

## METHODS AND RESOURCES

# Systematic exploration of *Escherichia coli* phage–host interactions with the BASEL phage collection

Enea Maffei<sup>1</sup>✉, Aisyulu Shaidullina<sup>1</sup>✉, Marco Burkholder<sup>1</sup>, Yannik Heyer<sup>1</sup>, Fabienne Estermann<sup>1</sup>, Valentin Druelle<sup>1</sup>, Patrick Sauer<sup>1</sup>, Luc Willi<sup>1</sup>, Sarah Michaelis<sup>2</sup>, Hubert Hilbi<sup>1</sup>, David S. Thaler<sup>1,3</sup>, Alexander Harms<sup>1</sup>\*

**1** Biozentrum, University of Basel, Basel, Switzerland, **2** Institute of Medical Microbiology, University of Zürich, Zürich, Switzerland, **3** Program for the Human Environment, Rockefeller University, New York City, New York, United States of America

✉ These authors contributed equally to this work.

\* [alexander.harms@unibas.ch](mailto:alexander.harms@unibas.ch)



## OPEN ACCESS

**Citation:** Maffei E, Shaidullina A, Burkholder M, Heyer Y, Estermann F, Druelle V, et al. (2021) Systematic exploration of *Escherichia coli* phage–host interactions with the BASEL phage collection. PLoS Biol 19(11): e3001424. <https://doi.org/10.1371/journal.pbio.3001424>

**Academic Editor:** Jeremy Barr, Monash University School of Science, AUSTRALIA

**Received:** March 14, 2021

**Accepted:** September 27, 2021

**Published:** November 16, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pbio.3001424>

**Copyright:** © 2021 Maffei et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** 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 3E,

## Abstract

Bacteriophages, the viruses infecting bacteria, hold great potential for the treatment of multi-drug-resistant bacterial infections and other applications due to their unparalleled diversity and recent breakthroughs in their genetic engineering. However, fundamental knowledge of the molecular mechanisms underlying phage–host interactions is mostly confined to a few traditional model systems and did not keep pace with the recent massive expansion of the field. The true potential of molecular biology encoded by these viruses has therefore remained largely untapped, and phages for therapy or other applications are often still selected empirically. We therefore sought to promote a systematic exploration of phage–host interactions by composing a well-assorted library of 68 newly isolated phages infecting the model organism *Escherichia coli* that we share with the community as the BASEL (BActeriophage SElection for your Laboratory) collection. This collection is largely representative of natural *E. coli* phage diversity and was intensively characterized phenotypically and genetically alongside 10 well-studied traditional model phages. We experimentally determined essential host receptors of all phages, quantified their sensitivity to 11 defense systems across different layers of bacterial immunity, and matched these results to the phages' host range across a panel of pathogenic enterobacterial strains. Clear patterns in the distribution of phage phenotypes and genomic features highlighted systematic differences in the potency of different immunity systems and suggested the molecular basis of receptor specificity in several phage groups. Our results also indicate strong trade-offs between fitness traits like broad host recognition and resistance to bacterial immunity that might drive the divergent adaptation of different phage groups to specific ecological niches. We envision that the BASEL collection will inspire future work exploring the biology of bacteriophages and their hosts by facilitating the discovery of underlying molecular mechanisms as the basis for an effective translation into biotechnology or therapeutic applications.

5F, 6E, 7E, 8D, 9F, 10D, 10G, and 12B–12H as well as all calculations underlying the summary data in these figures are compiled in [S1 Data](#). Genome sequences of BASEL collection phages are available at NCBI GenBank (accessions MZ501046–MZ501113).

**Funding:** This work was supported by Swiss National Science Foundation (SNSF) Ambizione Fellowship PZ00P3\_180085 (to AH), SNSF National Center of Competence in Research (NCCR) AntiResist, and a generous contribution of the Stiftung Emilia-Guggenheim-Schnurr of the Naturforschende Gesellschaft in Basel (NGiB; to AH). AS and VD were supported by Biozentrum PhD Fellowships. S.M. and H.H. were supported by the Institute of Medical Microbiology, University of Zürich. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** Abi, abortive infection; BASEL, BAActeriophage SElection for your Laboratory; EAEC, enteroaggregative *E. coli*; ECA, enterobacterial common antigen; EOP, efficiency of plating; ICTV, International Committee on the Taxonomy of Viruses; LB, Lysogeny Broth; LPS, lipopolysaccharide; PBS, phosphate-buffered saline; RBP, receptor-binding protein; RM, restriction-modification; UPEC, uropathogenic *E. coli*.

## Introduction

Bacteriophages, the viruses infecting bacteria, are the most abundant biological entities on earth with key positions in all ecosystems and carry large part of our planet's genetic diversity in their genomes [1–3]. Out of this diversity, a few phages infecting *Escherichia coli* became classical models of molecular biology with roles in many fundamental discoveries and are still major workhorses of research today [4]. The most prominent of these are the seven “T phages” T1 to T7 ([5]; reviewed in [6]) and bacteriophage lambda [7]. Like most known phages, these classical models are tailed phages or Caudovirales that use characteristic tail structures to bind host surface receptors and to eject their genomes from the virion head into the host cell. Three major virion morphotypes of Caudovirales are known, myoviruses with a contractile tail, siphoviruses with a long and flexible tail, and podoviruses with a very short, stubby tail [2]. While the T phages are all so-called virulent phages that kill their host to replicate at each infection event, lambda is a temperate phage and can either kill the host to directly replicate or decide to integrate into the host's genome as a prophage for transient passive replication by vertical transmission in the so-called lysogen [2,8]. These 2 alternative lifestyles as obligately lytic or as temperate phages have major implications for viral ecology and evolution: While virulent phages have primarily been selected to overcome host defenses and maximize virus replication, temperate phages characteristically encode genes that increase the lysogens' fitness, e.g., by providing additional bacterial immunity systems to fight other phages [8,9].

The ubiquity of phage predation has driven the evolution of a vast arsenal of bacterial immunity systems targeting any step of phage infection [9,10]. Inside host cells, phages encounter 2 lines of defense of which the first primarily comprises restriction-modification (RM) systems or CRISPR-Cas, the bacterial adaptive immunity, that directly attack viral genomes [11]. A second line of defense is formed by diverse abortive infection (Abi) systems that protect the host population by triggering an altruistic suicide of infected cells when sensing viral infections [9–12]. While RM systems and CRISPR-Cas are highly abundant and have been successfully adapted for biotechnology (honored with Nobel Prizes in 1978 and 2020, respectively), the molecular mechanisms underlying the function of collectively abundant, but each individually rare, Abi systems have remained elusive with few exceptions [9–12].

Research on phages has expanded at breathless pace over the last decade with a focus on biotechnology and on clinical applications against bacterial infections (“phage therapy”) [13,14]. Besides or instead of the few traditional model phages, many researchers now employ comparably poorly described, newly isolated phages that are often only used for a few studies and available only in their laboratory. The consequence of this development is a rapidly growing amount of very patchy data such as, e.g., the currently more than 15,000 available unique phage genomes for which largely no linked phenotypic data are available [15]. Despite the value of proof-of-principle studies and of a rich genome database, this lack of systematic, inter-linked data in combination with the diversity of bacteriophages makes it very difficult to gain a mechanistic understanding of phage biology or to uncover patterns in the data that would support the discovery of broad biological principles beyond individual models.

As an example, phage isolates for treating bacterial infections are necessarily chosen largely empirically due to the lack of systematic data about relevant phage properties. Currently, the selection of native phages and their engineering for therapeutic applications primarily focus on a virulent lifestyle, a broad host range, and very occasionally on biofilm- or cell wall-degrading enzymes that are comparably well understood genetically and mechanistically [13,14,16,17]. However, the molecular mechanisms and genetic basis underlying other desired features such as resistance to bacterial immunity systems and, in general, the distribution of all these features across different groups of phages have remained understudied. Given the notable incidence of

treatment failure in phage therapy [18–20], a better understanding of the links between phage taxonomy, genome sequence, and phenotypic properties seems timely to select more effective native phages for therapeutic applications and to expand the potential of phage engineering.

In this work, we therefore present the BASEL (Bacteriophage Selection for your Laboratory) collection as a reference set of 68 newly isolated virulent phages that infect the laboratory strain *E. coli* K-12 and make it accessible to the scientific community. We provide a systematic phenotypic and genomic characterization of these phages alongside 10 classical model phages regarding host receptors, sensitivity and resistance to bacterial immunity, and host range across diverse enterobacteria. Our results highlight clear phenotypic patterns between and within taxonomic groups of phages that suggest strong trade-offs between important bacteriophage traits. These findings greatly expand our understanding of bacteriophage ecology, evolution, and their interplay with bacterial immunity systems. We therefore anticipate that our work will not only establish the BASEL collection as a reference point for future studies exploring fundamental bacteriophage biology but also promote a rational application of phage therapy based on an improved selection and engineering of bacteriophages.

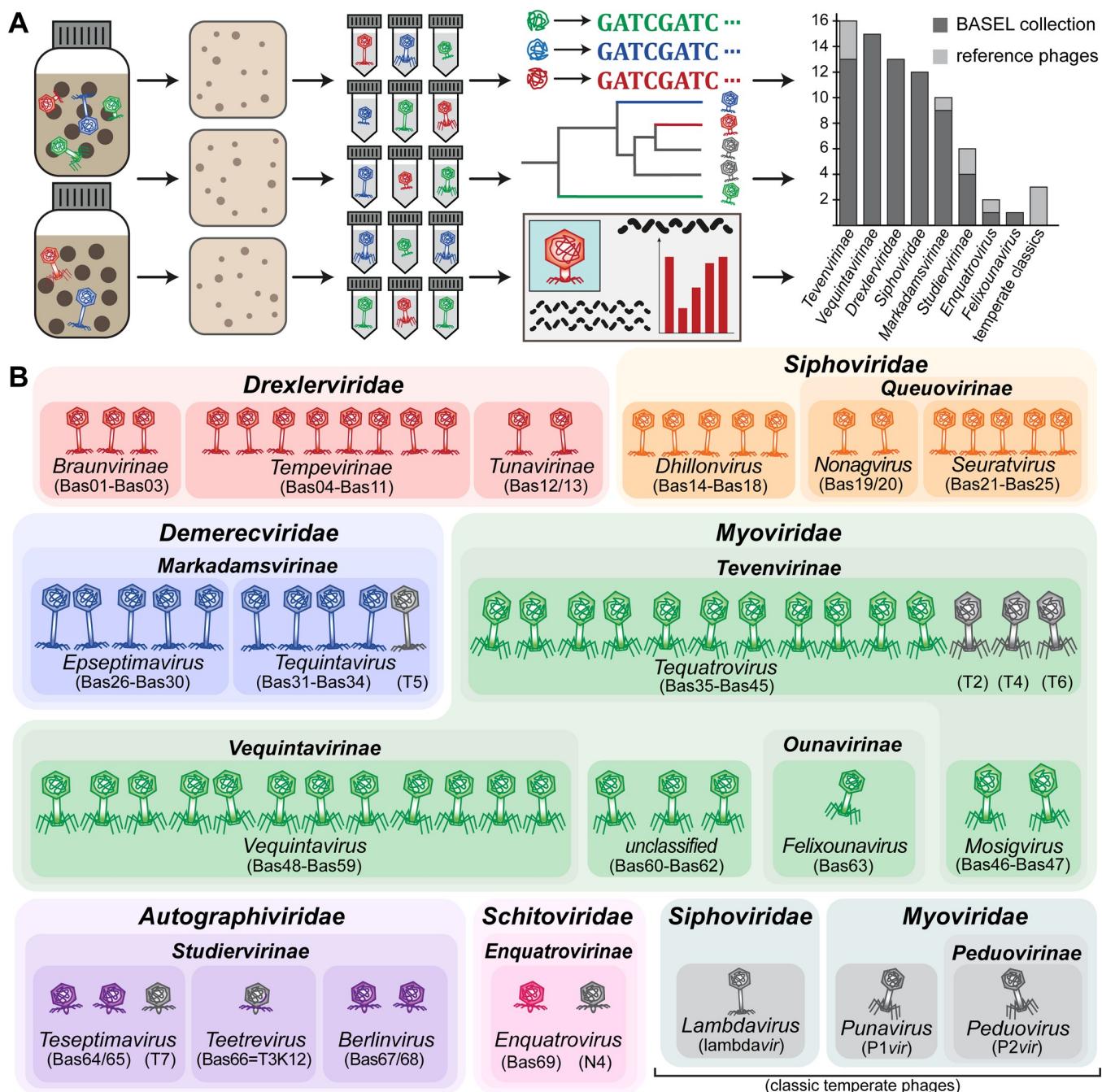
## Results

### Composition of the BASEL collection

The first aim of our study was to generate a collection of new phage isolates infecting the ubiquitously used *E. coli* K-12 laboratory strain that would provide representative insight into the diversity of virulent tailed phages by covering all major groups and containing a suitable selection of minor ones. We therefore started by generating a derivative of *E. coli* K-12 without any of the native barriers that might limit or bias phage isolation unfavorably. This strain, *E. coli* K-12 MG1655 ΔRM, therefore lacks the O-antigen glycan barrier (see below), all restriction systems, as well as the RexAB and PifA Abi systems (see [Materials and methods](#) as well as [S1 Text](#)). *E. coli* K-12 ΔRM was subsequently used to isolate hundreds of phages from environmental samples such as river water or compost, but mostly from the inflow of different sewage treatment facilities ([Fig 1A](#) and [Materials and methods](#)).

Previous bacteriophage isolation studies had already provided deep insight into the diversity of tailed, lytic *E. coli* phages in samples ranging from sewage over diverse natural environments to infant guts or blood and urine of patients in tertiary care [21–27]. Despite the wide diversity of known *E. coli* phages [28], nearly all phage isolates reported in these studies belonged to 5 major groups: Among myoviruses, the so-called “T-even” phages (relatives of T2, T4, and T6; Myoviridae subfamily Tevenvirinae) dominated together with relatives of *E. coli* O157:H7 typing phage rV5 (Myoviridae subfamily Vequintavirinae and related phages). Frequently isolated groups of siphoviruses were relatives of phage T5 (Demerecviridae subfamily Markadamsvirinae), the highly diverse kin of phage T1 (Drexlerviridae family), and the genera *Dhillonvirus*, *Seuratvirus*, and *Nonagvirus* of Siphoviridae. Podoviruses of any kind were rarely reported, and if, then were mostly relatives of T3 and T7 (Autographiviridae subfamily Studiervirinae) isolated using enrichment cultures that are known to greatly favor such fast-growing phages [22,29]. This pattern does not seem to be strongly biased by any given strain of *E. coli* as isolation host because a large, very thorough study using diverse *E. coli* strains reported essentially the same composition of taxonomic groups [21].

Whole-genome sequencing of around 120 different phage isolates from our isolation experiments largely reproduced this pattern, which suggests that several intrinsic limitations of our approach did not strongly affect the spectrum of phages that we sampled ([S2 Text](#)). The virion morphology of representative isolates of each taxonomic group was studied by transmission electron microscopy to confirm the sequence-based classification. After eliminating closely



**Fig 1. Overview of the BASEL collection.** (A) Illustration of the workflow of bacteriophage isolation, characterization, and selection that resulted in the BASEL collection (details in Materials and methods; the bar diagram includes phi92-like phages in Vequintavirinae for simplicity). (B) Taxonomic overview of the bacteriophages included in the BASEL collection and their unique Bas## identifiers. Newly isolated phages are colored by current taxonomic classification, while reference phages are shown in gray. BASEL, BActeriophage SElection for your Laboratory.

<https://doi.org/10.1371/journal.pbio.3001424.g001>

related isolates, the BASEL collection was formed as a set of 68 new phage isolates (Fig 1B; see also [Materials and methods](#) and [S5 Table](#)). We deliberately did not give full proportional weight to highly abundant groups like T-even phages so that the BASEL collection is not truly representative in a narrow quantitative sense. Instead, we included as many representatives of rarely isolated groups as possible to increase the biological diversity that could be an important asset for studying the genetics of phage/host interactions or for unraveling genotype/phenotype relationships. Besides these 68 new isolates (serially numbered as Bas01, Bas02, etc.; Fig 1 and [S5 Table](#)), we included a panel of 10 classical model phages in our genomic and phenotypic characterization and view them as an accessory part of the BASEL collection. Beyond the T phages (except T1 that is a notorious laboratory contaminant [30]), we included well-studied podovirus N4 and obligately lytic mutants of the 3 most commonly studied temperate phages lambda, P1, and P2 [5,7,31,32] (Fig 1; see also [S5 Table](#)).

### Overview: Identification of phage surface receptors

The infection cycle of most tailed phages begins with host recognition by reversible adsorption to a first “primary” receptor on host cells (often a sugar motif on surface glycans) followed by irreversible binding to a terminal or “secondary” receptor directly on the cell surface, which results in DNA injection [33]. Importantly, the adsorption to many potential hosts is blocked by long O-antigen chains on the lipopolysaccharide (LPS) and other exopolysaccharides that effectively shield the cell surface unless they can be degraded or specifically serve as the phage’s primary receptor [33–36]. Tailed phages bind to primary and secondary receptors using separate structures at the phage tail, which display the dedicated receptor-binding proteins (RBPs) in the form of tail fibers or central tail tips (comprehensively reviewed in [33]). Most commonly, surface glycans such as the highly variable O-antigen chains or the enterobacterial common antigen (ECA) are bound as primary receptors, and this interaction may or may not be required for infectivity. Secondary receptors on gram-negative hosts are near-exclusively porin-family outer membrane proteins for siphoviruses and core LPS sugar structures for podoviruses, while myoviruses were found to use either one of the two depending on phage subfamily or genus [33,37,38]. Notably, no host receptor is known for the vast majority of phages that have been studied, but understanding the genetic basis and molecular mechanisms underlying host recognition as the major determinant of phage host range is a crucial prerequisite for host range engineering or a rational application of phage therapy [39]. Since tail fibers and other host recognition modules are easily identified in phage genomes, we used phage receptor specificity as a model to demonstrate the usefulness of systematic phenotypic data with the BASEL collection as a key to unlock information hidden in the genome databases.

We therefore first experimentally determined the essential host receptor(s) of all phages in the BASEL collection before analyzing the phages’ genomes for the mechanisms underlying receptor specificity. Briefly, the dependence on surface proteins was assessed by plating each phage on a set of more than 50 single-gene mutants of *E. coli* K-12 ([S1 Table](#)) or whole-genome sequencing of spontaneously resistant bacterial clones (see [Materials and methods](#)). The role of host surface glycans was quantified with *waaG* and *waaC* mutants that display different truncations of the LPS core, a *wbbL*(+) strain with restored O16-type O-antigen expression, and a *wecB* mutant that is specifically deficient in production of the ECA [40–42] (Fig 2A, [S1 Table](#), and [Materials and methods](#)).

### Overview: Phenotyping of sensitivity/resistance to bacterial immunity systems

Each bacterial strain encodes a unique repertoire of a few very common and a larger number of rarer, strain-specific immunity systems [11], but it is unknown how far this diversity

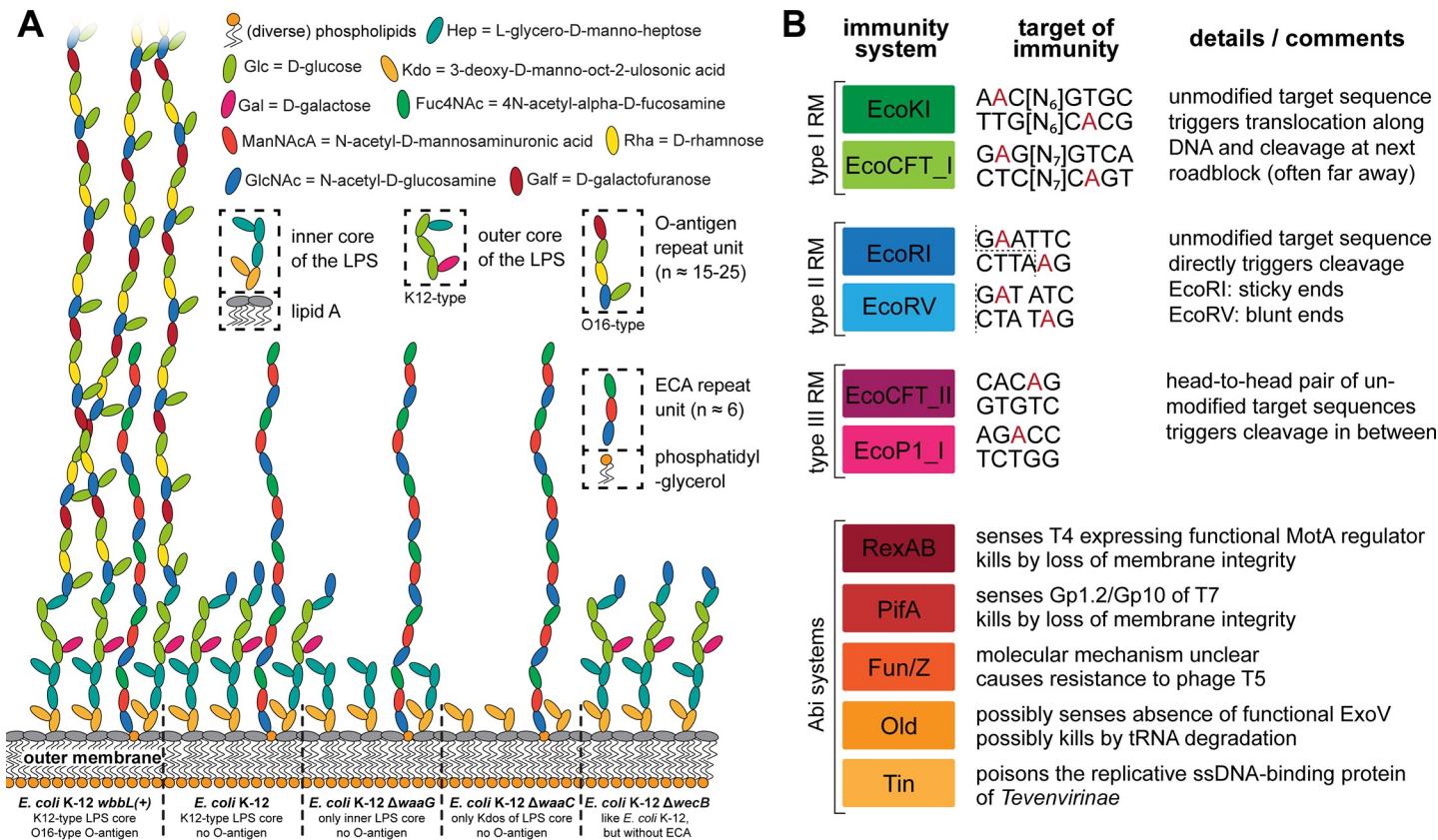
impacts the isolation of phages or their efficacy in therapeutic applications. A systematic view on the sensitivity and resistance of different phage groups to different immunity systems might therefore enable us to select or engineer phages with a higher and more reliable potency for phage therapy or biotechnology. As an example, previous work showed that T-even phages or the Queuovirinae subfamily of Siphoviridae (*Seuratvirus* and *Nonagvirus* genera) exhibit broad resistance to RM systems due to the hypermodification of cytosines or guanosines in their genomes, respectively, which could be an interesting target for phage engineering [43,44]. However, it is unknown whether this mechanism of RM resistance (or any other viral anti-immunity function) actually results in a measurably broader phage host range.

We therefore systematically quantified the sensitivity of all phages of the BASEL collection against a panel of 11 different immunity systems and scored their infectivity on a range of pathogenic enterobacteria that are commonly used as model systems (see [Materials and methods](#) and [Fig 2B](#)). Briefly, we tested 6 RM systems by including each two type I, type II, and type III systems that differ in the molecular mechanisms of DNA modification and cleavage [10,45–47] ([Fig 2B](#)). Besides these, we included the most well-studied Abi systems of *E. coli*, RexAB of the lambda prophage, PifA of the F-plasmid, as well as the Old, Tin, and Fun/Z systems of the P2 prophage ([Fig 2B](#)). Previous work suggested that RexAB and PifA sense certain proteins of phages T4 and T7, respectively, to trigger host cell death by membrane depolarization [10]. Conversely, Old possibly senses the inhibition of RecBCD/ExoV during lambda infections and might kill by tRNA degradation, Fun/Z abolishes infections by phage T5 via an unknown mechanism, and Tin poisons the replicative ssDNA binding protein of T-even phages [48]. Beyond *E. coli* K-12, we quantified the infectivity of the BASEL collection on uropathogenic *E. coli* (UPEC) strains UTI89 and CFT073, enteroaggregative *E. coli* (EAEC) strain 55989, alternative laboratory strain *E. coli* B REL606, and *Salmonella enterica* subsp. *enterica* serovar Typhimurium strains 12023s and SL1344 (see [Materials and methods](#) and [S1 Table](#)).

## Properties of the Drexlerviridae family

Phages of the Drexlerviridae family (previously also known as T1 superfamily [28]) are small siphoviruses with genome sizes of ca. 43 to 52 kb ([Fig 3A–3E](#)). The BASEL collection contains 13 new Drexlerviridae isolates that are broadly spread out across the various subfamilies and genera of this family ([Fig 3D](#)). Though it has never been directly demonstrated, it seems likely that the Drexlerviridae use their lateral tail fibers in the same way as their larger cousins of the Markadamsvirinae (see below) to contact specific O-antigen glycans as primary receptors without depending on this interaction for host recognition [49] ([Fig 3A](#)). Consistently, the locus encoding the lateral tail fibers is very diverse among Drexlerviridae, and the proteins forming these tail fibers are only highly related at the far N-terminus of the distal subunits where these are likely attached to the virion ([S1C Fig](#)). Among the Drexlerviridae that we isolated, only JakobBernoulli (Bas07) shows robust plaque formation on *E. coli* K-12 MG1655 with restored O16-type O-antigen expression, suggesting that its lateral tail fibers can use this O-antigen as primary receptor ([Fig 3A](#) and [3E](#)).

