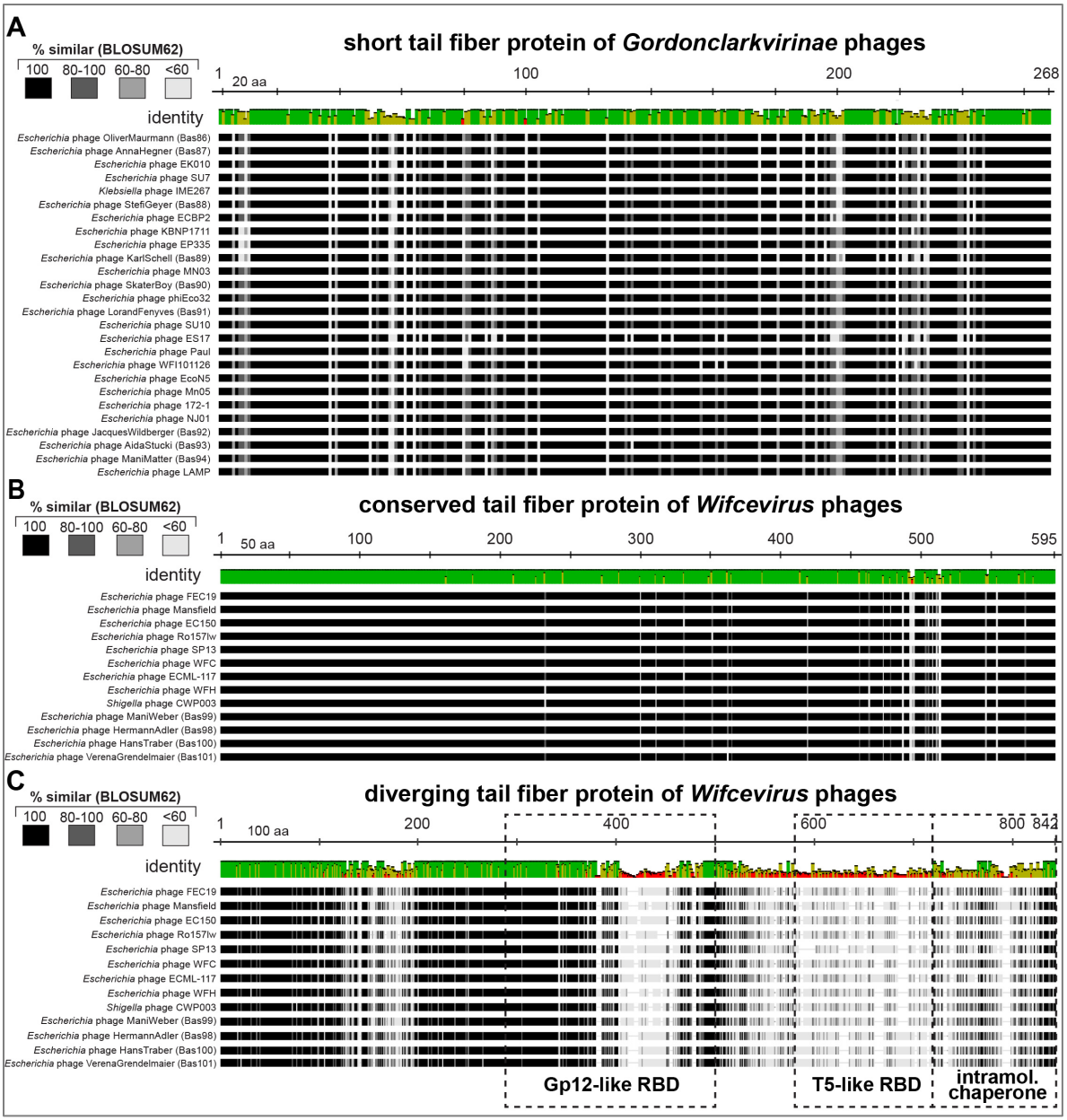
S6 Figure. Supplemental data for Fig. 10. (A) The illustration shows a Maximum-Likelihood phylogeny (left) and a sequence alignment (right) of Gp1.2 orthologs in all tested *Autographiviridae* phages in the complete BASEL collection. Experimental details are explained in *Materials and Methods*. Resistance to the PifA defense system (see Fig. 10E and our previous work [10]) correlates with one specific clade of Gp1.2 homologs (the middle one). **(B)** Plaque morphology of our four *Lederbergvirus* isolates infecting *E. coli* K-12 with restored O16-type O-antigen streaked on *E. coli* K-12 ΔRM *wbbL*(+). While CurroJimenez (Bas106) forms clear plaques of variable size, the other three isolates form highly turbid plaques that are indicative of lysogen formation.



