



Research article

Complete genome sequence of the novel virulent phage PMBT24 infecting *Enterocloster bolteae* from the human gut

Sabrina Sprotte^{*}, Erik Brinks, Horst Neve¹, Charles M.A.P. Franz

Department of Microbiology and Biotechnology, Max Rubner-Institut, Federal Research Institute of Nutrition and Food, Hermann-Weigmann-Str. 1, 24103, Kiel, Germany

ARTICLE INFO

Keywords:

Novel phage PMBT24
Enterocloster bolteae
Complete genome
Human gut

ABSTRACT

PMBT24, the first reported virulent bacteriophage infecting the anaerobic human gut bacterium *Enterocloster bolteae* strain MBT-21, was isolated from a municipal sewage sample and its genome was sequenced and analysed. Transmission electron microscopy revealed a phage with an icosahedral head and a long, non-contractile tail. The circularly permuted, 99,962-bp dsDNA genome of the *pac*-type phage has a mol% G + C content of 32.1 and comprises 173 putative ORFs. Using amino acid sequence-based phylogeny, phage PMBT24 showed similarity to other, hitherto non-published phage genomes in the databases. Our data suggested phage PMBT24 to present the type phage of a novel genus (proposed name Kielvirus) and novel family of phages (proposed name Kielviridae).

1. Introduction

Enterocloster bolteae [1] (formerly *Clostridium bolteae* [1,2]) is an opportunistic pathogen that occurs in the human gut. This species is obligately anaerobic and capable of forming spores [2]. It was found to be present with increased abundance in the intestines of autistic children [3–6] and was also associated with bacteremia, intra-abdominal infections and abscesses [5]. Frequently described antibiotic resistances of *E. bolteae* strains include ampicillin, erythromycin, lincomycin, ciprofloxacin, doxycycline and colistin [3]. Because of these multidrug resistance phenotypes, a supportive or alternative treatment for infections with *E. bolteae* besides treatment with antibiotics would be a desirable development. Bacteriophages (phages) are among the most promising alternative tools to combat bacterial infections. They are gaining increasing recognition as potential modulators of the gut ecosystem due to their ability to affect bacterial communities [7,8], e.g. by fecal microbial transplantation, as was previously investigated for *Clostridium difficile* [9,10]. So far, only six complete genome sequences of phages infecting *E. bolteae* have been deposited in the NCBI genome database. In this study, a novel virulent phage PMBT24 infecting *E. bolteae* strain MBT-21 was characterized and its genome was analysed to determine the phage's potential for use as a possible biotherapy tool for controlling infections with this opportunistic pathogen.

2. Materials and methods

Phage isolation under anaerobic conditions were performed out in a Whitley A45 anaerobic workstation (Meintrup DWS

^{*} Corresponding author.

E-mail address: sabrina.sprotte@mri.bund.de (S. Sprotte).

¹ retired from MRI.

Laborgeräte GmbH, Herzlake, Germany) maintained with an anaerobic gas mixture (ANO2: 10% hydrogen, 10% carbon dioxide and 80% nitrogen) was performed as described previously, using dehydrated Anaerobe Basal Broth (ABB; Thermo Fisher Scientific) and sewage samples from a municipal sewage treatment plant near to Kiel, Germany [11]. The host bacterium *E. bolteae* MBT-21 was isolated from human feces and cultivated in ABB under anaerobic conditions. The phage lysate was concentrated from the supernatant of 1 L of a phage PMBT24 infected *E. bolteae* MBT-21 culture using cesium chloride (CsCl) density gradient ultracentrifugation [12] and was subsequently analysed by transmission electron microscopy (TEM), as described elsewhere [11].

For genome sequencing, the bacterial DNA was prepared using the Quick-DNA™ Fungal/Bacterial MiniPrep Kit (ZymoResearch, Freiburg, Germany) and the genome was sequenced using the MiSeq Reagent Nano Kit v2 (Illumina, Munich, Germany) according to the manufacturers' instructions and sequencing was performed on a MiSeq sequencer (Illumina, Munich, Germany). The MiSeq® Reporter software was used for base calling, demultiplexing and adapter trimming directly on the MiSeq sequencer. The reads were *de novo* assembled using SPAdes version 3.10.0 (Bacterial and Viral Bioinformatics Resource Center, BV-BRC) [13]. Automatic annotation of the bacterial genome was performed with RASTtk (BV-BRC) [14,15] and the bacterium was identified by BLASTN comparison to bacterial genome sequences in the GenBank database.

