

# Diversity and phage sensitivity to phages of porcine enterotoxigenic *Escherichia coli*

Michela Gambino,<sup>1,2</sup> Simran Krishnakant Kushwaha,<sup>3,4</sup> Yi Wu,<sup>4</sup> Pauline van Haastrecht,<sup>3</sup> Victor Klein-Sousa,<sup>5</sup> Veronika T. Lutz,<sup>1</sup> Semeh Bejaoui,<sup>1</sup> Christoffer Moeskjær C. Jensen,<sup>1</sup> Martin S. Bojer,<sup>1</sup> Wenchen Song,<sup>6</sup> Minfeng Xiao,<sup>6</sup> Nicholas M. I. Taylor,<sup>5</sup> Franklin L. Nobrega,<sup>3</sup> Lone Brøndsted<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 19.

**ABSTRACT** Enterotoxigenic *Escherichia coli* (EPEC) is a diverse and poorly characterized *E. coli* pathotype that causes diarrhea in humans and animals. Phages have been proposed for the veterinary biocontrol of EPEC, but effective solutions require understanding of porcine EPEC diversity that affects phage infection. Here, we sequenced and analyzed the genomes of the PHAGEBio EPEC collection, gathering 79 diverse EPEC strains isolated from European pigs with post-weaning diarrhea (PWD). We identified the virulence factors characterizing the pathotype and several antibiotic resistance genes on plasmids, while phage resistance genes and other virulence factors were mostly chromosome encoded. We experienced that EPEC strains were highly resistant to *Enterobacteriaceae* phage infection. It was only by enrichment of numerous diverse samples with different media and conditions, using the 41 EPEC strains of our collection as hosts, that we could isolate two lytic phages that could infect a large part of our diverse EPEC collection: vB\_EcoP\_ETEP21B and vB\_EcoS\_ETEP102. Based on genome and host range analyses, we discussed the infection strategies of the two phages and identified components of lipopolysaccharides (LPS) as receptors for the two phages. Our detailed computational structural analysis highlights several loops and pockets in the tail fibers that may allow recognition and binding of EPEC strains, also in the presence of O-antigens. Despite the importance of receptor recognition, the diversity of the EPEC strains remains a significant challenge for isolating EPEC phages and developing sustainable phage-based products to address EPEC-induced PWD.

**IMPORTANCE** Enterotoxigenic *Escherichia coli* (EPEC)-induced post-weaning diarrhea is a severe disease in piglets that leads to weight loss and potentially death, with high economic and animal welfare costs worldwide. Phage-based approaches have been proposed, but available data are insufficient to ensure efficacy. Genome analysis of an extensive collection of EPEC strains revealed that phage defense mechanisms were mostly chromosome encoded, suggesting a lower chance of spread and selection by phage exposure. The difficulty in isolating lytic phages and the molecular and structural analyses of two EPEC phages point toward a multifactorial resistance of EPEC to phage infection and the importance of extensive phage screenings specifically against clinically relevant strains. The PHAGEBio EPEC collection and these two phages are valuable tools for the scientific community to expand our knowledge on the most studied, but still enigmatic, bacterial species—*E. coli*.

**KEYWORDS** phages, enterotoxigenic *E. coli*, phage defense mechanisms, phage receptors

Enterotoxigenic *Escherichia coli* (EPEC) is an intestinal opportunistic pathogen causing diarrhea in humans, cattle, and pigs with a substantial impact on human and animal

**Editor** Charles M. Dozois, INRS Armand-Frappier Sante Biotechnologie Research Centre, Laval, Canada

Address correspondence to Michela Gambino, mgam@kglakademi.dk.

The authors declare no conflict of interest.

See the funding table on p. 19.

**Received** 15 May 2024

**Accepted** 5 June 2024

**Published** 28 June 2024

Copyright © 2024 Gambino et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

health, society, and the economy. In pigs, ETEC is the main pathogen responsible for post-weaning diarrhea (PWD) (1, 2), a disease occurring within 10 days post-weaning, leading to dehydration, severe weight loss, and potentially death. The disease affects 20%–50% of weaned piglets, representing a major productive and economic loss in pig production worldwide (3, 4). Despite the societal impact, ETEC is one of the least studied pathotypes of *E. coli*. ETEC is defined by two host-specific virulence factors, fimbriae and enterotoxins, and porcine ETEC are thus different from human ETEC strains (5–7). Notably for the ecology of ETEC, both fimbriae and enterotoxins are often encoded by plasmids and other mobile genetic elements. Because of the crucial role of mobile genetic elements in their pathogenicity, ETEC strains represent a highly diverse pathotype and even strains belonging to the same phylogroup express diverse virulence factors, including toxins and colonization factors. ETEC infecting humans, for example, exhibit a remarkable sequence type and serotype diversity, and 25 different colonization factors have been reported so far (8, 9).

In porcine ETEC, fimbriae (commonly F4 and F18, but also F5, F6, and F17) recognize receptors exposed by the enterocytes in the small intestine of pigs after weaning. This binding subsequently triggers the release of enterotoxins (heat-labile toxin LT and heat-stable toxins ST1 and ST2) in the enterocytes, leading to fluid secretion and decreased water absorption in the gut lumen (2, 10). Since fimbriae and enterotoxins are often encoded by plasmids, porcine ETEC strains belong to several phylogenetic groups, including phylogroups A, B1, C, D, and E (11). Still, an association between fimbriae, serotype, and plasmid type has been observed. For example, in a collection of Danish ETEC strains, F4 strains were mainly O149 and O6 serotypes and harbored IncFII, IncFIB, and IncFIC plasmids, whereas F18 strains were O8 and O147 serotypes and harbored IncI1 and IncX1 plasmids (12). Recent findings also showed that the acquisition of virulence factors like fimbriae F4 and F18 is associated with antimicrobial resistance genes (specifically neomycin), suggesting co-transfer or co-location on the same plasmid (13). This observation is particularly worrisome since ETEC-induced PWD is one of the main reasons for antimicrobial use in livestock (14).

While genotyping, feeding, vaccines, and management strategies can reduce the impact of ETEC-induced PWD on pig farms (15–19), options for treatment are limited, especially after the restrictions on colistin and zinc oxide use for veterinary applications (14, 20). Alternatives are urgently needed, but efficacy in the long term and on a broad scale requires a deep understanding of the target organism. For phage therapy targeting porcine ETEC, this can be translated into characterizing a broad collection of porcine ETEC strains and defining the molecular mechanisms used by phages (short for bacteriophages, viruses that infect bacteria) to recognize and infect their host. Phages have been proposed and, in some cases, already used as alternatives to antimicrobials to treat or prevent several human and animal infections (21, 22).

Phages infecting *E. coli* are among the most isolated and best-characterized phages, including some still used as models for molecular biology studies (23). Nevertheless, most coliphages have been isolated on laboratory strains and, being phages extremely specific for their hosts, do not necessarily infect wild-type or pathogenic strains (24). To the best of our knowledge, only 30 phages have been reported to infect porcine ETEC strains (25–35). Unfortunately, the data and information regarding these phages and their hosts are fragmentary. For example, phage genome sequences are publicly available only for 10 lytic phages, and only two studies report the virulence factors that classify the infected strains as ETEC (27, 31). This incomplete data set fails to link the molecular aspects of phage infection to the ETEC strains, which is urgently needed to propose an efficient phage product on an international scale.

Here, we present a well-characterized collection of porcine ETEC strains intended as a reference for designing novel alternative antibacterial solutions and studying *E. coli* diversity, phages, and mobile genetic elements. In addition, we isolated two novel virulent phages infecting strains of the ETEC collection and characterized the receptor

binding apparatus of the phages, allowing us to explain their ability to infect porcine ETEC.

## RESULTS

### The PHAGEBio collection of porcine ETEC strains

To propose a relevant collection of porcine ETEC strains, we gathered 79 hemolytic strains, isolated from diverse pig farms in Europe (Denmark, Italy, Greece, Belgium, and Netherlands) between 2014 and 2020, from fecal or gut samples of pigs clinically affected by PWD. Using multiplex PCR (7), the strains were confirmed to encode for fimbriae but not shigatoxins or verotoxins that would assign the strains to other *E. coli* pathotypes. F4 fimbriae were identified in the majority of the ETEC strains ( $n = 47$ ), followed by F18 ( $n = 28$ ), whereas only four strains encoded other fimbria types (i.e., F5, F5/F41, or F6). Except for seven strains for which no enterotoxins have been identified, enterotoxins are present alone ( $n = 12$ ) or in different combinations ( $n = 60$ ), with the most common combination being ST2, LT ( $n = 29$ ) (Fig. 1; supplemental file S1A).

All ETEC genomes are available on NCBI (BioProject ID: PRJNA770188) as the PHAGEBio ETEC collection. The considerable variation in genome sizes (from 4.9 to 6.2 Mb) and the small size of the core genome (13% of the total genes) confirmed the polyphyletic origin of the ETEC strains. We determined the phylogroups of the ETEC strains (Fig. 1) and compared with the *E. coli* strains in the ECOR collection (supplemental file S1B) that represents the diversity of *E. coli*, and it is still used nowadays as a reference (36). The ETEC strains fell into phylogroups A, B1, C, D, and E, thus confirming the diverse genetic background of ETEC strains, as previously observed (9). The majority of the ETEC strains were assigned to phylogroup A (56 strains, 70.8%), while 13 belonged to phylogroup D (16.4%), 4 to phylogroup B1 (5.1%), the other 4 to C (5.1%), and 2 to E (2.5%).

The genetic diversity was also reflected in surface variation of these strains, as a total of 29 different serotypes were identified, resulting from a combination of 19 O-antigens and 17 H-antigens (Fig. 1; supplemental file S1A and B), with O149:H10 being by far the most common serotype (32 strains, 40.5%), followed by serotypes O23:H37 (6 strains) and O138:H14 (5 strains). All the strains with serotype O149:H10 fell into phylogroup A, where we also found the most common virotypes F4, ST2, and LT (25 strains, 31.6%). This collection well represents the extreme diversity of porcine ETEC strains in terms of genome, fimbriae, toxins, and O- and H- antigens (supplemental file S2).

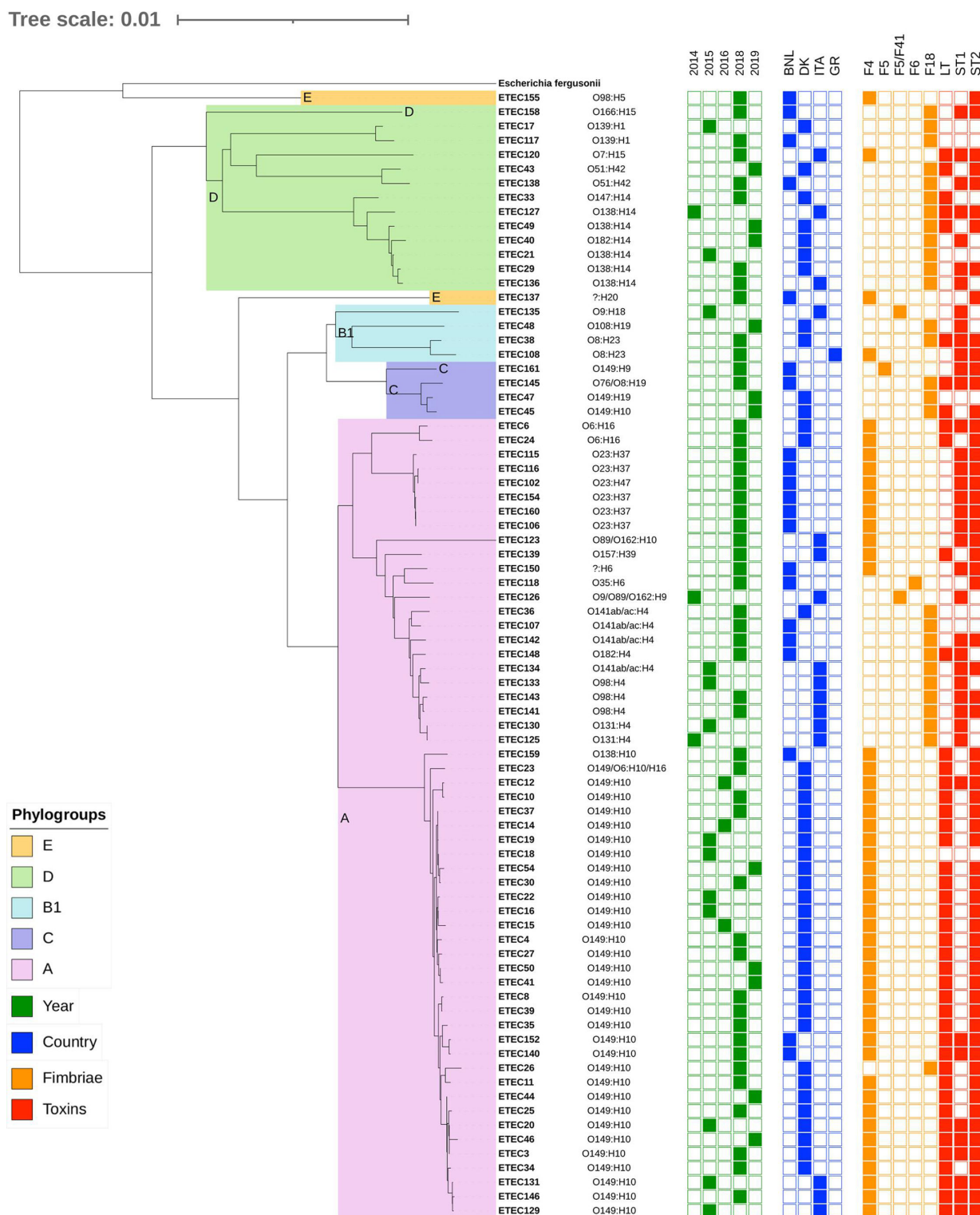
Further bioinformatic analysis revealed that the contribution of MGEs to the diversity of the ETEC collection was considerable (supplemental files S1 and S2). For each ETEC strain, 15 to 82 contigs were predicted to belong to plasmids and 2 to 16 prophages. Virulence, metal, and antibiotic resistance genes were widespread, with virulence factors characterizing the pathotype and several antibiotic resistance genes on plasmids.