Just like the much larger Markadamsvirinae siphoviruses (see below), Drexlerviridae phages use a small set of outer membrane porins as their secondary/final receptors for irreversible adsorption and DNA injection. To the best of our knowledge, previous work had only identified the receptors of T1 and IME18 (FhuA), IME347 (YncD), IME253 (FepA), and of LL5 as well as TLS (TolC), while for an additional phage, RTP, no protein receptor could be identified [38,50–52]. Using available single-gene mutants, we readily determined the terminal receptor of 11 out of our 13 Drexlerviridae phages as FhuA, BtuB, YncD, and TolC ([Fig 3C](#) and [3D](#)) but failed for 2 others, AugustePiccard (Bas01) and JeanPiccard (Bas02). However, whole-genome

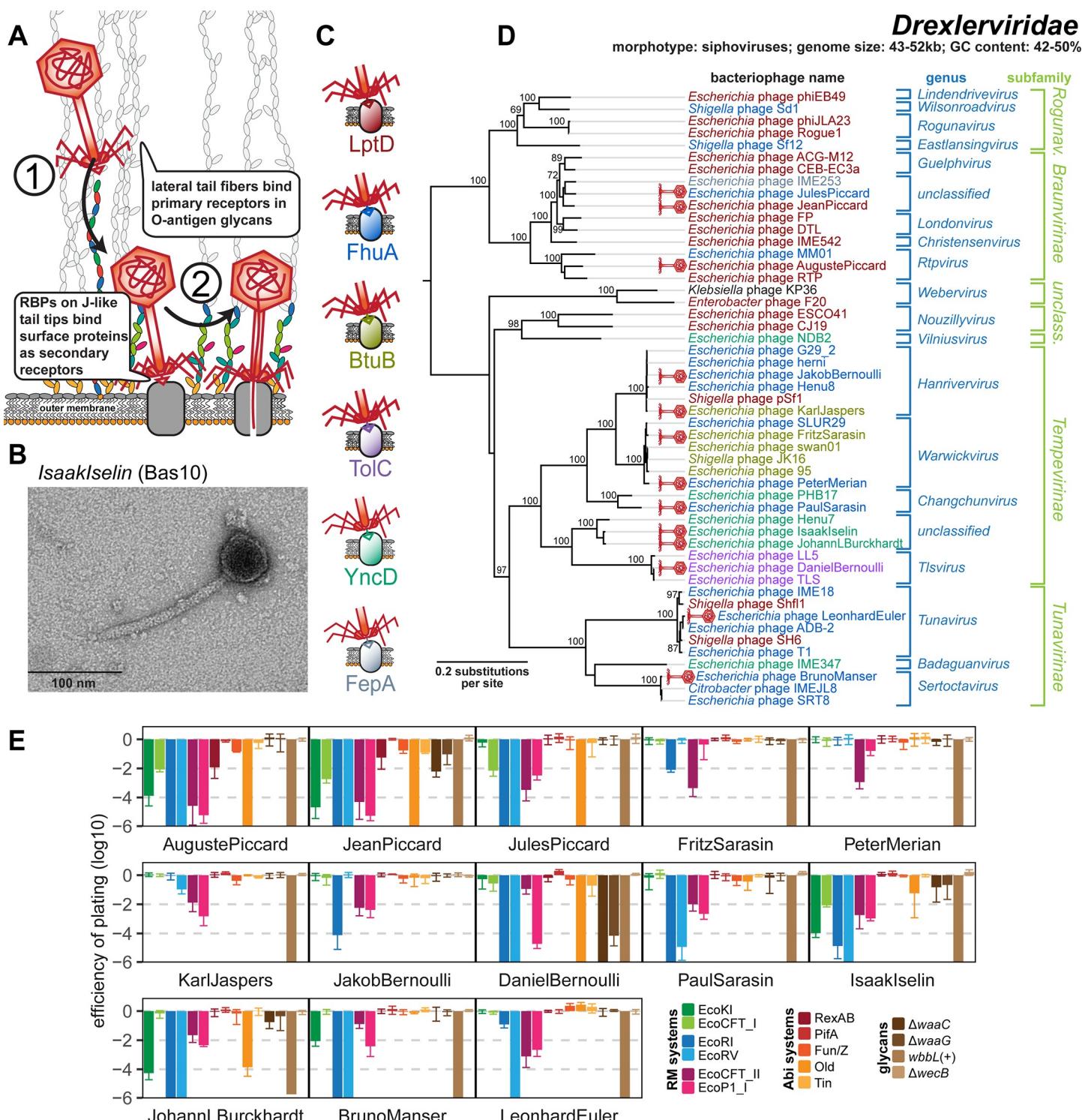


**Fig 2. Overview of *E. coli* surface glycan variants and the immunity systems used in this study.** (A) The surface glycans of different *E. coli* K-12 MG1655 variants are shown schematically (details in running text and Materials and methods). Note that the *E. coli* K-12 MG1655 laboratory wild type does not merely display the K12-type core LPS (classical rough LPS phenotype) but also the most proximal D-glucose of the O16-type O-antigen. (B) Key features of the 6 RM systems (each 2 of type I, type II, and type III) and the 5 Abi systems used for the phenotyping of this study are summarized schematically. Recognition sites of RM systems have either been determined experimentally or were predicted in REBASE (red nucleotides: methylation sites; dotted lines: cleavage sites) [45–47,129]. The Abi systems have been characterized to very different extent but constitute the most well-understood representatives of these immunity systems of *E. coli* [10,48]. Abi, abortive infection; ECA, enterobacterial common antigen; LPS, lipopolysaccharide; RM, restriction-modification.

<https://doi.org/10.1371/journal.pbio.3001424.g002>

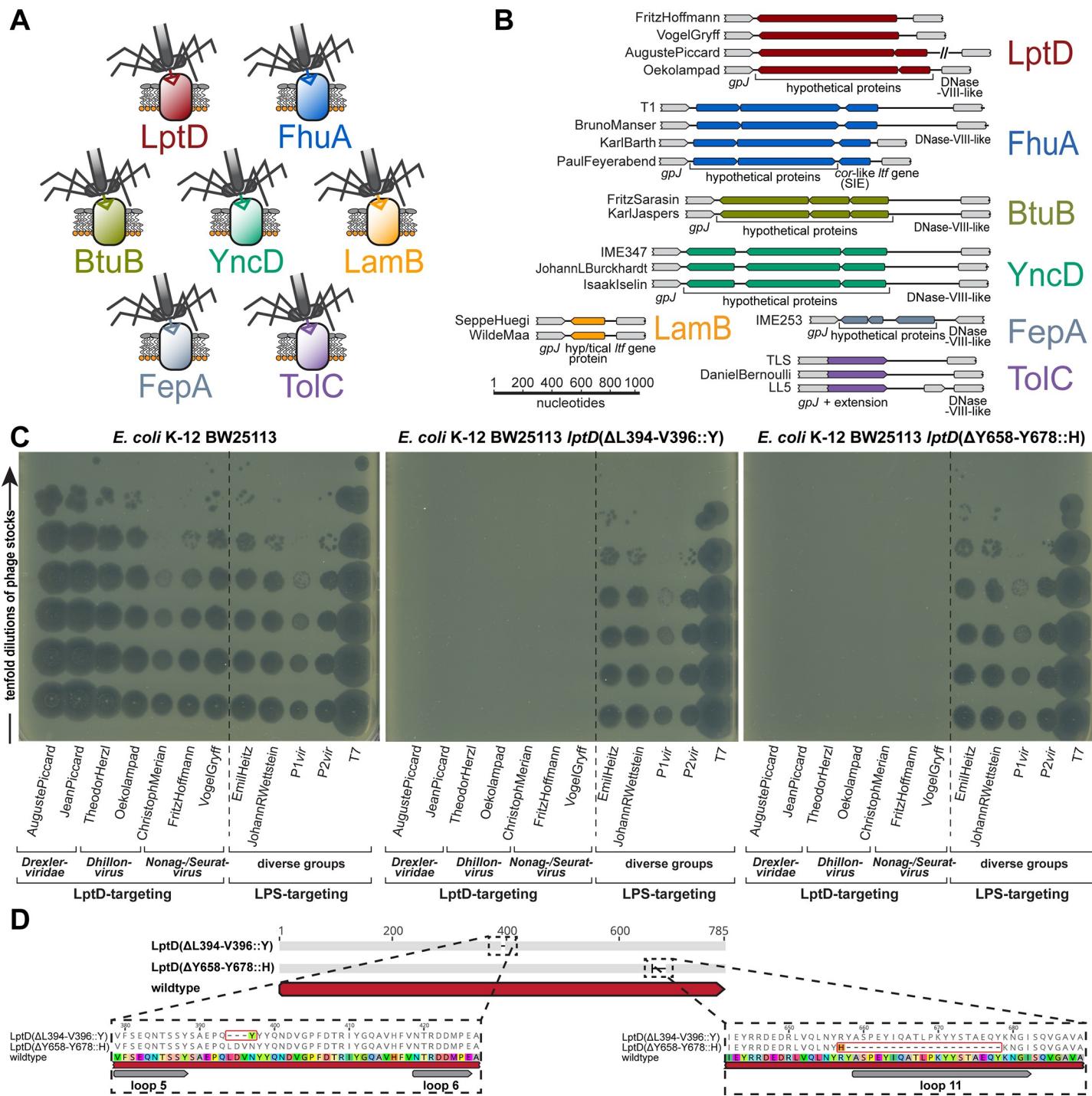
sequencing of spontaneously resistant *E. coli* mutants showed that resistance was linked to mutations in the gene coding for LptD, the LPS export channel [34], strongly suggesting that this protein was the terminal receptor of these phages (Fig 4 and Materials and methods).

Similar to the central tail fibers of Markadamsvirinae, the RBPs of Drexerviridae are thought to be displayed at the distal end of a tail tip protein related to well-studied GpJ of bacteriophage lambda [49,52,53]. The details of this host recognition module had remained elusive, though it was suggested that (like for T5 and unlike for lambda) dedicated RBPs are noncovalently attached to the J-like protein and might be encoded directly downstream of the gpJ homologs together with cognate superinfection exclusion proteins [52,53]. By comparing the genomes of all Drexerviridae phages with experimentally determined surface protein receptors, we were able to match specific allelic variants of these bona fide RBP loci to each known surface receptor of this phage family (Fig 4A and 4B). Notably, for the 3 known phages targeting TolC including DanielBernoulli (Bas08), the RBP locus is absent and apparently functionally replaced by a C-terminal extension of the GpJ-like tail tip protein that probably directly mediates receptor specificity like GpJ of bacteriophage lambda (Fig 4B) [54,55]. Interestingly, similar RBP loci with homologous alleles are also found at the same genomic locus in small Siphoviridae of *Dhillonvirus*, *Nonagvirus*, and *Seuratvirus* genera (see below) where they



**Fig 3. Overview of Drexerviridae phages.** (A) Schematic illustration of host recognition by Drexerviridae. (B) Representative TEM micrograph of phage IsaakIselin (Bas10). (C) Color code of terminal receptor specificity. (D) Maximum-Likelihood phylogeny of Drexerviridae based on several core genes with bootstrap support of branches shown if >70/100. Newly isolated phages of the BASEL collection are highlighted by red phage icons, and the determined or proposed terminal receptor specificity is highlighted at the phage names using the color code highlighted in (C). The phylogeny was rooted based on a representative phylogeny including *Dhillonvirus* sequences as outgroup (S1A Fig). (E) The results of quantitative phenotyping experiments with Drexerviridae phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as EOP. Data points and error bars represent average and standard deviation of at least 3 independent experiments. Raw data and calculations are available in S1 Data. BASEL, BActeriophage SElection for your Laboratory; EOP, efficiency of plating; RBP, receptor-binding protein; TEM, transmission electron microscopy.

<https://doi.org/10.1371/journal.pbio.3001424.g003>



**Fig 4. Shared bona fide RBP loci of small siphoviruses mediate specificity to terminal receptors including LptD.** (A) The 7 identified receptors of small siphoviruses are shown with a color code that is also used to annotate demonstrated or predicted receptor specificity in the phylogenies of Figs 3D, 5D, and 5E. (B) We show representative bona fide RBP loci that seem to encode the receptor specificity of the small siphoviruses studied in this work (same color code as in (A)). Note that the loci linked to each receptor are very similar while the genetic arrangement differs considerably between loci linked to different terminal host receptors (see also S1B Fig). (C) Whole-genome sequencing of bacterial mutants exhibiting spontaneous resistance to 7 small siphoviruses with no previously known receptor revealed different mutations or small deletions in the essential gene *lptD* that encodes the LptD LPS export channel. Top agar assays with 2 representative mutants in comparison to the ancestral *E. coli* K-12 BW25113 strain were performed with serial 10-fold dilutions of 12 different phages (undiluted high-titer stocks at the bottom and increasingly diluted samples toward the top). Both mutants display complete resistance to the 7 small siphoviruses of diverse genera within Drexlerviridae and Siphoviridae families that share the same bona fide RBP module (S1B Fig) while no other phage of the BASEL collection was affected. We excluded indirect effects, e.g., via changes in the LPS composition in the *lptD* mutants, by confirming that 5 LPS-targeting phages of diverse families (see below) showed full infectivity on all strains. (D) The amino acid

sequence alignment of wild-type LptD with the 2 mutants highlighted in (A) shows that resistance to LptD-targeting phages is linked to small deletions in or adjacent to regions encoding extracellular loops as defined in previous work [150], suggesting that they abolish the RBP–receptor interaction. BASEL, BActeriophage SElection for your Laboratory; LPS, lipopolysaccharide; RBP, receptor-binding protein.

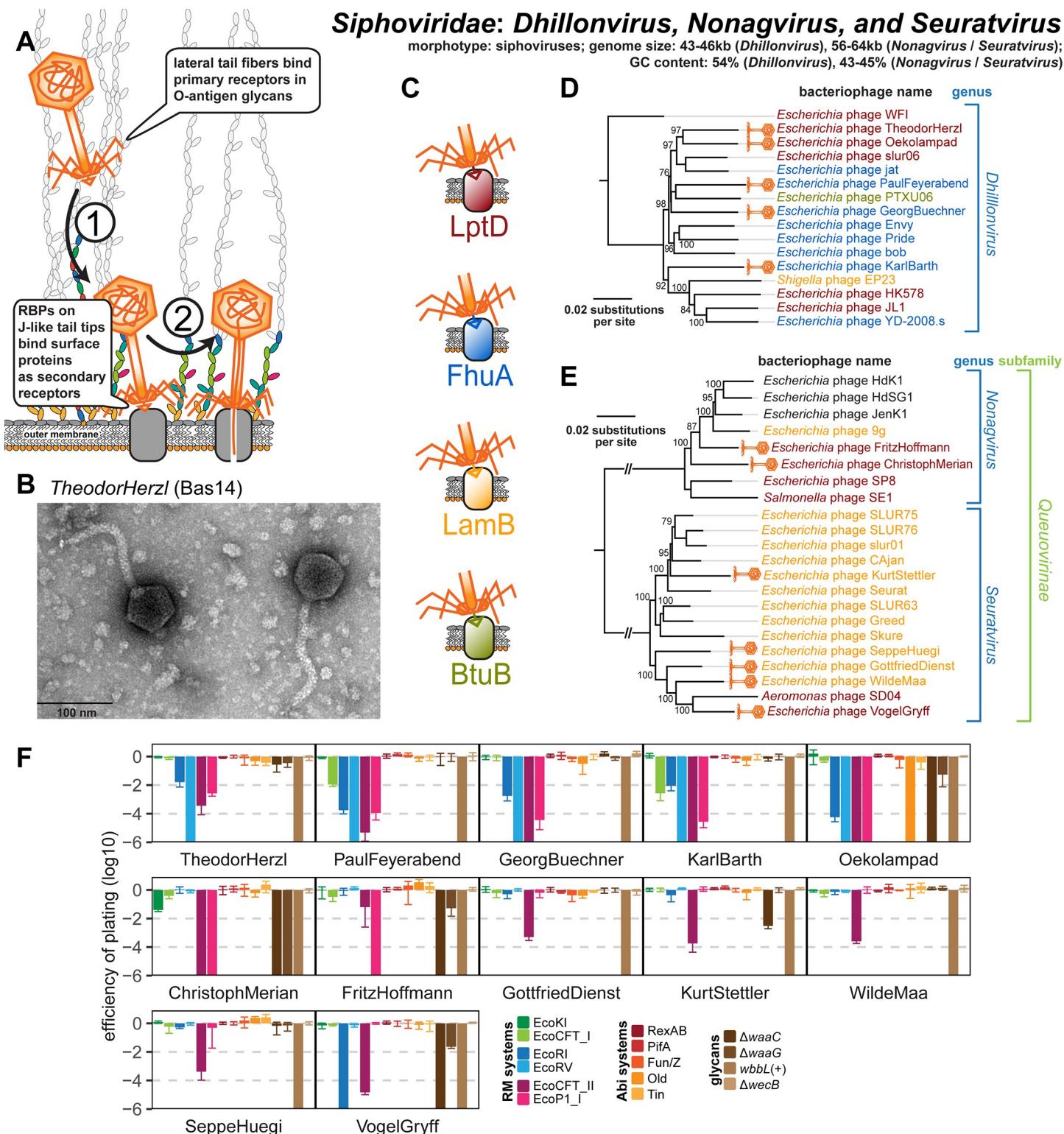
<https://doi.org/10.1371/journal.pbio.3001424.g004>

also match known receptor specificity without exception (Figs 4, 5, and S1B). Pending experimental validation, we therefore conclude that these distantly related groups of small siphoviruses share a common, limited repertoire of RBPs that enables receptor specificity of most representatives to be predicted in silico. As an example, it seems clear that phage RTP targets LptD as terminal receptor (Figs 3D and S1D). The poor correlation of predicted receptor specificity with the phylogenetic relationships of small siphoviruses (Figs 3D, 5D, and 5E) is indicative of frequent horizontal transfer of these RBP loci. This highlights the modular nature of this host recognition system, which might enable the targeted engineering of receptor specificity to generate “designer phages” for different applications as previously shown for siphoviruses infecting *Listeria* [14,56]. LptD would be a particularly attractive target for such engineering because it is strictly essential under all conditions, highly conserved, and heavily constrained due to multiple interactions that are critical for its functionality [34].

Once inside the host cell, Drexlerviridae phages are highly diverse in their sensitivity or resistance to diverse bacterial immunity systems (Fig 3E). While few representatives like Fritz-Sarasin (Bas04) or PeterMerian (Bas05) are highly resistant, most Drexlerviridae are very sensitive to any kind of RM systems sometimes including the type I machineries that are largely unable to inhibit any other phage that we tested (Fig 3E). Previous work on specific representatives suggested that the N6-adenine and/or C5-cytosine methyltransferases abundantly encoded by Drexlerviridae might counteract host restriction systems [22]. While this would be intuitive, we found no simple pattern linking methyltransferase genes or other genomic features to restriction resistance/sensitivity, suggesting that differences of this phenotype are driven by unknown factors or nuances in the activity or specificity of the DNA methyltransferases.

### Properties of Siphoviridae genera *Dhillonvirus*, *Nonagvirus*, and *Seuratvirus*

Phages of the *Dhillonvirus*, *Seuratvirus*, and *Nonagvirus* genera within the Siphoviridae family are small siphoviruses that are superficially similar to Drexlerviridae and have genomes with characteristic size ranges of 43 to 46 kb (*Dhillonvirus*), 56 to 61 kb (*Seuratvirus*), and 56 to 64 kb (*Nonagvirus*; see Fig 5A–5F). Our 12 isolates included in the BASEL collection are spread out broadly across the phylogenetic ranges of these genera (Fig 5D and 5E). Like Drexlerviridae (and Markadamsvirinae; see below), we suggest that they also recognize glycan motifs at the O-antigen as their primary receptor on different host strains (Fig 5A). This notion is strongly supported by the remarkable variation exhibited by the different genomes at the lateral tail fiber locus (S2A Fig), as observed previously [21,22]. None of these phages can infect *E. coli* K-12 MG1655 with restored O16-type O-antigen expression, but (like for Drexlerviridae) some require an intact LPS core for infectivity (Fig 5F). Experimental identification of the terminal receptor of all small Siphoviridae confirmed that receptor specificity seems to be encoded by the same system of bona fide RBP loci downstream of gpJ as for the Drexlerviridae (Figs 3C, 3D, 4A, 4B, and 5C–5E). Many of these phages target LptD or FhuA as terminal receptors, while others, unlike any Drexlerviridae, bind to LamB (Fig 5C–5E). Notably, 3 related *Nonagvirus* phages encode a distinct bona fide RBP module that could not be matched to any known terminal receptor (S2B Fig), suggesting that these phages are targeting a different protein.



**Fig 5. Overview of Siphoviridae genera Dhillonvirus, Nonagvirus, and Seuratvirus.** (A) Schematic illustration of host recognition by small siphoviruses. (B) Representative TEM micrograph of phage TheodorHerzl (Bas14). (C) Color code of terminal receptor specificity (same as in Figs 3 and 4). (D) Maximum-Likelihood phylogeny of the Dhillonvirus genus based on a whole-genome alignment with bootstrap support of branches shown if >70/100. Newly isolated phages of the BASEL collection are highlighted by orange phage icons and the determined or proposed terminal receptor specificity is highlighted at the phage names using the color code highlighted in (C). The phylogeny was rooted between phage WFI and all others based on a representative phylogeny including Drexlerviridae sequences as outgroup

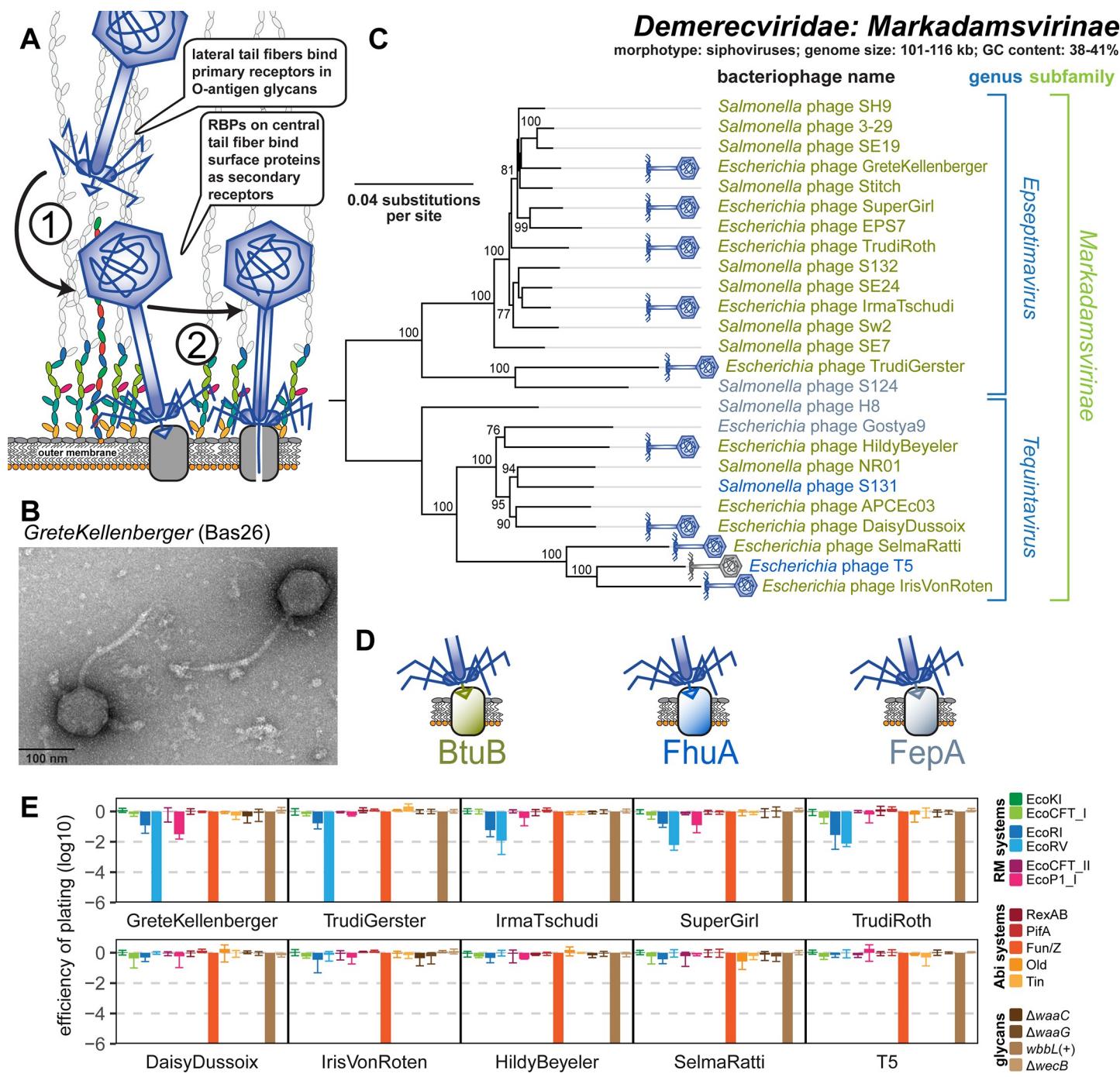
(S1A Fig). (E) Maximum-Likelihood phylogeny of the *Nonagvirus* and *Seuratvirus* genera based on a whole-genome alignment with bootstrap support of branches shown if >70/100. Newly isolated phages of the BASEL collection are highlighted by orange phage icons, and the determined or proposed terminal receptor specificity is highlighted at the phage names using the color code highlighted in (C) (see S2B Fig for the 3 phages without coloring). The phylogeny was rooted between the 2 genera that belong both to the Queuovirinae subfamily of Siphoviridae. (F) The results of quantitative phenotyping experiments with *Dhillonvirus*, *Nonagvirus*, and *Seuratvirus* phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as EOP. Data points and error bars represent average and standard deviation of at least 3 independent experiments. Raw data and calculations are available in S1 Data. Abi, abortive infection; BASEL, BActeriophage SElection for your Laboratory; EOP, efficiency of plating; RBP, receptor-binding protein; RM, restriction-modification; TEM, transmission electron microscopy.

<https://doi.org/10.1371/journal.pbio.3001424.g005>

In difference to *Dhillonvirus* phages that are relatives of Drexlerviridae [21] (see also S1A Fig), *Nonagvirus* and *Seuratvirus* phages are closely related genera in the newly formed Queuovirinae subfamily with genomes that are 10 to 15 kb larger than those of the other small siphoviruses in the BASEL collection (S5 Table) [57,58]. This difference in genome size is largely due to genes encoding their signature feature, the 7-deazaguanine modification of 2'-deoxyguanosine (dG) in their genomes into 2'-deoxy-7-cyano-7-deazaguanosine (dpreQ<sub>0</sub>) for *Seuratvirus* phages and into 2'-deoxyarchaosine (dG<sup>+</sup>) for *Nonagvirus* phages [43] (S2C Fig). These modifications were shown to provide considerable (dG<sup>+</sup>) or at least moderate (dpreQ<sub>0</sub>) protection against restriction of genomic DNA of *Seuratvirus* phage CAjan and *Nonagvirus* phage 9g in vitro, although chemical analyses showed that only around one-third (CAjan) or one-fourth (9g) of the genomic dG content is modified [43] (S2C Fig). Consistently, we found that Queuovirinae phages are remarkably resistant to type I and type II RM systems in our infection experiments, particularly if compared to other small siphoviruses of *Dhillonvirus* or Drexlerviridae groups (Figs 3E and 5F). Among *Nonagvirus* and *Seuratvirus* phages, only Christoph-Merian (Bas19, EcoKI) and VogelGryff (Bas25, EcoRI) show some sensitivity to these RM systems. However, all Queuovirinae phages are highly sensitive to type III RM systems (Fig 5F), possibly as a consequence of the much larger number of type III RM recognition sites in their genomes compared to those of the other types of RM systems (S5 Table; see also Fig 2B).