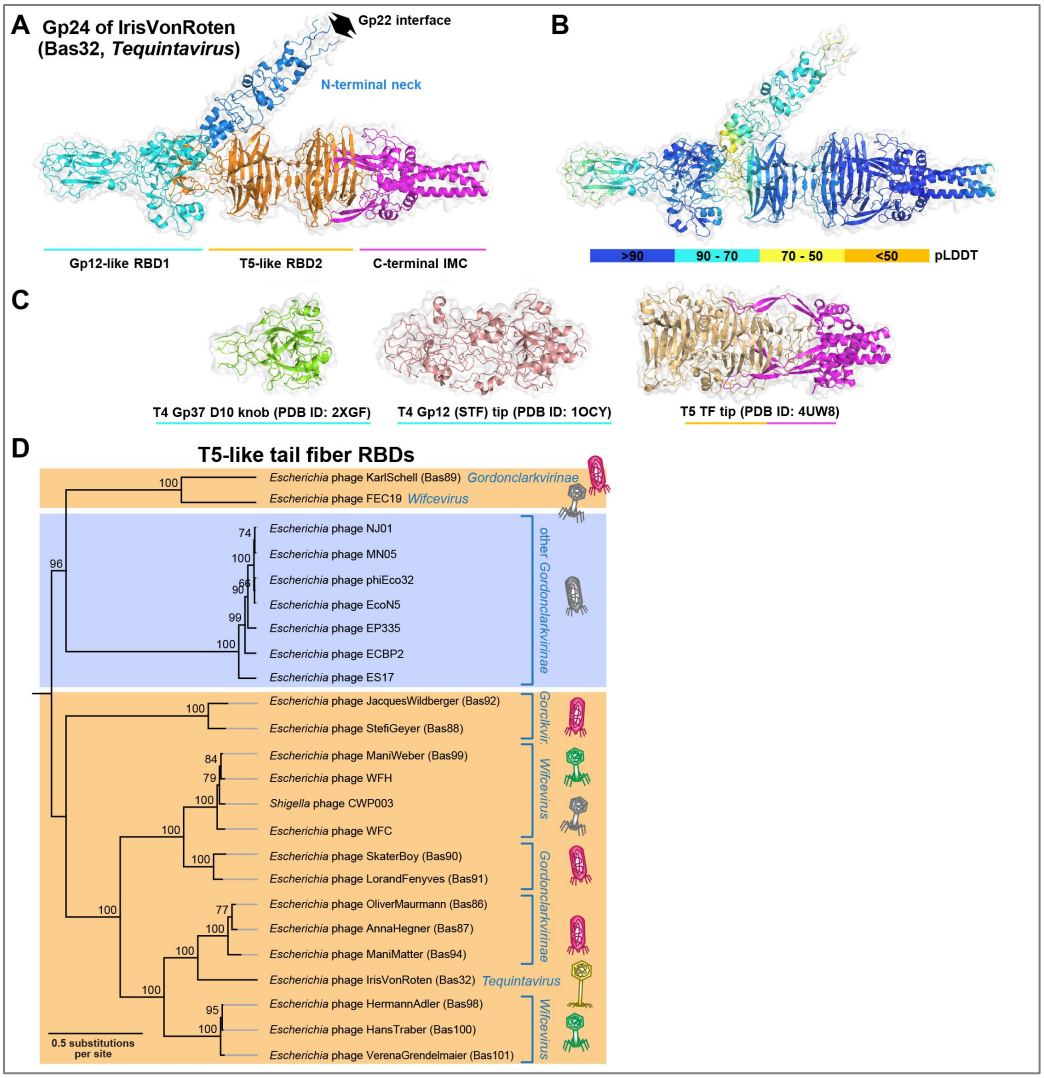
S7 Figure. Supplemental data for Figs. 11 and 12. AlphaFold-Multimer [198] predicted structures of additional tail fibers from selected BASEL phages with a single receptor-binding domain (RBD). Tail fibers of *Teseptimavirus* phages MeretOppenheim (Bas102) and EmilieLinder (Bas103) are formed by a homotrimer of Gp58 or Gp59, respectively. Note that the tail fiber of MeretOppenheim contains a C-terminal intramolecular chaperone (IMC) domain. Conversely, EmilieLinder encodes a separate Mu-like tail fiber assembly protein (Tfa, Gp60; green) downstream of the Gp59 tail fiber that are predicted to form a hexameric complex as observed for the tail fiber of phage Mu [207]. The homotrimeric tail fibers of *Xuquatrovirus* HelliStehle (Bas95) and *Jilinvirus* HeidiAbel (Bas97) feature comparably short fiber stalks. HeidiAbel also encodes a Tfa protein (Gp16; green) that is predicted to form a complex with the tail fiber similar to the complex predicted for EmilieLinder. In addition, HeidiAbel also encodes a separate tail fiber with a distal tip that is similar, but truncated, to the T4 Gp37 long tail fiber and the Gp32 tail fiber of *Wifcevirus* ManiWeber [37] (see Fig. 12C). Based on the size and genetic arrangement, we speculate that the Gp15 fiber of HeidiAbel serves as a lateral