### ETECs are equipped with a large variety of phage defense mechanisms

Phage defense systems influence phage susceptibility, and PADLOC and DefenseFinder were used to predict their presence in the PHAGEBio ETEC collection. Most of the defense systems were on the chromosomes, and each strain encoded its own combination and types, thus reflecting ETEC genomic diversity (Fig. 2; supplemental file S2). The number of defense systems per genome varied widely from a minimum of four in several ETEC genomes to a maximum of 10 in ET134. There was a slight overrepresentation of defense systems in phylogroups D and E, compared with the other phylogroups, and the difference was maintained even when considering only the chromosome-encoded ones (supplemental file S1C through G).

The most widespread systems were known to defend the host by degradation of nucleic acids, such as CRISPR-Cas (with CRISPR-Cas I-E in 68 strains) and restriction modification systems (RM type I in 45 strains, RM type II in 41 strains, RM type IIG in 18 strains, RM type III in 3 strains, and RM type IV in 45 strains), or by abortive infection,

Downloaded from https://journals.asm.org/journal/aem on 27 May 2026 by 152.78.0.24.

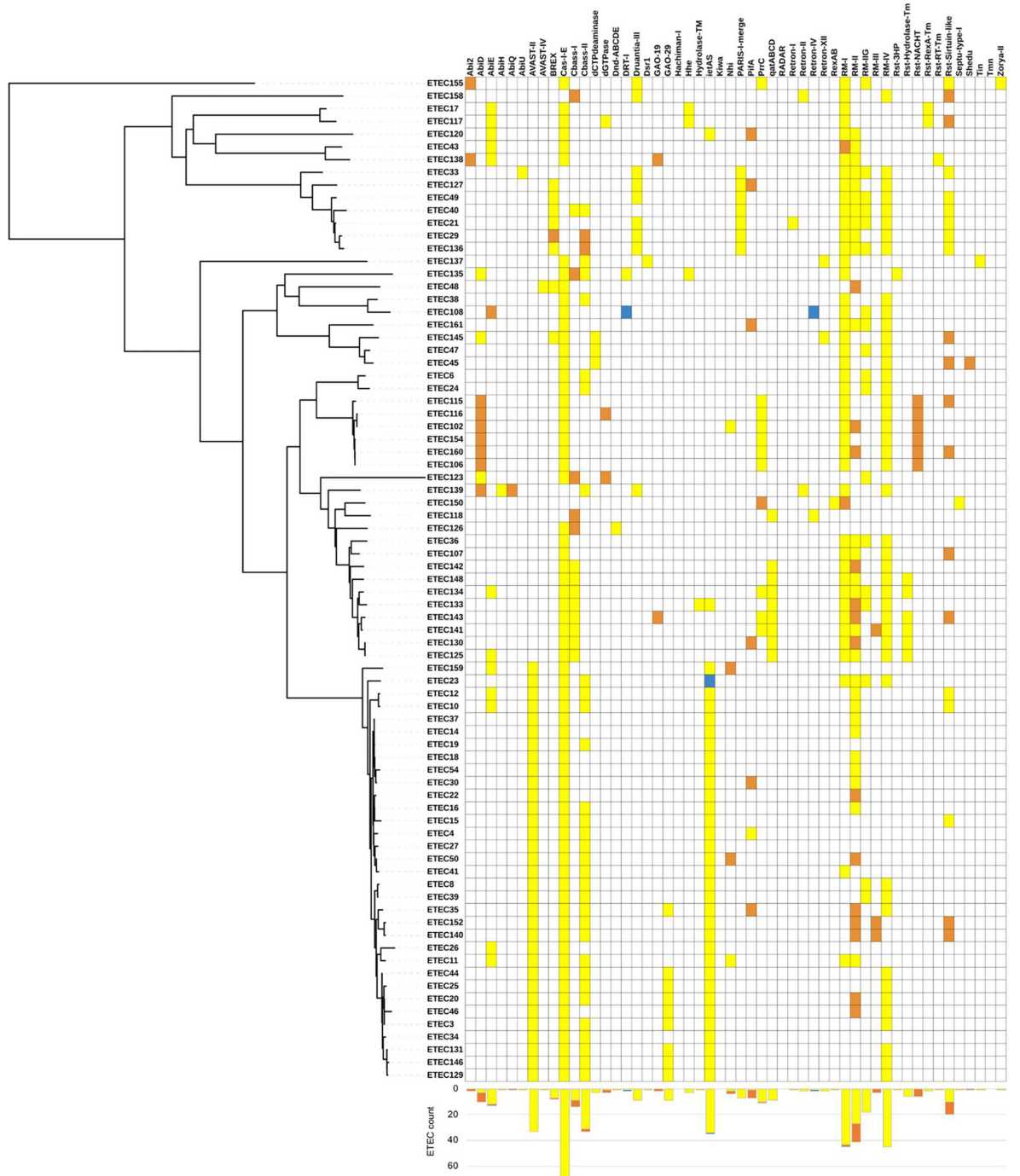


**FIG 1** Phylogenetic tree built with the core genomes of the ETEC collection. Colors on the clades indicate the phylogroups, determined by submitting the ECOR genome sequences to EzClermont (<https://ezclermont.hutton.ac.uk/>). From left to right, we report the serotype (O- and H-type), year (green), country (blue), fimbriae (orange), and toxins (red).

such as CBASS (CBASS type I in 14 strains, CBASS type II in 33 strains) and probably AVAST (AVAST type II in 33 strains and AVAST type I in 1 strain) (Fig. 2). The mechanism of action of other similarly widespread systems is still unknown, such as in the case of *ietAS*

Tree scale: 0.01

■ chromosome ■ plasmid ■ prophage



**FIG 2** Phage defense mechanisms predicted in the ETEC collection. Colors indicate the predicted position of the gene: yellow on chromosomes, orange on plasmids, blue on prophages. The histograms at the bottom represent the count of ETEC strains where each defense system was detected.

(35 strains). Interestingly, while the CRISPR-Cas I-E system alone was detected in most strains, the RM types I and II presence complemented each other, occasionally with types IIG and III, and they overlapped in most strains with type IV. Moreover, other systems were mainly present in specific ETEC clades. For example, the three systems AVAST type

Downloaded from https://journals.asm.org/journal/aem on 27 May 2026 by 152.78.0.24.

II, *ietAS*, and CBASS type II seemed specific for a closely related clade in phylogroup A (supplemental file S2).

The only few defense systems predicted to be also on plasmids were many of the abortive infection systems (*Abi2*, *AbiD*, *AbiE*, *AbiQ*, CBASS-I, CBASS-II, *PifA*, and *PrrC*) and some of the ones degrading nucleic acids (*Nhi*, RM-I, RM-II, RM-III, and *Shedu*, 40), dGTPase and BREX preventing phage DNA replication (37), and the uncharacterized GAO-19, *Rst NACHT*, and *Rst Sirtuin-like*. In only three cases, the defense systems were predicted to be prophage encoded, i.e., *ietAS* in ET23 and DRT I and Retron IV in ET108 (Fig. 2). We can conclude that the diversity of phage resistance mechanisms and distribution in ETEC were linked more to chromosomal diversity of the strains than to MGEs.

### Two lytic phages infected 44% of the ETEC collection

To identify phages infecting ETEC, phage sensitivity of the ETEC strains has been tested by spotting previously isolated phages, such as 42 *Salmonella* phages (38, 39) and up to 200 *E. coli* phages, including the well-studied coliphages T3, T7, and CBA120, and others isolated on *E. coli* MG1655 and ECOR04. While spotting undiluted phage stocks, lysis spots were observed for many phage-host combinations. Still, none of the phages produced single plaques on any ETEC strains, except for phage T7 that infected ET12 (Fig. 3; supplemental file S3).

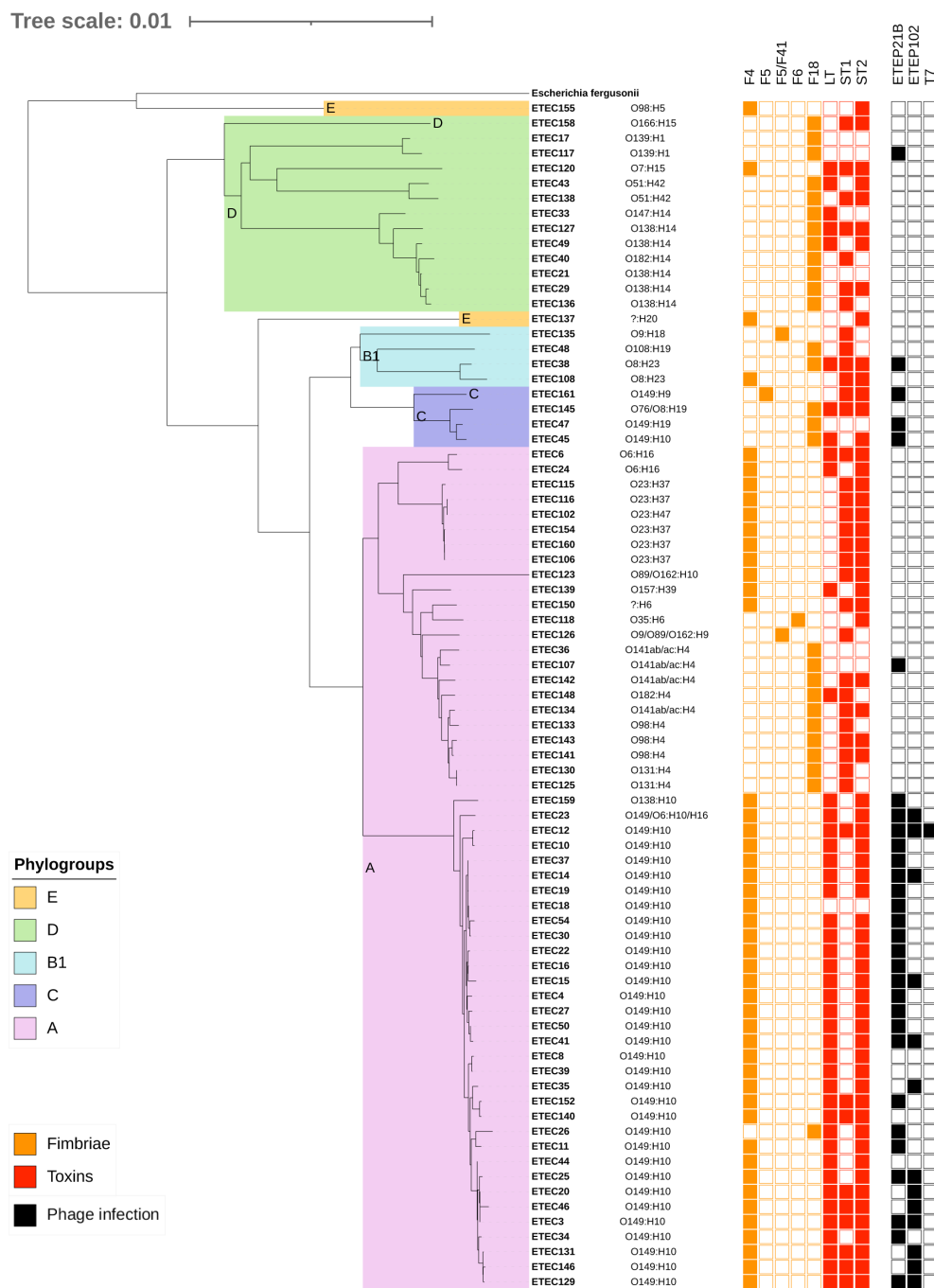
To isolate phages specifically infecting ETEC, 78 different samples (wastewater, pig manure, gut samples from pigs dying from ETEC diarrhea, and feces from healthy and ETEC-affected pigs) were screened for plaque formation by spotting directly and after enrichment using various media (LB, BHI, with or without mucin) on 41 Danish ETEC strains. More than 100 phage stocks were prepared, but most of the isolated phages most likely were temperate, as hypothesized by the turbidity of plaques. Indeed, sequencing a subset of them confirmed their lysogenic potential, since a simple RAST annotation revealed several genes predicted to encode for integrases, transposases, and repressor and derepressor proteins (an example in supplemental file S4). In addition, they resulted highly similar to a prophage in the ETEC strain UMNK88 [CP002729.1; 99% identity and query coverage (11)]. Given the envisioned future veterinary application targeting porcine PWD, we concentrated on the only two lytic phages isolated: ETEP21B and ETEP102. ETEP21B was isolated from a gut sample that also was the source of an ETEC strain (ET23) using ET14 as the isolation host and LB with mucin as media. ETEP102 was isolated from a gut sample that was the source of ETEC strain ET25 using ET03 as the isolation host and LB as media.

To unravel the mechanism used by phages ETEP21B and ETEP102 to infect ETEC, a host range analysis was performed, determining the ability to form plaques on lawns of the established ETEC collection, and found that 35 strains (44%) were sensitive to at least one of the phages (Fig. 3). Phage ETEP21B infected 30 strains (24 from Denmark, 6 from other countries), mostly of virotypes F4, ST2, LT, and O-type O149, but also seven F18 strains with other virotypes and O-types. Phage ETEP102 infected 13 ETEC strains (10 from Denmark, 3 from other countries), all F4 strains, including 5 additional strains that are resistant to ETEP21B infection (i.e., ET20, ET35, ET46, ET131, and ET146) (supplemental file S3). By comparing the host range of the two ETEC phages with the data collected on the ETEC strains (country, O-type, fimbriae, toxins, virulence factors, antibiotic and metal resistance genes, and phage defense mechanisms), no single variable responsible for phage infectivity or resistance could be identified. Thus, a combination of several factors most likely determines the host range of the phages.