## Properties of Demerecviridae: Markadamsvirinae

Phages of the Markadamsvirinae subfamily of the Demerecviridae family (commonly known as T5-like phages [28]) are large siphoviruses with genomes of 101 to 116 kb length (Fig 6A–6E). Our 9 new isolates in the BASEL collection (plus well-studied phage T5) are well representative of the 2 major genera *Eseptimavirus* and *Tequintavivirus* and their subclades (Fig 6C). These phages characteristically use their lateral tail fibers to bind each one or a few types of O-antigen very specifically as their primary host receptor, but this interaction is not essential for hosts like *E. coli* K-12 laboratory strains that do not express an O-antigen barrier [49] (Fig 6A). As observed previously [49,59], the diversity of O-antigen glycans is reflected in a high genetic diversity at the lateral tail fiber locus of the Markadamsvirinae of the BASEL collection (S3A Fig). Only one of these phages, IrisVonRoten (Bas32), can infect *E. coli* K-12 with restored expression of O16-type O-antigen, suggesting that it can recognize this glycan as its primary receptor (Fig 6E). However, several other isolates can infect different *E. coli* strains with smooth LPS or even *Salmonella*. Similar to what was proposed for the different small siphoviruses described above, the terminal receptor specificity of T5 and relatives is determined by dedicated RBPs attached noncovalently to the tip of a straight central tail fiber [49]. Compared to their smaller relatives, the repertoire of terminal receptors among Markadamsvirinae seems limited—the vast majority target BtuB (shown for, e.g., EPS7 [60] and S132 [61]), and each a few others bind FhuA (T5 itself [49] and probably S131 [61]) or FepA (H8 [62] and probably S124 [61]). Sequence analyses of the RBP locus allowed us to predict the terminal receptor of all remaining Markadamsvirinae, confirming that nearly all target BtuB (Figs 6C, 6D, and S3B).



**Fig 6. Overview of Demerecviridae subfamily Markadamsvirinae.** (A) Schematic illustration of host recognition by T5-like siphoviruses. (B) Representative TEM micrograph of phage GreteKellenberger (Bas26). (C) Maximum-Likelihood phylogeny of the Markadamsvirinae subfamily of Demerecviridae based on several core genes with bootstrap support of branches shown if >70/100. Phages of the BASEL collection are highlighted by little phage icons, and the determined or proposed terminal receptor specificity is highlighted at the phage names using the color code highlighted in (D). The phylogeny was rooted between the Epsseptimavirus and Tequintavirius genera. (D) Color code of terminal receptor specificity (same as in Figs 3–5). (E) The results of quantitative phenotyping experiments with Markadamsvirinae phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as EOP. Data points and error bars represent average and standard deviation of at least 3 independent experiments. Raw data and calculations are available in S1 Data. Abi, abortive infection; BASEL, BAteriophage SElection for your Laboratory; EOP, efficiency of plating; RBP, receptor-binding protein; RM, restriction-modification; TEM, transmission electron microscopy.

<https://doi.org/10.1371/journal.pbio.3001424.g006>

Previous work showed that phage T5 is largely resistant to RM systems, and this was thought to be due to elusive DNA protection functions encoded in the early-injected genome region largely shared by T5 and other Markadamsvirinae [63]. Surprisingly, we find that this resistance to restriction is a shared feature only of the *Teqvintavirus* genus, while their sister genus *Epseptimavirus* shows detectable yet variable sensitivity particularly to the type II RM system EcoRV (Fig 6E). It seems likely that this difference between the genera is due to the largely different number of EcoRV recognition sites in *Epseptimavirus* genomes (70 to 90 sites; S5 Table) and *Teqvintavirus* genomes (4 to 18 sites; S5 Table). This observation suggests that efficient DNA ligation and RM site avoidance [64] play important roles in the restriction resistance of T5 and relatives besides their putative DNA protection system. Remarkably, all Markadamsvirinae are invariably sensitive to the Fun/Z Abi system as shown for T5 previously [48].

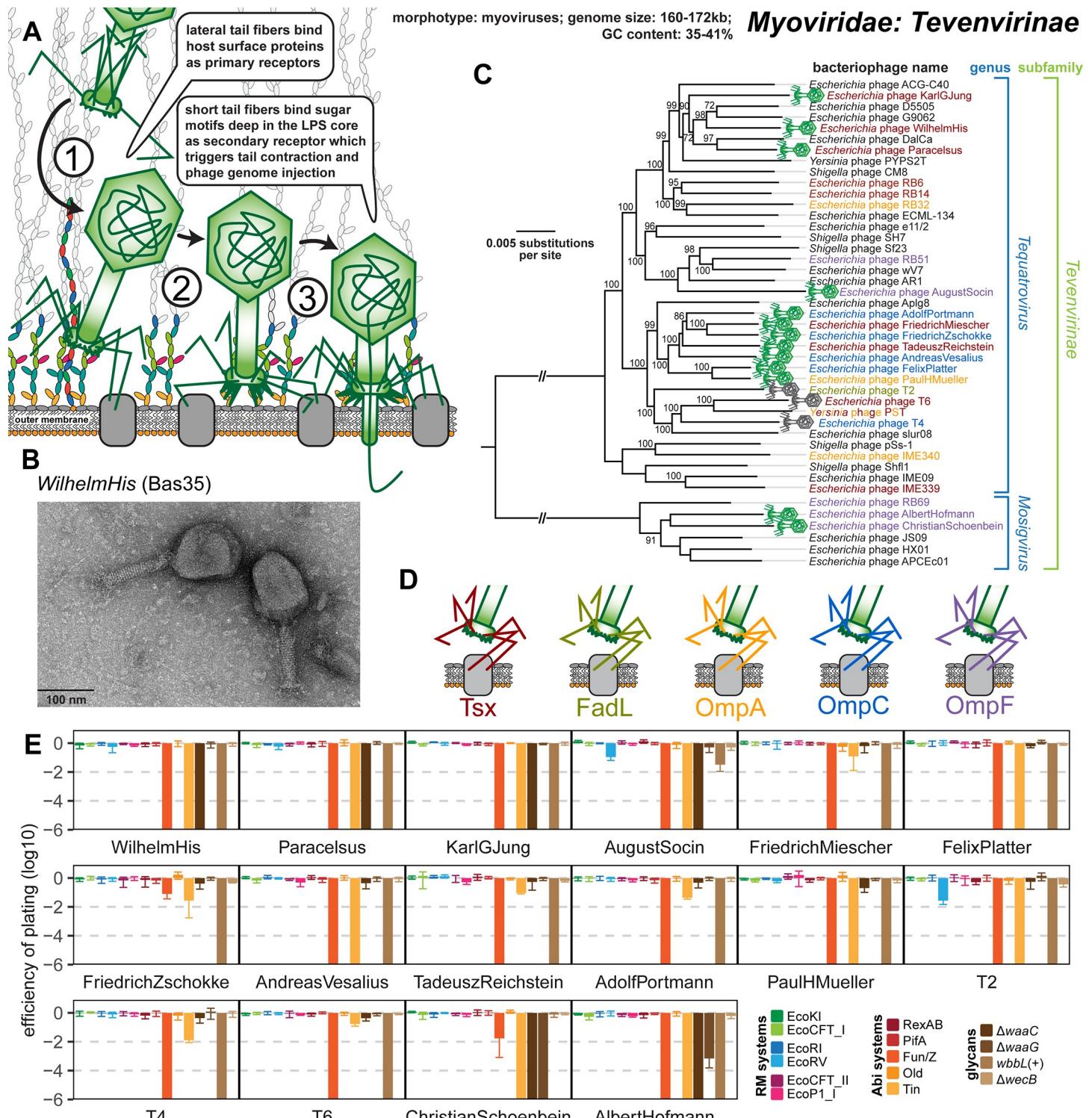
### Properties of Myoviridae: Tevenvirinae

Phages of the Tevenvirinae subfamily within the Myoviridae family are large myoviruses with characteristic prolate capsids and genomes of 160 to 172 kb size that infect a wide variety of gram-negative hosts [28] (Fig 7A–7E). The kinked lateral tail fibers of these phages contact primary receptors on the bacterial cell surface that are usually surface proteins like OmpC, Tsx, and FadL for the original “T-even” phages T4, T6, and T2, respectively, but can also be sugar motifs in the LPS like for T4 when OmpC is not available [65] (Fig 7A). Robust interaction with the primary receptor unpins the short tail fibers from the myovirus baseplate, which enables their irreversible adsorption to terminal receptors in the LPS core. Consequently, the tail sheath contracts like a syringe and the phage tail penetrates the cell envelope to eject the viral genome [33] (Fig 7A).

The 13 new T-even isolates in the BASEL collection mostly belong to different groups of the *Tequatrovirus* genus that also contains well-studied reference phages T2, T4, and T6, but 2 isolates were assigned to the distantly related *Mosigvirus* genus (Fig 7C). As expected, the infectivity of all tested Tevenvirinae shows strong dependence on each one of a small set of *E. coli* surface proteins that have previously been described as primary receptors of this group, i.e., Tsx, OmpC, OmpF, OmpA, and FadL [65] (Fig 7C and 7D). Interestingly, while for some of these phages the absence of the primary receptor totally abolished infectivity, others still showed detectable yet greatly reduced plaque formation (S4A Fig). It seems likely that these differences are caused by the ability of some Tevenvirinae lateral tail fibers to contact several primary receptors such as, e.g., OmpC and the truncated *E. coli* B LPS core in case of T4 [65,66].

Specificity for the secondary receptor depends on the short tail fibers that, in case of T4, target the lipid A–Kdo region deep in the enterobacterial LPS core [67]. Given that this region is still present in the *waaC* mutant, the most deep-rough mutant of *E. coli* K-12 that is viable (Fig 2A), it is unsurprising that T4 and some other Tevenvirinae did not seem to show a dependence on the LPS core in our experiments (Fig 7E). However, some *Tequatrovirus* isolates and all tested *Mosigvirus* isolates required an intact inner core of the host LPS for infectivity (Fig 7E). This phenotype is correlated with an alternative allele of the short tail fiber gene that varies between Tevenvirinae phages irrespective of their phylogenetic position (S4B Fig). We therefore suggest that those phages encoding the alternative allele express a short tail fiber that targets parts of the LPS core above the lipid A–Kdo region (S4B and S4C Fig; see also Fig 2A).

Besides receptor specificity, the phenotypes of our diverse Tevenvirinae phages were highly homogeneous. The hallmark of this Myoviridae subfamily is a high level of resistance to DNA-targeting immunity like RM systems because of cytosine hypermodification (hydroxymethyl-



**Fig 7. Overview of the Myoviridae subfamily Tevenvirinae.** (A) Schematic illustration of host recognition by T-even myoviruses. (B) Representative TEM micrograph of phage WilhelmHis (Bas35). (C) Maximum-Likelihood phylogeny of the Tevenvirinae subfamily of Myoviridae based on a curated whole-genome alignment with bootstrap support of branches shown if  $>70/100$ . The phylogeny was rooted between the *Tequattrovirus* and *Mosigvirus* genera. Phages of the BASEL collection are highlighted by little phage icons and experimentally determined primary receptor specificity is highlighted at the phage names using the color code highlighted in (D). Primary receptor specificity of T-even phages depends on RBPs expressed either as a C-terminal extension of the distal half fiber (T4 and other OmpC-targeting phages) or as separate small fiber tip adhesins [65], but sequence analyses of the latter remained ambiguous. We therefore only annotated experimentally determined primary receptors (see also S4A Fig) [51,65]. (D) Color code of primary receptor specificity. (E) The results of quantitative phenotyping experiments with Tevenvirinae phages

regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as EOP. Data points and error bars represent average and standard deviation of at least 3 independent experiments. Raw data and calculations are available in [S1 Data](#). Abi, abortive infection; BASEL, BActeriophage SElection for your Laboratory; EOP, efficiency of plating; LPS, lipopolysaccharide; RBP, receptor-binding protein; RM, restriction-modification; TEM, transmission electron microscopy.

<https://doi.org/10.1371/journal.pbio.3001424.g007>

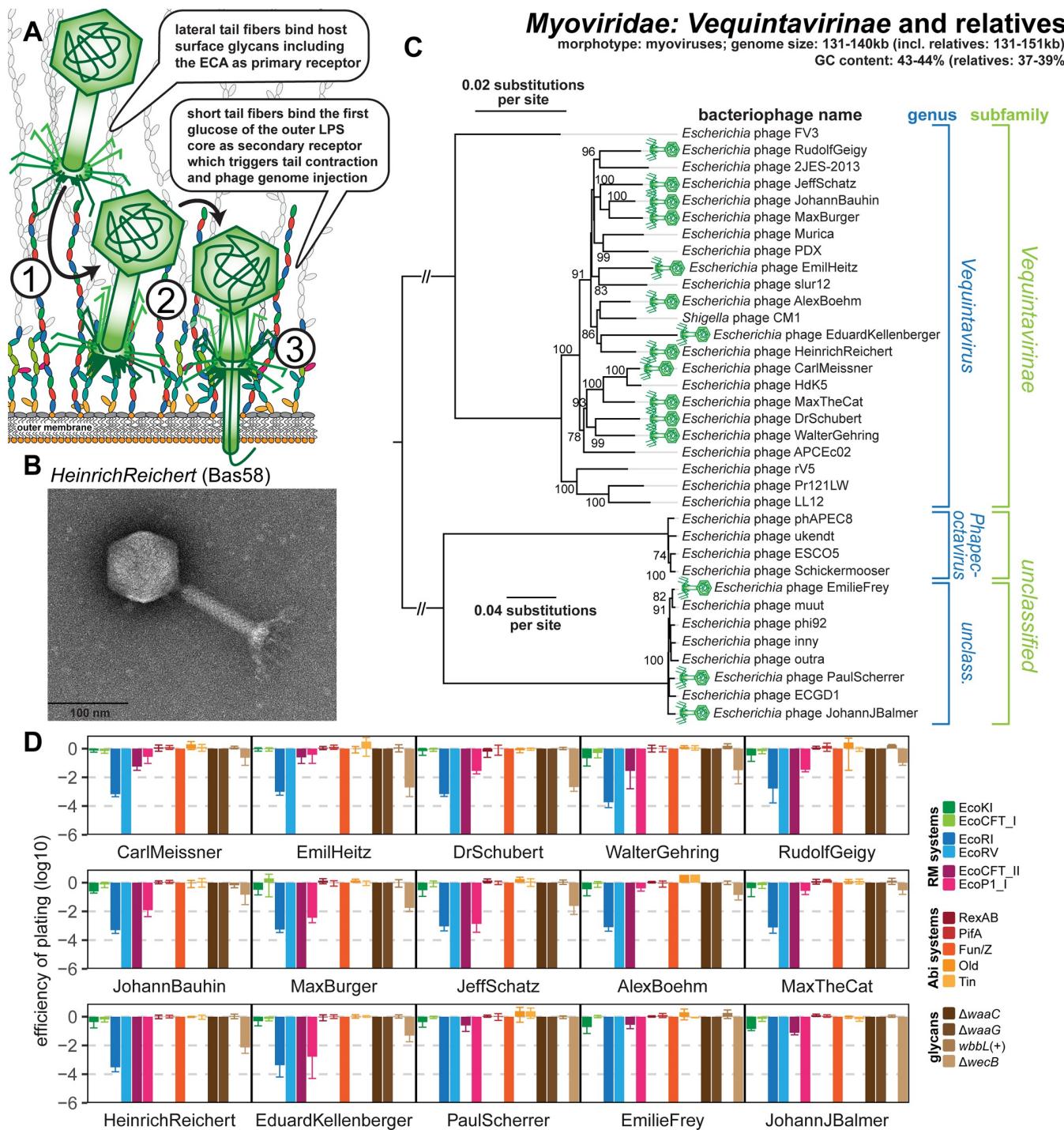
glucosylated for *Tequatrovirus* and hydroxymethyl-arabinosylated for *Mosigvirus*) [44,68]. Consistently, no or only very weak sensitivity to any of the 6 RM systems was detected for any tested T-even phage ([Fig 7E](#)). The weak sensitivity to EcoRV observed for AugustSocin (Bas38) and phage T2 does, unlike for Markadamsvirinae (see above), not correlate with a higher number of recognition sites and therefore possibly depends on another feature of these phages such as, e.g., differences in DNA ligase activity ([S5 Table](#)). Conversely, all tested T-even phages were sensitive to the Fun/Z and Tin Abi systems ([Fig 7E](#)). Tin was previously shown to specifically target the DNA replication of T-even phages [48], and we found that indeed all Tevenvirinae but no other tested phage are sensitive to this Abi system ([Fig 7E](#)). None of the tested T-even phages was sensitive to RexAB, the iconic Abi system targeting *rIIA/B* mutants of phage T4 [10], suggesting that RexAB is generally unable to inhibit wild-type T-even phages ([Fig 7E](#); see [S5A](#) and [S5B Fig](#) for the RexAB sensitivity of a T4 *rIIA/B* mutant).

## Properties of Myoviridae: Vequintavirinae and relatives

Phages of the Vequintavirinae subfamily within the Myoviridae family are myoviruses with genomes of 131 to 140 kb size that characteristically encode 3 different sets of lateral tail fibers [21,69] ([Fig 8A–8D](#)). Besides Vequintavirinae of the *Vequintavirus* genus, this feature is shared by 2 groups of phages that are closely related to these Vequintavirinae sensu stricto: One group forms a cluster around phage phAPEC8 (genus *Phaecoctavivirus* within Myoviridae), the other group is their sister clade including phage phi92 [70] and has so far remained unclassified ([Fig 8C](#)). Despite not being Vequintavirinae by current taxonomic classification, we are covering them together due to their considerable similarities and propose to classify the phi92-like phages (including 3 new isolates) as *Nonagintaduovirus* genus within the Vequintavirinae.

Three different, coexpressed lateral tail fibers have been directly visualized in the cryogenic electron microscopy structure of phi92, and large orthologous loci are found in all Vequintavirinae and relatives [21,69,70]. The considerable repertoire of glycan-hydrolyzing protein domains in these tail fiber proteins has been compared to a “nanosized Swiss army knife” and might be responsible for the exceptionally broad host range of these phages and their ability to infect even diverse capsulated strains of enterobacteria [21,50,69,70]. However, it is far from clear which genes code for which components of the different lateral tail fibers. Three large lateral tail fiber genes of Vequintavirinae exhibit clear allelic variation between genomes and might therefore encode proteins at the distal end of the 3 tail fibers that contact variable host structures ([S5C Fig](#)). However, most of the lateral tail fiber locus is highly conserved including multiple genes encoding proteins with sugar-binding or glycan-hydrolyzing domains ([S5C Fig](#)). This suggests that good part of these tail fibers might not have evolved to counter host variability but probably target conserved glycans on the host surface. Consistently, all tested Vequintavirinae and relatives seem to use the ECA as their shared primary receptor because their infectivity is compromised on a *wecB* knockout ([Fig 8A and 8D](#)). If true, the Vequintavirinae and relatives would be effectively monovalent on *E. coli* K-12 and might primarily use their “nanosized Swiss army knife” of tail fibers to bind and overcome capsules or other more specialized exopolysaccharides on diverse enterobacteria [69,70].

The terminal receptor of Vequintavirinae has been unraveled genetically for phage LL12, a close relative of rV5 ([Fig 8C](#)), and seems to be at the first, heptose-linked glucose of the LPS outer core, which is shared by all *E. coli* core LPS types [50,71] ([Fig 8A](#)). Consistently, we



**Fig 8. Overview of the Myoviridae subfamily Vequintavirinae and relatives.** (A) Schematic illustration of host recognition by Vequintavirinae and related myoviruses. (B) Representative TEM micrograph of phage HeinrichReichert (Bas58). (C) Maximum-Likelihood phylogeny of the Vequintavirinae subfamily of Myoviridae and relatives based on a curated whole-genome alignment with bootstrap support of branches shown if >70/100. The phylogeny was rooted between the Vequintavirinae sensu stricto and the 2 closely related, formally unclassified groups at the bottom. Newly isolated phages of the BASEL collection are highlighted by green phage icons. (D) The results of quantitative phenotyping experiments with Vequintavirinae and their phi92-like relatives regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as EOP. Data points and error bars represent average and standard deviation of at least 3 independent experiments. Raw data and calculations are available in S1 Data. Abi, abortive infection; BASEL, BActeriophage SElection for your Laboratory; ECA, enterobacterial common antigen; EOP, efficiency of plating; LPS, lipopolysaccharide; RM, restriction-modification; TEM, transmission electron microscopy.

<https://doi.org/10.1371/journal.pbio.3001424.g008>

found that all Vequintavirinae in the BASEL collection are completely unable to infect the *waaC* and *waaG* mutants with core LPS defects that result in the absence of this sugar (Fig 8D; see also Fig 2A). This observation would be compatible with the idea that all tested Vequintavirinae including the phi92-like phages use the same secondary receptor. Indeed, the two very similar short tail fiber paralogs of Vequintavirinae sensu stricto do not show any considerable allelic variation between genomes (S5C Fig; unlike, e.g., the Teenvirinae short tail fibers presented in S4B and S4C Fig) and are also closely related to the single short tail fiber protein of phi92-like phages or *Phapcoctavirus* (S5D Fig).

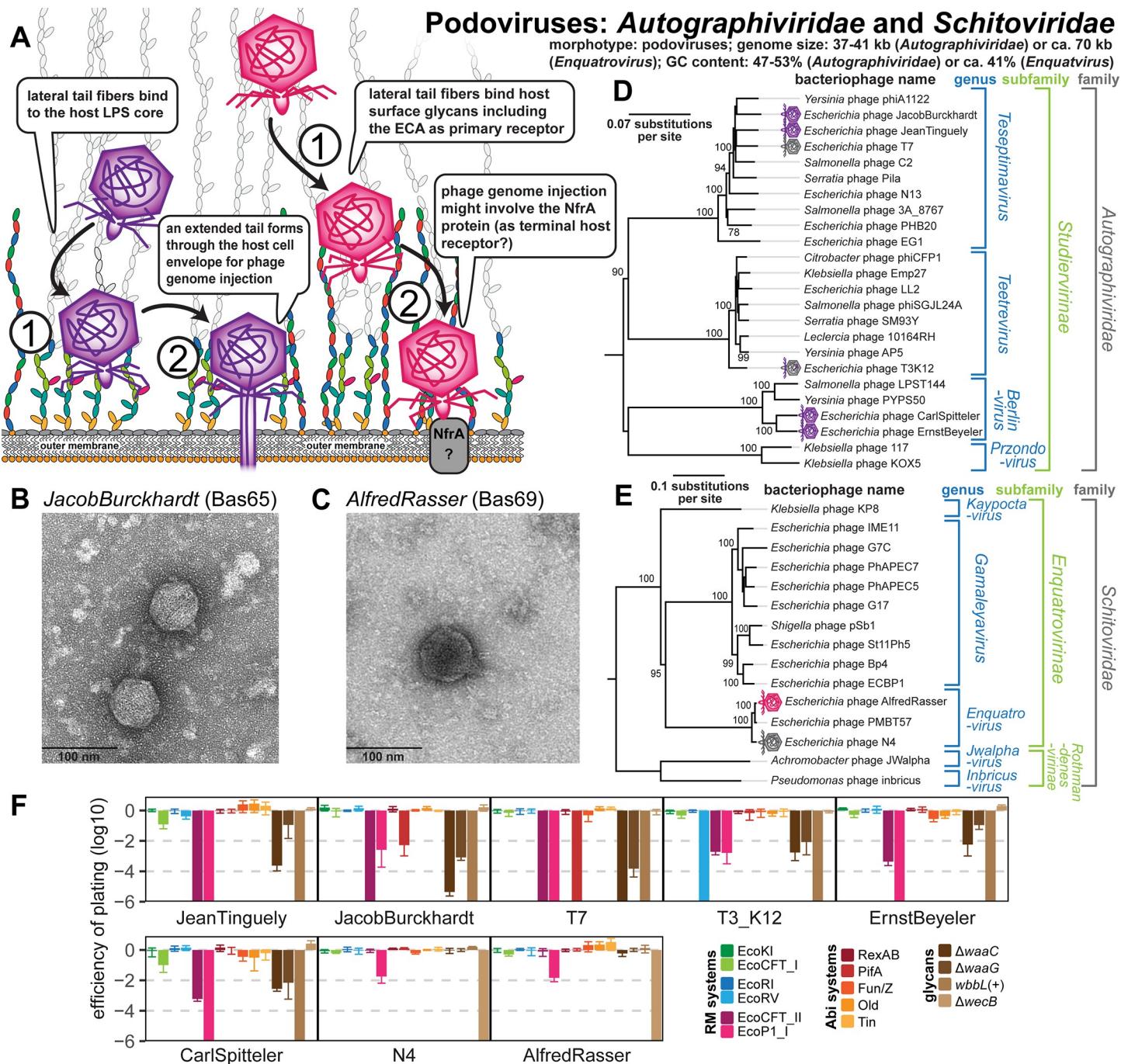
Unlike large myoviruses of the Teenvirinae subfamily, the Vequintavirinae and relatives are exceptionally susceptible RM systems, albeit with minor differences in the pattern of sensitivity and resistance from phage to phage (Fig 8D). At least under laboratory conditions, these phages therefore seem to lack any effective protection against this highly common type of bacterial immunity despite, e.g., a putative dTDP glycosylation machinery encoded by the phi92-like phages [22].

### Properties of Autographiviridae: Studiervirinae and Schitoviridae: Enquattrovirinae

Podoviruses of the Autographiviridae family like T3 and T7 or the Schitoviridae family like N4 have characteristic, stubby tails that contact the phages' terminal receptors on the cell surface for irreversible adsorption and DNA injection after initial host recognition by lateral tail fibers (Fig 9A–9F).

**Autographiviridae: Studiervirinae—T3, T7, and relatives.** Phages T3 and T7 belong to the large Studiervirinae subfamily of Autographiviridae that typically carry genomes of ca. 37 to 41 kb in size and target LPS receptors on the host surface, which has been most well studied for iconic T7. The lateral tail fibers of this phage target the rough LPS of *E. coli* K-12 in ways that are not fully understood, possibly because several alternative and overlapping sugar motifs can be bound [72–74] (Fig 9A). Conformational changes in the stubby tail tube triggered by this receptor interaction then initiate the ejection of the phage genome from the virion [74,75]. Remarkably, at first, several internal virion proteins are ejected and then fold into an extended tail that spans the full bacterial cell envelope, which, in a second step, enables DNA injection into the host cytosol [76] (Fig 9A). Notably, our knowledge of this process does not allow the distinction between a “primary” receptor for host recognition and a “secondary” terminal receptor for irreversible adsorption and DNA injection. However, other Autographiviridae follow this classical scheme more closely and feature enzymatic domains at their tail fibers or tail spikes that likely mediate attachment to surface glycans as primary receptors followed by enzyme-guided movement toward the cell surface [36,77].