Phage DNA was isolated from high titer phage lysate obtained from CsCl density gradient using the same kit with some previously described modifications [11]. For phage genome sequencing, the TruSeq DNA Library Preparation Kit was used according to the manufacturers' instructions. Sequencing and quality control of the phage genome was performed as described before. The reads were *de novo* assembled using Unicycler v0.4.8 (BV-BRC) [15]. Annotation of the phage genome was performed automatically with VIGOR4 in bacteriophage mode (BV-BRC) [14,15], followed by manual comparison of predicted proteins with proteins in the databases using BLASTP [16] and HHPred [17]. BLASTP results were only accepted with E-values less than 1×10^{-5} , with a query coverage of $\geq 60\%$ and an identity of $\geq 35\%$. Using HHPred, proteins were compared to the PDB, UniProt, Pfam and SMART databases (accessed on 15 March 2023). Results were only accepted with E-values less than 1×10^{-3} and a minimum probability in the hitlist of 20%. Putative host and phage promoters in intergenic regions were identified using PhagePromoter [18] and putative rho-independent terminators occurring in intergenic regions were determined using ARNold [19], which computes the free energy of the predicted terminator stem-loops. For identification of phage termini and analysis of the DNA packaging mechanism, PhageTerm 1.0.12 [20] was used. The phage genome was analysed for the presence of tRNA genes with tRNAscan-SE version 2.0 [21,22]. The PHACTS software (Galaxy Version 0.3) [23] was applied to predict the phage lifestyle (i.e., whether lysogenic or lytic). ResFinder 4.1 [20,24–27] was used for the prediction of antibiotic resistance genes and VirulenceFinder 2.0 [28,29] for the prediction of bacterial virulence genes on the phage genome. The phage genome map was generated using Geneious version 9.1.8.

A comparison of the phage PMBT24 genome to other phage genomes deposited in the databases of the National Center for Biotechnology Information (NCBI) and the European Nucleotide Archive (ENA) was performed at the nucleotide level with megablast on the BLASTN platform [16] (accessed on 2 May 2023). The resulting identity and coverage values were used to calculate the sequence similarity according to Turner et al. [30]. The identity between phage PMBT24 and CB457P1 (acc. no. OP172753.1) at the amino acid level was plotted according to tBLASTx using the genome comparison visualiser Easy Fig 2.2.3 [31]. In addition, the phage PMBT24 tape measure protein was compared to those of the closest related phages CB457P1 and cteZU1 (acc. no. BK021559.1) using BLASTP. The Virus Classification and Tree Building Online Resource (VICTOR) was used for a whole amino acid sequence-based phylogeny and classification. In this analysis all pairwise comparisons of the amino acid sequences of phage PMBT24 and of the 22 closest related phages from the megablast analysis were conducted using the Genome-BLAST Distance Phylogeny (GBDP) method [32], with settings recommended for prokaryotic viruses [33]. The evolution tree was reconstructed with the formula D6, rooted at the midpoint [34] and visualised with gtree [35]. Taxon boundaries at the species, genus and family level were estimated automatically by VICTOR [33,36,37]. A list of the analysed phages is shown in Supplementary Table S1. In addition, VIRIDIC (Virus Intergenomic Distance Calculator) [38] was used to calculate the pairwise intergenomic similarities amongst these phages. It uses the traditional algorithm used by the International Committee on Taxonomy of Viruses (ICTV), Bacterial and Archaeal Viruses Subcommittee for