### The two virulent ETEC phages belonged to different families

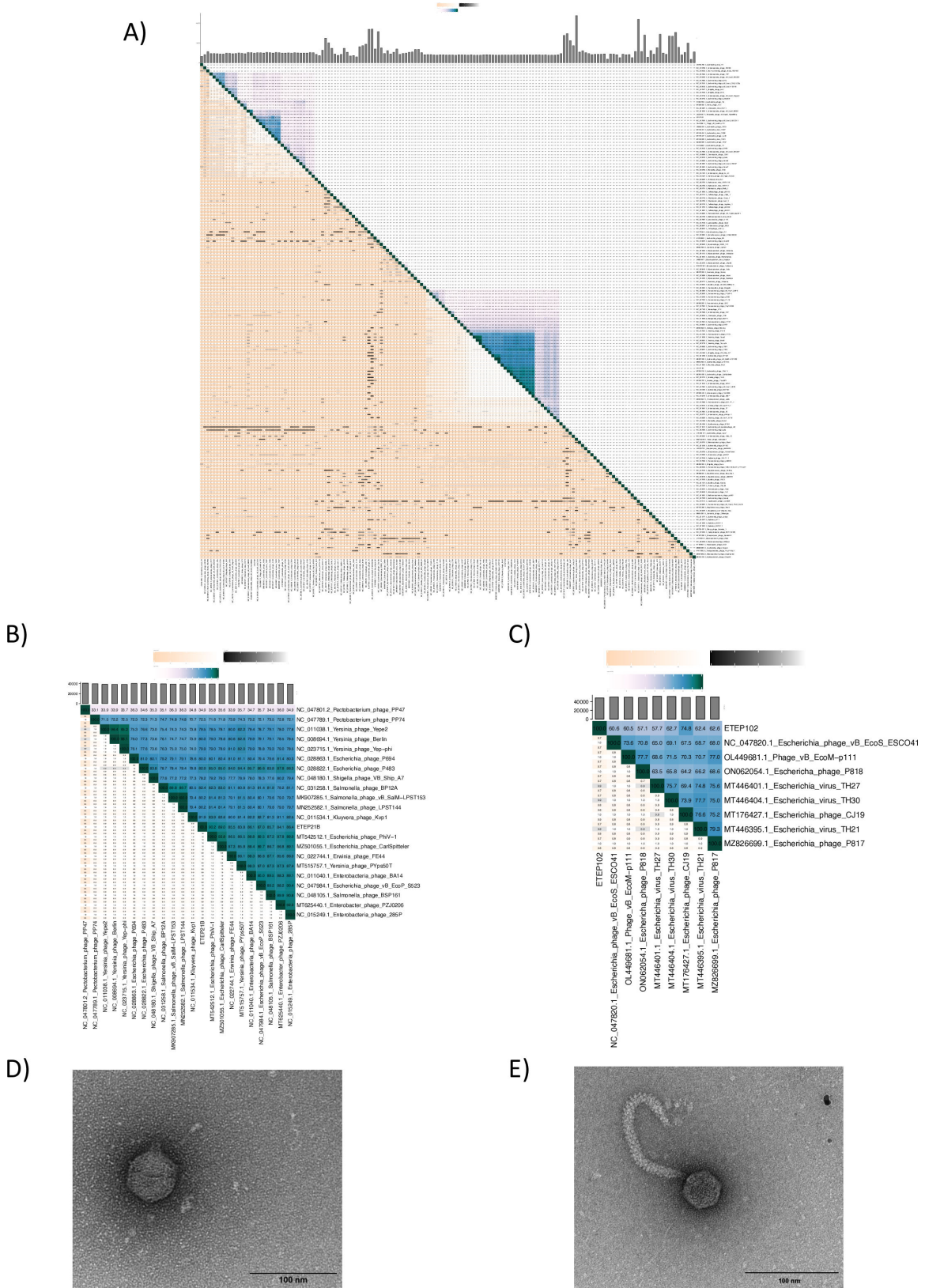
Sequencing the genomes of ETEP21B and ETEP102 showed that both were dsDNA phages. While ETEP21B carried a genome of 39,424 bp with 49.1% GC content and encoded no tRNA, the genome of ETEP102 was 48,422 bp with 46.1% GC content and 1 tRNA binding L-Arginine. Comparison to the closest relatives to the two phage genomes

Downloaded from https://journals.asm.org/journal/aem on 27 May 2026 by 152.78.0.24.



**FIG 3** Host range analysis of the ETEC strains. For each strain, the phylogroup, the O-type, the virulence profile, and the infection by phages ETEP21B, ETEP102, and T7 are indicated.

using vCONTACT (supplemental file S5C) confirmed that the two phages were diverse and belonged to two different families (Fig. 4A). Phage ETEP21B was taxonomically assigned to the *Autographiviridae* family, *Studiervirinae* subfamily, *Berlinvirus* genus (Fig. 4B). ETEP21B closest relatives were other phages infecting *Enterobacteriaceae* (*Escherichia*, *Yersinia*, *Salmonella*, *Enterobacter*, and *Kluyvera* spp.) assigned to the *Berlinvirus* genus (supplemental file S5A), confirmed by ETEP21B sharing 87.36% average nucleotide identity (ANI) with *Yersinia* phage Berlin, reference for the *Berlinvirus* genus. In contrast, phage ETEP102 was assigned to the *Drexelviriidae* family (Fig. 4C), *Nouzillyvirus* genus, as it fell within the same vCONTACT subcluster as coliphage ESCO41. ETEP102 closest



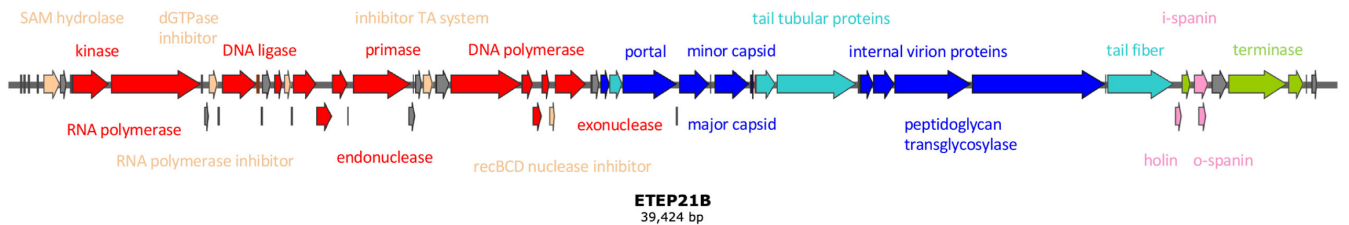
**FIG 4** Two diverse phages infect the ETEC strains. (A) Intergenomic similarity between ETEP21B, ETEP102, and other 158 genomes (their closest relatives, one genome for each of the genera within the same family, and one genome for each of the other families as listed in ICTV). Close up of (B) ETEP21B and (C) ETEP102 clusters. For each graph: on the right side, darker colors indicate higher intergenomic similarity and the numbers represent the similarity values for each genome pair; on the left side, for each genome pair are the aligned fraction genome versus other genomes in the same row, in the same column and the genome length ratio. Transmission electron micrographs of (D) phage ETEP21B and (E) ETEP102. The bar represents 100 nm, as indicated.

relatives were other phages infecting *E. coli*, likewise assigned to the *Nouzillyvirus* genus (supplemental file S5B). Despite the separate genus assigned by VIRIDIC clustering, ETEP102 shared 82.41% ANI with ESCO41 and was thus classified as a *Nouzillyvirus* phage. Transmission electron micrographs confirmed that the phages were diverse, showing two different morphotypes, with ETEP21B being a podovirus with an icosahedral head ( $59.7 \pm 1.5$  nm) and very short tail ( $12.0 \pm 2.6$  nm; Fig. 4D), while ETEP102 was a siphovirus with an icosahedral head ( $55.8 \pm 9.0$  nm and  $63.4 \pm 1.9$  nm) and a long, flexible tail ( $158.5 \pm 8.4$  nm; Fig. 4E). Thus, we isolated two diverse phages, belonging to different families, yet both infecting porcine ETEC.

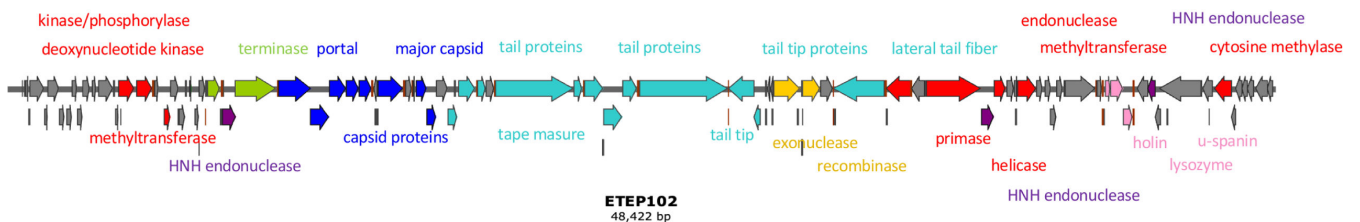
### The identification of the primary receptor of phage ETEP21B partially explained its host range

The ETEP21B genome was predicted to encode 46 open-reading frames (ORFs), all but 3 (initiating with GUG) starting with the AUG codon. Few ORFs (10 ORFs, corresponding to 22%) were annotated as hypothetical proteins, with no similarity to characterized proteins (Fig. 5A; supplemental file S6A). ETEP21B had the classic genome structure of the well-described *E. coli* phages T7 and T3 (40), with early (ORFs 01 to 07) and middle genes (ORFs 08 to 24) for DNA metabolism and host takeover and late genes for phage morphogenesis (ORFs 25 to 38) and lysis and packaging (ORFs 39 to 46). Based on the similarities with T7, ETEP21B was expected to initiate transcription using the *E. coli* RNA polymerase via the three strong promoters predicted near the left genome end. Transcription was expected to continue using the ETEP21B-encoded RNA polymerase initiated at phage-specific promoters, predicted at positions homologous to T7 promoters. In addition to two terminators downstream of the DNA ligase (ORF07) and the capsid genes (ORF31) also present in T7, an additional terminator was predicted downstream of the small terminase subunit gene (ORF40), suggesting a more complex regulation of the packaging process (supplemental file S6A). As in T7 and T3, the ETEP21B genome encoded most of the proteins needed for transcription and replication

A)



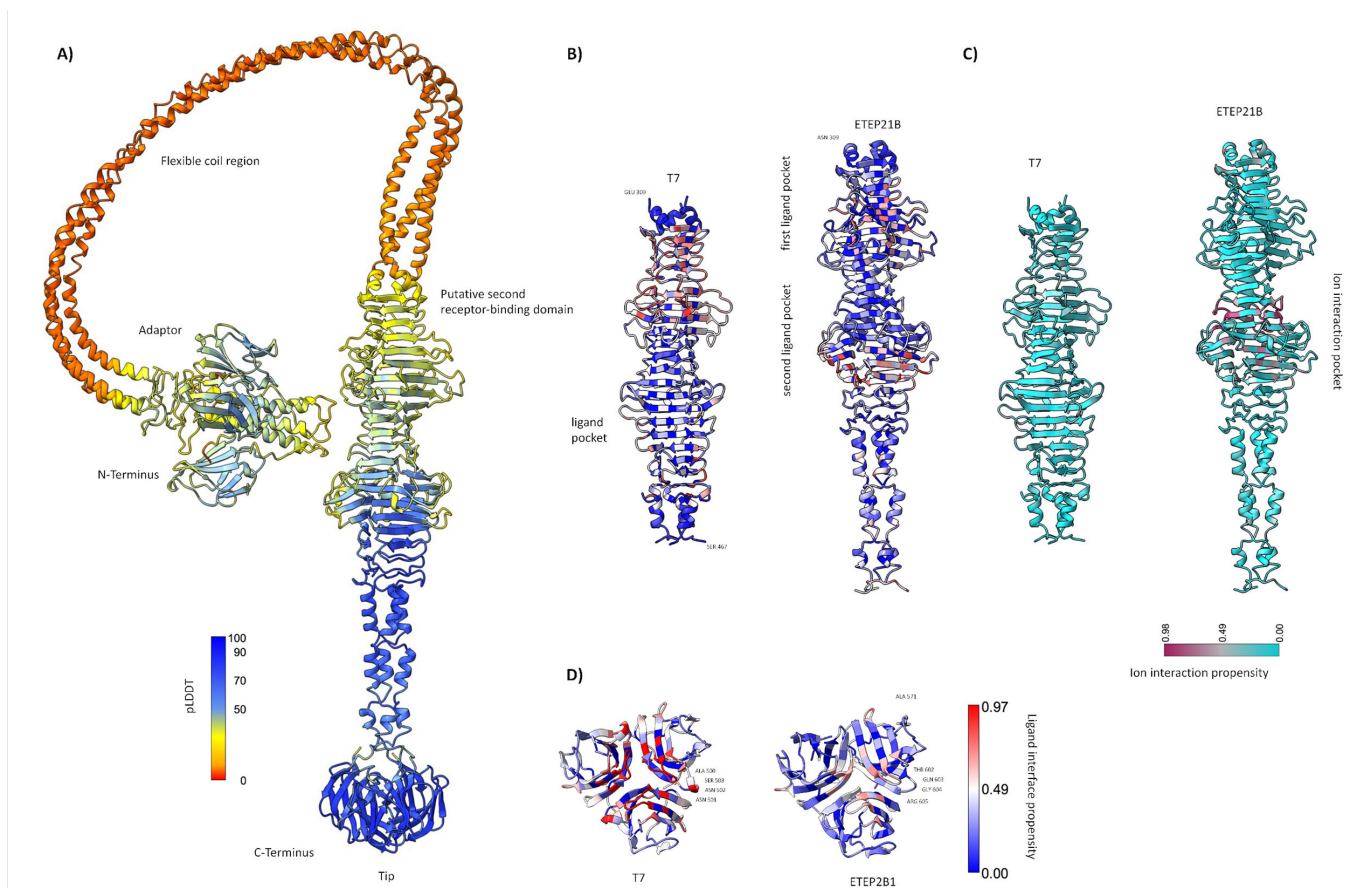
B)



**FIG 5** Genomic organization of phages ETEP21B (A) and ETEP102 (B). Genes are indicated as arrows: capsid morphogenesis in blue, tail morphogenesis in light-blue, DNA packaging in green, lysis-associated genes in pink, genes involved in DNA manipulation in red, morons and homing endonucleases in purple, recombination in yellow, and hypothetical in gray. Putative promoters are indicated as orange lines and terminators as brown lines.

during phage propagation, with the early transcribed genes involved in evading the host defense systems, shutting off the host transcription, and switching to the viral transcription. In addition, sites used for methyltransferase modification were underrepresented in ETEP21B (two Dam sites and no Dcm sites), as well as in its closest similar phages (supplemental file S6B) and other *Autographiviridae* phages (40).