The Autographiviridae are named with reference to their single-subunit T3/T7-type RNA polymerase that plays key roles during phage infection but has also become a ubiquitous tool in biotechnology [78]. All Autographiviridae studied in this work belong to 3 different genera of the Studiervirinae subfamily that contain phages infecting diverse enterobacteria, suggesting frequent host switches (Fig 9D). Because phage T3 recognizes the truncated R1-type LPS core of *E. coli* B (see also below) and cannot infect K-12 strains, we generated a T3(K12) chimera that encodes the lateral tail fiber gene of T7 as was reported previously by others (see Materials and methods) [79]. As expected, all tested Autographiviridae use core LPS structures as host receptor and show impaired plaque formation on *waaC* and *waaG* mutants, but in almost all cases, some infectivity is retained even on the *waaC* mutant (Fig 9F). This suggests that, as postulated for T7 [72–74], these phages are not strictly dependent on a single glycan motif but might recognize a broader range of target structures at the LPS core.



**Fig 9. Overview of Autographiviridae subfamily Studiervirinae and Schitoviridae subfamily Enquattrovirinae.** (A) Schematic illustration of host recognition by the studied Autographiviridae and Schitoviridae phages. (B) Representative TEM micrograph of T7-like Autographiviridae phage JacobBurckhardt (Bas65). (C) Representative TEM micrograph of Enquattrovirus AlfredRasser (Bas69). (D) Maximum-Likelihood phylogeny of the Studiervirinae subfamily of Autographiviridae based on several core genes with bootstrap support of branches shown if >70/100. The phylogeny was midpoint-rooted between the clade formed by *Teseptimavirus* and *Teetrevirus* and the other genera. Phages of the BASEL collection are highlighted by little phage icons. (E) Maximum-Likelihood phylogeny of the *Enquattrovirus* genus and related groups of Enquattrovirinae and other Schitoviridae based on several core genes with bootstrap support of branches shown if >70/100. The phylogeny was midpoint-rooted between the distantly related *Jwalphavirus* and *Inbricaviridae* genera and the other shown Schitoviridae. Phages of the BASEL collection are highlighted by little phage icons. (F) The results of quantitative phenotyping experiments with Autographiviridae and Schitoviridae phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as EOP. Data points and error bars represent average and standard deviation of at least 3 independent experiments. Raw data and calculations are available in S1 Data. Abi, abortive infection; BASEL, BActeriophage SElection for your Laboratory; ECA, enterobacterial common antigen; EOP, efficiency of plating; LPS, lipopolysaccharide; RM, restriction-modification; TEM, transmission electron microscopy.

<https://doi.org/10.1371/journal.pbio.3001424.g009>

The overall pattern of restriction sensitivity and resistance is similar for all tested Autographiviridae. While type I and type II RM systems are largely ineffective, type III RM systems show remarkable potency across all phage isolates. Good part of the resistance to type II RM systems is likely due to the near-complete absence of EcoRI and EcoRV recognition sites in the genomes of these phages as described previously [64] ([S5 Table](#)) with the exception of 10 EcoRV sites in T3(K12), which, consequently, cause massive restriction ([Fig 9F](#)). The relatively few recognition sites for type I RM systems do not result in considerable sensitivity for any phage, because at the *gp0.3* locus, all of them either encode an Ocr-type DNA mimic type I restriction inhibitor like T7 (*Teseptimavirus*) or, like T3, an S-adenosylmethionine (SAM) hydrolase that deprives these RM systems of their substrate (all other genera) [9,80,81]. Notably, phages JacobBurckhardt (Bas65) and its close relative T7 are the only phages tested in our study that show any sensitivity to the PifA Abi system encoded on the *E. coli* K-12 F-plasmid ([Fig 9F](#)). Previous work showed that sensitivity of T7 to PifA immunity—as opposed to T3, which is resistant—depends on the dGTPase Gp1.2 of the phage, though the major capsid protein Gp10 also seems to play some role in sensitivity [82,83]. Consequently, we find that phages T7 and JacobBurckhardt, but not closely related *Teseptimavirus* JeanTinguely (Bas64), encode a distinct variant of dGTPase Gp1.2 that likely causes their sensitivity to PifA ([S5E Fig](#)).

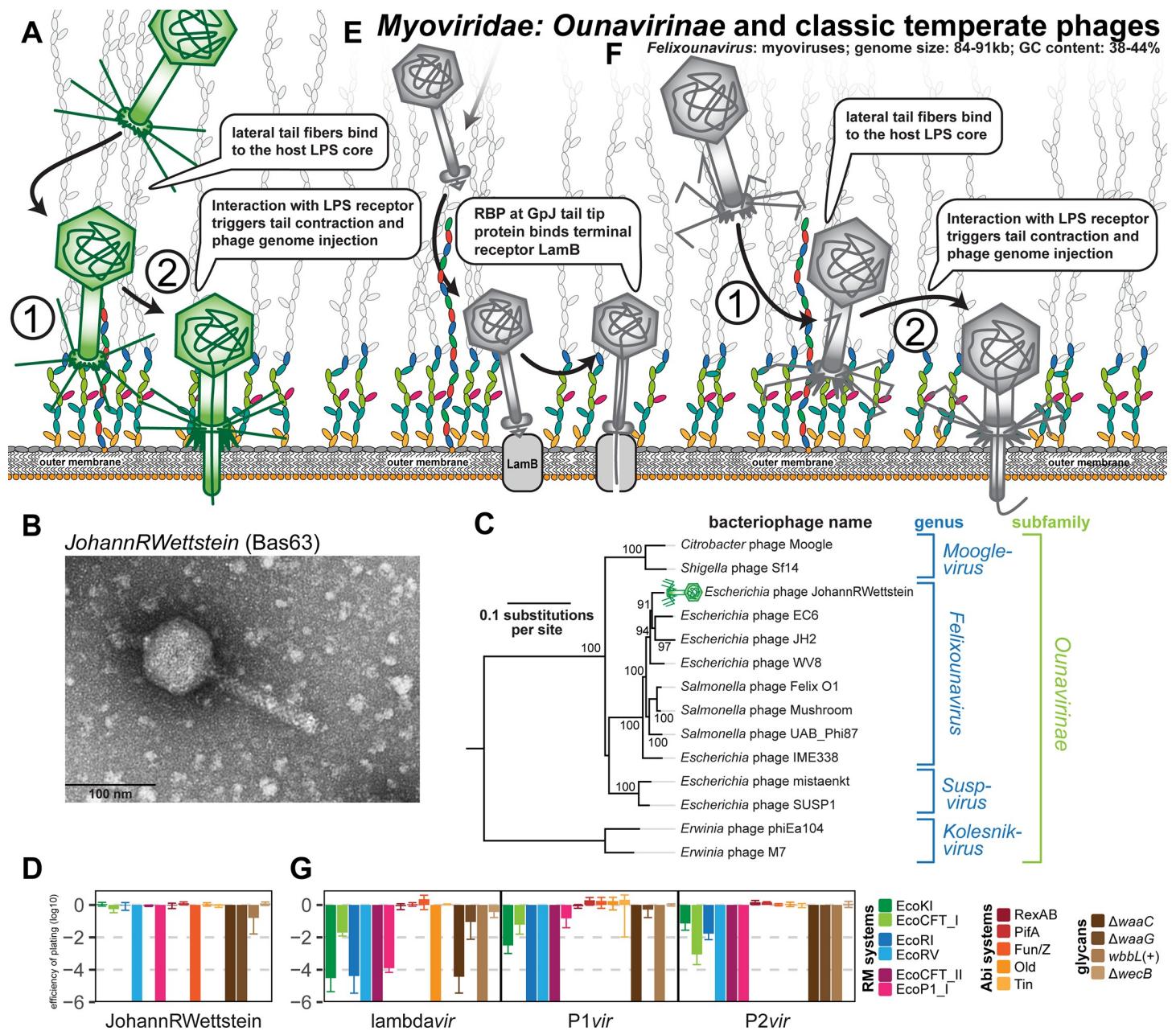
**Schitoviridae: Enquattrovirinae—N4 and relatives.** Bacteriophage N4 and its close relative AlfredRasser (Bas69) are phages of the *Enquattrovirus* genus within the newly formed Schitoviridae family that typically carry genomes of ca. 68 to 74 kb and are hallmark by a large, virion-encapsidated RNA polymerase used for the transcription of their early genes [31,58] ([Fig 9A, 9C, and 9E](#)). Based primarily on genetic evidence, phage N4 is thought to initiate infections by contacting the host's ECA with its lateral tail fibers [72,84,85] ([Fig 9A](#)). We indeed confirmed a remarkable dependence of N4 and AlfredRasser on *wecB* ([Fig 9F](#)) and detected homology between the glycan deacetylase domain of the N4 lateral tail fiber and a lateral tail fiber protein of Vequintavirinae, which also seem to use the ECA as their primary receptor ([S5F Fig](#); see [Fig 8](#)). Like the O-antigen deacetylase tail fiber of its *Gamaleayavirus* relative G7C ([Fig 9E](#)), the enzymatic activity of the N4 tail fiber might provide a directional movement toward the cell surface that is essential for the next steps of infection [36,86] ([Fig 9A](#)). At the cell surface, the elusive outer membrane porin NfrA was suggested to be the terminal receptor of phage N4 based on the findings that it is required for N4 infection and interacts with the stubby tail of this phage [72,84,87]. To the best of our knowledge, this would make phage N4 the first podovirus with an outer membrane protein as terminal receptor [37,38], though the absence of homologs to the tail extension proteins of Autographiviridae indeed suggests a somewhat different mode of DNA injection ([Fig 9A](#)).

*Enquattrovirus* phages are highly resistant to any tested antiviral defenses with exception of a slight sensitivity to the EcoP1\_I type III RM system ([Fig 9F](#)). While their resistance to tested type II RM systems is due to the absence of EcoRI or EcoRV recognition sites ([S5 Table](#)), the genetic basis for their only slight sensitivity to type III RM systems is not known. We note that *Enquattrovirus* phages encode an *rIIAB* locus homologous to the long-known yet poorly understood *rIIAB* locus found in many large myoviruses, which, in case of phage T4, provides resistance to the RexAB Abi system of phage lambda [88] ([S5A and S5B Fig](#)). Both phage N4 and AlfredRasser are indifferent to the presence of (O16-type) O-antigen or the tested truncations of the K-12 LPS core at the host cell surface ([Fig 9F](#)), suggesting that LPS structures play no role for their infection process.

## Properties of Myoviridae: Ounavirinae

The *Felixounavirus* genus in the Ounavirinae subfamily of Myoviridae comprises a group of phages with genomes of 84 to 91 kb and characteristically straight lateral tail fibers of which

*Salmonella* phage Felix O1 has been most well studied [28,89] (Fig 10A–10D). Previous work aimed at the isolation of *E. coli* phages differed greatly in the reported abundance of *Felixounavirus* phages, ranging from no detection [21] over a moderate number of isolates in most



**Fig 10. Overview of Myoviridae: Ounavirinae and classic temperate phages.** (A) Schematic illustration of host recognition by Ounavirinae: *Felixounavirus* phages. Note that the illustration shows short tail fibers simply in analogy to Tevenvirinae or Vequintavirinae (Figs 7A and 8A), but any role for such structures has not been explored for Felix O1 and relatives. (B) Representative TEM micrograph of phage JohannRWettstein (Bas63). (C) Maximum-Likelihood phylogeny of the Ounavirinae subfamily of Myoviridae based on several core genes with bootstrap support of branches shown if >70/100. The phylogeny was midpoint-rooted between *Kolesnikvirus* and the other genera. Our new isolate JohannRWettstein is highlighted by a green phage icon. (D) The results of quantitative phenotyping experiments with JohannRWettstein regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as EOP. (E, F) Schematic illustration of host recognition by classic temperate phages lambda, P1, and P2. Note the absence of lateral tail fibers due to a mutation in lambda PaPa laboratory strains [93]. (G) The results of quantitative phenotyping experiments with the 3 classic temperate phages regarding sensitivity to altered surface glycans and bacterial immunity systems are presented as EOP. Data points and error bars in (D) and (G) represent average and standard deviation of at least 3 independent experiments. Raw data and calculations are available in S1 Data. Abi, abortive infection; EOP, efficiency of plating; LPS, lipopolysaccharide; RBP, receptor-binding protein; RM, restriction-modification; TEM, transmission electron microscopy.

<https://doi.org/10.1371/journal.pbio.3001424.g010>

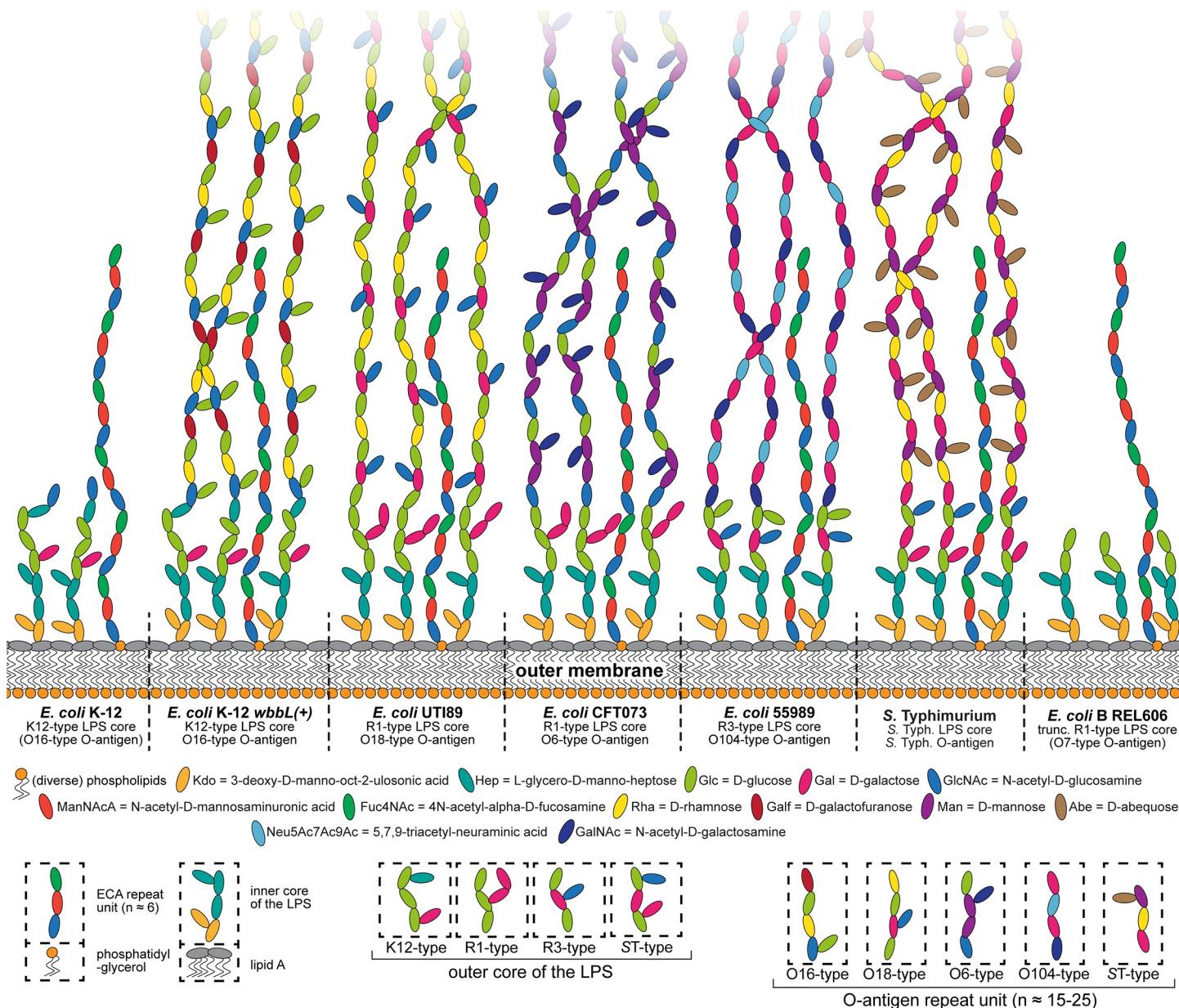
studies [23,90] to around one-fourth of all [22]. Across our phage sampling experiments, we found a single *Felixounavirus* phage, JohannRWettstein (Bas63), which is rather distantly related to Felix O1 within the eponymous genus (Fig 10C). Prototypic phage Felix O1 has been used for decades in *Salmonella* diagnostics because it lyses almost every *Salmonella* strain but only very few other Enterobacteriaceae (reviewed in [89]). The mechanism of its host recognition and the precise nature of its host receptor(s) have remained elusive, but it is generally known to target bacterial LPS and not any kind of protein receptors [89]. While many Ounavirinae seem to bind O-antigen glycans of smooth LPS, the isolation of multiple phages of this subfamily on *E. coli* K-12 with rough LPS and on *E. coli* B with an even further truncated LPS indicate that the functional expression of O-antigen is not generally required for their host recognition [22,89,91]. Phage JohannRWettstein is only slightly inhibited on *E. coli* K-12 with restored O16-type O-antigen, suggesting that it can either use or bypass these glycans and totally depends on an intact LPS core (Fig 10D). In addition, it is remarkably sensitive to several tested RM systems and shares sensitivity to the Fun/Z Abi system with all other tested Myoviridae and the Markadamsvirinae (Fig 10D).

### Properties of classical temperate phages lambda, P1, and P2

Besides the lytic T phages, temperate phages lambda, P1, and P2 have been extensively studied both as model systems for fundamental biology questions as well as regarding the intricacies of their infection cycle [7,32,48] (Fig 10E and 10F). Phage lambda was a prophage encoded in the original *E. coli* K-12 isolate and forms siphovirus particles that display the GpJ tail tip to contact the LamB porin as the terminal receptor for DNA injection [7,92], while the lateral tail fibers (thought to contact OmpA as primary receptor) are missing in most laboratory strains of this phage due to a mutation [7,93] (Fig 10E). We reproduced the dependence of our lambdavir variant on LamB and, as reported previously, found that an intact inner core of the LPS was required for infectivity ([72] and literature cited therein) (Fig 10G). Myoviruses P1 and P2 were both shown to require a rough LPS phenotype to contact their receptors in the LPS core, but the exact identity of these receptors has not been unraveled [38,48,72,94]. While the molecular details of the infection process after adsorption have not been well studied for P1, previous work showed that an interaction of the P2 lateral tail fibers with the LPS core of K-12 strains triggers penetration of the outer membrane by the tail tip [38,48,95] (Fig 10F). Consistently, our results confirm that mutations compromising the integrity of the K-12 core LPS abolished bacterial sensitivity to P1vir and P2vir [48,72,94] (Fig 10G). The quantification of these phages' sensitivity to different immunity systems revealed that they are remarkably sensitive to all tested RM systems (Fig 10G), a property only shared with a few Drexlerviridae (Fig 3E). As expected from the considerably lower number of restriction sites, the sensitivity was less pronounced for type I RM systems (Fig 10G and S5 Table). Phage lambdavir was additionally specifically sensitive to the Old Abi system of the P2 prophage (Fig 10G), as shown previously [48].

### Host range across pathogenic enterobacteria and laboratory wild-type *E. coli* B

The host range of bacteriophages is a critical feature for phage therapy because broad infectivity can enable phages to be used against different strains of the same pathogen without repeated sensitivity testing “analogous to the use of broad-spectrum antibiotics” [16]. Intuitively, across strains of the same host species, the infectivity of phages depends on their ability to successfully bind the variable surface structures of host cells and to overpower or evade strain-specific bacterial immunity [16,96] (Fig 11; see also Fig 2). While simple qualitative tests

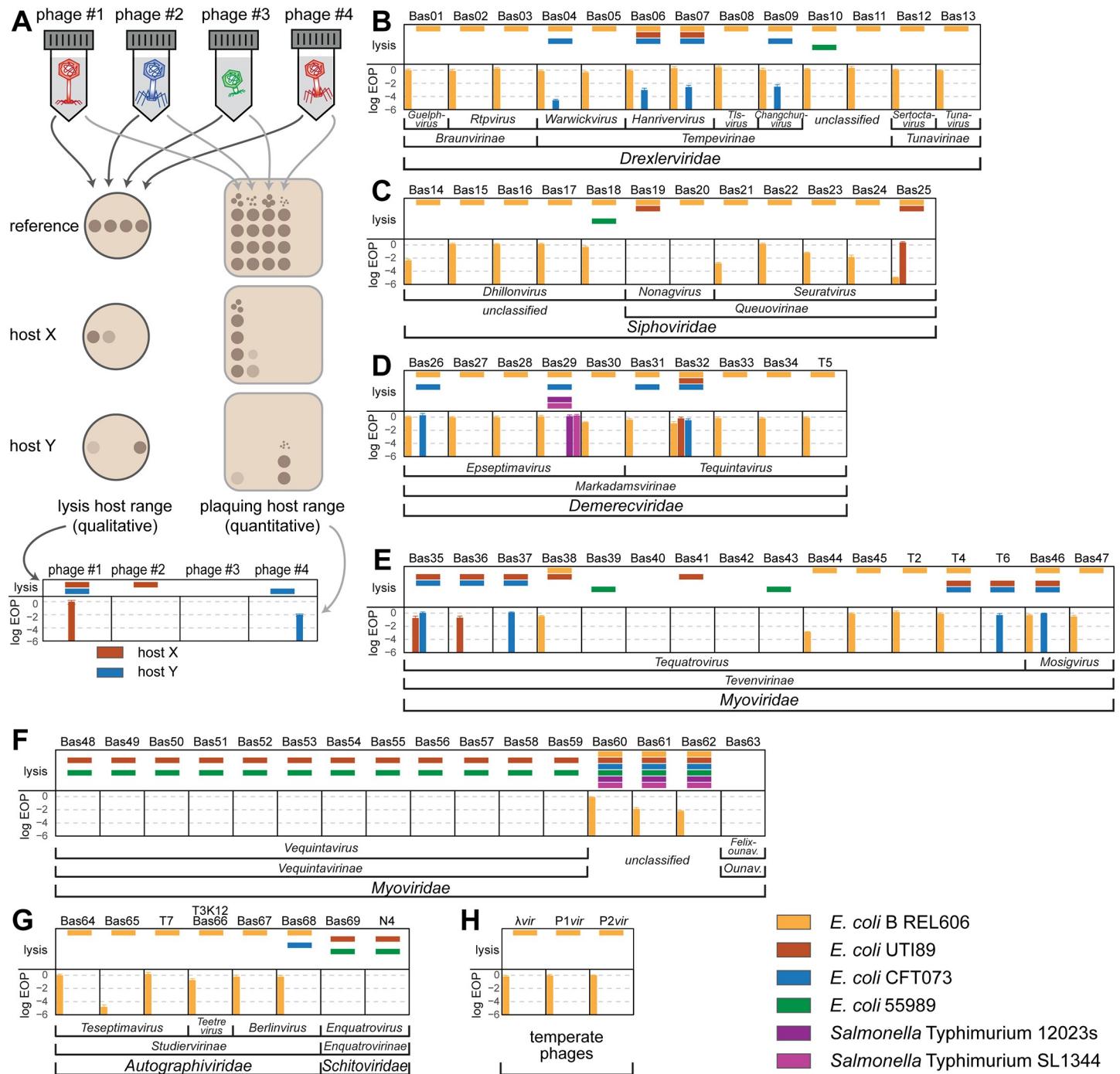


**Fig 11. Surface glycans of the enterobacterial strains used in this work.** The known surface glycans of different enterobacterial strains used in this work were drawn in comparison to highlight the diversity and barrier function of their O-antigen as well as the slight differences between LPS core types (see Materials and methods for details on how the illustration was composed). ECA, enterobacterial common antigen; LPS, lipopolysaccharide.

<https://doi.org/10.1371/journal.pbio.3001424.g011>

assessing the lysis host range primarily inform about host recognition alone, the more laborious detection of robust plaque formation as a sign of full infectivity is the gold standard of host range determination [16,96] (Fig 12A). We therefore challenged a panel of commonly used pathogenic enterobacterial strains with the BASEL collection and recorded the phages' lysis host range as well as plaque formation separately to gain insight into their host recognition as well as their ability to overcome immunity barriers inside host cells (Fig 12; see Materials and methods).

**Host range across pathogenic *E. coli* and *Salmonella Typhimurium*.** As expected from previous work, Vequintavirinae phages showed a broad lysis host range [21,23,69,70] and all



**Fig 12. Host range of phages in the BASEL collection.** (A) Schematic overview of qualitative spot test assays with high-titer phage stocks ( $>10^9$  pfu/ml; left) and quantitative EOP assays using serial dilutions of phage stocks (right) to determine the lysis host range and the plaquing host range of bacteriophages, respectively (see Materials and methods for details of the methodology). The graph at the bottom highlights how the results of the 4 fictional phages in this example would be plotted (with colorful bars at the top indicating the lysis host range and a bar diagrams for the EOP at the bottom that shows the plaquing host range). (B–H) The lysis host range and the plaquing host range of all phages in the BASEL collection across different enterobacteria were determined and visualized as shown in (A). Data points and error bars of the EOP plots represent average and standard deviation of at least 3 independent experiments. Raw data and calculations are available in S1 Data. BASEL, BAacteriophage SElection for your Laboratory; EOP, efficiency of plating.

<https://doi.org/10.1371/journal.pbio.3001424.g012>

invariably infected *E. coli* UTI89, *E. coli* 55989, and *E. coli* K-12 with restored O16-type O-antigen (Figs 8D and 12F). This exact same lysis host range is shared by *Enquattrovirus* phages (Figs 9F and 12G), suggesting that this phenotype is driven by their shared recognition of the highly conserved ECA as primary receptor using a homologous RBP (S5F Fig). Recognition of the ECA molecules among smooth LPS by this glycan deacetylase tail fiber might enable Vequintavirinae and *Enquattrovirus* to bypass the O-antigen barrier and move to the cell surface along this glycan chain (see Fig 11) as previously described for *Salmonella* phage P22 and the N4 relative G7C that target O-antigen chains directly [36,86] (S5F Fig). Interestingly, the phi92-like relatives of Vequintavirinae have the broadest lysis host range of all tested phages and lysed all tested smooth *E. coli* and *Salmonella* hosts (Bas60-Bas62 in Fig 12G), consistent with previous impressive results for phi92 itself [70]. Given the remarkable lysis host range of all these bona fide ECA-targeting phages, it would be valuable to explore the molecular basis of their host recognition in future studies to use this knowledge for phage host range engineering [14].