tail fiber for initial host recognition while Gp17 forms a short tail fiber for irreversible adsorption triggering viral genome injection (analogous to the much larger T-even myoviruses [10, 73, 74]). In all predicted structures, N-terminal anchor domains and fiber shafts are colored blue with the RBDs colored red. A ~5 nm vertical scale bar is provided for reference and viral morphotypes are annotated by phage icons analogous to Figs. 11A and 11B.



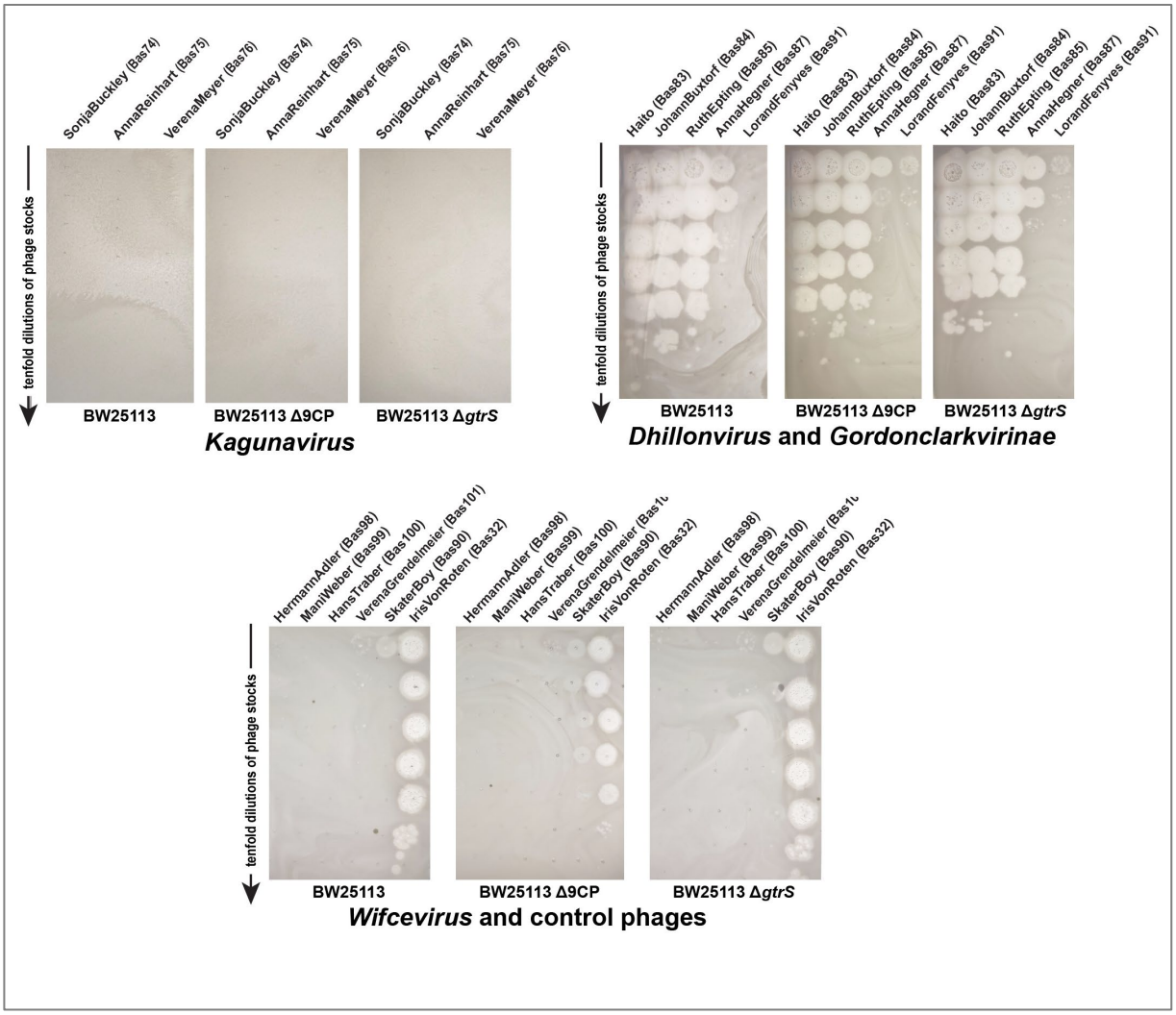
S8 Figure. Supplemental data for Figs. 11 and 12. The illustration shows a sequence alignment of the tailspike locus of different *Gamaleyavirus* phages (compare Fig. 8E; see *Materials and Methods*). Colors in the sequence identity graph above the alignment indicate the sequence identity at each individual position with green representing 100% identity, greenish brown 30%-99% identity, and red <30% identity. The alignment shows that the most 5' parts of the two tailspike loci – encoding the N-terminal ends of the (proximal subunits of the) tailspikes – are highly conserved, likely encoding conserved structural elements that anchor them at the virion. The receptor-binding domains of all tailspikes vary greatly. These observations are further discussed in the main text.



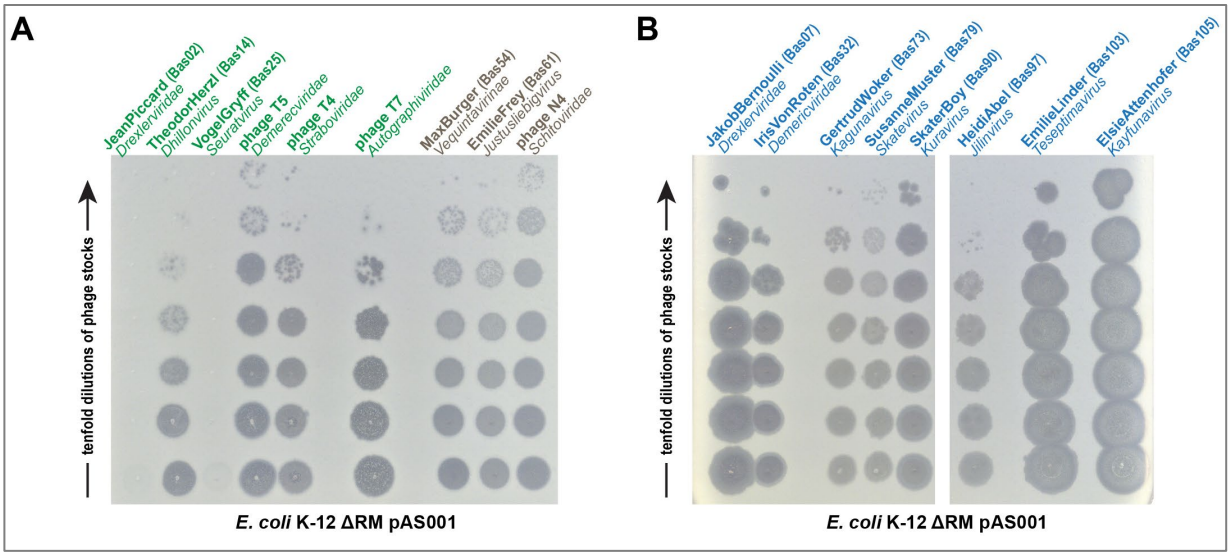
S9 Figure. Supplemental data for Figs. 11 and 12. (**A**) The illustration shows an amino acid sequence alignment of the short tail fiber proteins of all *Gordonclarkvirinae* included in Fig. 7C in the same order (see *Materials and Methods*). It is apparent that short tail fiber proteins are highly conserved across the *Gordonclarkvirinae* subfamily with only minor differences. Across all sequences there are 163 of 267 fully identical sites (60.8%). **(B)** Amino acid sequence alignment of the conserved tail fiber of *Wifcevirus* phages encoded with their tail fiber locus (Gp32 of ManiWeber / Bas99, see Fig. 12C). It is clear from the alignment that all orthologs are very similar in amino acid sequence, suggesting a conserved and invariable function. **(C)** Amino acid sequence alignment of the T-shaped tail fibers of *Wifcevirus* phages (see Figs. 11B, 12B, and S10D). The alignment shows that the N-terminus of all orthologs is highly conserved while the two receptor-binding domains (RBDs) as well as – to lesser extent – the intramolecular chaperone of the T5-like RBD2 vary a lot between sequences. In (A) to (C), colors in the sequence identity graph above the alignment indicate the sequence identity at each individual position with green representing 100% identity, greenish brown 30%-99% identity, and red <30% identity.