Considering that the O-antigen is known to inhibit the related phage T7 binding to the receptor (40), it was striking that ETEP21B infects many ETEC strains carrying an O-antigen (Fig. 3). Since receptor recognition in T7 relies on the tail fiber, we further compared the tail fiber structures of the two phages. In ETEP21B, ORF38 was predicted to encode for the tail fiber due to its similarity to the T7 tail fiber. We predicted the structure of the ETEP21B tail fiber (ETEP21B-TF) and compared it with an AlphaFold2 prediction of T7 gp17 (T7-TF; supplemental file S7A). Importantly, the T7-TF AlphaFold2 prediction was highly similar to the experimentally determined structures of the C-terminal region (41) and electron micrographs of the complete fiber (42). ETEP21B-TF contains four major structural domains (Fig. 6A): an adaptor region at the N-terminal, a flexible coiled-coil region, a region with a pyramid domain, and a tail tip containing a receptor-binding domain at the C-terminal. Full alignments of T7-TF and ETEP21B-TF resulted in 42.3% overall identity but showed a lower identity in the pyramid domain (only 26.6%) and higher in the tip domain (52%). Compared with T7-TF, the ETEP21B-TF pyramid region exhibited additional interface regions (a complete beta-strand with some affinity for ligand binding and a pocket) with ligands and ions, suggesting it may function as a putative second receptor-binding domain (Fig. 6B and C; supplemental file S7B) and that distinct ligands could be accommodated. Moreover, two additional short alpha-helices



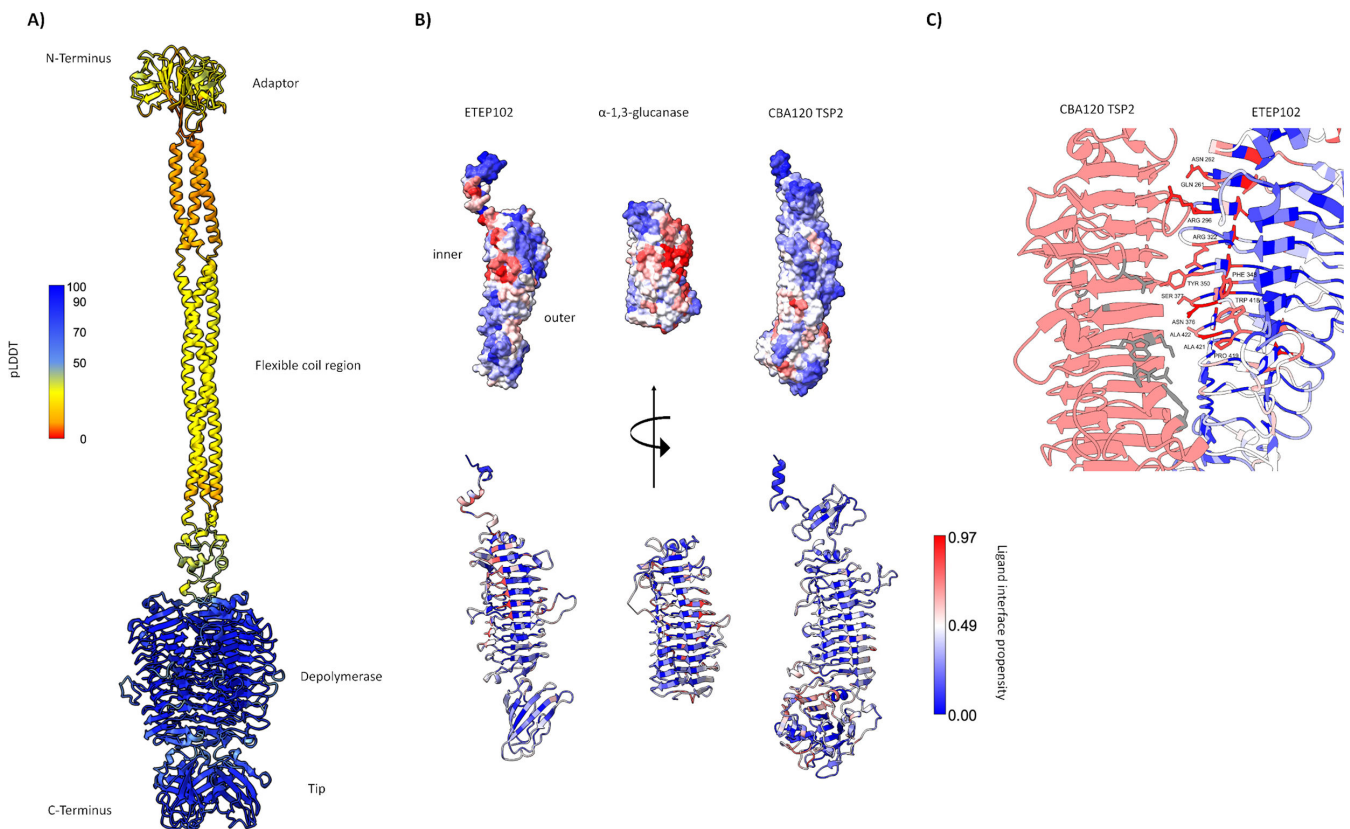
**FIG 6** (A) AlphaFold2-multimer model of ETEP21B-TF, colored by pLDDT confidence scores. (B) Putative second receptor-binding domain in T7-TF and ETEP21B-TF colored by ligand interface propensity and (C) ion interaction propensity. (D) Tips of T7-TF and ETEP21B-TF colored by ligand interface propensity.

connected the pyramid domain to the tip in the C-terminal. The four external loops in the T7-TF tip were known to be responsible for receptor recognition in phage T7 (41). The ETEP21B-TF tip had a similar architecture to the T7-TF tip, except for loops 2, 3, and 4 (Fig. 6D). The reduced ligand interface propensity score of ETEP21B-TF loop 2 (THR602 and ARG605) suggested a reduced interaction with the receptor compared with T7-TF. Instead, loop 3 exhibited a propensity to bind ligands such as sugars, and the residue ARG643 (one residue before loop 4) was predicted to have a high propensity for protein binding. Overall, the ability of ETEP21B to infect strains carrying O-antigens could originate from the differences highlighted in the loops, especially loop 3, and/or in the putative second receptor-binding domain that potentially could host two different receptors.

We also noticed that ETEP21B could infect rough strains (ECOR04, MG1655; supplemental S3), showing that the O-antigen was not essential for phage infection. Thus, ETEP21B infection was tested on a collection of ET54 strains deleted for common coliphage receptors, including some lacking the last sugars of the outer core and the O-antigen ( $\Delta waaC$ ) or only the O-antigen ( $\Delta waaR$ ). Phage ETEP21B infected all mutants with the same efficiency as the wild type, except for  $\Delta waaC$ , where a 3-log reduction indicated that sugars in the inner core or the first sugars of the outer core were needed for ETEP21B infection but not the O-antigen (supplemental file S6C). Double receptor mutants prepared in the  $\Delta waaC$  background did not reduce infection further, leaving a potential secondary receptor for ETEP21B unknown. In addition, ETEP21B infected a few strains with different serotypes, namely, ET107 (O141), ET117 (O139), ET119 (O139), and ET159 (O138). The different structures of these four O-antigens [O149 has a linear three-residue backbone, O138 a linear four-residue backbone, O141 a branched three-residue backbone, and O139 a branched six-residue backbone; reviewed in reference (43)] and the ability of ETEP21B to still infect *waaC* mutants suggested that ETEP21B recognized multiple sugars in the core. Alternatively, an unidentified second receptor interacting with the putative second binding pocket highlighted by the ETEP21B-TF structural analysis may play a role in ETEP21B infection.

### The phage ETEP102 tail apparatus suggested a two-step infection process, starting with a depolymerase activity

The ETEP102 genome was predicted to encode 81 ORFs, all but five (initiating with GCG and four with GUG) starting with the AUG codon. Most ORFs (44, corresponding to 55%) were predicted as hypothetical proteins, with no similarity to characterized proteins (Fig. 5B; supplemental file S8), despite the genome synteny with the type coliphage T1 (44) and other well-characterized phages such as coliphage TLS (45) and Rtp (46). Despite the similarity with phages T1 and TLS indicating a similar infection process, more than half of the predicted ORFs had no suggested function and these might largely affect the ETEP102 host range. Our host range analysis (Fig. 3) showed that ETEP102 infects strains with diverse O-antigens, and we thus studied the ETEP102 tail apparatus to understand host recognition. The tail apparatus of ETEP102 was very similar to the *E. coli* phage Rtp, which has leaf-like lateral tail fibers (46), as suggested by the ETEP102 TEM picture (Fig. 4E). In *Drexlerviridae* phages, the lateral tail fibers are thought to reversibly bind O-antigen glycans as primary receptors. The AlphaFold2 model of the ETEP102 tail fiber (ETEP102-TF) exhibits a four-domain structure that includes an adaptor, a flexible coil region, a depolymerase domain, and a tip domain (Fig. 7A). The depolymerase domain was predicted with high confidence, and it shared high structural similarity with glycoside hydrolases produced by *Paenibacillus* sp. [13.9% sequence identity and 164 FoldSeek score with PDB-ID 6KFN (47) and 10.9% sequence identity and 126 FoldSeek score with PDB-ID 6k0n (48)], as well as with the bacteriophage CBA120 tail spike [TSP2 (49); 11.4% sequence identity and 120 FoldSeek score]. Despite their similar architecture, these proteins exhibited distinct ligand propensity regions (Fig. 7B), which suggests a different hydrolase activity of phages ETEP102 and CBA120 and might explain why phage CBA120 did not infect any of the tested ETEC strains (supplemental file S3).



**FIG 7** (A) AlphaFold2 model for the ETEP102 tail fiber colored by pLDDT confidence scores. (B) Comparison of ETEP102 fiber depolymerase with the  $\alpha$ -1,3-glucanase of *Paenibacillus* sp. (PDB-ID 6k0n) and CBA120 TSP2 (PDB-ID 5W6P), colored by ligand interface propensity. (C) Alignment of ETEP102-TF (right) with CBA120 TSP2 (left), showing that residues with high ligand interface propensity in ETEP102 are located closer to the N-terminus compared with the residues known to bind O-157 in TSP2 (in gray). The root mean square deviation between the two structures is 1.12 angstroms, considering 90 pruned atom pairs.

After binding to O-antigen glycans, a second protein receptor was recognized by *Drexlerviridae* for irreversible binding and DNA injection by the tail tip. The ETEP102 tail tip genes (gpJ-like corresponding to ORF45 and following genes) were predicted to be structurally similar to the ones of *E. coli* phage AugustePiccard shown to bind the essential lipopolysaccharide export channel LptD (23). Despite the synteny of the tail tip genes, the relatively low sequence identity between AugustePiccard and ETEP102 genes (55% amino acid identity) left space for uncertainty. In the attempt to experimentally verify the predicted receptor, ET03 spontaneous mutants resistant to ETEP102 were isolated and sequenced. A few mutations with high frequency were identified (equal to 65% or higher; supplemental file S9), but none of them was in *lptD*: a silent mutation in the sugar ABC transporter ATP-binding protein YphE, a missense mutation in a gene expected to encode a D-alanyl-D-alanine carboxypeptidase, involved in the peptidoglycan biosynthesis pathway, and other missense mutations in a gene with high similarity to *tral*, a secreted DNA helicase and relaxase, essential for conjugation (50–52). Deletion mutants of each of these three targets were constructed, but all were infected by ETEP102 at a similar level as the wild type (supplemental file S9), thus leaving LptD as the only predicted ETEP102 secondary receptor.

## DISCUSSION

Porcine ETEC-induced PWD is a severe threat with substantial economic costs and impact on animal welfare. The treatment used so far has relied on the prophylactic use of antibiotics and zinc oxide shown to contribute to the spread of antibiotic resistance (53–55), thus becoming ineffective and unsustainable and, for this reason, banned in the

European Union (20). As for many other bacterial infections in humans and animals, no efficient and sustainable solutions are available on the market, and phages have been proposed as a valuable resource. Conscious of the careless use of antibiotics in the past, the European Union expects new interventions to comply with the One Health policy and thus to demonstrate efficacy and safety for humans, animals, and the environment. This requires a detailed knowledge of the pathogen to target and its interaction and response to the intervention.