Besides the bona fide ECA-targeting phages that can apparently bypass the O-antigen barrier, the lysis host range of only around a third of the other tested phages (24/61) included at least 1 host strain with smooth LPS, while the O16-type O-antigen of *E. coli* K-12 blocked infections by all but 5 of these phages scattered across the different taxonomic groups (5/61; Fig 12; see also Figs 3 and 5–10). In particular, the lysis and plaquing host ranges of tested siphoviruses were consistently narrow across taxonomic groups with a few exceptions especially among the T5-like phages (Markadamsvirinae; Fig 12B–12D). One of these isolates, SuperGirl (Bas29), is the only phage in the BASEL collection that shows robust plaque formation on *Salmonella* Typhimurium (Fig 12D). Taken together, these results confirm an important role of the O-antigen as a formidable barrier to bacteriophage infection that can only be overcome by specific recognition with tail fibers to breach it, e.g., via enzymatic activities [33].

A broad lysis host range was also observed for some T-even phages (Fig 12E), in line with previous work [21], though less homogeneous than for the bona fide ECA-targeting phages. For both groups of phages, it is astonishing how poorly their broad host recognition is reflected by actual plaque formation (Fig 12E–12G). Despite robust “lysis-from-without” in some cases [97], no reliable plaque formation on any other host than *E. coli* K-12 was observed for the bona fide ECA-targeting phages, and the range of plaque formation of Tevenvirinae was also severely contracted relative to the lysis host range (Fig 12E–12G). This observation highlights the potency of bacterial immunity systems as well as the well-known limitations of simple top agar spot tests to determine bacteriophage host range [16,96].

**Sensitivity and resistance of *E. coli* B to the BASEL collection phages.** The *E. coli* B lineage of laboratory strains are significant as main models of early bacteriophage research and as the original hosts of the T phages [5,98]. Like *E. coli* K-12, the B strains have lost O-antigen expression and even part of their R1-type LPS core during domestication [98]. The lack of this barrier enables most phages of the BASEL collection to lyse the typical *E. coli* B laboratory strain REL606 (Figs 11 and 12B–12H; see also S2 Text). Curiously, most bona fide ECA-targeting phages are unable to infect *E. coli* B (Fig 12F and 12G), possibly due to differences in ECA expression and/or the truncated R1-type LPS core. It seems likely that resistance of *E. coli* B to Vequintavirinae is the reason why no representative of this highly abundant group was sampled as one of the seven T phages. For different reasons, *E. coli* B strains also lack expression of *ompC*, and the REL606 strain additionally lacks functional *tsx* that both encode common primary receptors of T-even phages (Fig 7A–7D) [65,99]. Though phage T4 itself can use the *E. coli* B LPS core as a primary receptor if OmpC is absent [66], the inability of several other Tevenvirinae to infect *E. coli* B REL606 corresponds well to their dependency on OmpC and Tsx as primary receptors (compare Figs 7C and 12E). A diverse set of phages including, e.g.,

the *Nonagvirus* genus of small siphoviruses or the T7-like phage JacobBurckhardt (Bas65) can lyse *E. coli* B REL606 but show no or only poor plaque formation ([Fig 12B–12H](#)). Based on previous work, we suggest that this phenotype is caused by the specific repertoire of cryptic prophages of *E. coli* B strains that differ considerably from those of K-12 strains and are known to harbor active immunity systems [99].

## Discussion

### The BASEL collection as a representative and practical collection of *E. coli* phages

Assorted collections of model organisms like the BASEL collection exist, e.g., for *E. coli* in form of the ECOR collection of 72 *E. coli* strains that are commonly used to study how this species varies in certain traits or can deal with/evolves under certain conditions [100,101]. Nothing comparable had so far been available for bacteriophages, though the seven T phages had originally been chosen in the early 1940s with the explicit aim of providing a reference point that would enable comparative, systematic research on *E. coli* phages (reviewed in [6]). However, while these and the other classical model phages such as lambda have been invaluable to uncover many fundamental principles of molecular biology, their number and diversity was clearly too limited for a systematic view on bacteriophage biology. The BASEL collection closes this gap because it is broad enough for combined genomic and phenotypic studies aimed at uncovering molecular mechanisms of phage–host interactions as we show, e.g., for the terminal receptor specificity of small siphoviruses (Figs 3–5) and Tevenvirinae ([S4B](#) and [S4C Fig](#)). Diversity and size of the collection allowed us to uncover meaningful patterns of phage sensitivity or resistance to bacterial immunity systems within and across taxonomic groups (Figs 3 and 5–10). We envision that this approach will be helpful to understand how phages can overcome bacterial immunity (also, e.g., in clinical settings) but also for studies exploring the molecular biology of newly discovered bacterial immunity systems.

The composition of the BASEL collection mirrors the diversity of all common groups of *E. coli* phages as determined by previous work and also includes representatives of several rare groups to maximize the biological diversity, which could be probed for phage–host interactions ([Fig 1](#); see Composition of the BASEL collection above). Any such collection of model organisms has inherent limitations that we discuss for the BASEL collection in [S2 Text](#), but we feel that these do not affect the practical value of our phages. Most importantly, the absence of an O-antigen barrier from the *E. coli* K-12 MG1655 ΔRM sampling strain excluded any phages that depend on this glycan as an essential host receptor (like iconic *Salmonella* phage P22 [36]). To determine the impact of this limitation, we quantified the proportion of such O-antigen–dependent phages across sewage from 4 different treatment facilities and consistently found that O-antigen dependency is a very rare phenomenon (only a few percent of isolates; [S6 Fig](#)). Though it would be interesting to study the biology of these phages in more detail, their rarity and practical considerations (given that all used *E. coli* laboratory strains lack O-antigen) strongly suggest that their absence is not a major limitation of the BASEL collection.

### Remarkable patterns of bacteriophage receptor specificity

Our results regarding the terminal receptor specificity of different groups of siphoviruses inherently present the question why all these phages target less than 10 of the more than 150 outer membrane proteins of *E. coli* K-12 of which around 2 dozen are porins [37,38,102] (Figs 4A and 6D). Similarly, T-even phages use only 5 different outer membrane proteins as primary receptors ([Fig 7D](#)) [37,38]. Previous work suggested that this bias might be linked to the

abundance of the targeted proteins, both via a preference for particularly numerous proteins (favoring phage adsorption) or very scarce ones (avoiding competition among phages) [103]. However, we feel that this stark bias in phage preference might be largely driven by functional constraints. Notably, the large siphoviruses of Markadamsvirinae bind exactly a subset of those terminal receptors targeted by the small siphoviruses despite using nonhomologous RBPs, while there is no overlap between these proteins and the primary receptors of the Tevenvirinae [37,38] (Figs 3–7). It seems therefore likely that the receptors targeted by siphoviruses need to have certain properties, e.g., favoring DNA injection, that greatly limit the repertoire of suitable candidates. We envision that future studies will unravel the molecular mechanisms of how siphoviruses recognize and use outer membrane proteins as terminal receptors on gram-negative hosts similar to, e.g., how it has been studied for myoviruses and podoviruses that directly puncture the outer membrane [67,76]. Currently, it is even still unknown if the phage DNA is truly injected through the porin channel (as it is often drawn in the field [104] and also in our illustrations Figs 3A, 5A, and 6A) or rather passes the outer membrane adjacent to it through a phage-derived channel (as suggested by some experimental evidence (reviewed in [105])). Another open question amenable to combined genomic and phenotypic studies is whether the known specificity of siphoviruses for certain inner membrane channels (like ManYZ by phage lambda or PtsG for HK97 [7]) has a genetic basis similar to the proposed mix and match of RBPs and outer membrane proteins. Gaining a deeper understanding of bacteriophage host recognition would not only expand our knowledge of molecular mechanisms underlying bacteriophage ecology and evolution but could also help to apply this knowledge for biotechnology and therapeutic applications, e.g., in the context of “phage steering” [14,106].

### Bacteriophage sensitivity and resistance to host immunity

Our systematic analysis of the sensitivity and resistance profiles across the different taxonomic groups of phages of the BASEL collection revealed several strong patterns that are informative about underlying molecular mechanisms. Some groups of phages like Tevenvirinae or Queuovirinae are highly homogeneous in their profiles of sensitivity or resistance (Figs 5F and 7E), probably largely driven by conserved DNA modifications. Others are more heterogeneous, possibly due to a wide variety of known and unknown genetic factors such as restriction site avoidance or allelic differences that affect detection by Abi systems (see above, e.g., for Markadamsvirinae and Studiervirinae).

Overall, we observed that phages with bigger genomes such as Tevenvirinae, Vequintavirinae, and Markadamsvirinae are commonly targeted by Abi systems, while smaller siphoviruses or podoviruses are rarely sensitive to this type of bacterial immunity (Figs 3 and 5–10; see also S2 Text). Similarly, we confirmed the intuitive expectation that the potency of RM systems is linked to the number of recognition sites in each phage genome unless they are masked by DNA modifications. Consequently, large phages with many recognition sites (like Vequintavirinae and relatives) are exceptionally sensitive to RM systems (Fig 8D), while type I RM systems perform very poorly against most phages because their long recognition sites are rare (Fig 2B and S5 Table; see also Figs 3 and 5–10). For several groups of phages like Markadamsvirinae and some Tevenvirinae, the EcoRV system is the only RM system with significant impact (Figs 6E and 7E). We suggest that this might be due to the blunt-cutting activity of EcoRV that is much more difficult to seal for DNA repair than the sticky ends introduced by the other tested RM systems (Fig 2B), though phage DNA ligases can also differ in their efficiency on different kinds of DNA breaks [107].

One surprising result of our phenotyping was the remarkably poor target range of the most well-studied Abi systems of *E. coli* K-12, RexAB and PifA, that each affected only 2 or 3 phages

including all those previously known to be sensitive (Figs 3 and 5–10 as well as S5A and S5B Fig) [12]. Conversely, the three Abi systems of phage P2 protected each against a considerable number of phages, especially Fun/Z that inhibited all Teenvirinae, Vequintavirinae, and Marmadamsvirinae plus the *Felixounavirus JohannRWettstein* (Figs 3 and 5–11). One possible explanation could be that RexAB and PifA have actually not evolved to serve as bacterial immunity systems and instead primarily have other biological functions like the recently proposed role of RexAB in the lysogeny/lysis switch of phage lambda [108]. Temperate phages themselves seem generally to be highly sensitive to any kind of host immunity (Fig 10G), possibly because their evolution is less driven by selection to overcome host defenses but rather by optimizing the lysogens' fitness, e.g., by providing additional bacterial immunity systems [8,9].

Though the technicalities of our immunity phenotyping are subject to a few minor caveats (see Materials and methods and S2 Text), our results demonstrate how the BASEL collection can be used to explore the biology of bacterial immunity systems and the underlying molecular mechanisms. Several open questions remain: To which extent do the different layers of bacterial immunity limit bacteriophage host range? Why are some groups of phages like the Teenvirinae so strongly targeted by seemingly unrelated Abi systems and other groups of phages apparently not at all? And how well would it be possible, based on systematic phenotypic and genomic data of phages and hosts, to predict the sensitivity of newly isolated phages to diverse immunity systems or even bacteriophage lysis and plaquing host ranges just from phage and host genome sequences?

### Royal families instead of Darwinian Demon: Trade-offs and ecological niches in phage evolution

Our results show that remarkable phage traits such as the broad host recognition of Vequintavirinae and their relatives or the massive resistance of T-even phages to RM systems do not generally confer these phages an exceptionally large plaquing host range (Figs 7E, 8D, 12E, and 12F). This observation could be caused by strong trade-offs between different phage fitness traits as proposed already in earlier work [109]. Unlike the hypothetical “Darwinian Demon” that can maximize all fitness traits simultaneously and would eventually dominate its ecosystem alone [110], the limits imposed by these trade-offs would instead drive the adaptation of different phage groups toward specific niches that enable their long-term coexistence [109]. In marine environments, the coexistence of a few specialized, highly successful phage groups was proposed to be stable in space and time because extinction of individual phages would usually result in their replacement by relatives from the same, highly adapted group [111]. This “royal family model” could elegantly explain why the same groups of *E. coli* phages sampled in our work for the BASEL collection have also been found again and again in previous studies that sampled in distant geographic locations and specific other environments [21–27] (Fig 1). Notably, the molecular mechanisms underlying a possible divergent adaptation of phage groups into “functional types” with different ecological strategies have largely remained elusive so far [112]. However, the results of our systematic study highlight clear and consistent phenotypic differences between the most common groups of *E. coli* phages and suggest that the BASEL collection could be a useful tool to explore the fundamental ecology of bacteriophages.

### Concluding remarks

A recent landmark study systematically explored the genetic profile of host requirements for several *E. coli* phages and uncovered multiple unexpected and exciting twists of phage biology such as the dependence of phage N4 on high levels of the second messenger cyclic di-GMP [72]. Our current study based on the BASEL collection represents a complementary approach

to explore the biology of *E. coli* phages not with elaborate genome-wide screens on the host side but rather by combining phenotypic and genomic analyses of a well-assorted set of bacteriophages. With this strategy, we achieved significant advances in bacteriophage biology regarding the recognition of host receptors, the sensitivity or resistance of phages to host immunity, and how these factors come together to determine bacteriophage host range. Our work therefore establishes the BASEL collection as a powerful tool to explore new aspects of bacteriophage biology by unraveling links between phage phenotypes and their genome sequences. Furthermore, our extensive characterization of the most abundant lineages of *E. coli* phages also provides a useful field guide for teaching and outreach activities analogous to the successful SEA-PHAGES initiative [113] and the diverse student and high school projects that have enabled our study (see Acknowledgments and S5 Table).

## 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<sub>2</sub>O and sterilized by autoclaving. LB agar plates were prepared by supplementing LB medium with agar at 1.5% w/v before autoclaving. The M9RICH culture medium was conceived as a variant of regular M9 medium [114] supplemented with trace elements and 10% v/v LB medium (prepared without NaCl) to promote the growth of diverse enterobacterial strains. It was prepared from sterilized components by mixing (for 50 ml) 33.75 ml Milli-Q H<sub>2</sub>O, 10 ml 5× M9 salts solution, 5 ml LB medium without NaCl, 500 µl 40% w/v D-glucose solution, 100 µl 1 M MgSO<sub>4</sub>, and 5 µl 1 M CaCl<sub>2</sub> using sterile technique. Unless indicated otherwise, all components were sterilized by filtration (0.22 µm).

Phosphate-buffered saline (PBS) was prepared as a solution containing 8 g/l NaCl, 0.2 g/l KCl, 1.44 g/l Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, and 0.24 g/l KH<sub>2</sub>PO<sub>4</sub> 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<sub>4</sub>, and 0.05 M Tris (pH 7.5) using sterile technique.

### Bacterial handling and culturing

*E. coli* and *Salmonella* Typhimurium strains were routinely cultured in LB medium at 37°C in glass culture tubes or Erlenmeyer flasks with agitation at 170 rpm. For all phenotyping assays, the bacteria were instead grown in M9RICH, which supports robust growth of all strains to high cell densities. LB agar plates were routinely used as solid medium. 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 [115] with a top agar prepared as LB agar with only 0.5% w/v agar supplemented with 20 mM MgSO<sub>4</sub> and 5 mM CaCl<sub>2</sub>. 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. We routinely used the improvements of classical phage techniques published by Kauffman and Polz [116].

High-titer stocks of bacteriophages were generated using the plate overlay method. Briefly, top agar plates were set up to grow almost confluent plaques of a given phage and then covered with 12 ml of SM buffer. After careful agitation for 24 hours at 4°C, the suspension on each plate was pipetted off and centrifuged at 8,000 g for 10 minutes. Supernatants were sterilized

with few drops of chloroform and stored in the dark at 4°C. For archiving, bacteriophages were stored as virocells at -80°C [117].

## Bacterial strains and strain construction

All bacterial strains used in this work are listed in [S1 Table](#), all oligonucleotide primers in [S2 Table](#), and all plasmids in [S3 Table](#). The identity of all strains of the KEIO collection used in this study was confirmed by diagnostic PCRs [118] ([S7 Fig](#)). The phenotypes of all bacterial mutants used in this study could be complemented using ectopic expression of the gene that was knocked out or mutated to exclude polar effects and other genetic artifacts ([S8–S13 Figs](#); see technical details below at the section Plasmid construction).

***Escherichia coli* K-12 MG1655 ΔRM.** The standard laboratory strain *E. coli* K-12 MG1655 (CGSC #6300) was engineered into a more permissive host for bacteriophage isolation by knocking out the EcoKI type I RM system (encoded by *hsdRMS*) and the McrA, Mrr, and McrBC type IV restriction systems using lambda red recombineering [119] (see [S1 Text](#) and our previous work for more technical details [120]). The resulting strain lacks all known *E. coli* K-12 restriction systems and was therefore called *E. coli* K-12 ΔRM. To enable isolation of bacteriophages with tropism for the sex pilus of the F-plasmid conjugation system, we supplied *E. coli* K-12 MG1655 ΔRM with a variant of the F-plasmid in which the *pifA* immunity gene (which might otherwise have interfered with bacteriophage isolation) had been replaced with a zeocin resistance cassette by one-step recombineering (see [S1 Text](#) for details). The strain was additionally transformed with plasmid pBR322\_ΔPtet of reference [121] as an empty vector control for experiments with plasmid-encoded immunity systems (see below).

***Escherichia coli* K-12 MG1655 ΔRM mutants with altered surface glycans.** To test the effect of altered surface glycans of bacteriophage infection, we generated derivatives of *E. coli* K-12 MG1655 ΔRM in which genes linked to specific glycans were knocked out. The *waaC* and *waaG* genes of the LPS core biosynthesis pathway were knocked out to generate mutants displaying a deep-rough or extremely deep-rough phenotype ([Fig 2A](#); see also the considerations in [S2 Text](#)). Expression of the O16-type O-antigen of *E. coli* K-12 was restored by precisely removing the IS5 element disrupting *wbbL* [42] ([Fig 2A](#)). These mutants were generated by two-step recombineering (see details in [S1 Text](#) and a list of all strains in [S1 Table](#)). A strain specifically lacking the ECA was obtained as the *wecB* mutant of the KEIO collection [118]. This strain is deficient in WecB, the enzyme synthesizing UDP-ManNAcA (UDP-N-acetyl-D-mannosaminuronic acid), which is thought to be specifically required for the synthesis of the ECA but no other known glycan of *E. coli* K-12 [41].

***Escherichia coli* K-12 BW25113 *btuB* and *tolC* knockout mutants.** *btuB* and *tolC* knockout mutants isogenic with the KEIO collection were generated to use these strains lacking known bacteriophage receptors for qualitative top agar assays (see below). Details of the strain construction are provided in [S1 Text](#).

**Other enterobacterial strains used for bacteriophage phenotyping.** *E. coli* B REL606 is a commonly used laboratory strain with a parallel history of domestication to *E. coli* K-12 MG1655 [98]. *E. coli* UTI89 (ST95, O18:K1:H7) and *E. coli* CFT073 (ST73, O6:K2:H1; we used the variant with restored *rpoS* described previously [122]) are commonly used as model strains for UPEC and belong to phylogroup B2 [123]. *E. coli* 55989 (phylogroup B1, ST678, O104:H4) is commonly used as a model strain for EAEC and closely related to the Shiga toxin-producing *E. coli*, which caused the 2011 outbreak in Germany [124,125]. *Salmonella enterica* subsp. *enterica* serovar Typhimurium strains 12023s (also known as ATCC 14028) and SL1344 are both commonly used in laboratory experiments but exhibit phylogenetic and biological differences [126].

## Plasmid construction

Plasmid vectors were generally cloned following the method of Gibson and colleagues (“Gibson Assembly”) [127] in which 2 or more linear fragments (usually PCR products) are ligated directionally guided by short 25 bp overlaps. Initially, some plasmids were also constructed using classical restriction-based molecular cloning. Briefly, a PCR-amplified insert and the vector backbone were each cut with appropriate restriction enzymes (New England Biolabs, Ipswich, Massachusetts, USA). After dephosphorylation of the backbone (using FastAP dephosphorylase; Thermo Scientific, Waltham, Massachusetts, USA), insert and backbone were ligated using T4 DNA ligase (Thermo Scientific). Local editing of plasmid sequences was performed by PCR with partially overlapping primers as described by Liu and Naismith [128]. *E. coli* strain EC100 *pir*(+) was used as host for all clonings. The successful construction of every plasmid was confirmed by Sanger Sequencing. A list of all plasmids used in this study is found in [S3 Table](#), and their construction is summarized in [S4 Table](#). The sequences of all oligonucleotide primers used in this study are listed in [S2 Table](#).

For the series of plasmids encoding the 11 different bacterial immunity systems studied in this work (see [Fig 2B](#)), we used the EcoRI and EcoRV constructs of Pleška, Qian and colleagues [121] as templates and, consequently, the corresponding empty vector pBR322\_ΔPtet as general cloning backbone and experimental control. The different immunity systems were generally cloned together with their own transcriptional promoter region. However, the *rexAB* genes are transcribed together with the *cI* repressor gene of the lambda prophage [7]. We therefore cloned them directly downstream of the *Ptet* promoter of the pBR322 backbone and obtained a functional construct (validated in [S5A Fig](#)). The EcoKI, EcoRI, EcoRV, and EcoP1\_I RM systems have been studied intensively in previous work [45–47]. Besides the type IA RM system EcoKI, we also cloned EcoCFT\_I that is nearly identical to the well-characterized type IB RM system EcoAI but encoded in the genome of *E. coli* CFT073 that was available to us [45,129]. Similarly, the EcoCFT\_II system was identified as a type III RM system in the *E. coli* CFT073 genome using REBASE [47,129]. Due to problems with toxicity, some immunity systems were cloned not into pBR322\_ΔPtet but rather into a similar plasmid carrying a low-copy SC101 origin of replication (pAH186SC101e [122]; see [S3](#) and [S4](#) Tables as well as the considerations in [S2 Text](#)). Since we failed to obtain any functional construct for the ectopic expression of *pifA* (as evidenced by lack of immunity against T7 infection), we instead replaced the F(*pifA::zeoR*) plasmid in *E. coli* K-12 MG1655 ΔRM with a wild-type F plasmid that had merely been tagged with kanamycin resistance and encodes a functional *pifA* (pAH200e).

For the series of plasmids generated to complement the different *E. coli* mutants used in this study, the genes that were knocked out or mutated (*wecB*, *btuB*, *tolC*, *fhuA*, *yncD*, *lamB*, *tsx*, *fadL*, *ompA*, *ompC*, *ompF*, *lptD*, *waaC*, *waaG*) were each cloned downstream of the *P<sub>lac</sub>* promoter of low-copy plasmid pAH186SC101e [122] (see [S3 Table](#) for a list of plasmids and [S4 Table](#) for the technical details of plasmid construction). Genes were generally cloned with their native promoter as identified using RegulonDB ([130]; <http://regulondb.ccg.unam.mx/>) or simply including ca. 200 bp of upstream region for solitary genes when the promoter has remained undescribed (*yncD*). Four genes (*wecB*, *lamB*, *waaC*, *waaG*) are encoded in operons and were therefore cloned merely with their own bona fide ribosome-binding site by including ca. 20 to 25 bp upstream of the start codon. All plasmids complemented the phenotypes that we had observed for the mutant strains ([S8–S13 Figs](#)).

## Bacteriophage isolation

**Basic procedure.** Bacteriophages were isolated from various different samples between July 2019 and March 2021 on *E. coli* K-12 MG1655 ΔRM carrying F(*pifA::zeoR*) as the host (see [S5 Table](#) for details) using a protocol similar to common procedures in the field [16].

Phage isolation was generally performed without an enrichment step to avoid biasing the isolation toward fast-growing phages (but see below).

For aqueous samples, we directly used 50 ml, while samples with major solid components (like soil or compost) were agitated overnight at 4°C in 50 ml of PBS to release viral particles. Subsequently, all samples were centrifuged at 8,000 g for 15 minutes to spin down particles larger than viruses. The supernatants were sterilized treated with 2.5% v/v chloroform, which safely inactivates any bacteria as well as enveloped viruses but will generally leave most Caudovirales intact [16]. Subsequently, viral particles were precipitated by adding 1 ml of a 2-M ZnCl<sub>2</sub> solution per 50 ml of sample, mixing shortly by inversion, and incubating the suspension at 37°C without agitation for 15 minutes [131]. After precipitation, the samples were centrifuged again at 8,000 g for 15 minutes, and the supernatant was discarded. The pellets were carefully resuspended in each 500 µl of SM buffer by agitation at 4°C for 15 minutes. Subsequently, the suspensions were cleared quickly using a tabletop spinner and mixed with 500 µl of bacterial overnight culture (resuspended in fresh LB medium to induce resuscitation). After incubation at room temperature for 15 minutes to promote phage adsorption, each mixture was added to 9 ml of top agar and poured onto a prewarmed square LB agar plate (ca. 12 cm × 12 cm). After solidification, the plates were incubated at 37°C for up to 24 hours.

**Isolation of bacteriophage clones.** Bacteriophages were visible as plaques forming in the dense bacterial growth of the top agar. For isolation of bacteriophage clones, they were picked from clearly separated plaques of diverse morphologies with sterile toothpicks and propagated at least 3 times via single plaques on top agars of the isolation host strain *E. coli* K-12 MG1655 ΔRM. To avoid isolating temperate phages or phages that are poorly adapted to *E. coli* hosts, we only picked clear plaques (indicative of virulent phages) and discarded isolates that showed poor plaque formation [16].

**Isolation of Autographiviridae using enrichment cultures.** The direct plating procedure outlined above never resulted in the isolation of phages belonging to the Autographiviridae like iconic T phages T3 and T7. Given that these phages are known for fast replication and high burst sizes [132], we therefore performed a series of enrichment culture isolation experiments to obtain phages forming the characteristically large, fast-growing plaques of Autographiviridae. For this purpose, we prepared M9 medium using a 5× M9 salts solution and chloroform-sterilized sewage plant inflow instead of water (i.e., containing ca. 40 ml of sewage plant inflow per 50 ml of medium) and supplemented it with 0.4% w/v D-glucose as carbon source. About 50 ml cultures were set up by inoculating these media with each 1 ml of an *E. coli* K-12 MG1655 ΔRM overnight culture and agitated the cultures at 37°C for 24 hours. Subsequently, the cultures were centrifuged at 8,000 g for 15 minutes, and each 50 µl of supernatant was plated with the *E. coli* K-12 MG1655 ΔRM isolation strain in a top agar on one square LB agar plate. After incubation at 37°C for 3 or 4 hours, the first Autographiviridae plaques characteristically appeared (before most other plaques) and were picked and propagated as described above. Using this procedure, we isolated 4 different new Autographiviridae phages (see S5 Table and Fig 9D).