S10 Figure. Supplemental data for Figs. 11 and 12. (A) The AlphaFold-Multimer [198] predicted structure of the T-shaped tail fiber of IrisVonRoten (Gp24) was colored to highlight the domain boundaries of this unique dual RBD tail fiber. **(B)** The same structure as in (A) was colored by predicted local distance difference test (pLDDT) confidence scores as generated by AlphaFold to demonstrate the high level of confidence (overall average = 72.6) in the predicted T-shaped tail fiber structure. **(C)** The DALI server [201] identified structural similarities between the Gp12-like RBD1 with crystal structures of the distal knob domain (D10) of the T4 long tail fiber Gp37 [118] (Z-score 8.8, RMSD 3.2, 100 residues) and unsurprisingly with the Gp12 short tail fiber distal tip itself [118] (Z-score 6.2, RMSD 4.1, 127 residues) as well as similarity between the T5-like RBD2 with the distal binding tip and IMC domain of the phage T5 lateral tail fiber [119] (Z-score 6.3, RMSD 4.8, 179 residues). **(D)** Maximum-Likelihood phylogeny of T5-like receptor-binding domains of T-shaped tail fibers together with their intramolecular chaperone domain as identified by structure prediction (see *Materials and Methods*). The phylogeny was midpoint-rooted between a clade of distant *Gordonclarkvirinae* tail fibers at the top (blue) together with two other sequences and all other RBDs mostly belonging to phages presented in this study (orange). Note how RBDs of *Wifcevirus*, *Gordonclarkvirinae*, and *Tequintavirus* isolates are scattered over the phylogeny in small clusters, suggesting frequent horizontal gene transfer as often observed for viral receptor-binding domains [10, 33, 115]. Phage AidaStucki (Bas93, *Gordonclarkvirinae*) was not included due to a large truncation in its intramolecular chaperone domain.



S11 Figure. Supplemental data for Fig. 14. The same serial dilutions as in Fig. 14B were spotted on variants of *E. coli* K-12 BW25113 without restored O-antigen expression (left) and with either deletion of the GtrS glucosyltransferase (middle) or the nine cryptic prophages including CPS-53 (right). In difference to the results shown in Fig. 14B with restored O-antigen expression, the three *Dhillonvirus* phages Bas83-85 readily infect all of the strains lacking O-antigen expression.



S12 Figure. Supplemental data for Materials and Methods. The same serial dilutions as in Figs. 1B (for (A)) and 2B (for (B)) were spotted on the *E. coli* K-12 ΔRM strain with partially restored O16-type O-antigen due to ectopic expression of *wbbL* from plasmid pAS001 (see *Materials and Methods*). Most but not all phages inhibited by a full O16-type O-antigen barrier (green) show robust growth on this host (compare Fig. 1B) while those that can bypass it (brown; compare Fig. 1B) or that depend on it (blue; compare Fig. 2B) grow very well. Notably, a weak inhibition of phage growth had already been observed previously for intermediate restoration of the *E. coli* K-12 O16-type O-antigen [30].

S13 Figure. Supplemental data for Materials and Methods. The illustration shows the genome of phage HeidiAbel (Bas97; see Fig. 8F) with annotations highlighted in different colors. Blue annotations had been predicted by Pharokka in a similar functional context but could be improved by GAPS while red annotations had only been predicted as hypothetical proteins by Pharokka [179]. The genes highlighted in grey had not been predicted by Pharokka in this setup but were inferred by manual analysis of a whole-genome alignment between HeidiAbel and related phages (see *Materials and Methods*). A tRNA is highlighted in pink. The GAPS output for HeidiAbel based on the Pharokka-annotated genome as input is included as S2 Data.