To obtain detailed knowledge on ETEC causing PWD in piglets, we established and characterized the PHAGEBio ETEC collection. ETEC strains in the collection are available upon request, and the genomes can be found on NCBI for researchers to work on clinically relevant ETEC strains, expanding our understanding of the diversity and evolution of *E. coli*, the most classic yet still poorly understood bacterial species. It has previously been observed that ETEC strains are genetically highly diverse (9, 11). Indeed, the strains in the PHAGEBio ETEC collection fall into the most common phylogroups (A, B1, and D), with mostly generalist and commensal strains. Still, some also belong to the rarest phylogroups (C and E), usually consisting of intestinal and extraintestinal pathogenic strains (56). This genetic diversity mirrors the need and ability for *E. coli* to adapt to multiple and very different niches—from water to mammal gut. As in other *E. coli* strains, this flexibility is supported by high genetic plasticity: niche-specific genes, such as virulence genes that allow animal and human infection that are easily transferred by MGEs, making the *E. coli* species a perfect example of an open pangenome with a high rate of horizontal gene transfer (6, 56, 57). ETEC is no exception, especially since the virulence genes that define the pathotype are often MGE encoded (2), as observed in our collection, particularly for the few strains in phylogroups C and E (supplemental file S2). The genetic diversity is also revealed by the diverse surface, with the 19 different O-antigens identified in our collection that were commonly reported for porcine ETEC strains before (1, 12, 58, 59) and the predominance of the O149 antigen for the F4 strains. At the same time, the O-antigen of F18 varied between O138, O139, O141, O149, and a few others. The extreme diversity of ETEC genomes and O-antigens previously suggested that any *E. coli* may acquire virulence gene-encoding plasmids and become ETEC. Nevertheless, a strong association was found in human ETEC strains between phylogenetic lineages, virulence genes, plasmids, and O-antigens (60), suggesting that specific chromosome traits or lineages might be needed for maintaining plasmids encoding ETEC virulence genes (61, 62). A similar scenario can be imagined for porcine ETEC strains, as we and others [i.e., references (11, 12)] observed a co-occurrence of serotypes, virulence factors, and lineages (e.g., O149:H10, with F4, ST2, and LT in phylogroup A). The *in silico* analysis of our ETEC collection demonstrated that plasmid maintenance is driven by PWD-specific virulence factors (fimbriae and toxins) and resistance to antibiotics and metals. Metals have been reported to co-select for antibiotic resistance genes in several environments and concentrations (63), and, specifically for porcine ETEC (12), antibiotic resistance genes were predicted to co-locate on plasmids with metal resistance genes or virulence factors, thus implying that the extended and indiscriminate use of metals and antibiotics will be detrimental in the long term.

In contrast, the internal antiphage resistance mechanisms were predicted to be chromosome encoded. From the point of view of phage therapy and biocontrol, the observation suggests a reduced risk of co-selection of phage resistance genes with virulence genes and antibiotic or metal resistance genes. From an evolutionary point of view, phage resistance genes in our ETEC collection were not horizontally transferred, as elsewhere reported (64). Still, they coevolved with the bacterial chromosome, with a pace and contribution from gene loss and gain similar to any other chromosomal genes (65). In each ETEC genome, we detected several defense systems. As previously reported for *E. coli* and most bacterial species (66), the RM systems are the most prevalent phage defense mechanisms, followed by the CRISPR-Cas and several abortive infection systems. In our collection, while the CRISPR-Cas type I-E system alone is present in most strains, several RM types complement each other and overlap, especially with the RM type IV,

which targets modified DNA contrary to the other RM systems (67). In addition, the presence of AVAST type II, CBASS type II, and *ietAS* in a clade of closely related strains of the phylogroup A is a peculiar observation. The different defense systems might act synergistically to defend the cell from the infection of diverse phages (68), to increase the defense efficiency against one phage (64), or to build several consecutive lines of defense to avoid phage immune evasion (69).

Literature data on phage-based products against ETEC-induced PWD are very fragmented, and only two phages (GJ1 from 24 and 25 and CJ19 from 29) have been demonstrated to be lytic and infecting porcine *E. coli* strains with fimbriae and toxins that define ETEC. Here, we isolated several phages and characterized in detail the two lytic phages, ETEP21B and ETEP102, isolated from pig feces samples. These samples were also used to isolate ETEC strains, supporting the notion that phages infect hosts isolated in the same space and time more efficiently (e.g., 38, 70, 71). These two phages infected 44% of the PHAGEBio ETEC collection, mainly O149: F4, ST2, and LT strains, but not exclusively. Despite the overlap of host range, the two phages are diverse. ETEP21B was classified as a *Berlinvirus*, a podovirus belonging to the same *Autographiviridae* family as the well-studied coliphages T3 and T7 (40). Thanks to this similarity, a function for most genes could be proposed, including the early genes involved in evading the host defense systems, such as the S-adenosylmethionine hydrolase that overcomes the action of RM type I (72) or the dGTPase inhibitor that might allow ETEP21B to infect ET117, a strain carrying a dGTPase as a phage defense mechanism (supplemental file S2). ETEP102 is a siphovirus that belongs to the *Nouzillyvirus* genus, part of the *Drexlerviridae* family, and despite some similarity with the coliphages T1 and TLS, the genome annotation was more challenging, leaving half of the genes with no predicted function. Further analysis suggested that the phages employ diverse strategies for infection and could thus work synergistically in a phage cocktail. For example, ETEP21B is expected to be independent of the host for DNA replication and transcription, whereas ETEP102 most probably relies on the host machinery since it does not encode an RNA polymerase, similar to phage T1 (73). Notably, as demonstrated by the host range, both phages are partially resistant to RM systems, widespread in the PHAGEBio ETEC collection. Common ways to escape DNA-targeting RM systems are the methylation of the viral DNA and the underrepresentation of restriction sites. Restriction sites are underrepresented in ETEP21B (only two Dam and no Dcm sites), as typical for *Autographiviridae* phages (40). In addition, ETEP102 encodes for its own Dam and Dcm, thus suggesting its DNA is methylated, as in other *Drexlerviridae* phages, such as T1 and TLS (73, 74).

Analyzing the host range, the most evident observation was the infection of many O149 strains by both phages, which suggests a crucial role in receptor recognition and binding. Since it is well established that the presence of the O-antigen inhibits T7 infection (40) and that DNA injection follows the binding of tail fibers to sugars in the inner core of the LPS on rough *E. coli* strains (75), we compared T7's and ETEP21B's tail fibers to understand the ETEP21B infection process of O149 strains. The four external loops at the tip of the T7 tail fiber are essential for recognizing the receptor (41). Our data suggest that loop 3 might contribute more significantly to a difference in the host range because of its predicted propensity to bind sugars. Nevertheless, ETEP21B's TF structure is more complex than T7's, with a possible second ligand domain that could host an additional or different receptor than T7, contributing to the ability to infect O149 strains. Based on host ranges of mutants lacking known phage receptors, it was established that ETEP21B recognizes sugars in the LPS core as a receptor and that it can still infect the *waaC* mutant at low EOP, as previously observed for most *Autographiviridae* phages (23). To explain this observation, it was previously speculated that these phages might recognize multiple sugar combinations (76) or a second receptor, such as OmpA and OmpF porins (77, 78). For ETEP21B, it was not possible to identify a second receptor, and the infection of additional strains with other O-antigens supports the idea that binding does not depend on a specific O-antigen glycan.

Phages in the *Drexlerviridae* family, such as ETEP102, infect their host via a two-step process with the reversible binding of LPS first by the tail fibers and the recognition of a second receptor by the distal end of the tail tip protein (46, 79). Given the recognition of mainly O149 strains, we suggest that ETEP102 tail fibers can bind and depolymerize glycans of this specific O-antigen. The second and irreversible receptor is usually an outer membrane protein, and it can be predicted by comparison with the different gene variants encoding for the receptor binding proteins (RBP) in the locus downstream of the *gpJ* homolog (23). The RBP locus in ETEP102 encodes proteins that are structurally similar to phage Rtp, which uses the essential LPS export channel LptD as receptor (23). Nevertheless, the low sequence similarity raised some doubts about assigning LptD as receptor of phage ETEP102, thus leaving ETEP102's second receptor to be confirmed with additional experiments.

## Conclusion

ETEC is a heterogenous pathotype both in the virulence genes encoded and in their phylogeny. Antibiotic resistance genes and pathotype virulence factors are mostly plasmid encoded, while phage resistance genes and other virulence factors are mostly on chromosomes, thus reducing the risk of transfer to other strains and co-selection with virulence genes and antibiotic or metal resistance genes. Because of the high number of encoded resistance mechanisms, ETEC strains seem to be resistant to *Enterobacteriaceae* phages in a complex and multifactorial way. Nevertheless, we could isolate two lytic phages infecting ETEC strains and covering a large part of our diverse ETEC collection. By studying the genome of the two ETEC phages, it was observed that they have different strategies to propagate in the bacterial host and overcome the defense systems. In addition, the receptors for the two phages were identified as part of the LPS and speculated that specific loops and pockets in the tail fiber structures of the two phages allow recognition and binding of ETEC strains despite the presence of O-antigens. With this work, we contribute to the sustainable use of phage products specifically targeting the ETEC strains that are currently causing PWD in Europe, and we hope that the scientific community will use the collection to expand our knowledge from laboratory to clinically relevant *E. coli* strains.

## MATERIALS AND METHODS

### ETEC collection and other bacterial strains

Seventy-nine ETEC isolates were collected by isolating hemolytic *E. coli* strains from pigs with PWD in the period from 2014 to 2019 from different farms in Denmark (via the surveillance institute Danish Pig Research Center, SEGES), Belgium-Netherlands, Greece, and Italy (via the surveillance institute Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna, IZSLER) (supplemental file S1). Fimbriae and toxins were identified at isolation by multiplex quantitative PCR (qPCR) by SEGES (80) and by multiplex PCR by IZLER (81) and then further confirmed once received at the University of Copenhagen (7).

ETEC strains were streaked on blood agar plates (BA, 5% calf blood in blood agar base, Thermo Fischer, CM0055) to ensure the absence of contaminants, grown in LB (Lysogeny Broth, Merck) at 37°C, with agitation, overnight (ON), and stored at –80°C as frozen stocks in 15% glycerol. Where stated, the strains have been grown in LB with mucin (LB M; 0.1 g/mL porcine mucin, M1778-10G, Sigma-Aldrich) or BHI broth (Brain Heart Infusion broth, CM1135, Oxoid).

The laboratory strains ECOR04, MG1655, and NCTC12900, used for the host range and for propagation of phages not isolated in this work, were grown in LB at 37°C, with agitation, ON, and stored at –80°C as frozen stocks in 15% glycerol.

## DNA extraction and sequencing

Bacterial DNA was extracted from ON cultures of the 79 ETEC strains with DNeasy Blood & Tissue Kit (Qiagen). Except for ET54 that was sequenced with MiSeq (Illumina, San Diego, CA, USA), the remaining strains were sequenced by paired-end libraries with an insert size of 200–400 bp and further sequenced on the MGISEQ-2000 (MGI, BGI-Shenzhen) platform to obtain about >100× clean reads data. Quality control for clean reads was performed by filtering with SOAPnuke and fastp (82), and clean reads were assembled with SPAdes v3.13.0 (83).

## Assembly and genome analysis

All ETEC genomes are available on GenBank (BioProject ID: [PRJNA770188](https://ncbi.nlm.nih.gov/bioproject/PRJNA770188)) as the PHAGEBio ETEC collection. Sequence quality has been checked by BUSCO. The genomes varied widely in size from 4.9 to 6.2 Mb (average 5.5 Mb), with a density of protein-coding sequences of the range 82%–88% (average 86%) and GC content of the range 50%–50.9%. The pan-genome contains 20,909 gene families, and the core genome contains 2779 genes (13%).

Antibiotic resistance, metal resistance, and virulence factor genes in the bacterial genomes were predicted using tools ABRicate v1.0.0 (<https://github.com/tseemann/abricate>) and NCBI AMRFinder Plus v3.10.30 (84) using default parameters. The defense systems in these bacterial genomes were identified using the tools PADLOC v1.1.0 (85) and DefenseFinder v2.1.1 (66, 86) using standard parameters. The plasmid-borne contigs (not closed plasmids) were identified using the tool Platon v1.6 (87). The prophage regions were detected using the tool Phigaro v2.3.0 (88). The anti-microbial resistance, virulence, and defense system genes were mapped on plasmid and prophage regions using customized Python scripts (supplemental file S2).

The ETEC phylogroup was established by submitting the ETEC genomes to EzClermont (<https://ezclermont.hutton.ac.uk/>; 89) and further confirmed by building a phylogenetic tree with the ETEC strains, the 72 strains of the *Escherichia coli* Reference Collection (ECOR; 90) and *Escherichia fergusonii* (NC\_011740.1) as an outgroup using Population Partitioning Using Nucleotide K-mers (PopPUNK) v 2.5.0 (91). The core and the accessory genome distance between the strains was estimated for K-mer sizes 15 to 31 with a step size of 2. The phylogeny created by PopPUNK was visualized using Microreact (92) and iTOL v6.7 (93) (supplemental file S2).