**Composition of the BASEL collection.** Bacteriophages were mostly isolated and characterized from randomly picked plaques in direct selection experiments, but we later adjusted the procedure to specifically isolate Autographiviridae, which were the only major group of phages previously shown to infect *E. coli* K-12 initially missing from our collection (see above). After every set of 10 to 20 phages that had been isolated, we performed whole-genome sequencing and preliminary phylogenetic analyses to keep an overview of the growing collection (see below). In total, more than 120 different bacteriophages were sequenced and analyzed of which we selected 68 tailed, lytic phage isolates to compose the BASEL collection (see Fig 1 and S5 Table for details). Phages closely related to other isolates were deliberately excluded

unless they displayed obvious phenotypic differences such as, e.g., a different host receptor. In addition to the 68 newly isolated bacteriophages, 10 classical model phages were included for genomic and phenotypic characterization, and we view these phages as an accessory part of the BASEL collection. These 10 phages were 6 of the 7 T phages (excluding T1 because it is a notorious laboratory contaminant [30]), phage N4, and obligately lytic mutants of the 3 well-studied temperate phages lambda, P1, and P2 [5,7,31,32] (Fig 1; see also S5 Table). To generate the T3(K12) chimera, the 3' end of *gp17* lateral tail fiber gene of phage T7 was cloned into low-copy plasmid pUA139 with flanking regions exhibiting high sequence similarity to the phage T3 genome (generating pUA139\_T7(*gp17*); see S4 Table). Phage T3 was grown on *E. coli* B REL606 transformed with this plasmid and then plated on *E. coli* K-12 MG1655 ΔRM to isolate recombinant clones. Successful exchange of the parental T3 *gp17* allele with the variant of phage T7 was confirmed by Sanger Sequencing.

All 68 newly isolated bacteriophages and the T3(K12) chimera are shared via the German Collection of Microorganisms and Cell Cultures (DSMZ) with DSM identifiers as listed in S5 Table.

### Qualitative top agar assays

The lysis host range of isolated bacteriophages on different enterobacterial hosts and their ability to infect strains of a set of KEIO collection mutants lacking each one surface protein (or isogenic mutants that were generated in this study; see S1 Table and S1 Text) were tested by qualitative top agar assays. For this purpose, top agars were prepared for each bacterial strain on LB agar plates. For regular round Petri dishes (ca. 9.4 cm diameter), we used 3 ml of top agar supplemented with 100 μl of bacterial overnight culture, while for larger square Petri dishes (ca. 12 cm × 12 cm), we used 9 ml of top agar supplemented with 200 μl of overnight culture. After solidification, each 2.5 μl of undiluted high-titer stocks of all tested bacteriophages (>10<sup>9</sup> pfu/ml) were spotted onto the top agar plates and dried into the top agar before incubation at 37°C for at least 24 hours. If lysis zones on any enterobacterial host besides our *E. coli* K-12 ΔRM reference strain were observed, we quantified phage infectivity in efficiency of plating (EOP) assays (see below). Whenever a phage failed to show lysis on a mutant strain lacking a well-known phage receptor, we interpreted this result as indicating that the phage infecting depends on this factor as host receptor.

### Efficiency of plating assays

The infectivity of a given bacteriophage on a given host was quantified by determining the EOP, i.e., by quantitatively comparing its plaque formation on a certain experimental host to plaque formation on reference strain *E. coli* K-12 MG1655 ΔRM carrying F(*pifA::zeoR*) and pBR322\_ΔP<sub>tet</sub> [133]. Experimental host strains were identical to the reference strain with the difference that they either carried plasmids encoding a certain bacterial immunity system (S3 Table) or had a chromosomal modification changing surface glycan expression (Fig 2A and S1 Table).

For quantitative phenotyping, top agars were prepared for each bacterial strain on LB agar plates by overlaying them with top agar (LB agar containing only 0.5% agar and additionally 20 mM MgSO<sub>4</sub> as well as 5 mM CaCl<sub>2</sub>; stored at 60°C) supplemented with a suitable bacterial inoculum. For regular round Petri dishes (ca. 9.4 cm diameter), we used 3 ml of top agar supplemented with 100 μl of bacterial overnight culture, while for larger square Petri dishes (ca. 12 cm × 12 cm), we used 9 ml of top agar supplemented with 200 μl of overnight culture.

While the top agars were solidifying, serial dilutions of bacteriophage stocks (previously grown on *E. coli* K-12 MG1655 ΔRM to erase any EcoKI methylation) were prepared in sterile PBS.

Subsequently, each 2.5 µl of all serial dilutions were spotted on all top agar plates and dried into the top agar before incubation at 37°C for at least 24 hours. Plaque formation was recorded repeatedly throughout this time (starting after 3 hours of incubation for fast-growing phages). The EOP of a given phage on a certain host was determined by calculating the ratio of plaques obtained on this host over the number of plaques obtained on the reference strain *E. coli* K-12 MG1655 ΔRM carrying F(*pifA::zeoR*) and pBR322\_ΔPtet [133]. Plaque counts and the calculation of EOP of all replicates of all experiments reported in this study (Figs 3E, 5F, 6E, 7E, 8D, 9F, 10D, 10G, and 12B–12H) as well as their summary statistics are compiled in S1 Data.

When no plaque formation could be unambiguously recorded by visual inspection, the EOP was determined to be below detection limit even if the top agar showed lysis from without (i.e., lysis zones caused by bacterial cell death without phage infection at an efficiency high enough to form plaques [97]). However, for all non-K12 strains of *E. coli* as well as *Salmonella* Typhimurium, we determined the lysis host range (i.e., the range of hosts on which lysis zones were observed) besides the numerical determination of EOP (see the illustration in Fig 12A). Occasionally, we found that certain phage/host pairs were on the edge between merely strong lysis from without and very poor plaque formation. Whenever in doubt, we recorded the result conservatively as an EOP below detection limit (e.g., for phage FriedrichMiescher on *E. coli* 55989, *Enquatrovirus* phages and EmilHeitz on *E. coli* UTI89, and all Vequintavirinae and relatives on the host expressing the Fun/Z Abi system).

## Bacteriophage genome sequencing and assembly

Genomic DNA of bacteriophages was prepared from high-titer stocks using the Norgen Biotek Phage DNA Isolation Kit according to the manufacturer's guidelines and sequenced at the Microbial Genome Sequencing Center (MiGS) using the Illumina NextSeq 550 platform. Trimmed sequencing reads were assembled using the Geneious Assembler implemented in Geneious Prime 2021.0.1 with a coverage of typically 50 to 100× (S5 Table). Usually, circular contigs (indicating a complete assembly due to the fusion of characteristically repeated sequence at the genome ends [134]) were easily obtained using the “Medium Sensitivity/Fast” setting. Consistently incomplete assemblies or local ambiguities were solved by PCR amplification using the high-fidelity polymerase Phusion (New England Biolabs) followed by Sanger Sequencing. For annotation and further analyses, sequences were linearized with the 5' end set either to the first position of the small terminase subunit gene or the first position of the operon containing the small terminase subunit gene.

## Bacteriophage genome annotation

A preliminary, automated annotation of the genes in all genomes was generated using MultiPhate [135] and then manually refined. For this purpose, whole-genome alignments of all new isolates within a given group of phages and well-studied and/or well-annotated references were generated using MAFFT v7.450 [136] implemented in Geneious 2021.0.1 and used to inform the annotation based on identified orthologs. Bona fide protein-coding genes without clear functional annotation were translated and analyzed using the blastp tool on the NCBI server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), the InterPro protein domain signature database [137], as well as the Phyre2 fold recognition server ([138]; <http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>). tRNA genes were predicted using tRNA-ScanSE ([139]; <http://lowelab.ucsc.edu/tRNAscan-SE/>), and spanins were annotated with help of the Spanin DataBase tool [140]. While endolysin genes were easily recognized by homology to lysozyme-like proteins and other peptidoglycan hydrolases, holins were more difficult to identify when

no holins were annotated in closely related bacteriophage genomes. In these cases, we analyzed all small proteins (<250 amino acids) for the presence of transmembrane helices—a prerequisite for the functionality of known holins [141]—and studied their possible relationships to previously described holins using blastp and InterPro. In most but not all cases, bona fide holins encoded close to endolysin and/or spanin genes could be identified. The annotation of our genomes in a comparative genomics setup made it easily possible to identify the boundaries of introns associated with putative homing endonucleases and to precisely identify inteins (see also S5 Table). The annotated genome sequences of all 68 newly isolated phages are available at the NCBI GenBank database under accession numbers listed in S5 Table.

### Bacteriophage naming and taxonomy

Newly isolated bacteriophages were named according to rules and conventions in the field and classified in line with the rules of the International Committee on the Taxonomy of Viruses (ICTV) [58,142] (S5 Table). As a first step of taxonomic classification, phages were roughly sorted by family and genus based on whole-genome blastn searches against the nonredundant nucleotide collection database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>; see also [143]). For each such broad taxonomic group, we selected a set of reference sequences from NCBI GenBank that would cover the diversity of this group and that always included all members, which had already previously been studied intensively. Subsequently, we generated whole-genome alignments of these sets of sequences generated using MAFFT v7.450 [136] implemented in Geneious 2021.0.1. These alignments or merely regions encoding highly conserved marker genes such as the terminase subunits, portal protein gene, major capsid protein gene, or other suitable loci were then used to generate Maximum-Likelihood phylogenies to unravel the evolutionary relationships between the different bacteriophage isolates and database references (see S3 Text for details). Care was taken to avoid genome regions that are recognizably infested with homing endonucleases or that showed obvious signs of having recently moved by horizontal gene transfer (e.g., an abrupt shift in the local sequence identity to the different other genomes in the alignment). Clusters of phage isolates observed in these phylogenies generally correlated very well with established taxonomy as inferred from NCBI taxonomy (<https://www.ncbi.nlm.nih.gov/taxonomy>), the ICTV (<https://talk.ictvonline.org/taxonomy/>; [144]), and other reports in the literature [28]. To classify our phage isolates on the species level, we generated pairwise whole-genome alignments generated using MAFFT v7.450 [136] implemented in Geneious 2021.0.1 with genomes of related phages as identified in our phylogenies. From these alignments, the nucleotide sequence identity was determined as the query coverage multiplied by the identity of aligned segments [22]. When the genomes were largely syntenic and showed >95% nucleotide sequence identity, we classified our isolates as the same species as their close relative [143], at least if this phage has been assigned to a species by the ICTV [144].

Most phages were named in honor of scientists and other historically relevant personalities with links to the city of Basel, Switzerland, where the majority of phages had been isolated. Despite efforts to name many phages after female personalities, a considerable gender imbalance remains as a consequence of biases in how science and history have been made and recorded before the 20th century.

### Sequence alignments and phylogenetic analyses

The NCBI GenBank accession numbers of all previously published genomes used in this study are listed in S6 Table. Sequence alignments of different sets of homologous genes were generated using MAFFT v7.450 implemented in Geneious Prime 2021.0.1 [136]. Whenever

required, poor or missing annotations in bacteriophage genomes downloaded from NCBI GenBank were supplemented using the ORF finder tool of Geneious Prime 2021.0.1 guided by orthologous sequence parts of related genomes. Alignments were set up using default settings typically with the fast FFT-NS-2 algorithm and 200 PAM/k = 2 or BLOSUM62 scoring matrices for nucleotide and amino acid sequences, respectively. Subsequently, alignments were curated manually to improve poorly aligned sequence stretches and to mask nonhomologous parts.

For phylogenetic analyses, sequence alignments of orthologous stretches from different genomes (genes, proteins, or whole genome) were used to calculate Maximum-Likelihood phylogenies with PhyML 3.3.20180621 implemented in Geneious Prime 2021.0.1 [145]. Phylogenies were calculated with the HYK85 substitution model for nucleotide sequences and with the LG substitution model for amino acid sequences. For the inference of phylogenetic relationships between phage genomes, we sometimes used curated whole-genome alignments, but the infestation with homing endonucleases and the associated gene conversion made this approach impossible for all larger genomes. Instead, we typically used curated sequence alignments of several conserved core genes (on nucleotide or amino acid level, depending on the distance between the genomes) as the basis for Maximum-Likelihood phylogenies. The detailed procedures for each phylogeny shown in this article are described in S3 Text.

### LPS structures of enterobacterial strains

*Escherichia* K-12 MG1655 (serogroup A) codes for a K12-type LPS core and an O16-type O-antigen, but functional expression of the O-antigen was lost during laboratory adaptation due to inactivation of *wbbL* by an IS5 insertion so that only a single terminal GlcNAc is attached to the LPS core [42,71] (see Fig 2A). *E. coli* strains UTI89 (serogroup B2), CFT073 (serogroup B2), and 55989 (serogroup B1) express O-antigens of the O18, O6, and O104 types, respectively [123,124]. Their core LPS types were determined to be R1 (UTI89 and CFT073) and R3 (55989) by BLAST searches with diagnostic marker genes similar to the PCR-based approached described previously [71]. For the illustration in Fig 11, the structures of LPS cores and linked O-antigen polysaccharides were drawn as described in the literature with typical O-antigen chain lengths of 15 to 25 repeat units [71,146,147]. For the O18-type of O-antigen polysaccharides, different subtypes with slight differences in the repeat unit have been described [147], but, to the best of our knowledge, it has remained elusive which subtype is expressed by *E. coli* UTI89. In Fig 11, we therefore chose an O18A type as exemplary O-antigen polysaccharide for *E. coli* UTI89. Similar to the laboratory adaptation of *E. coli* K-12 MG1655, strains of the *E. coli* B lineage like *E. coli* B REL606 lack O-antigen expression due to an IS1 insertion in *wbbD* but have an additional truncation in their R1-type LPS core due to a second IS1 insertion in *waaT* that leaves only 2 glucoses in the outer core (see Fig 11) [66,98]. *S. enterica* subsp. *enterica* serovar Typhimurium strains 12023s (also known as ATCC 14028) and SL1344 share the common *S. Typhimurium* LPS core and *Salmonella* serogroup B/O4 O-antigen [40,148] (see Fig 11).

### Bacterial genome sequencing and analyses to identify host receptors

While top agar assays with *E. coli* mutants carrying defined deletions in genes coding for different surface proteins or genes involved in the LPS core biosynthesis readily identified the host receptor of most bacteriophages (see above), few small siphoviruses of the Drexlerviridae family and the *Dhillonvirus*, *Nonagvirus*, and *Seuratvirus* genera of the Siphoviridae family could not be assigned to any surface protein as secondary receptor. This was the case for phages AugustePiccard (Bas01) and JeanPiccard (Bas02) of Drexlerviridae, TheodorHerzl

(Bas14) and Oekolampad (Bas18) of *Dhillonvirus*, ChristophMerian (Bas19) and FritzHoffmann (Bas20) of *Nonagyrirus*, and VogelGryff (Bas25) of *Seuratvirus*. However, as siphoviruses infecting gram-negative bacteria, it seemed highly likely that these phages would use an outer membrane porin as their secondary receptor [38]. We therefore isolated resistant mutants of *E. coli* K-12 BW25113 by plating bacteria on LB agar plates, which had been densely covered with high-titer lysates of each of these bacteriophages. While attempting to verify that the resistance of isolated bacterial clones was specific to the phage on which they had been isolated, we found that many clones were fully resistant specifically but without exception to all these small siphoviruses with yet no known host receptor, suggesting that they all targeted the same receptor.

For several of these phage-resistant clones, genomic DNA was prepared using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, St. Louis, Missouri, USA) according to the manufacturer's guidelines and sequenced at the Microbial Genome Sequencing Center (MiGS) using the Illumina NextSeq 550 platform. Sequencing reads were assembled to the *E. coli* K-12 MG1655 reference genome (NCBI GenBank accession U000096.3) using Geneious Prime 2021.0.1 with a coverage of typically 100 to 200 $\times$  to identify the mutations underlying their phage resistance. These analyses uncovered a diversity of in-frame deletions in *lptD* for the different mutant clones (Fig 4).

### Morphological analyses by transmission electron microscopy

The virion morphology of representative bacteriophages of the different taxonomic groups included in the BASEL collection was analyzed by transmission electron microscopy following common procedures in the field [149]. Briefly, 5  $\mu$ l drops of high-titer lysates were adsorbed to 400 mesh carbon-coated grids, which were rendered hydrophilic using a glow-discharger at low vacuum conditions. They were subsequently stained on 5  $\mu$ l drops of 2% (w/v) uranyl acetate. Samples were examined using an FEI Tecnai G2 Spirit transmission electron microscope (FEI Company, Hillsboro, Oregon, USA) operating at 80-kV accelerating voltage. Images were recorded with a side-mounted Olympus Veleta CCD camera 4k using EMSIS RADIUS software at a nominal magnification of typically 150,000 $\times$ .

### Quantification and statistical analysis

Quantitative data sets were analyzed by calculating mean and standard deviation of at least 3 independent biological replicates for each experiment. Detailed information about replicates and statistical analyses for each experiment is provided in the figure legends.

### Supporting information

**S1 Table. List of all bacterial strains used in this study.** 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  $\mu$ g/ml, 50  $\mu$ g/ml, and 100  $\mu$ g/ml, respectively. The following mutants of the KEIO collection were used for qualitative top agar assays but are not included in the strain list because no phage showed any growth phenotype on them: *ompW::kanR*, *phoE::kanR*, *flgG::kanR*, *fepA::kanR*, *hofQ::kanR*, *cirA::kanR*, *fluE::kanR*, *fiu::kanR*, *ompN::kanR*, *pgaA::kanR*, *chiP::kanR*, *ompL::kanR*, *yddB::kanR*, *fecA::kanR*, *uidC::kanR*, *nanC::kanR*, *yfaZ::kanR*, *bglH::kanR*, *bcsC::kanR*, *cusC::kanR*, *gfcE::kanR*, *mdtP::kanR*, *ompG::kanR*, *ompX::kanR*, *yfeN::kanR*, *csgF::kanR*, *wza::kanR*, *flu::kanR*, *nmpC::kanR*, *eaeH::kanR*, *ydiY::kanR*, *yiaT::kanR*, *yaiO::kanR*, *mdtQ::kanR*, *pgaB::kanR*, *mipA::kanR*, *pldA::kanR*, *yzcX::kanR*, *ydeT::kanR*, *blc::kanR*, *gspD::kanR*, *yjgL::kanR*.  
(DOCX)

**S2 Table. List of all oligonucleotide primers used in this study.**  
(DOCX)

**S3 Table. List of all plasmids used in this study.** 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 µg/ml, respectively.  
(DOCX)

**S4 Table. Construction of all plasmids generated in this study.**  
(XLSX)

**S5 Table. List of all phages used in this study.** The table lists all details regarding the taxonomic classification, isolation/source, host receptors, and genomic features of all phages used in this study. As Caudovirales, all these phages are classified as Viruses/Duplodnaviria/Heung-gongvirae/Uroviricota/Caudoviricetes/Caudovirales, so only the classification on the level of family and below as defined by the current viral taxonomy of the ICTV (<https://talk.ictvonline.org/taxonomy/>; [58,144]) is listed. For the reference phages, genome sizes and estimation of RM system recognition sites are based on the reference genomes in NCBI GenBank as indicated. Note that for phage T5, the reference genome includes the 10,219 bp terminal repeat twice, unlike all other *Markadamsvirinae* genomes that we listed. The P2vir phage had been isolated as a spontaneous mutant of a P2 *cox3* prophage (supposedly unable to excise [151]) and was analyzed by whole-genome sequencing (as described in Materials and methods), revealing few single-nucleotide differences to the reference genome, the expected *cox3* mutation, and a 1-bp insertion in the lysogeny repressor gene C that results in a premature stop codon. Recognition sites of RM systems were identified using Geneious Prime 2021.0.1. The numbers include the 2 orientations of palindromic recognition sequences separately and do not account for the fact that the counts are not fully comparable, e.g., because type III RM systems require 2 unmodified recognition sites in head-to-head orientation for cleavage [47]. Given the high number of type III RM recognition sites in all genomes because of their short length (5 nt in case of both EcoCFT\_II and EcoP1\_I, see Fig 2B), this limitation is largely theoretical. ICTV, International Committee on the Taxonomy of Viruses; RM, restriction-modification.  
(XLSX)

**S6 Table. List of all phage genomes used in the in silico analyses.**  
(XLSX)

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

**S1 Text. Construction of bacterial mutant strains.**  
(DOCX)

**S2 Text. Different considerations regarding the composition of the BASEL collection and the phenotypic analyses to characterize it.** BASEL, BActeriophage SElection for your Laboratory.  
(DOCX)

**S3 Text. Generation of the Maximum-Likelihood phylogenies shown in this article.**  
(DOCX)

**S1 Fig. Supporting information for Figs 3 and 4.** (A) Maximum-Likelihood phylogeny of Drexlerviridae and the *Dhillonvirus* genus of Siphoviridae based on several core genes with bootstrap support of branches shown if >70/100. It is clearly apparent that Drexlerviridae are split into 2 major clades, one formed by Braunvirinae and Rogunavirinae and another one formed by Tempevirinae, Tunavirinae, plus a few other groups. Given that the phylogenies strongly agree on all major branches, the root of the Drexlerviridae phylogeny shown in Fig 3D was placed between these 2 major clades. (B) The bona fide RBP loci downstream of the *gpJ* homolog are shown for all small siphoviruses (Drexlerviridae and Siphoviridae of *Dhillonvirus*, *Nonagvirus*, and *Seuratvirus* genera) where we had experimentally determined the terminal receptor (together with selected representatives where previous work had determined the receptor specificity). (C) The locus encoding lateral tail fibers was analyzed in a sequence alignment of the 13 Drexlerviridae phage genomes of the BASEL collection (see Materials and methods). It is clearly visible that the upstream and downstream regions (encoding genes involved in recombination as well as primase/helicase proteins for genome replication) are highly conserved and fully syntenic, with exception of small insertions in a few sequences. Conversely, only the most 5' end of the largest lateral tail fiber protein gene is very similar among all analyzed genomes (green circle), while the rest shows neither synteny nor clear homology across all genomes. (D) The bona fide RBP locus of *E. coli* phage RTP was aligned to the homologous locus of phage AugustePiccard (Bas01) as described in Materials and methods. For the region comprising *rtp44* and *rtp45* of phage RTP, the pairwise identity of the 2 nucleotide sequences is ca. 93%. BASEL, BActeriophage SElection for your Laboratory; RBP, receptor-binding protein.

(TIF)

**S2 Fig. Supporting information for Fig 5.** (A) The locus encoding lateral tail fibers was analyzed in sequence alignments of the 5 *Dhillonvirus* phage genomes of the BASEL collection (top) and the 7 Queuvirinae phage genomes (genera *Nonagvirus* + *Seuratvirus*; bottom) of the BASEL collection as described in Materials and methods. In both cases, 2 clear dips in overall sequence similarity are obvious, once at the bona fide RBP locus and then at the lateral tail fiber locus downstream of the far 5' end of its first gene. (B) Schematic comparison of representative bona fide RBP loci as shown in S1B Fig to the corresponding allele of *E. coli* phages JenK1, HdK1, and HdsG1 that does not clearly match any of them. (C) *Nonagvirus* and *Seuratvirus* phages share a core 7-deazaguanosine biosynthesis pathway involving FolE, QueD, QueE, and QueC, which synthesizes dPreQ<sub>0</sub> that is inserted into their genomes by DpdA. In *Nonagvirus* phages, the fusion of QueC with a Gat domain to Gat-QueC results in the modification with dG<sup>+</sup> instead of dPreQ<sub>0</sub> [43]. BASEL, BActeriophage SElection for your Laboratory; Gat, glutamate amidotransferase; RBP, receptor-binding protein.

(TIF)

**S3 Fig. Supporting information for Fig 6.** (A) The locus encoding lateral tail fibers was analyzed in a sequence alignment of the Demerecviridae: Markadamsvirinae phage genomes of the BASEL collection as described in Materials and methods. Sequence identity is high upstream and downstream of the lateral tail fiber locus (with exception of presence/absence of a few putative homing endonucleases) but drops considerably at the lateral tail fiber genes. Note that, as described previously, the lateral tail fibers can either be composed of a single large polypeptide or by 2 (or more) separate proteins [49,59]. The same diversity in architecture of lateral tail fibers can also be seen at the corresponding loci of small siphoviruses (S1C and S2A Figs). (B) The illustration shows a phylogeny of the RBPs of all Markadamsvirinae phages shown in Fig 6C. Briefly, the RBP genes of all genomes (invariably encoded directly

upstream of the terminase genes) were translated, aligned, and then used to generate a phylogeny as described in Materials and methods. Three clear clusters emerge, one including all phages known to bind BtuB (left), one including all phages known to bind FepA (top right), and one including all phages known to bind FhuA (bottom right). Based on similar analyses by others [61], we conclude that the position of RBPs in these clusters is predictive of terminal receptor specificity of the phages encoding them. BASEL, BActeriophage SElection for your Laboratory; RBP, receptor-binding protein.

(TIF)

**S4 Fig. Supporting information for Fig 7.** (A) Top agar assays with different surface protein mutants of the KEIO collection in comparison to the ancestral *E. coli* K-12 BW25113 strain were performed with serial 10-fold dilutions of all Tevenvirinae phages used in this study (undiluted phage lysates at the bottom and increasingly diluted samples toward the top). The phages show impaired growth on each one of the mutants, which identifies the primary receptor of each phage (also indicated by the color code highlighted on the right). Note that growth inhibition on the primary receptor mutants is rarely total and often still enables strongly reduced, heterogeneous plaque formation (arrows), especially after prolonged incubation of the top agar plates. (B) The Maximum-Likelihood phylogeny of Tevenvirinae short tail fiber proteins reveals 2 homologous, yet clearly distinct, clusters that correlate with the absence (variant #1, like T4) or presence (variant #2) of detectable LPS core dependence as shown in Fig 7E. (C) The results of (B) indicate that variant #1, as shown for T4, binds the deep lipid A-Kdo region of the enterobacterial LPS core, while variant #2 binds a more distal part of the (probably inner) core. LPS, lipopolysaccharide.