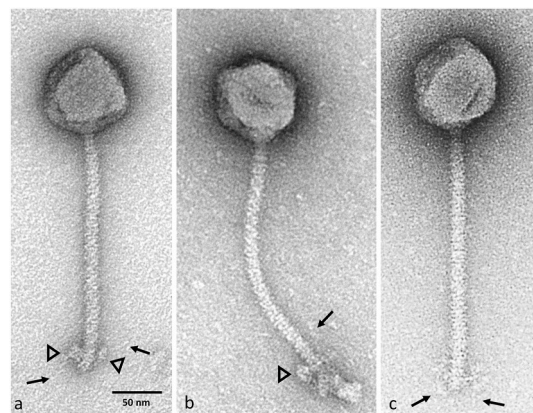


Fig. 1. Electron micrographs of *E. bolteae* phage PMBT24 stained with 1% (w/v) uranyl acetate. The PMBT24 tail is coated by thin spiral structures. Globular structures at the distal tail end are indicated by the triangles (a–b). The arrows indicate thin tail fibers visible at the tail tip (a–c).

calculation. All phage genomes were also used as queries to build a whole-genome tree with the Viral Proteomic Tree server (ViPTree) version 3.6 [39]. ViPTree calculates the sequence similarities of the whole genomes using tBLASTx. Furthermore, the amino acid sequence of the putative terminase large subunit from phage PMBT24 was compared in coverage and identity to the most closely related proteins using BLASTP.

3. Results

Electron micrographs of the CsCl gradient purified phage PMBT24 revealed a siphovirus morphotype with an icosahedral head of 84 ± 1.3 nm ($n = 7$) in diameter and a 238 ± 6.5 nm ($n = 7$) long, flexible and non-contractile tail (Fig. 1). Globular appendages were visible at the distal tail ends and thin tail fibers at the tail tip. Regarding plaque formation, phage PMBT24 formed small turbid plaques with a diameter of 0.5–1 mm on an *E. bolteae* MBT-21 lawn (data not shown).

All raw reads from the host bacterium *E. bolteae* MBT-21 were deposited in the NCBI SRA with accession number PRJNA1078606. Phage genome sequencing generated 258,014 paired-end reads (2×251 bp). The mean read length averaged 235 bp. The reads were *de novo* assembled into a single contig with a total length of 99,962 bp and a mol% G + C content of 32.1. The assembled genome showed a high 416-fold coverage. Amongst the predicted 173 open reading frames (ORFs), 48 ORFs (28%) could be assigned a putative function, while 125 ORFs (72%) encoding hypothetical proteins were unclassified (Supplementary Table S2). The ORFs for which a putative function could be assigned were categorized into five functional groups (Fig. 2): *i*) DNA packaging: large terminase subunit (ORF36), portal protein (ORF37); *ii*) structural proteins: major capsid protein (ORF43), major tail protein (ORF50), tape measure protein (ORF60), baseplate hub proteins (ORF64, ORF66); *iii*) DNA replication/modification and transcription regulation: dUTPase (ORF83), ribonucleotide reductase (anaerobic) (ORF87, ORF88), exodeoxyribonuclease V (ORF92), RNA polymerase sigma factor SigK (ORF93), DNA polymerase III alpha subunit (ORF95), HNH homing endonuclease (ORF97), single-stranded-DNA-specific exonuclease (ORF98), DNA primase (ORF106), host-nuclease inhibitor protein Gam (ORF110), DNAB-like replicative helicase (ORF112), RecA (ORF116), site-specific recombinase XerD (ORF122), methyltransferase Dmt (ORF126), restriction alleviation protein Lar (ORF130); *iv*) host cell lysis: holin (ORF55), lysin (ORF77) and *v*) other functions: e.g., lipoprotein (ORF21), lambda protein ninF (ORF22) and ferritin (ORF39). 97 ORFs were located on the positive and 76 ORFs were located on the negative strand. Putative host and phage promoters and putative *rho*-independent terminators occurring in intergenic regions were identified (both Supplementary Tables S3 and S4). According to the PhageTerm Method, phage PMBT24 contains a circularly permuted genome with redundant ends used to circularize the genome. As a significant peak with $\tau > 0.1$ was found only on the reverse strand, the phage was considered a *pac*-type phage, which uses the headful DNA packaging mechanism initiated at the *pac* site. The *pac* site was selected as the start of the phage PMBT24 genome map (Fig. 2). No tRNAs could be identified in the phage genome. Phage lifestyle was predicted to be strictly lytic. Further analysis revealed the absence of antibiotic resistance and bacterial virulence genes in the genome of phage PMBT24.