## Receptor mutant generation

A series of knockout mutants in genes coding for known or potential phage receptors was generated in ETEC strains ET54 and ET3 using Lambda-Red-mediated mutagenesis. A list of used primers is available as supplemental file S10. Allelic replacement was facilitated by the recombination protein expression plasmid pRedET from the Quick & Easy *E. coli* Gene Deletion Kit (Gene Bridges) and based on the FRT-PGK-gb2-neo-FRT cassette also supplied in the kit. All deletion fragments were synthesized by PCR; the majority were amplified with gene-specific primers on mutants recently generated in *E. coli* ECOR04 (24), while the remaining were made with resistance cassette-specific primers equipped with 50-bp extensions matching the targeted gene. For ET54 double mutants, a *waaC* mutant was first generated with a deletion fragment encoding chloramphenicol resistance, amplified from pKD3, and the secondary mutations were then introduced by the kanamycin resistance-encoding mutations already established in single mutants. Target strains were first transformed with pRedET by heat shock using a standard CaCl<sub>2</sub> chemically competent cell protocol and selected on 100 mg/L ampicillin at 30°C. Transformants were subsequently grown in liquid culture (LB, 100 mg/L ampicillin, 30°C, 180 rpm) until an optical density at 600 nm (OD<sub>600</sub>) of 0.3, followed by induction of plasmid expression with 0.35% L-arabinose (final concentration) for 1 hour at 37°C. Cells were washed twice with one volume, once with 1/2 volume, and once with 1/4 volume and finally resuspended in 1/100 vol of ice-cold water. Fifty microliters of

competent cells was mixed with 2  $\mu\text{L}$  concentrated, PCR-generated deletion fragment in a pre-chilled electroporation cuvette and pulsed in a MicroPulser Electroporator (Bio-Rad) employing the Ec2 setting. Transformed cells were recovered in 1 mL LB and incubated at 37°C for 2–3 hours before selection on LA with 15 mg/L kanamycin or 25 mg/L chloramphenicol at 37°C ON. Target gene knockout was confirmed with gene-specific primers.

### Sample processing and phage isolation

Wastewater ( $n = 3$ ), pig manure ( $n = 4$ ), gut samples from pigs dying from ETEC diarrhea ( $n = 23$ ), feces from healthy pigs ( $n = 16$ ), and feces from ETEC-affected pigs ( $n = 31$ ) have been processed and screened for phages infecting ETEC. Solid samples were diluted (1:10 wt/vol) in sterile SM buffer (0.1 M NaCl, 8 mM  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.01% gelatine, 50 mM Tris-HCl, pH 7.5), stomached, and centrifuged (18,000  $\times g$  for 10 min). Supernatants were filtered twice through 0.45  $\mu\text{m}$  filters and stored at 4°C.

The 77 supernatants were screened for phages infecting ETEC by direct spotting and enrichment on the 41 Danish ETEC strains (from ET03 to ET54). For the enrichments, 400  $\mu\text{L}$  of each supernatant was added to 1 mL of liquid medium (LB, LB + M, BHI) and 400  $\mu\text{L}$  of a 4-hour culture of each of the selected 41 ETEC strains and incubated ON at 37°C in static or agitation. The day after the enrichments were centrifuged for 5 min at 10,000  $\times g$ , and three drops of 10  $\mu\text{L}$  each of the enriched and not enriched supernatants were spotted on bacterial lawns of the respective ETEC strains, along with the not-enriched supernatants. For preparing bacterial lawns, 330  $\mu\text{L}$  of a 4-hour cultures (LB, agitation, 37°C) was mixed with 11 mL of molten overlay agar (LBov: LB with 0.6% Agar bacteriological no.1, Oxoid; BHlov: BHI with 0.5% Agar bacteriological no.1, Oxoid) and spread on square 12-cm plates (LA: LB with 1.2% agar, or BHIA: BHI with 1.2% agar) plates. After settling for 5 min, lawns were dried in a laminar hood for 35 min and used immediately thereafter. For any spot with an inhibition halo, the spotted supernatant (enriched or not) was 10-fold diluted in SM buffer and spotted on a lawn of the same ETEC strain in order to observe plaques. Visible plaques were transferred to 500  $\mu\text{L}$  of SM buffer (0.1 M NaCl, 8 mM  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 50 mM Tris-HCl, pH 7.5), vortexed, and 10-fold diluted. A 100- $\mu\text{L}$  aliquot of selected dilutions was mixed with 100  $\mu\text{L}$  of the isolation strain in 3.2 mL LBov or BHlov and spread on LA or BHIA plates. Each single plaque was purified by repeating this procedure for at least three rounds. Single plaques from the final purification steps were used for phage propagation on the isolation strain, and phage stocks were prepared by the plate lysis methods as adopted from Carlson (94).

### Phage host range

The host range of the isolated ETEC infecting phages (ETEP21B, ETEP102), of 42 *Salmonella* phages (38, 39) and up to 200 *E. coli* phages, including the T3, T7, and CBA120, and others isolated on *E. coli* MG1655 and ECOR04 have been verified by spotting assay. Tenfold serial dilutions (up to  $10^{-7}$ – $10^{-9}$ ) of the stocks in SM buffer were prepared, and 10- $\mu\text{L}$  aliquots were spotted on bacterial lawns of the ETEC strains, let adsorb, and incubated ON. The day after, plaques were counted for each phage and strain and plaque-forming units per mL (PFU/mL) and the efficiency of plating compared with the propagation host (EOP) were calculated. Log of the calculated efficiency of plating is reported in supplemental file S3. At least two independent replicates confirmed the host range.

### Phage DNA extraction, sequencing, assembly, and annotation

DNA was extracted from  $10^9$  PFU/mL stocks of phages ETEP21B and ETEP102 with a modified phenol-chloroform protocol (95). Briefly, phage stocks were exposed to RNase (10  $\mu\text{g}/\text{mL}$ , Thermo Fisher Scientific, Waltham, MA, USA) and DNase (20  $\mu\text{g}/\text{mL}$ , Thermo Fisher Scientific) activity at 37°C for 3 hours in a thermo-shaker (500 rpm, Eppendorf, Germany). Following the addition of EDTA (20 mM) and proteinase K (50  $\mu\text{g}/\text{mL}$ , Thermo

Fisher Scientific) and incubation at 56°C for 4 hours, phenol (Fluka), phenol-chloroform-isoamyl alcohol (25:24:1, Ambion), and three rounds of chloroform-isoamyl alcohol (24:1) treatment were performed. To precipitate the DNA, 0.1 vol of 3 M sodium acetate (pH 5.5), glycogen (final concentration of 0.05 µg/µL, Thermo Fisher Scientific), and 2.5 volumes of ice-cold ethanol (99.9%) were added. After incubation at –20°C for 72 hours, precipitated DNA was centrifuged at 31,000 × *g* for 20 min, washed three times with 70% ice-cold ethanol, and dissolved in 10 mM Tris-HCl (pH 8.0). DNA concentration was measured using Qubit (Thermo Fisher Scientific). DNA libraries were prepared using the Nextera XT v.3 (Illumina, San Diego, CA, USA) Kit. Next-generation sequencing was performed using the MiSeq (Illumina) platform with the paired-end (2 × 250 bp) operating mode.

Sequencing reads were *de novo* assembled using CLC Genomics Workbench 9.5.3 (Qiagen, Aarhus, Denmark). A consensus sequence was obtained with a minimum of 30-fold coverage. Analysis and annotation of the phage genome were performed using tools in Galaxy (cpt.tamu.edu/galaxy -pub) and Web Apollo, hosted by the Center for Phage Technology at Texas A&M University (CPT Galaxy) (<https://cpt.tamu.edu/galaxy-pub/>). Gene calling was performed using GLIMMER 3.0 and MetaGeneAnnotator 1.0 within the structural workflow, while for the genome annotation, the functional workflow was used, by interrogating the databases of UniProtKB Swiss-Prot/TrEMBL, Canonical Phages, and HHMER. The promoters were identified with PhagePromoter (96), selecting *E. coli* as host, and manually curated. Terminators were predicted with TransTermHP, available on CPT Phage Galaxy, and manually curated (97). Putative spanins have been identified with the spanin tools in CPT Phage Galaxy and curated as suggested by Kongari (98). Phage genomes were visualized with the SnapGene software ([www.snapgene.com](http://www.snapgene.com)). The presence of encoded tRNA genes in the phage was checked using Aragorn (99), and the phage codon usage was established with the Sequence Manipulation Suite (100). In addition, homolog detection and structural prediction by HHPRED (101), SWISS-MODEL (102), and, to some extent, PHYRE2 (103) were used to investigate further the function of the predicted structural proteins in the two phages.

### Phage taxonomy and genome comparison

To taxonomically classify ETEP21B and ETEP102, VIRIDIC (104) and vConTACT v2.0 0.9.19 (105) were run to identify the genus cluster the two ETEC phages belong to. The two ETEC phage genomes and other 158 genomes (their closest relatives, one genome for each of the genera within the same family, and one genome for each of the other families as listed in the ICTV\_Master\_Species\_List\_2021\_v2.xlsx) were used to build a phylogenetic tree with VIRIDIC for final taxonomic classification. ANI has been calculated with an ANI calculator on EZBioCloud ([www.ezbiocloud.net](http://www.ezbiocloud.net)). Phage genome alignment with others within their genera was run and visualized with Clinker (106).

### Transmission electron micrographs

For visualization of phage ETEP21B and ETEP102, samples were prepared as previously described (107). Briefly, high-titer phage stocks were pelleted at 12,000 × *g* for 60 min, at 4°C, and then washed three times with ammonium acetate (0.1 M, pH 7). For imaging, 200 mesh copper-coated grids (Ted Pella Inc.) were glow discharged using a Leica Coater ACE 200 for 30 sec at 10 mA. Grids were incubated with phage samples for 30 sec at room temperature. Liquid in excess was blotted off, and grids were stained with 2% uranyl acetate for 30 sec. Imaging was performed using a CM100 electron microscope with a Bio TWIN objective lens and a LaB6 emitter. Images were taken using an Olympus Veleta camera. Phage morphology features were measured using ImageJ.

### AlphaFold2 prediction and structure analysis

The tail fiber models for ETEP21B, ETEP102, and T7 were predicted using AlphaFold2-multimer v3.2.1 (108, 109) on an Nvidia Quadro RTX 8000 GPU. The full database search

was employed for the multiple sequence alignment. The multimer mode utilized two seeds per model, while the side chain relaxation step was omitted, with a maximum of five recycles. The best model for each fiber was selected based on the ipTM+ pLDDT scores (Predicted Aligned Error Matrices in supplemental file S7C). Structural homologs were identified using the online version of FoldSeek (110).

For structure analysis, we utilized ChimeraX-v1.5 software (111) for tasks such as protein superposition, model handling, and coloring. Protein interaction interfaces were predicted using the online version of PeSTo (112). The protein topology diagram was generated using the online tool PDBSum (113).

## ACKNOWLEDGMENTS

For providing ETEC strains and/or samples for phage isolation, we thank Svend Haugegaard and collaborators from SEGES Landbrug & Fødevarer (DK), Andrea Luppi from Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna (IT), Inge Larsen from University of Copenhagen (DK), and Jan Lorenzen from Danish Technological Institute Århus (DK).

This work was supported by the Innovation Fund Denmark (IFD; BioPIGLET project, 7076-00038B). V.K.-S. acknowledges the Novo Nordisk Foundation Copenhagen PhD Programme for grant NNF0069780.

We acknowledge the Core Facility for Integrated Microscopy, Faculty of Health and Medical Sciences, University of Copenhagen. We acknowledge the use of the IRIDIS High-Performance Computing Facility and associated support services at the University of Southampton.

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg, Denmark

<sup>2</sup>Institute of Conservation, The Royal Danish Academy, Copenhagen, Denmark

<sup>3</sup>School of Biological Sciences, Faculty of Environmental & Life Sciences, University of Southampton, Southampton, United Kingdom

<sup>4</sup>Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Rajasthan, India

<sup>5</sup>Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

<sup>6</sup>BGI-Shenzhen, Shenzhen, China

## AUTHOR ORCIDs

Michela Gambino  <http://orcid.org/0000-0001-7098-3549>

## FUNDING

Funder	Grant(s)	Author(s)
Innovationsfonden (IFD)	7076-00038B	Michela Gambino Lone Brøndsted
<a href="#">Novo Nordisk Foundation Center for Basic Metabolic Research (NovoNordisk Foundation Center for Basic Metabolic Research)</a>	NNF0069780	Victor Klein-Sousa

## AUTHOR CONTRIBUTIONS

Michela Gambino, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing | Simran Krishnakant Kushwaha, Formal analysis, Software, Visualization | Yi Wu, Formal analysis, Software, Visualization | Pauline

van Haastrecht, Formal analysis, Software, Visualization | Victor Klein-Sousa, Formal analysis, Investigation, Visualization, Writing – review and editing | Veronika T. Lutz, Investigation | Semeh Bejaoui, Investigation | Christoffer Moeskjær C. Jensen, Investigation | Martin S. Bojer, Investigation, Writing – review and editing | Wenchen Song, Investigation, Software | Minfeng Xiao, Software | Nicholas M. I. Taylor, Software | Franklin L. Nobrega, Conceptualization, Formal analysis, Software, Visualization | Lone Brøndsted, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review and editing

## DATA AVAILABILITY

ETEC genomes have been deposited in the GenBank database with the BioProject ID [PRJNA770188](https://doi.org/10.26434/chemrxiv-2024-00000) and the phage genome sequences under accession numbers [OR979722](https://doi.org/10.26434/chemrxiv-2024-00000) (ETEP21B) and [OR979723](https://doi.org/10.26434/chemrxiv-2024-00000) (ETEP102). ETEC strains are available upon request to Lone Brøndsted, Department of Veterinary and Animal Sciences, University of Copenhagen, Stigbøjlen 4, 1870, Frederiksberg C, Denmark.

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Supplemental material part I (AEM00807-24-s0001.pdf).** Files S1 to S5.