(TIF)

**S5 Fig. Supporting information for Figs 7–10.** (A) Top agar assays of reference strain *E. coli* K-12 ΔRM carrying empty vector pBR322\_ΔPtet or pAH213\_rexAB were performed with serial 10-fold dilutions of phage T4 wild type, a T4 variant encoding an apparently hypomorphic *rIIAB* fusion (see (B)), and phage T5 (as control). The *rIIAB* mutant of phage T4 is unable to form plaques on the *rexAB*-expressing host and shows only “lysis from without” [97], while the T4 wild type and phage T5 are not affected. (B) A T4 phage mutant was erroneously obtained from a culture collection instead of the wild type and encoded a peculiar *rII* allele with fusion of the *rIIA* and *rIIB* open reading frames (shown as orange arrows; genome sequence determined by whole-genome sequencing as described in Materials and methods). Since such a mutant seems unlikely to arise spontaneously during shipping, we find it likely that this phage strain is related to the *rIIAB* fusion mutants employed for discovery of the triplet nature of the genetic code that were once commonly used (concisely reviewed in [152,153]). Notably, position and size of the deletion fusing *rIIA* and *rIIB* are indistinguishable from the sketch drawn by Benzer and Champe for the *rIIAB* fusion mutant *r1589*, which was used in the aforementioned work [154]. Unlike T4 wild type, the *rIIAB* fusion mutant is susceptible to *rexAB* when ectopically expressed from a plasmid vector (see (A)) and therefore validates functionality of the *rexAB* construct. (C) The illustration shows a sequence alignment of the locus encoding lateral tail fiber genes in phage rV5 and new isolates DrSchubert, Alex-Boehm, and JeffSchatz that broadly cover the phylogenetic range of this genus (Fig 8C). It extends from the tail tape measure protein of phage rV5 (*gp49*) to the last large lateral tail fiber gene (*gp27*) [69,70]. Note that most genes are highly conserved including the lateral tail fiber component with sugar deacetylase domain (see (F); around position 8,000 in this alignment) or the short tail fiber locus (compare (D)). Only the 3 largest lateral tail fiber genes show considerable allelic variation (highlighted by arrows) that is very strong for 2 of them (positions 10,000 to 14,000 and 16,000 to 19,000) and moderate for another one (positions 29,000 to

34,000). (D) Vequintavirinae sensu stricto (represented by rV5 and Jeff Schatz) encode 2 paralogous short tail fiber proteins and a tail fiber chaperone that are homologous to the corresponding locus in phi92-like phages like PaulScherrer and, ultimately, to short tail fiber GpS and chaperone GpU of Mu(+) (which targets a different glucose in the K12-type LPS core GpS [94,95]). (E) Amino acid sequence alignment and Maximum-Likelihood phylogeny of Gp1.2 orthologs in all tested Autographiviridae phages. Phage JeanTinguely (Bas64) belongs to the *Teseptimavirus* genus but encodes an allele of *gp1.2* that is closely related to those of the *Berlinvirus* genus, possibly explaining its resistance to PifA (see Fig 9F). (F) Amino acid sequence alignment of the lateral tail fiber Gp64 of phage N4 and a lateral tail fiber conserved among Vequintavirinae and relatives (representatives shown). The proteins share a predicted poly-GlcNAc deacetylase domain as identified by Phyre2 [138]. LPS, lipopolysaccharide.

(TIF)

**S6 Fig. Quantification of bacteriophages that depend on the O16-type O-antigen.** (A) Scheme showing how the fraction of bacteriophages depending on the O16-type O-antigen for each sewage treatment facility was quantified. Briefly, per facility, we first sampled 144 new phage isolates on *E. coli* K-12 MG1655 ΔRM *wbbL*(+) that expresses the O16-type O-antigen (Fig 2A). Subsequently, we performed qualitative top agar spot assays of these isolates on the same strain (growth control) and the parental *E. coli* K-12 MG1655 ΔRM that lacks O16-type O-antigen expression (Fig 2A). O-antigen-dependent isolates were identified as those phages able to grow on the *wbbL*(+) host but not at all on the parental strain without O-antigen. (B) The fraction of bacteriophages that require the O16-type O-antigen for infectivity was very low for all sewage treatment facilities. Precise counts were 7.6% for ARA Basel (11/144), 6.3% for ARA Canius (9/144), and 2.1% for each REAL Luzern and ARA Triengen (3/144 in both cases).

(TIF)

**S7 Fig. Confirming the identity of KEIO collection strains and isogenic mutants.** (A, B) The identity of all KEIO collection strains and isogenic mutants (*btuB* and *tolC* KOs; see Materials and methods and S1 Text) was probed by diagnostic PCR over the gene that should have been replaced with a kanamycin resistance cassette [118] and subsequent agarose gel electrophoresis of the PCR products. Panel (A) shows the agarose gel with all PCR products and, for reference, the “1 kb DNA ladder” of New England Biolabs (right). Panel (B) lists the oligonucleotide primers used for all PCRs (sequences compiled in S2 Table) and the expected sizes of PCR products for the OS PCRs of WT and KO alleles. In cases where WT and KO alleles were expected to have a similar size, we performed an additional PCR with 1 INT primer that anneals inside the kanamycin resistance cassette and 1 primer at the target locus. These PCRs can only result in a PCR product if the correct mutant strain was used as PCR template. The agarose gel in (A) confirms the identity of all KEIO collection KOs and isogenic strains that we used in our study. INT, internal; KO, knockout; OS, overspanning; WT, wild-type.

(TIF)

**S8 Fig. Complementation of all *E. coli* mutants used for bacteriophage receptor identification—*E. coli* K-12 BW25113 wild-type control.** Thirteen phages were chosen to confirm the phenotypes of all *E. coli* mutants used for bacteriophage receptor identification by genetic complementation in qualitative top agar spot assays (S9–S13 Figs). These were N4 (bona fide ECA-targeting phage dependent on *wecB*), TrudiRoth (Bas30; targets BtuB), DanielBernoulli (Bas08; targets TolC), T5 (targets FhuA), IsaakIselin (Bas10; targets YncD), KurtStettler (Bas22; targets LamB), WilhelmHis (Bas35; targets Tsx), T2 (targets FadL), PaulHMueller (Bas45; targets OmpA), FelixPlatter (Bas40; targets OmpC), AlbertHofmann (Bas47; targets

OmpF), VogelGryff (Bas25; targets LptD), and EmilieFrey (Bas61; bona fide ECA-targeting phage dependent on *wecB*). Several phages additionally depend on the LPS core and therefore on *waaC* and sometimes also *waaG* for their infectivity (see Fig 2A). These were DanielBernoulli (Bas08), WilhelmHis (Bas35), AlbertHofmann (Bas47), VogelGryff (Bas25), and EmilieFrey (Bas61). The top agar plate shown in this figure was generated with the parental *E. coli* K-12 BW25113 wild-type control that was readily infected by all phages. LPS, lipopolysaccharide. (TIF)

**S9 Fig. Complementation of *btuB*, *tolC*, and *fhuA* mutants.** Top agar spot assays with the 13 phages highlighted in S8 Fig were performed with the *btuB* (top), *tolC* (middle), and *fhuA* (bottom) mutant strains of *E. coli* K-12 BW25113. The plates on the left side were generated with strains carrying the respective complementation plasmid (see Materials and methods) and the plates on the right side with *E. coli* carrying the ev control. The plates show that the *btuB* dependency of TrudiRoth (Bas30; top), the *tolC* dependency of DanielBernoulli (Bas08; middle), and the *fhuA* dependency of T5 (bottom) could be fully complemented. ev, empty vector. (TIF)

**S10 Fig. Complementation of *yncD*, *lamB*, and *tsx* mutants.** Top agar spot assays with the 13 phages highlighted in S8 Fig were performed with the *yncD* (top), *lamB* (middle), and *tsx* (bottom) mutant strains of *E. coli* K-12 BW25113. The plates on the left side were generated with strains carrying the respective complementation plasmid (see Materials and methods) and the plates on the right side with *E. coli* carrying the ev control. The plates show that the *yncD* dependency of IsaakIselin (Bas10; top), the *lamB* dependency of KurtStettler (Bas33; middle), and the *tsx* dependency of WilhelmHis (Bas35; bottom) could be fully complemented. ev, empty vector.

(TIF)

**S11 Fig. Complementation of *ompA*, *ompC*, and *waaC* mutants.** Top agar spot assays with the 13 phages highlighted in S8 Fig were performed with the *ompA* (top) and *ompC* (middle) mutant strains of *E. coli* K-12 BW25113 as well as the *waaC* mutant of *E. coli* K-12 MG1655 ΔRM (bottom). The plates on the left side were generated with strains carrying the respective complementation plasmid (see Materials and methods) and the plates on the right side with *E. coli* carrying the ev control. The plates show that the *ompA* dependency of PaulHMueller (Bas45; top) and the *ompC* dependency of FelixPlatter (Bas40; middle) could be fully complemented. Similarly, the loss of infectivity of DanielBernoulli (Bas08), WilhelmHis (Bas35), AlbertHofmann (Bas47), VogelGryff (Bas25), and EmilieFrey (Bas61) on the *waaC* mutant could be fully complemented by genetic complementation (bottom). ev, empty vector. (TIF)

**S12 Fig. Complementation of *waaG*, *wecB*, and *fadL* mutants.** Top agar spot assays with the 13 phages highlighted in S8 Fig were performed with the *waaG* mutant of *E. coli* K-12 MG1655 ΔRM (top) and the *wecB* (middle) as well as *fadL* (bottom) mutants of *E. coli* K-12 BW25113. The plates on the left side were generated with strains carrying the respective complementation plasmid (see Materials and methods) and the plates on the right side with *E. coli* carrying the ev control. The plates show that the total loss of infectivity of EmilieFrey (Bas61) as well as the clearly reduced infectivity of AlbertHofmann (Bas47) and VogelGryff (Bas25) on the *waaG* mutant could be fully complemented by genetic complementation (top). Similarly, the *wecB* dependency of N4 (middle) and the *fadL* dependency of T2 (bottom) could be fully complemented. Unlike for all other complementation constructs, full complementation with the *wecB* plasmid required induction of expression driven by the *P<sub>lac</sub>* on pAH186SC101 with 1 mM of IPTG during the overnight culture. ev, empty vector; IPTG, isopropyl β-D-

1-thiogalactopyranoside.  
(TIF)

**S13 Fig. Complementation of *ompF* and *lptD* mutants.** Top agar spot assays with the 13 phages highlighted in S8 Fig were performed with the *ompF* knockout (top) as well as the *lptDΔ(Y658-Y678)::H* (middle) and *lptDΔ(L394-V396)::Y* mutants (middle and bottom plates; see also Fig 4C and 4D). The plates on the left side were generated with strains carrying the respective complementation plasmid (see Materials and methods) and the plates on the right side with *E. coli* carrying the ev control. The plates show that the *ompF* dependency of AlbertHofmann (Bas47; top) and the *lptD* dependency of VogelGryff (Bas25; middle and bottom plates) could be fully complemented. Curiously, complementation of the *ompF* mutant abolished infections by FelixPlatter (Bas40; compare left and right plates of the top panel). A similar phenomenon was previously observed by others who reported that *ompF* overexpression inhibited infections by bacteriophage T4, possibly by indirectly interfering with expression of its primary receptor OmpC that is also the primary receptor of FelixPlatter (Bas40; see Figs 7C and 7D and S4A) ([72] and literature cited therein). ev, empty vector.  
(TIF)

## Acknowledgments

The authors are grateful to Dr. Julie Sollier, Prof. Martin Loessner, Prof. Urs Jenal, and Prof. Marek Basler for valuable input and critical reading of the manuscript. We are deeply indebted to Prof. Urs Jenal and Prof. Christoph Dehio for sharing multiple different *E. coli* strains and plasmids (S1 and S3 Tables). Prof. Calin Guet is acknowledged for sharing plasmids (S3 Table). The authors thank high school students participating in the Biozentrum Summer Science Academy 2019, Julia Harms, and Dr. Thomas Schubert for assistance with the isolation of multiple bacteriophages described in this study (S5 Table). *Salmonella Typhimurium* strains 12023s and SL1344 were obtained from Prof. Dirk Bumann and Prof. Mederic Diard, respectively, while *E. coli* B REL606 was obtained from Dr. Jenna Gallie. Phage T4D and the P1 lysogen *E. coli* EB1484 were obtained from Prof. Kenneth Kreuzer. Phage P2vir was generously shared by Dr. Nicolas Wenner and Prof. Jay Hinton. The Coli Genetic Stock Center (CGSC) and the German Collection of Microorganisms and Cell Cultures (DSMZ) are acknowledged for sharing different strains of *E. coli* or different bacteriophages, respectively (S1 and S5 Tables). We thank Dr. Mohamed Chami and Carola Alampi of the BioEM lab of the Biozentrum, University of Basel, for their support with transmission electron microscopy. The authors are grateful to Dr. Nikoline Olsen and Prof. Bert van den Berg for helpful comments regarding the genetic basis of phage resistance to host immunity and regarding the use of porin channels as bacteriophage receptors, respectively. Dr. Christian Beisel and Dr. Geoffrey Fucile are acknowledged for advice regarding DNA sequencing. The authors are grateful to ARA Basel, ARA Canius (Lenzerheide), ARA Limeco (Dietikon), ARA Triengen, and REAL Luzern (Emmen) for providing samples of sewage plant inflow and to Universitätsspital Basel for hospital sewage.

## Author Contributions

**Conceptualization:** Alexander Harms.

**Data curation:** Alexander Harms.

**Formal analysis:** Enea Maffei, Aisyulu Shaidullina, Marco Burkholter, Yannik Heyer, Fabienne Estermann, Valentin Druelle, Patrick Sauer, Luc Willi, Sarah Michaelis, Hubert Hilbi, David S. Thaler, Alexander Harms.

**Funding acquisition:** Aisyulu Shaidullina, Valentin Druelle, Hubert Hilbi, Alexander Harms.

**Investigation:** Enea Maffei, Aisyulu Shaidullina, Marco Burkholter, Yannik Heyer, Fabienne Estermann, Valentin Druelle, Patrick Sauer, Luc Willi, Sarah Michaelis, Hubert Hilbi, David S. Thaler, Alexander Harms.

**Methodology:** Enea Maffei, Aisyulu Shaidullina, Marco Burkholter, Yannik Heyer, Fabienne Estermann, Valentin Druelle, Patrick Sauer, Luc Willi, Sarah Michaelis, Hubert Hilbi, David S. Thaler, Alexander Harms.

**Project administration:** Alexander Harms.

**Resources:** Enea Maffei, Aisyulu Shaidullina, Marco Burkholter, Yannik Heyer, Fabienne Estermann, Valentin Druelle, Patrick Sauer, Luc Willi, Sarah Michaelis, Hubert Hilbi, David S. Thaler, Alexander Harms.

**Software:** Valentin Druelle, Alexander Harms.

**Supervision:** Enea Maffei, Fabienne Estermann, Hubert Hilbi, David S. Thaler, Alexander Harms.

**Validation:** Enea Maffei, Aisyulu Shaidullina, Marco Burkholter, Yannik Heyer, Fabienne Estermann, Valentin Druelle, Patrick Sauer, Luc Willi, Sarah Michaelis, Hubert Hilbi, David S. Thaler, Alexander Harms.

**Visualization:** Fabienne Estermann, Alexander Harms.

**Writing – original draft:** Alexander Harms.

**Writing – review & editing:** Enea Maffei, Aisyulu Shaidullina, Marco Burkholter, Yannik Heyer, Fabienne Estermann, Valentin Druelle, Patrick Sauer, Luc Willi, Sarah Michaelis, Hubert Hilbi, David S. Thaler, Alexander Harms.

## References

1. Mahmoudabadi G, Phillips R. A comprehensive and quantitative exploration of thousands of viral genomes. *elife*. 2018; 7. <https://doi.org/10.7554/elife.31955> PMID: 29624169; PubMed Central PMCID: PMC5908442.
2. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol*. 2020; 18(3):125–38. Epub 2020/02/06. <https://doi.org/10.1038/s41579-019-0311-5> PMID: 32015529.
3. 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. <https://doi.org/10.1146/annurev-virology-100114-054952> PMID: 27741409.
4. Keen EC. A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays*. 2015; 37(1):6–9. <https://doi.org/10.1002/bies.201400152> PMID: 25521633; PubMed Central PMCID: PMC4418462.
5. Demerec M, Fano U. Bacteriophage-Resistant Mutants in *Escherichia Coli*. *Genetics*. 1945; 30(2):119–36. Epub 1945/03/01. <https://doi.org/10.1093/genetics/30.2.119> PMID: 17247150; PubMed Central PMCID: PMC1209279.
6. Abedon ST. The murky origin of Snow White and her T-even dwarfs. *Genetics*. 2000; 155(2):481–6. Epub 2000/06/03. <https://doi.org/10.1093/genetics/155.2.481> PMID: 10835374; PubMed Central PMCID: PMC1461100.
7. Casjens SR, Hendrix RW. Bacteriophage lambda: Early pioneer and still relevant. *Virology*. 2015; 479–480:310–30. <https://doi.org/10.1016/j.virol.2015.02.010> PMID: 25742714; PubMed Central PMCID: PMC4424060.

8. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J. 2017; 11(7):1511–20. <https://doi.org/10.1038/ismej.2017.16> PMID: 28291233; PubMed Central PMCID: PMC5520141.
9. Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their phage foes. Nature. 2020; 577(7790):327–36. Epub 2020/01/17. <https://doi.org/10.1038/s41586-019-1894-8> PMID: 31942051.
10. Dy RL, Richter C, Salmond GP, Fineran PC. Remarkable Mechanisms in Microbes to Resist Phage Infections. Annu Rev Virol. 2014; 1(1):307–31. <https://doi.org/10.1146/annurev-virology-031413-085500> PMID: 26958724.
11. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource. Nat Rev Microbiol. 2020; 18(2):113–9. Epub 2019/11/07. <https://doi.org/10.1038/s41579-019-0278-2> PMID: 31695182.
12. 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. <https://doi.org/10.1146/annurev-virology-011620-040628> PMID: 32559405.
13. Schmidt C. Phage therapy's latest makeover. Nat Biotechnol. 2019; 37(6):581–6. Epub 2019/05/10. <https://doi.org/10.1038/s41587-019-0133-z> PMID: 31068679.
14. Kilcher S, Loessner MJ. Engineering Bacteriophages as Versatile Biologics. Trends Microbiol. 2019; 27(4):355–67. Epub 2018/10/17. <https://doi.org/10.1016/j.tim.2018.09.006> PMID: 30322741.
15. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, et al. INfrastructure for a PHAge REference Database: Identification of large-scale biases in the current collection of phage genomes. bioRxiv. 2021:2021.05.01.442102. <https://doi.org/10.1101/2021.05.01.442102>
16. Hyman P. Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. Pharmaceuticals (Basel). 2019; 12(1). Epub 2019/03/14. <https://doi.org/10.3390/ph12010035> PMID: 30862020; PubMed Central PMCID: PMC6469166.
17. Gordillo Altamirano FL, Barr JJ. Phage Therapy in the Postantibiotic Era. Clin Microbiol Rev. 2019; 32(2). Epub 2019/01/18. <https://doi.org/10.1128/CMR.00066-18> PMID: 30651225; PubMed Central PMCID: PMC6431132.
18. 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. <https://doi.org/10.1016/j.chom.2019.01.014> PMID: 30763536.
19. 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. <https://doi.org/10.1093/ofid/ofaa389> PMID: 33005701; PubMed Central PMCID: PMC7519779.
20. Pires DP, Costa AR, Pinto G, Meneses L, Azeredo J. Current challenges and future opportunities of phage therapy. FEMS Microbiol Rev. 2020; 44(6):684–700. Epub 2020/05/31. <https://doi.org/10.1093/femsre/fuaa017> PMID: 32472938.
21. 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). <https://doi.org/10.3390/v11050454> PMID: 31109012; PubMed Central PMCID: PMC6563267.
22. 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. <https://doi.org/10.3390/v12090986> PMID: 32899836; PubMed Central PMCID: PMC7552041.
23. Mathieu A, Dion M, Deng L, Tremblay D, Moncaut E, Shah SA, et al. Virulent coliphages in 1-year-old children fecal samples are fewer, but more infectious than temperate coliphages. Nat Commun. 2020; 11(1):378. Epub 2020/01/19. <https://doi.org/10.1038/s41467-019-14042-z> PMID: 31953385; PubMed Central PMCID: PMC6969025.
24. Sørensen PE, Van Den Broeck W, Kiil K, Jasinskyte D, Moodley A, Garmyn A, et al. New insights into the biodiversity of coliphages in the intestine of poultry. Sci Rep. 2020; 10(1):15220. Epub 2020/09/18. <https://doi.org/10.1038/s41598-020-72177-2> PMID: 32939020; PubMed Central PMCID: PMC7494930.
25. Michniewski S, Redgwell T, Grigonyte A, Rihtman B, Aguiló-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. <https://doi.org/10.1111/1462-2920.14590> PMID: 30884081; PubMed Central PMCID: PMC6563131.
26. Smith R, O'Hara M, Hobman JL, Millard AD. Draft Genome Sequences of 14 *Escherichia coli* Phages Isolated from Cattle Slurry. Genome Announc. 2015; 3(6). Epub 2016/01/02. <https://doi.org/10.1128/genomeA.01364-15> PMID: 26722010; PubMed Central PMCID: PMC4698387.

27. Pacifico C, Hilbert M, Sofka D, Dinhopl N, Pap IJ, Aspock C, et al. Natural Occurrence of *Escherichia coli*-Infecting Bacteriophages in Clinical Samples. *Front Microbiol.* 2019; 10:2484. Epub 2019/11/19. <https://doi.org/10.3389/fmicb.2019.02484> PMID: 31736918; PubMed Central PMCID: PMC6834657.
28. 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. <https://doi.org/10.1016/j.virol.2014.08.024> PMID: 25240328; PubMed Central PMCID: PMC4301999.
29. Olsen NS, Hendriksen NB, Hansen LH, Kot W. A New High-Throughput Screening Method for Phages: Enabling Crude Isolation and Fast Identification of Diverse Phages with Therapeutic Potential. *PHAGE.* 2020; 1(3):137–48. <https://doi.org/10.1089/phage.2020.0016>
30. Drexler H. Bacteriophage T1. In: Calendar R, editor. *The Bacteriophages.* Boston, MA: Springer US; 1988. p. 235–58.
31. Wittmann J, Turner D, Millard AD, Mahadevan P, Kropinski AM, Adriaenssens EM. From Orphan Phage to a Proposed New Family—the Diversity of N4-Like Viruses. *Antibiotics (Basel).* 2020; 9(10). Epub 2020/10/04. <https://doi.org/10.3390/antibiotics9100663> PMID: 33008130; PubMed Central PMCID: PMC7650795.
32. Bertani G. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J Bacteriol.* 2004; 186(3):595–600. Epub 2004/01/20. <https://doi.org/10.1128/JB.186.3.595-600.2004> PMID: 14729683; PubMed Central PMCID: PMC321500.
33. 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. <https://doi.org/10.1038/s41579-018-0070-8> PMID: 30104690.
34. 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. <https://doi.org/10.1146/annurev-micro-102215-095308> PMID: 27359214.
35. Letarov AV, Kulikov EE. Adsorption of Bacteriophages on Bacterial Cells. *Biochemistry (Mosc).* 2017; 82(13):1632–58. Epub 2018/03/11. <https://doi.org/10.1134/S0006297917130053> PMID: 29523063.
36. 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. <https://doi.org/10.1111/mmi.13729> PMID: 28618013.
37. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett.* 2016; 363(4). Epub 2016/01/13. <https://doi.org/10.1093/femsle/fnw002> PMID: 26755501.
38. 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. <https://doi.org/10.1093/femsle/fnaa013> PMID: 32009155.
39. 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. <https://doi.org/10.1016/j.copbio.2020.10.002> PMID: 33202354.
40. Bertani B, Ruiz N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus.* 2018; 8(1). Epub 2018/08/02. <https://doi.org/10.1128/ecosalplus.ESP-0001-2018> PMID: 30066669; PubMed Central PMCID: PMC6091223.
41. Rai AK, Mitchell AM. Enterobacterial Common Antigen: Synthesis and Function of an Enigmatic Molecule. *mBio.* 2020; 11(4). Epub 2020/08/14. <https://doi.org/10.1128/mBio.01914-20> PMID: 32788387; PubMed Central PMCID: PMC7439462.
42. Liu D, Reeves PR. *Escherichia coli* K12 regains its O antigen. *Microbiology.* 1994; 140(Pt 1):49–57. <https://doi.org/10.1099/13500872-140-1-49> PMID: 7512872.
43. 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. <https://doi.org/10.1038/s41467-019-13384-y> PMID: 31784519; PubMed Central PMCID: PMC6884629.
44. 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. <https://doi.org/10.1021/acs.chemrev.6b00114> PMID: 27319741.
45. Loenen WA, Dryden DT, Raleigh EA, Wilson GG. Type I restriction enzymes and their relatives. *Nucleic Acids Res.* 2014; 42(1):20–44. Epub 2013/09/27. <https://doi.org/10.1093/nar/gkt847> PMID: 24068554; PubMed Central PMCID: PMC3874165.