With a coverage of 77% and an identity of 96.55%, the assembled genome showed closest similarity (i.e., 74.34% calculated by % identity multiplied by % coverage according to Turner et al. [30]) to the genome of *E. bolteae* phage CB457P1, which has a slightly smaller size of 96,610 bp. Furthermore, the phage PMBT24 genome showed 67.41% similarity to the partial 72,822 bp-genome fragment of a *Caudoviricetes* sp. isolate cteZU1 from a metagenome assembled genome [40]. Genome comparison at the amino acid level of phages PMBT24 and CB457P1 showed a similar genome organisation (Fig. 3a). More precisely, the genomes showed a higher identity in the 5' end where the packaging proteins and the structural protein genes are located than in the middle region and 3' end where genes for replication, modification and transcription regulation proteins are located. The predicted structural proteins, i.e. the

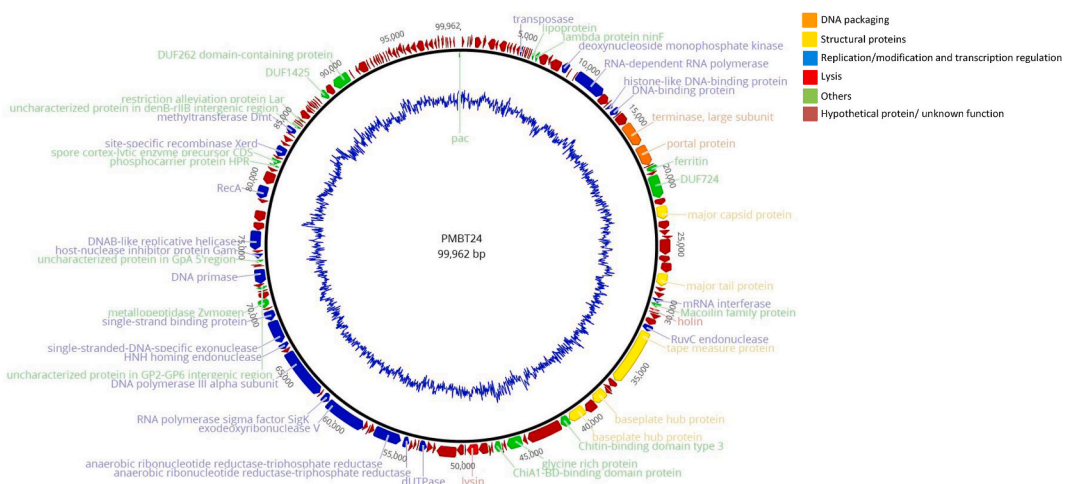


Fig. 2. Circular map of the phage PMBT24 genome with structural and functional annotations and mol% GC content (inner circle) created by Geneious. Arrows indicate the predicted ORFs, which are labeled and color-coded based on their predicted function (see colour legend for details). The *pac*-site is shown in green at the top of the map. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

major capsid protein, major tail protein and baseplate hub protein I and II showed 90–95% identity between the two phages. In contrast, the tape measure proteins were of different organization. Thus, homologous sequences of the PMBT24 tape measure protein's (1508 aa) C- and N-termini were found in two different ORFs (ORF33 with 388 aa and ORF35 with 672 aa) on the genome of phage CB457P1, and these are separated by ORF34 (157 aa) which shows no similarity to any sequence found in PMBT24. A 1511 aa protein with 78.40% sequence identity to the tape measure protein of phage PMBT24 was furthermore identified in the closely related phage cteZU1 using BLASTP (data not shown).

Analysis of phage PMBT24 and its 22 closest related phages from the megablast analysis using VICTOR suggested that phage PMBT24 represents a new species in a yet unassigned genus together with 12 other phages (Fig. 3b). The VIRIDIC heatmap showed 80.4% and 76.2% intergenomic similarity between phage PMBT24 and cteZU1 (partial genome) and CB457P1, respectively, and PMBT24 belongs to the same unassigned genus as phage cteZU1. In contrast to the VICTOR analysis, phage PMBT24 and cteZU1 were assigned to another genus cluster as the remaining 21 phages (Fig. 4). However, the analysis with VIRIDIC revealed that two phages (CB457P2 and CB473P2) within this group belong to the same species, while the analysis with VICTOR revealed four phages (CB473P1, CB473P2, CB473P3, CB457P2). In a ViPTree, which only showed related phage genomes with sequence similarities $S_G > 0.02$, phage PMBT24 and 12 of the 22 analysed phages were connected by deep branches between 0.1 and 0.5 (Fig. 5). They formed a distinct cluster near the *Herelleviridae* phage family, which includes *Bacillus*, *Lactobacillus*, *Staphylococcus* and *Listeria* phages. The remaining 10 analysed phages clustered elsewhere. A comparison of the predicted terminase large subunit protein of PMBT24 with that from the 10 closest related *Enterocloster* phages showed that these proteins of phage PMBT24 and cteZU1 were identical (100% coverage and identity), while the corresponding proteins of PMBT24 and CB457P1 and the remaining proteins showed only low identity (data not shown).