**Supplemental material part II (AEM00807-24-s0002.pdf).** Files S6 to S10.

## REFERENCES

- Fairbrother JM, Nadeau E, Gyles CL. 2005. *Escherichia coli* in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. *Anim Health Res Rev* 6:17–39. <https://doi.org/10.1079/ahr2005105>
- Dubreuil JD, Isaacson RE, Schifferli DM. 2016. Animal enterotoxigenic *Escherichia coli*. *EcoSal Plus* 7:1–80. <https://doi.org/10.1128/ecosalplus.ESP-0006-2016>
- Luppi A, Gibellini M, Gin T, Vangroenweghe F, Vandenbroucke V, Bauerfeind R, Bonilauri P, Labarque G, Hidalgo Á. 2016. Prevalence of virulence factors in enterotoxigenic *Escherichia coli* isolated from pigs with post-weaning diarrhoea in Europe. *Porc Health Manag* 2:1–6. <https://doi.org/10.1186/s40813-016-0039-9>
- Sjölund M, Zorix M, Wallgren P. 2014. Financial impact of disease on pig production. Part III. Gastrointestinal disorders. The 6th European Symposium of Porcine Health Management 189
- Fleckenstein JM, Hardwidge PR, Munson GP, Rasko DA, Sommerfelt H, Steinsland H. 2010. Molecular mechanisms of enterotoxigenic *Escherichia coli* infection. *Microbes Infect* 12:89–98. <https://doi.org/10.1016/j.micinf.2009.10.002>
- Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebahia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J. 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* 190:6881–6893. <https://doi.org/10.1128/JB.00619-08>
- Zhang W, Zhao M, Ruesch L, Omot A, Francis D. 2007. Prevalence of virulence genes in *Escherichia coli* strains recently isolated from young pigs with diarrhea in the US. *Vet Microbiol* 123:145–152. <https://doi.org/10.1016/j.vetmic.2007.02.018>
- Madhavan TPV, Sakellaris H. 2015. Chapter five - Colonization factors of enterotoxigenic *Escherichia coli*, p 155–197. In *Advances in applied microbiology*. Academic Press.
- Denamur E, Clermont O, Bonaccorsi S, Gordon D. 2021. The population genetics of pathogenic *Escherichia coli*. *Nat Rev Microbiol* 19:37–54. <https://doi.org/10.1038/s41579-020-0416-x>
- Kopic S, Geibel JP. 2010. Toxin mediated diarrhea in the 21<sup>st</sup> century: the pathophysiology of intestinal ion transport in the course of ETEC, *V. cholerae* and rotavirus Infection. *Toxins (Basel)* 2:2132–2157. <https://doi.org/10.3390/toxins2082132>
- Shepard SM, Danzeisen JL, Isaacson RE, Seemann T, Achtman M, Johnson TJ. 2012. Genome sequences and phylogenetic analysis of K88- and F18- positive porcine enterotoxigenic *Escherichia coli*. *J Bacteriol* 194:395–405. <https://doi.org/10.1128/JB.06225-11>
- García V, Gambino M, Pedersen K, Haugegaard S, Olsen JE, Herrero-Fresno A. 2020. F4- and F18-positive enterotoxigenic *Escherichia coli* isolates from diarrhea of postweaning pigs: genomic characterization. *Appl Environ Microbiol* 86:e01913-20. <https://doi.org/10.1128/AEM.01913-20>
- Subramani P, Pirolo M, Haugegaard S, Skarbye AP, Conrady B, Pedersen KS, Guardabassi L, Damborg P. 2023. Neomycin resistance in clinical *Escherichia coli* from Danish weaner pigs is associated with recent neomycin use and presence of F4 or F18 fimbriae. *Prev Vet Med* 212:105852. <https://doi.org/10.1016/j.prevetmed.2023.105852>
- Rhouma M, Fairbrother JM, Beaudry F, Letellier A. 2017. Post weaning diarrhea in pigs: risk factors and non-colistin-based control strategies. *Acta Vet Scand* 59:31. <https://doi.org/10.1186/s13028-017-0299-7>
- Middelkoop A, Kettunen H, Guan X, Vuorenmaa J, Tichelaar R, Gambino M, Rydal MP, Molist F. 2024. Effect of dietary tall oil fatty acids and hydrolysed yeast in SNP2-positive and SNP2-negative piglets challenged with F4 enterotoxigenic *Escherichia coli*. *Sci Rep* 14:2060. <https://doi.org/10.1038/s41598-024-52586-3>
- Canibe N, Højberg O, Kongsted H, Vodolazska D, Lauridsen C, Nielsen TS, Schönherz AA. 2022. Review on preventive measures to reduce post-weaning diarrhoea in piglets. *Animals (Basel)* 12:2585. <https://doi.org/10.3390/ani12192585>
- Bonetti A, Tugnoli B, Piva A, Grilli E. 2021. Towards zero zinc oxide: feeding strategies to manage post-weaning diarrhea in piglets. *Animals (Basel)* 11:642. <https://doi.org/10.3390/ani11030642>
- Rydal M.P, Gambino M, Jørgensen CB, Poulsen LL, Brøndsted L, Nielsen JP. 2022. Pilot study on CHCF1 genotype in a pig challenge model for enterotoxigenic *Escherichia coli* F4ab/ac associated post-weaning diarrhea. *BMC Vet Res* 18:382. <https://doi.org/10.1186/s12917-022-03474-3>

19. Rydal MP, Jørgensen CB, Gambino M, Poulsen LL, Nielsen JP. 2023. Complete association between CHCF1 genotype and enterotoxigenic *Escherichia coli* F4ab-associated post-weaning diarrhea in a pig challenge trial. *Vet Microbiol* 282:109771. <https://doi.org/10.1016/j.vetmic.2023.109771>
20. European Commission. 2017. A European one health action plan against antimicrobial resistance (AMR). Brussels 24
21. Kortright KE, Chan BK, Turner PE. 2020. High-throughput discovery of phage receptors using transposon insertion sequencing of bacteria. *Proc Natl Acad Sci U S A* 117:18670–18679. <https://doi.org/10.1073/pnas.2001888117>
22. Nagel T, Musila L, Muthoni M, Nikolich M, Nakavuma JL, Clokie MR. 2022. Phage banks as potential tools to rapidly and cost-effectively manage antimicrobial resistance in the developing world. *Curr Opin Virol* 53:101208. <https://doi.org/10.1016/j.coviro.2022.101208>
23. Maffei E, Shaidullina A, Burkolter M, Heyer Y, Estermann F, Druelle V, Sauer P, Willi L, Michaelis S, Hilbi H, Thaler DS, Harms A. 2021. Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. *PLoS Biol* 19:e3001424. <https://doi.org/10.1371/journal.pbio.3001424>
24. Vitt AR, Sørensen AN, Bojer MS, Bortolaia V, Sørensen MCH, Brøndsted L. 2024. Diverse bacteriophages for biocontrol of ESBL- and AmpC- $\beta$ -lactamase-producing *E. coli*. *iScience* 27:108826. <https://doi.org/10.1016/j.isci.2024.108826>
25. Smith HW, Huggins MB. 1983. Effectiveness of phages in treating experimental *Escherichia coli* diarrhoea in calves, piglets and lambs. *Microbiology* 129:2659–2675. <https://doi.org/10.1099/00221287-129-8-2659>
26. Jamalludeen N, Kropinski AM, Johnson RP, Lingohr E, Harel J, Gyles CL. 2008. Complete genomic sequence of bacteriophage  $\phi$ EcoM-GJ1, a novel phage that has myovirus morphology and a podovirus-like RNA polymerase. *Appl Environ Microbiol* 74:516–525. <https://doi.org/10.1128/AEM.00990-07>
27. Jamalludeen N, Johnson RP, Shewen PE, Gyles CL. 2009. Evaluation of bacteriophages for prevention and treatment of diarrhea due to experimental enterotoxigenic *Escherichia coli* O149 infection of pigs. *Vet Microbiol* 136:135–141. <https://doi.org/10.1016/j.vetmic.2008.10.021>
28. Cha SB, Yoo AN, Lee WJ, Shin MK, Jung MH, Shin SW, Cho YW, Yoo HS. 2012. Effect of bacteriophage in enterotoxigenic *Escherichia coli* (ETEC) infected pigs. *J Vet Med Sci* 74:1037–1039. <https://doi.org/10.1292/jvms.11-0556>
29. Nguyen XH, Fang T, Qinqin B, Chanjuan S, Thanh TV, Chengping L. 2013. Isolation and characterization of bacteriophages against enterotoxigenic *Escherichia coli* strains causing swine diseases. *Afr J Microbiol Res* 7:2787–2793. <https://doi.org/10.5897/AJMR2013.5744>
30. Lee CY, Kim SJ, Park BC, Han JH. 2017. Effects of dietary supplementation of bacteriophages against enterotoxigenic *Escherichia coli* (ETEC) K88 on clinical symptoms of post-weaning pigs challenged with the ETEC pathogen. *J Anim Physiol Anim Nutr* 101:88–95. <https://doi.org/10.1111/jpn.12513>
31. Kim GH, Kim JW, Kim J, Chae JP, Lee JS, Yoon SS. 2020. Genetic analysis and characterization of a bacteriophage  $\phi$ CJ19 active against enterotoxigenic *Escherichia coli*. *Food Sci Anim Resour* 40:746–757. <https://doi.org/10.5851/kosfa.2020.e49>
32. Ferreira A, Oliveira H, Silva D, Almeida C, Burgan J, Azeredo J, Oliveira A. 2020. Complete genome sequences of eight phages infecting enterotoxigenic *Escherichia coli* in swine. *Microbiol Resour Announc* 9:5–8. <https://doi.org/10.1128/MRA.00858-20>
33. Wei B, Cong C, Zheng L, Chen L, Yan X. 2023. Isolation, characterization and whole genome analysis of the novel genus *Lederbergvirus*, phage vB\_EcoP\_E212 infecting enterotoxigenic *Escherichia coli* K88. *Virus Res* 331:199125. <https://doi.org/10.1016/j.virusres.2023.199125>
34. Navez M, Antoine C, Laforêt F, Goya-Jorge E, Douny C, Scippo ML, Vermeersch M, Duprez JN, Daube G, Mainil J, Taminiau B, Delcensier V, Thiry D. 2023. *In vitro* effect on piglet gut microbiota and *in vivo* assessment of newly isolated bacteriophages against F18 enterotoxigenic *Escherichia coli* (ETEC). *Viruses* 15:1053. <https://doi.org/10.3390/v15051053>
35. Lin Y, Zhou B, Zhu W. 2021. Pathogenic *Escherichia coli*-specific bacteriophages and polyvalent bacteriophages in piglet guts with increasing coliphage numbers after weaning. *Appl Environ Microbiol* 87:e0096621. <https://doi.org/10.1128/AEM.00966-21>
36. Ochman H, Selander RK. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* 157:690–693. <https://doi.org/10.1128/jb.157.2.690-693.1984>
37. Shen BW, Doyle LA, Werther R, Westburg AA, Bies DP, Walter SI, Luyten YA, Morgan RD, Stoddard BL, Kaiser BK. 2023. Structure, substrate binding and activity of a unique AAA+ protein: the BrxL phage restriction factor. *Nucleic Acids Res* 51:3513–3528. <https://doi.org/10.1093/nar/gkad083>
38. Gencay YE, Gambino M, Prüssing TF, Brøndsted L. 2019. The genera of bacteriophages and their receptors are the major determinants of host range. *Environ Microbiol* 21:2095–2111. <https://doi.org/10.1111/1462-2920.14597>
39. Gambino M, Nørgaard Sørensen A, Ahern S, Smyrlis G, Gencay YE, Hendrix H, Neve H, Noben J-P, Lavigne R, Brøndsted L. 2020. Phage S144, a new polyvalent phage infecting *Salmonella* spp. and *Cronobacter sakazakii*. *Int J Mol Sci* 21:5196. <https://doi.org/10.3390/ijms21155196>
40. Molineux I. 2006. The T7 group, p 277–301. In *The bacteriophages*. Oxford University Press.
41. Garcia-Doval C, van Raaij MJ. 2012. Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. *Proc Natl Acad Sci U S A* 109:9390–9395. <https://doi.org/10.1073/pnas.1119719109>
42. Steven AC, Trus BL, Maizel JV, Unser M, Parry DAD, Wall JS, Hainfeldt JF, Studier FW. 1988. Molecular substructure of a viral receptor-recognition protein: the gp17 tail-fiber of bacteriophage T7. *J Mol Biol* 200:351–365. [https://doi.org/10.1016/0022-2836\(88\)90246-x](https://doi.org/10.1016/0022-2836(88)90246-x)
43. Stenutz R, Weintraub A, Widmalm G. 2006. The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev* 30:382–403. <https://doi.org/10.1111/j.1574-6976.2006.00016.x>
44. Roberts MD, Martin NL, Kropinski AM. 2004. The genome and proteome of coliphage T1. *Virology* 318:245–266. <https://doi.org/10.1016/j.virol.2003.09.020>
45. Casjens SR, Gilcrease EB. 2009. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods Mol Biol* 502:91–111. [https://doi.org/10.1007/978-1-60327-565-1\\_7](https://doi.org/10.1007/978-1-60327-565-1_7)
46. Wietzorrek A, Schwarz H, Herrmann C, Braun V. 2006. The genome of the novel phage Rtp, with a rosette-like tail tip, is homologous to the genome of phage T1. *J Bacteriol* 188:1419–1436. <https://doi.org/10.1128/JB.188.4.1419-1436.2006>
47. Itoh T, Nakagawa E, Yoda M, Nakaichi A, Hibi T, Kimoto H. 2019. Structural and biochemical characterisation of a novel alginate lyase from *Paenibacillus* sp. str. FPU-7. *Sci Rep* 9:14870. <https://doi.org/10.1038/s41598-019-51006-1>
48. Itoh T, Intuy R, Suyotha W, Hayashi J, Yano S, Makabe K, Wakayama M, Hibi T. 2020. Structural insights into substrate recognition and catalysis by glycoside hydrolase family 87  $\alpha$ -1,3-glucanase from *Paenibacillus glycanilyticus* FH11. *FEBS J* 287:2524–2543. <https://doi.org/10.1111/febs.15161>
49. Plattner M, Shneider MM, Arbatsky NP, Shashkov AS, Chizhov AO, Nazarov S, Prokhorov NS, Taylor NMI, Buth SA, Gambino M, Gencay YE, Brøndsted L, Kutter EM, Knirel YA, Leiman PG. 2019. Structure and function of the branched receptor-binding complex of bacteriophage CBA120. *J Mol Biol* 431:3718–3739. <https://doi.org/10.1016/j.jmb.2019.07.022>
50. Lang S, Kirchberger PC, Gruber CJ, Redzej A, Raffl S, Zellnig G, Zangger K, Zechner EL. 2011. An activation domain of plasmid R1 Tral protein delineates stages of gene transfer initiation. *Mol Microbiol* 82:1071–1085. <https://doi.org/10.1111/j.1365-2958.2011.07872.x>
51. Ilangovan A, Kay CWM, Roier S, El Mkami H, Salvadori E, Zechner EL, Zanetti G, Waksman G. 2017. Cryo-EM structure of a relaxase reveals the molecular basis of DNA unwinding during bacterial conjugation. *Cell* 169:708–721. <https://doi.org/10.1016/j.cell.2017.04.010>
52. Zechner EL, Moncalián G, de la Cruz F. 2017. Relaxases and plasmid transfer in Gram-negative bacteria, p 93–113. In *Backert S, Grohmann E (ed), Type IV secretion in Gram-negative and Gram-positive bacteria*. Springer International Publishing, Cham.
53. Slifnerz MJ, Friendship R, Weese JS. 2015. Zinc oxide therapy increases prevalence and persistence of methicillin-resistant *Staphylococcus*