46. Pingoud A, Wilson GG, Wende W. Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.* 2014; 42(12):7489–527. Epub 2014/06/01. <https://doi.org/10.1093/nar/gku447> PMID: 24878924; PubMed Central PMCID: PMC4081073.
47. Rao DN, Dryden DT, Bheemanaik S. Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* 2014; 42(1):45–55. Epub 2013/07/19. <https://doi.org/10.1093/nar/gkt616> PMID: 23863841; PubMed Central PMCID: PMC3874151.
48. Christie GE, Calendar R. Bacteriophage P2. *Bacteriophage.* 2016; 6(1):e1145782. Epub 2016/05/05. <https://doi.org/10.1080/21597081.2016.1145782> PMID: 27144088; PubMed Central PMCID: PMC4836473.
49. Goulet A, Spinelli S, Mahony J, Cambillau C. Conserved and Diverse Traits of Adhesion Devices from *Siphoviridae* Recognizing Proteinaceous or Saccharidic Receptors. *Viruses.* 2020; 12(5). Epub 2020/05/10. <https://doi.org/10.3390/v12050512> PMID: 32384698; PubMed Central PMCID: PMC7291167.
50. Piya D, Lessor L, Koehler B, Stonecipher A, Cahill J, Gill JJ. Genome-wide screens reveal *Escherichia coli* genes required for growth of T1-like phage LL5 and V5-like phage LL12. *Sci Rep.* 2020; 10(1):8058. Epub 2020/05/18. <https://doi.org/10.1038/s41598-020-64981-7> PMID: 32415154; PubMed Central PMCID: PMC7229145.
51. Li P, Lin H, Mi Z, Xing S, Tong Y, Wang J. Screening of Polyvalent Phage-Resistant *Escherichia coli* Strains Based on Phage Receptor Analysis. *Front Microbiol.* 2019; 10:850. Epub 2019/05/21. <https://doi.org/10.3389/fmicb.2019.00850> PMID: 31105661; PubMed Central PMCID: PMC6499177.
52. Wietzorek 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. <https://doi.org/10.1128/JB.188.4.1419-1436.2006> PMID: 16452425; PubMed Central PMCID: PMC1367250.
53. Hamdi S, Rousseau GM, Labrie SJ, Tremblay DM, Kourda RS, Ben Slama K, et al. Characterization of two polyvalent phages infecting *Enterobacteriaceae*. *Sci Rep.* 2017; 7:40349. Epub 2017/01/17. <https://doi.org/10.1038/srep40349> PMID: 28091598; PubMed Central PMCID: PMC5238451.
54. Wang J, Hofnung M, Charbit A. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LambB, its receptor at the surface of *Escherichia coli* K-12. *J Bacteriol.* 2000; 182(2):508–12. Epub 2000/01/12. <https://doi.org/10.1128/JB.182.2.508-512.2000> PMID: 10629200; PubMed Central PMCID: PMC94303.
55. Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science.* 2012; 335(6067):428–32. Epub 2012/01/28. <https://doi.org/10.1126/science.1214449> PMID: 22282803; PubMed Central PMCID: PMC3306806.
56. Dunne M, Ruf B, Tala M, Qabrati X, Ernst P, Shen Y, et al. Reprogramming Bacteriophage Host Range through Structure-Guided Design of Chimeric Receptor Binding Proteins. *Cell Rep.* 2019; 29(5):1336–50 e4. Epub 2019/10/31. <https://doi.org/10.1016/j.celrep.2019.09.062> PMID: 31665644.
57. Sazinas P, Redgwell T, Rihtman B, Grigonyte A, Michniewski S, Scanlan DJ, et al. Comparative Genomics of Bacteriophage of the Genus *Seuratvirus*. *Genome Biol Evol.* 2018; 10(1):72–6. Epub 2017/12/23. <https://doi.org/10.1093/gbe/evx275> PMID: 29272407; PubMed Central PMCID: PMC5758909.
58. Krupovic M, Turner D, Morozova V, Dyall-Smith M, Oksanen HM, Edwards R, et al. Bacterial Viruses Subcommittee and Archaeal Viruses Subcommittee of the ICTV: update of taxonomy changes in 2021. *Arch Virol.* 2021. Epub 2021/08/22. <https://doi.org/10.1007/s00705-021-05205-9> PMID: 34417873.
59. Golomidova AK, Kulikov EE, Prokhorov NS, Guerrero-Ferreira Rcapital Es C, Knirel YA, Kostryukova ES, et al. Branched Lateral Tail Fiber Organization in T5-Like Bacteriophages DT57C and DT571/2 is Revealed by Genetic and Functional Analysis. *Viruses.* 2016; 8(1). Epub 2016/01/26. <https://doi.org/10.3390/v8010026> PMID: 26805872; PubMed Central PMCID: PMC4728585.
60. Hong J, Kim KP, Heu S, Lee SJ, Adhya S, Ryu S. Identification of host receptor and receptor-binding module of a newly sequenced T5-like phage EPS7. *FEMS Microbiol Lett.* 2008; 289(2):202–9. Epub 2008/11/26. <https://doi.org/10.1111/j.1574-6968.2008.01397.x> PMID: 19025561.
61. Gencay YE, Gambino M, Prussing TF, Brondsted L. The genera of bacteriophages and their receptors are the major determinants of host range. *Environ Microbiol.* 2019; 21(6):2095–111. Epub 2019/03/20. <https://doi.org/10.1111/1462-2920.14597> PMID: 30888719.
62. Rabsch W, Ma L, Wiley G, Najar FZ, Kaserer W, Schuerch DW, et al. FepA- and TonB-dependent bacteriophage H8: receptor binding and genomic sequence. *J Bacteriol.* 2007; 189(15):5658–74. Epub 2007/05/29. <https://doi.org/10.1128/JB.00437-07> PMID: 17526714; PubMed Central PMCID: PMC1951831.

63. Davison J. Pre-early functions of bacteriophage T5 and its relatives. *Bacteriophage*. 2015; 5(4):e1086500. Epub 2016/02/24. <https://doi.org/10.1080/21597081.2015.1086500> PMID: 26904381; PubMed Central PMCID: PMC4743489.
64. Rusinov IS, Ershova AS, Karyagina AS, Spirin SA, Alexeevskii AV. Avoidance of recognition sites of restriction-modification systems is a widespread but not universal anti-restriction strategy of prokaryotic viruses. *BMC Genomics*. 2018; 19(1):885. Epub 2018/12/12. <https://doi.org/10.1186/s12864-018-5324-3> PMID: 30526500; PubMed Central PMCID: PMC6286503.
65. 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. <https://doi.org/10.1093/gbe/evr059> PMID: 21746838; PubMed Central PMCID: PMC3157838.
66. Washizaki A, Yonesaki T, Otsuka Y. Characterization of the interactions between *Escherichia coli* receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. *Microbiology*. 2016; 5(6):1003–15. Epub 2016/06/09. <https://doi.org/10.1002/mbo3.384> PMID: 27273222; PubMed Central PMCID: PMC5221442.
67. 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. <https://doi.org/10.1073/pnas.1501064112> PMID: 26283379; PubMed Central PMCID: PMC4568249.
68. Thomas JA, Orwenyo J, Wang LX, Black LW. The Odd "RB" Phage-Identification of Arabinosylation as a New Epigenetic Modification of DNA in T4-Like Phage RB69. *Viruses*. 2018; 10(6). Epub 2018/06/13. <https://doi.org/10.3390/v10060313> PMID: 29890699; PubMed Central PMCID: PMC6024577.
69. Kropinski AM, Waddell T, Meng J, Franklin K, Ackermann HW, Ahmed R, et al. The host-range, genomics and proteomics of *Escherichia coli*O157:H7 bacteriophage rV5. *Virol J*. 2013; 10:76. Epub 2013/03/19. <https://doi.org/10.1186/1743-422X-10-76> PMID: 23497209; PubMed Central PMCID: PMC3606486.
70. 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. <https://doi.org/10.1128/JVI.00801-12> PMID: 22787233; PubMed Central PMCID: PMC3457257.
71. Amor K, Heinrichs DE, Frirdich E, Ziebell K, Johnson RP, Whitfield C. Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infect Immun*. 2000; 68(3):1116–24. Epub 2000/02/26. <https://doi.org/10.1128/IAI.68.3.1116-1124.2000> PMID: 10678915; PubMed Central PMCID: PMC97256.
72. Mutualik 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. <https://doi.org/10.1371/journal.pbio.3000877> PMID: 33048924; PubMed Central PMCID: PMC7553319 following competing interests: VKM, AMD, and APA consult for and hold equity in Felix Biotechnology, Inc.
73. Qimron U, Marintcheva B, Tabor S, Richardson CC. Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc Natl Acad Sci U S A*. 2006; 103(50):19039–44. Epub 2006/12/01. <https://doi.org/10.1073/pnas.0609428103> PMID: 17135349; PubMed Central PMCID: PMC1748173.
74. 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. <https://doi.org/10.1074/jbc.M114.614222> PMID: 25697363; PubMed Central PMCID: PMC4400320.
75. Cuervo A, Fabrega-Ferrer M, Machon C, Conesa JJ, Fernandez FJ, Perez-Luque R, et al. Structures of T7 bacteriophage portal and tail suggest a viral DNA retention and ejection mechanism. *Nat Commun*. 2019; 10(1):3746. Epub 2019/08/23. <https://doi.org/10.1038/s41467-019-11705-9> PMID: 31431626; PubMed Central PMCID: PMC6702177.
76. 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. <https://doi.org/10.1126/science.1231887> PMID: 23306440; PubMed Central PMCID: PMC3873743.
77. Scholl D, Merrill 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. <https://doi.org/10.1128/JB.187.24.8499-8503.2005> PMID: 16321955; PubMed Central PMCID: PMC1317022.
78. Adriaenssens EM, Sullivan MB, Knezevic P, van Zyl LJ, Sarkar BL, Dutill BE, et al. Taxonomy of prokaryotic viruses: 2018–2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Arch Virol*. 2020; 165(5):1253–60. Epub 2020/03/13. <https://doi.org/10.1007/s00705-020-04577-8> PMID: 32162068.

79. Ando H, Lemire S, Pires DP, Lu TK. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Syst.* 2015; 1(3):187–96. Epub 2016/03/15. <https://doi.org/10.1016/j.cels.2015.08.013> PMID: 26973885; PubMed Central PMCID: PMC4785837.
80. Studier FW, Movva NR. SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J Virol.* 1976; 19(1):136–45. Epub 1976/07/01. <https://doi.org/10.1128/JVI.19.1.136-145.1976> PubMed Central PMCID: PMC354840. PMID: 781304.
81. Bandyopadhyay PK, Studier FW, Hamilton DL, Yuan R. Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J Mol Biol.* 1985; 182(4):567–78. Epub 1985/04/20. [https://doi.org/10.1016/0022-2836\(85\)90242-6](https://doi.org/10.1016/0022-2836(85)90242-6) PMID: 2989534.
82. 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. <https://doi.org/10.1128/jb.173.4.1536-1543.1991> PMID: 1995595; PubMed Central PMCID: PMC207293.
83. 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. <https://doi.org/10.1073/pnas.81.5.1465> PMID: 6324192; PubMed Central PMCID: PMC344857.
84. Kiino DR, Rothman-Denes LB. Genetic analysis of bacteriophage N4 adsorption. *J Bacteriol.* 1989; 171(9):4595–602. Epub 1989/09/01. <https://doi.org/10.1128/jb.171.9.4595-4602.1989> PMID: 2670887; PubMed Central PMCID: PMC210256.
85. Kiino DR, Licudine R, Wilt K, Yang DH, Rothman-Denes LB. A cytoplasmic protein, NfrC, is required for bacteriophage N4 adsorption. *J Bacteriol.* 1993; 175(21):7074–80. Epub 1993/11/01. <https://doi.org/10.1128/jb.175.21.7074-7080.1993> PMID: 8226648; PubMed Central PMCID: PMC206835.
86. 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. <https://doi.org/10.1111/mmi.13710> PMID: 28513100.
87. McPartland J, Rothman-Denes LB. The tail sheath of bacteriophage N4 interacts with the *Escherichia coli* receptor. *J Bacteriol.* 2009; 191(2):525–32. Epub 2008/11/18. <https://doi.org/10.1128/JB.01423-08> PMID: 19011026; PubMed Central PMCID: PMC2620810.
88. Shinedling S, Parma D, Gold L. Wild-type bacteriophage T4 is restricted by the lambda rex genes. *J Virol.* 1987; 61(12):3790–4. Epub 1987/12/01. <https://doi.org/10.1128/JVI.61.12.3790-3794.1987> PubMed Central PMCID: PMC255994. PMID: 2960831.
89. Whichard JM, Weigt LA, Borris DJ, Li LL, Zhang Q, Kapur V, et al. Complete genomic sequence of bacteriophage felix o1. *Viruses.* 2010; 2(3):710–30. Epub 2010/03/01. <https://doi.org/10.3390/v2030710> PMID: 21994654; PubMed Central PMCID: PMC3185647.
90. Kaczorowska J, Casey E, Neve H, Franz C, Noben JP, Lugli GA, et al. A Quest of Great Importance—Developing a Broad Spectrum *Escherichia coli* Phage Collection. *Viruses.* 2019; 11(10). Epub 2019/09/29. <https://doi.org/10.3390/v11100899> PMID: 31561510; PubMed Central PMCID: PMC6832132.
91. Simoliunas E, Vilkaityte M, Kaliniene L, Zajanckauskaite A, Kaupinis A, Staniulis J, et al. Incomplete LPS Core-Specific FelixO1-Like Virus vB\_EcoM\_VpaE1. *Viruses.* 2015; 7(12):6163–81. Epub 2015/12/04. <https://doi.org/10.3390/v7122932> PMID: 26633460; PubMed Central PMCID: PMC4690856.
92. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. *Science.* 1997; 277(5331):1453–62. Epub 1997/09/05. <https://doi.org/10.1126/science.277.5331.1453> PMID: 9278503.
93. Hendrix RW, Duda RL. Bacteriophage lambda PaPa: not the mother of all lambda phages. *Science.* 1992; 258(5085):1145–8. <https://doi.org/10.1126/science.1439823> PMID: 1439823.
94. Sandulache R, Prehm P, Kamp D. Cell wall receptor for bacteriophage Mu G(+). *J Bacteriol.* 1984; 160(1):299–303. Epub 1984/10/01. <https://doi.org/10.1128/jb.160.1.299-303.1984> PMID: 6384194; PubMed Central PMCID: PMC214716.
95. 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 2019/06/19. <https://doi.org/10.1038/s41564-019-0477-7> PMID: 31209305.
96. Hyman P, Abedon ST. Bacteriophage host range and bacterial resistance. *Adv Appl Microbiol.* 2010; 70:217–48. Epub 2010/04/03. [https://doi.org/10.1016/S0065-2164\(10\)70007-1](https://doi.org/10.1016/S0065-2164(10)70007-1) PMID: 20359459.
97. Abedon ST. Lysis from without. *Bacteriophage.* 2011; 1(1):46–9. <https://doi.org/10.4161/bact.1.13980> PMID: 21687534; PubMed Central PMCID: PMC3109453.
98. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, et al. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J Mol Biol.* 2009; 394(4):644–52. Epub 2009/09/30. <https://doi.org/10.1016/j.jmb.2009.09.052> PMID: 19786035.

99. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *J Mol Biol.* 2009; 394(4):653–80. <https://doi.org/10.1016/j.jmb.2009.09.021> PMID: 19765592.
100. Ochman H, Selander RK. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol.* 1984; 157(2):690–3. <https://doi.org/10.1128/jb.157.2.690-693.1984> PMID: 6363394; PubMed Central PMCID: PMC215307.
101. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol.* 2010; 8(3):207–17. <https://doi.org/10.1038/nrmicro2298> PMID: 20157339.
102. Sueki A, Stein F, Savitski MM, Selkirk J, Typas A. Systematic Localization of *Escherichia coli* Membrane Proteins. *mSystems.* 2020; 5(2). Epub 2020/03/05. <https://doi.org/10.1128/mSystems.00808-19> PMID: 32127419; PubMed Central PMCID: PMC7055658.
103. Kortright KE, Chan BK, Turner PE. High-throughput discovery of phage receptors using transposon insertion sequencing of bacteria. *Proc Natl Acad Sci U S A.* 2020; 117(31):18670–9. Epub 2020/07/18. <https://doi.org/10.1073/pnas.2001888117> PMID: 32675236; PubMed Central PMCID: PMC7414163.
104. ViralZone SloB. Bacteriophage lambda replication cycle 2014 [25.07.2021]. Available from: <https://viralzone.expasy.org/4596>.
105. Braun V. FhuA (TonA), the career of a protein. *J Bacteriol.* 2009; 191(11):3431–6. Epub 2009/03/31. <https://doi.org/10.1128/JB.00106-09> PMID: 19329642; PubMed Central PMCID: PMC2681897.
106. Gurney J, Brown SP, Kaltz O, Hochberg ME. Steering Phages to Combat Bacterial Pathogens. *Trends Microbiol.* 2020; 28(2):85–94. Epub 2019/11/21. <https://doi.org/10.1016/j.tim.2019.10.007> PMID: 31744662; PubMed Central PMCID: PMC6980653.
107. Bauer RJ, Zhelkovsky A, Bilotti K, Crowell LE, Evans TC Jr., McReynolds LA, et al. Comparative analysis of the end-joining activity of several DNA ligases. *PLoS ONE.* 2017; 12(12):e0190062. Epub 2017/12/29. <https://doi.org/10.1371/journal.pone.0190062> PMID: 29284038; PubMed Central PMCID: PMC5746248.
108. Thomason LC, Schiltz CJ, Court C, Hosford CJ, Adams MC, Chappie JS, et al. Bacteriophage λ RexA and RexB Functions Assist the Transition from Lysogeny to Lytic Growth. *bioRxiv.* 2021:2021.04.19.439022. <https://doi.org/10.1111/mmi.14792> PMID: 34379857
109. Keen EC. Tradeoffs in bacteriophage life histories. *Bacteriophage.* 2014; 4(1):e28365. Epub 2014/03/13. <https://doi.org/10.4161/bact.28365> PMID: 24616839; PubMed Central PMCID: PMC3942329.
110. Law R. Optimal Life Histories Under Age-Specific Predation. *Am Nat.* 1979; 114(3):399–417.
111. Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. *Nat Microbiol.* 2018; 3(7):754–66. Epub 2018/06/06. <https://doi.org/10.1038/s41564-018-0166-y> PMID: 29867096.
112. Clokie MRJ, Blasdel BG, Demars BOL, Sicheritz-Pontén T. Rethinking Phage Ecology by Rooting it Within an Established Plant Framework. *PHAGE.* 2020; 1(3):121–36. <https://doi.org/10.1089/phage.2020.0015>
113. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, et al. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *mBio.* 2014; 5(1):e01051–13. Epub 2014/02/06. <https://doi.org/10.1128/mBio.01051-13> PMID: 24496795; PubMed Central PMCID: PMC3950523.
114. M9 minimal medium (standard). *Cold Spring Harb Protoc.* 2010; 2010(8):pdb.rec12295. <https://doi.org/10.1101/pdb.rec12295>
115. 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. [https://doi.org/10.1007/978-1-60327-164-6\\_7](https://doi.org/10.1007/978-1-60327-164-6_7) PMID: 19066811.
116. Kauffman KM, Polz MF. Streamlining standard bacteriophage methods for higher throughput. *MethodsX.* 2018; 5:159–72. Epub 2019/01/10. <https://doi.org/10.1016/j.mex.2018.01.007> PMID: 30622914; PubMed Central PMCID: PMC6318102.
117. Golec P, Dabrowski K, Hejnowicz MS, Gozdek A, Los JM, Wegrzyn G, et al. A reliable method for storage of tailed phages. *J Microbiol Methods.* 2011; 84(3):486–9. Epub 2011/01/25. <https://doi.org/10.1016/j.mimet.2011.01.007> PMID: 21256885.
118. 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. <https://doi.org/10.1038/msb4100050> PMID: 16738554; PubMed Central PMCID: PMC1681482.

119. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A. 2000; 97(12):6640–5. <https://doi.org/10.1073/pnas.120163297> PMID: 10829079; PubMed Central PMCID: PMC18686.
120. Harms A, Fino C, Sørensen MA, Semsey S, Gerdes K. Prophages and Growth Dynamics Confound Experimental Results with Antibiotic-Tolerant Persister Cells. mBio. 2017; 8(6). <https://doi.org/10.1128/mBio.01964-17> PMID: 29233898; PubMed Central PMCID: PMC5727415.
121. 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. <https://doi.org/10.1016/j.cub.2015.12.041> PMID: 26804559.
122. Fino C, Vestergaard M, Ingmer H, Pierrel F, Gerdes K, Harms A. PasT of *Escherichia coli* sustains antibiotic tolerance and aerobic respiration as a bacterial homolog of mitochondrial Coq10. Microbiology. 2020; 9(8):e1064. Epub 2020/06/20. <https://doi.org/10.1002/mbo3.1064> PMID: 32558363; PubMed Central PMCID: PMC7424257.
123. Wiles TJ, Kulesus RR, Mulvey MA. Origins and virulence mechanisms of uropathogenic *Escherichia coli*. Exp Mol Pathol. 2008; 85(1):11–9. Epub 2008/05/17. <https://doi.org/10.1016/j.yexmp.2008.03.007> PMID: 18482721; PubMed Central PMCID: PMC2595135.
124. Grad YH, Godfrey P, Cerquiera GC, Mariani-Kurdjian P, Gouali M, Bingen E, et al. Comparative genomics of recent Shiga toxin-producing *Escherichia coli* O104:H4: short-term evolution of an emerging pathogen. mBio. 2013; 4(1):e00452–12. Epub 2013/01/24. <https://doi.org/10.1128/mBio.00452-12> PMID: 23341549; PubMed Central PMCID: PMC3551546.
125. Bernier C, Gounon P, Le Bouguenec C. Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative *Escherichia coli* as a sensitive probe for detecting the AAF-encoding operon family. Infect Immun. 2002; 70(8):4302–11. Epub 2002/07/16. <https://doi.org/10.1128/IAI.70.8.4302-4311.2002> PMID: 12117939; PubMed Central PMCID: PMC128174.
126. Branchu P, Bawn M, Kingsley RA. Genome Variation and Molecular Epidemiology of *Salmonella enterica* Serovar Typhimurium Pathovariants. Infect Immun. 2018; 86(8). Epub 2018/05/23. <https://doi.org/10.1128/IAI.00079-18> PMID: 29784861; PubMed Central PMCID: PMC6056856.
127. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6(5):343–5. Epub 2009/04/14. <https://doi.org/10.1038/nmeth.1318> PMID: 19363495.
128. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. 2008; 8:91. <https://doi.org/10.1186/1472-6750-8-91> PMID: 19055817; PubMed Central PMCID: PMC2629768.
129. 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. <https://doi.org/10.1093/nar/gku1046> PMID: 25378308; PubMed Central PMCID: PMC4383893.
130. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sanchez-Perez M, Gomez-Romero L, Ledezma-Tejeida D, et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. Nucleic Acids Res. 2019; 47(D1):D212–D20. Epub 2018/11/06. <https://doi.org/10.1093/nar/gky1077> PMID: 30395280; PubMed Central PMCID: PMC6324031.
131. Czajkowski R, Ozymko Z, Lojkowska E. Application of zinc chloride precipitation method for rapid isolation and concentration of infectious *Pectobacterium* spp. and *Dickeya* spp. lytic bacteriophages from surface water and plant and soil extracts. Folia Microbiol (Praha). 2016; 61(1):29–33. Epub 2015/06/24. <https://doi.org/10.1007/s12223-015-0411-1> PMID: 26099750; PubMed Central PMCID: PMC4691450.
132. You L, Suthers PF, Yin J. Effects of *Escherichia coli* physiology on growth of phage T7 *in vivo* and *in silico*. J Bacteriol. 2002; 184(7):1888–94. Epub 2002/03/13. <https://doi.org/10.1128/JB.184.7.1888-1894.2002> PMID: 11889095; PubMed Central PMCID: PMC134924.
133. Abedon ST, Katsaounis TI. Basic Phage Mathematics. Methods Mol Biol. 2018; 1681:3–30. Epub 2017/11/15. [https://doi.org/10.1007/978-1-4939-7343-9\\_1](https://doi.org/10.1007/978-1-4939-7343-9_1) PMID: 29134583.
134. 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. <https://doi.org/10.1038/s41598-017-07910-5> PMID: 28811656; PubMed Central PMCID: PMC5557969.
135. Ecale Zhou CL, Malfatti S, Kimbrel J, Philipson C, McNair K, Hamilton T, et al. multiPhATE: bioinformatics pipeline for functional annotation of phage isolates. Bioinformatics. 2019; 35(21):4402–4. Epub 2019/05/16. <https://doi.org/10.1093/bioinformatics/btz258> PMID: 31086982; PubMed Central PMCID: PMC6821344.

136. 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. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690; PubMed Central PMCID: PMC3603318.
137. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019; 47(D1): D351–D60. Epub 2018/11/07. <https://doi.org/10.1093/nar/gky1100> PMID: 30398656; PubMed Central PMCID: PMC6323941.
138. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015; 10(6):845–58. <https://doi.org/10.1038/nprot.2015.053> PMID: 25950237; PubMed Central PMCID: PMC5298202.
139. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol*. 2019; 1962:1–14. Epub 2019/04/26. [https://doi.org/10.1007/978-1-4939-9173-0\\_1](https://doi.org/10.1007/978-1-4939-9173-0_1) PMID: 31020551; PubMed Central PMCID: PMC6768409.
140. Kongari R, Rajaura M, Cahill J, Rasche E, Mijalis E, Berry J, et al. Phage spanins: diversity, topological dynamics and gene convergence. *BMC Bioinformatics*. 2018; 19(1):326. Epub 2018/09/17. <https://doi.org/10.1186/s12859-018-2342-8> PMID: 30219026; PubMed Central PMCID: PMC6139136.
141. Cahill J, Young R. Phage Lysis: Multiple Genes for Multiple Barriers. *Adv Virus Res*. 2019; 103:33–70. Epub 2019/01/13. <https://doi.org/10.1016/bs.aivir.2018.09.003> PMID: 30635077; PubMed Central PMCID: PMC6733033.
142. Adriaenssens E, Brister JR. How to Name and Classify Your Phage: An Informal Guide. *Viruses*. 2017; 9(4). <https://doi.org/10.3390/v9040070> PMID: 28368359; PubMed Central PMCID: PMC5408676.
143. Tolstoy I, Kropinski AM, Brister JR. Bacteriophage Taxonomy: An Evolving Discipline. *Methods Mol Biol*. 2018; 1693:57–71. Epub 2017/11/10. [https://doi.org/10.1007/978-1-4939-7395-8\\_6](https://doi.org/10.1007/978-1-4939-7395-8_6) PMID: 29119432.
144. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res*. 2018; 46(D1):D708–D17. Epub 2017/10/19. <https://doi.org/10.1093/nar/gkx932> PMID: 29040670; PubMed Central PMCID: PMC5753373.
145. 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. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638.
146. Heinrichs DE, Yethon JA, Whitfield C. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol*. 1998; 30(2):221–32. Epub 1998/10/29. <https://doi.org/10.1046/j.1365-2958.1998.01063.x> PMID: 9791168.
147. Liu B, Furevi A, Perepelov AV, Guo X, Cao H, Wang Q, et al. Structure and genetics of *Escherichia coli* O antigens. *FEMS Microbiol Rev*. 2020; 44(6):655–83. Epub 2019/11/30. <https://doi.org/10.1093/femsre/fuz028> PMID: 31778182; PubMed Central PMCID: PMC7685785.
148. Micoli F, Ravenscroft N, Cescutti P, Stefanetti G, Londero S, Rondini S, et al. Structural analysis of O-polysaccharide chains extracted from different *Salmonella* Typhimurium strains. *Carbohydr Res*. 2014; 385:1–8. Epub 2014/01/05. <https://doi.org/10.1016/j.carres.2013.12.003> PMID: 24384528.
149. 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. [https://doi.org/10.1007/978-1-4939-7343-9\\_15\\_15](https://doi.org/10.1007/978-1-4939-7343-9_15_15). PMID: 29134597.
150. Storek KM, Chan J, Vij R, Chiang N, Lin Z, Bevers J 3rd, et al. Massive antibody discovery used to probe structure-function relationships of the essential outer membrane protein LptD. *eLife*. 2019; 8. Epub 2019/06/27. <https://doi.org/10.7554/eLife.46258> PMID: 31237236; PubMed Central PMCID: PMC6592684.
151. Eriksson JM, Haggard-Ljungquist E. The multifunctional bacteriophage P2 cox protein requires oligomerization for biological activity. *J Bacteriol*. 2000; 182(23):6714–23. Epub 2000/11/14. <https://doi.org/10.1128/JB.182.23.6714-6723.2000> PMID: 11073917; PubMed Central PMCID: PMC111415.
152. Silhavy TJ. Gene fusions. *J Bacteriol*. 2000; 182(21):5935–8. Epub 2000/10/13. <https://doi.org/10.1128/JB.182.21.5935-5938.2000> PMID: 11029410; PubMed Central PMCID: PMC94724.
153. Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. *Nature*. 1961; 192:1227–32. Epub 1961/12/30. <https://doi.org/10.1038/1921227a0> PMID: 13882203.
154. Benzer S, Champe SP. A change from nonsense to sense in the genetic code. *Proc Natl Acad Sci U S A*. 1962; 48:1114–21. Epub 1962/07/15. <https://doi.org/10.1073/pnas.48.7.1114> PMID: 13867417; PubMed Central PMCID: PMC220916.