- aureus* in pigs: a randomized controlled trial. *Zoonoses Public Health* 62:301–308. <https://doi.org/10.1111/zph.12150>
54. Becerra-Castro C, Machado RA, Vaz-Moreira I, Manaia CM. 2015. Assessment of copper and zinc salts as selectors of antibiotic resistance in Gram-negative bacteria. *Sci Total Environ* 530–531:367–372. <https://doi.org/10.1016/j.scitotenv.2015.05.102>
  55. Ciesinski L, Guenther S, Pieper R, Kalisch M, Bednorz C, Wieler LH. 2018. High dietary zinc feeding promotes persistence of multi-resistant *E. coli* in the swine gut. *PLoS One* 13:e0191660. <https://doi.org/10.1371/journal.pone.0191660>
  56. Touchon Marie, Perrin A, de Sousa JAM, Vangchhia B, Burn S, O'Brien CL, Denamur E, Gordon D, Rocha EP. 2020. Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *PLoS Genet* 16:e1008866. <https://doi.org/10.1371/journal.pgen.1008866>
  57. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, et al. 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5:e1000344. <https://doi.org/10.1371/journal.pgen.1000344>
  58. Garcillán-Barcia MP, Francia MV, de la Cruz F. 2009. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev* 33:657–687. <https://doi.org/10.1111/j.1574-6976.2009.00168.x>
  59. Frydendahl K. 2002. Prevalence of serogroups and virulence genes in *Escherichia coli* associated with postweaning diarrhoea and edema disease in pigs and a comparison of diagnostic approaches. *Vet Microbiol* 85:169–182. [https://doi.org/10.1016/s0378-1135\(01\)00504-1](https://doi.org/10.1016/s0378-1135(01)00504-1)
  60. von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, Rasko DA, Joffe E, Corander J, Pickard D, Wiklund G, Svennerholm A-M, Sjöling Å, Dougan G. 2014. Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat Genet* 46:1321–1326. <https://doi.org/10.1038/ng.3145>
  61. Steinsland H, Lacher DW, Sommerfelt H, Whittam TS. 2010. Ancestral lineages of human enterotoxigenic *Escherichia coli*. *J Clin Microbiol* 48:2916–2924. <https://doi.org/10.1128/JCM.02432-09>
  62. Johnson TJ, Nolan LK. 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol Mol Biol Rev* 73:750–774. <https://doi.org/10.1128/MMBR.00015-09>
  63. Nguyen CC, Hugie CN, Kile ML, Navab-Daneshmand T. 2019. Association between heavy metals and antibiotic-resistant human pathogens in environmental reservoirs: a review. *Front Environ Sci Eng* 13:46. <https://doi.org/10.1007/s11783-019-1129-0>
  64. Georjon H, Bernheim A. 2023. The highly diverse antiphage defence systems of bacteria. *Nat Rev Microbiol* 21:686–700. <https://doi.org/10.1038/s41579-023-00934-x>
  65. Puigbò P, Makarova KS, Kristensen DM, Wolf YI, Koonin EV. 2017. Reconstruction of the evolution of microbial defense systems. *BMC Evol Biol* 17:94. <https://doi.org/10.1186/s12862-017-0942-y>
  66. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A. 2022. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Commun* 13:2561. <https://doi.org/10.1038/s41467-022-30269-9>
  67. Tock MR, Dryden DTF. 2005. The biology of restriction and anti-restriction. *Curr Opin Microbiol* 8:466–472. <https://doi.org/10.1016/j.mib.2005.06.003>
  68. Wu Y, Garushyants SK, van den Hurk A, Aparicio-Maldonado C, Kushwaha SK, King CM, Ou Y, Todeschini TC, Clokie MRJ, Millard AD, Gençay YE, Koonin EV, Nobrega FL. 2024. Bacterial defense systems exhibit synergistic anti-phage activity. *Cell Host Microbe* 32:557–572. <https://doi.org/10.1016/j.chom.2024.01.015>
  69. Arias CF, Acosta FJ, Bertocchini F, Herrero MA, Fernández-Arias C. 2022. The coordination of anti-phage immunity mechanisms in bacterial cells. *Nat Commun* 13:7412. <https://doi.org/10.1038/s41467-022-35203-7>
  70. Piel D, Bruto M, Labreuche Y, Blanquart F, Goudenège D, Barcia-Cruz R, Chenivresse S, Le Panse S, James A, Dubert J, Petton B, Lieberman E, Wegner KM, Hussain FA, Kauffman KM, Polz MF, Bikard D, Gandon S, Rocha EPC, Le Roux F. 2022. Phage–host coevolution in natural populations. *Nat Microbiol* 7:1075–1086. <https://doi.org/10.1038/s41564-022-01157-1>
  71. Flores CO, Valverde S, Weitz JS. 2013. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *ISME J* 7:520–532. <https://doi.org/10.1038/ismej.2012.135>
  72. Hughes JA, Brown LR, Ferro AJ. 1987. Expression of the cloned coliphage T3 S-adenosylmethionine hydrolase gene inhibits DNA methylation and polyamine biosynthesis in *Escherichia coli*. *J Bacteriol* 169:3625–3632. <https://doi.org/10.1128/jb.169.8.3625-3632.1987>
  73. German G, Misra R, Kropinski A. M. The T1-like bacteriophages. In *The bacteriophages*. Oxford University Press.
  74. Auer B, Schweiger M. 1984. Evidence that *Escherichia coli* virus T1 induces a DNA methyltransferase. *J Virol* 49:588–590. <https://doi.org/10.1128/JVI.49.2.588-590.1984>
  75. Hu B, Margolin W, Molineux IJ, Liu J. 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* 339:576–579. <https://doi.org/10.1126/science.1231887>
  76. Qimron U, Marintcheva B, Tabor S, Richardson CC. 2006. Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc Natl Acad Sci U S A* 103:19039–19044. <https://doi.org/10.1073/pnas.0609428103>
  77. Zhao X, Cui Y, Yan Y, Du Z, Tan Y, Yang H, Bi Y, Zhang P, Zhou L, Zhou D, Han Y, Song Y, Wang X, Yang R. 2013. Outer membrane proteins Ail and OmpF of *Yersinia pestis* are involved in the adsorption of T7-related bacteriophage Yep-phi. *J Virol* 87:12260–12269. <https://doi.org/10.1128/JVI.01948-13>
  78. Parent KN, Erb ML, Cardone G, Nguyen K, Gilcrease EB, Porcsek NB, Pogliano J, Baker TS, Casjens SR. 2014. OmpA and OmpC are critical host factors for bacteriophage Sf6 entry in *Shigella*. *Mol Microbiol* 92:47–60. <https://doi.org/10.1111/mmi.12536>
  79. Goulet A, Spinelli S, Mahony J, Cambillau C. 2020. Conserved and diverse traits of adhesion devices from siphoviridae recognizing proteinaceous or saccharidic receptors. *Viruses* 12:1–21. <https://doi.org/10.3390/v12050512>
  80. Weber NR, Nielsen JP, Hjulsager CK, Jorsal SE, Haugegaard S, Hansen CF, Pedersen KS. 2017. Comparison of bacterial culture and qPCR testing of rectal and pen floor samples as diagnostic approaches to detect enterotoxigenic *Escherichia coli* in nursery pigs. *Prev Vet Med* 143:61–67. <https://doi.org/10.1016/j.prevetmed.2017.05.009>
  81. Casey TA, Bosworth BT. 2009. Design and evaluation of a multiplex polymerase chain reaction assay for the simultaneous identification of genes for nine different virulence factors associated with *Escherichia coli* that cause diarrhea and edema disease in swine. *J Vet Diagn Invest* 21:25–30. <https://doi.org/10.1177/104063870902100104>
  82. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7:1–6. <https://doi.org/10.1093/gigascience/gix120>
  83. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
  84. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S, Hsu CH, McDermott PF, Tadesse DA, Morales C, Simmons M, Tillman G, Wasilenko J, Folster JP, Klimke W. 2019. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* 63:1–19. <https://doi.org/10.1128/AAC.00483-19>
  85. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, Nobrega FL, Jackson SA. 2021. Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Res* 49:10868–10878. <https://doi.org/10.1093/nar/gkab883>
  86. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. 2014. MacSy-Finder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* 9:e110726. <https://doi.org/10.1371/journal.pone.0110726>
  87. Schwengers O, Barth P, Falgenhauer L, Hain T, Chakraborty T, Goesmann A. 2020. Platon: identification and characterization of

- bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. *Microbial Genomics* 6:1–12. <https://doi.org/10.1099/mgen.0.000398>
88. Starikova EV, Tikhonova PO, Prianichnikov NA, Rands CM, Zdobnov EM, Ilna EN, Govorun VM. 2020. Phigaro: high-throughput prophage sequence annotation. *Bioinformatics* 36:3882–3884. <https://doi.org/10.1093/bioinformatics/btaa250>
  89. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. 2018. Easily phylotyping *E. coli* via the EzClermont web app and command-line tool. *bioRxiv*. <https://doi.org/10.1101/317610>
  90. Patel IR, Gangiredla J, Mammel MK, Lampel KA, Elkins CA, Lacher DW. 2018. Draft genome sequences of the *Escherichia coli* reference (ECOR) collection. *Microbiol Resour Announc* 7:e01133-18. <https://doi.org/10.1128/MRA.01133-18>
  91. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. 2019. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 29:304–316. <https://doi.org/10.1101/gr.241455.118>
  92. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, Spratt BG, Aanensen DM. 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2:e000093. <https://doi.org/10.1099/mgen.0.000093>
  93. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>
  94. Carlson K. 2005. Working with bacteriophages: common techniques and methodological approaches. In *Bacteriophage biology and applications*, 1st ed. CRC Press, Boca Raton, FL, USA.
  95. Sambrook R. *Molecular cloning*. 3rd ed. Cold Spring Harbour Laboratory Press, New York.
  96. Sampaio M, Rocha M, Oliveira H, Dias O. 2019. Predicting promoters in phage genomes using PhagePromoter. *Bioinformatics* 35:5301–5302. <https://doi.org/10.1093/bioinformatics/btz580>
  97. Kingsford CL, Ayanbule K, Salzberg SL. 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 8:R22. <https://doi.org/10.1186/gb-2007-8-2-r22>
  98. Kongari R, Rajaure M, Cahill J, Rasche E, Mijalis E, Berry J, Young R. 2018. Phage spanins: diversity, topological dynamics and gene convergence. *BMC Bioinformatics* 19:326. <https://doi.org/10.1186/s12859-018-2342-8>
  99. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>
  100. Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Bioinformatics* 28:1102–1104. <https://doi.org/10.2144/00286ir01>
  101. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430:2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007>
  102. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303. <https://doi.org/10.1093/nar/gky427>
  103. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–858. <https://doi.org/10.1038/nprot.2015.053>
  104. Moraru C, Varsani A, Kropinski AM. 2020. VIRIDIC—a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses* 12:1–10. <https://doi.org/10.3390/v12111268>
  105. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol* 37:632–639. <https://doi.org/10.1038/s41587-019-0100-8>
  106. Gilchrist CLM, Chooi Y-H. 2021. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 37:2473–2475. <https://doi.org/10.1093/bioinformatics/btab007>
  107. Ackermann HW. 2009. Basic Phage Electron Microscopy. In *Bacteriophages*. *Methods in molecular biology*
  108. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
  109. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. 2022. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 50:D439–D444. <https://doi.org/10.1093/nar/gkab1061>
  110. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. 2024. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 42:243–246. <https://doi.org/10.1038/s41587-023-01773-0>
  111. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci* 30:70–82. <https://doi.org/10.1002/pro.3943>
  112. Krapp LF, Abriata LA, Cortés Rodríguez F, Dal Peraro M. 2023. PeSto: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nat Commun* 14:2175. <https://doi.org/10.1038/s41467-023-37701-8>
  113. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. 2018. PDBsum: structural summaries of PDB entries. *Protein Sci* 27:129–134. <https://doi.org/10.1002/pro.3289>