4. Discussion

So far, only six complete genome sequences of *E. bolteae* phages (acc. no. OP172753, OP172756, OP172754, OP172755, OP172757, OP074454) and several partial genomes stemming from metagenomic projects showing similarity to *Enterocloster* phages have been deposited in the NCBI genome database (accessed on 27 November 2023). However, detailed characterizations of these phages have not been published so far. The bacterium *E. bolteae* has been linked to autism, bacteremia, intra-abdominal infections and abscesses and some strains also exhibit antibiotic resistance. In this study, phage PMBT24 has been isolated under anaerobic conditions from sewage and was shown to infect the *E. bolteae* MBT-21 isolate from the human gut. The phage displayed a siphovirus morphotype with an icosahedral head and a long, non-contractile tail. Globular appendages are visible at the distal tail ends, which are also known to occur in other tailed phages [41]. Notably, the whole tail surface is covered by thin spiral structures originally described for

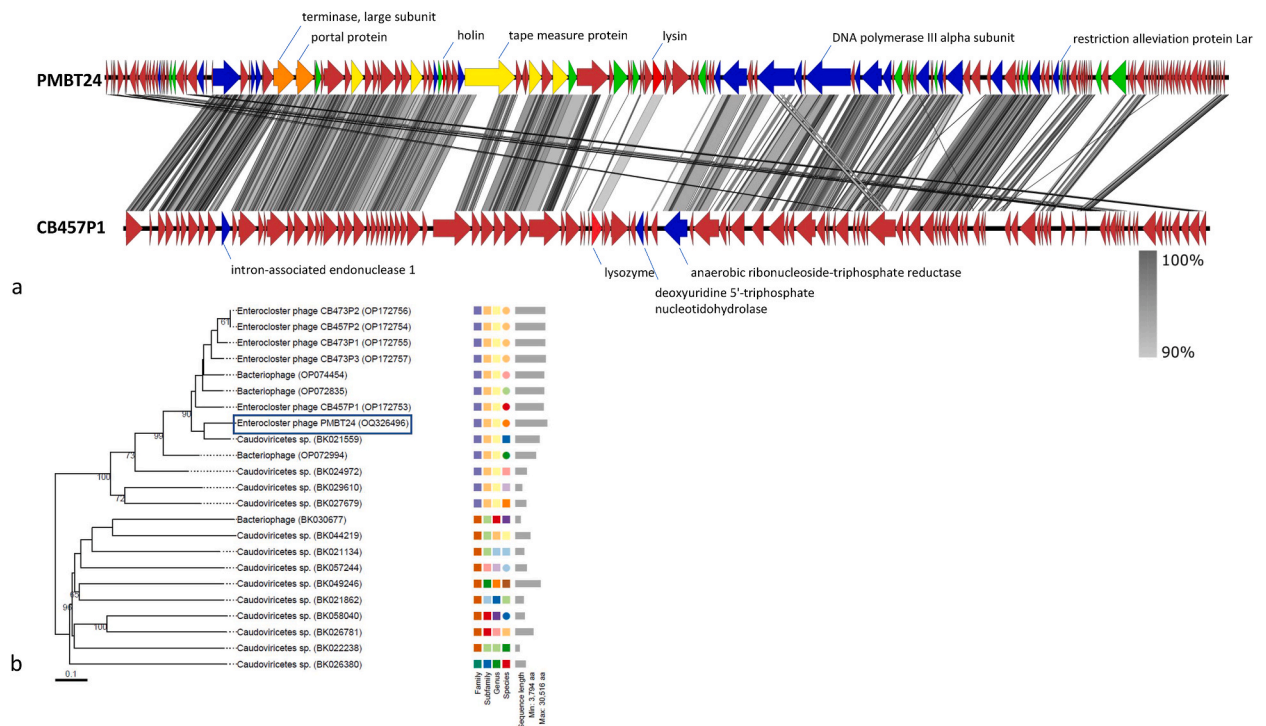


Fig. 3. a. Alignment of phages PMBT24 and CB457P1 genomes at the amino acid level created by Easyfig. b. Phylogenomic analysis of phage PMBT24 with the 22 closest related phages from the NCBI database at the amino acid level using VICTOR. The tree is based on the recommended formula D6 and shows GBDP pseudo-bootstrap support values from 100 replications above branches.

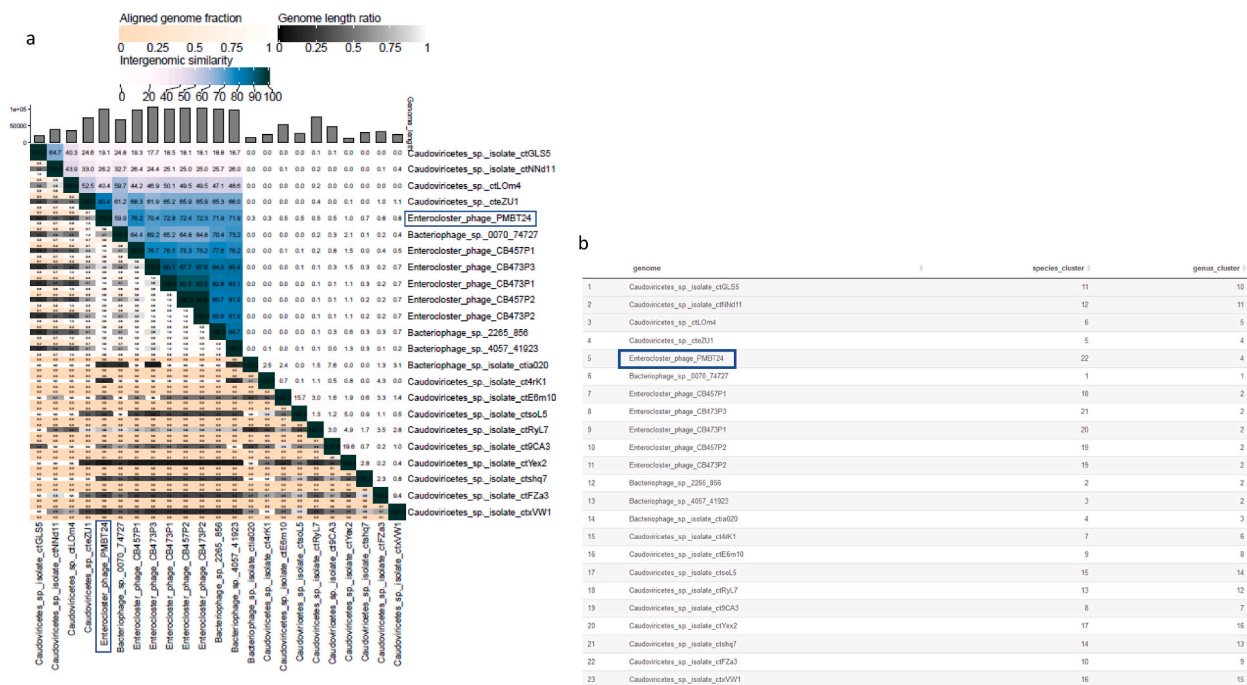


Fig. 4. VIRIDIC heatmap of phage PMBT24 and its 22 closest related phages (a) and corresponding species and genus cluster (b).

Lactococcus lactis Skunavirus phages [42,43].

Using the formula according to Turner et al. [30], who proposed that a phage represents the first isolate of a new species in an undefined or existing genus if its genome exhibits $<95\%$ but $\geq 70\%$ sequence similarity (calculated by % identity multiplied by % coverage) at the nucleotide level over the full genome length to other phages in the databases, phage PMBT24 showed sequence similarity under the 95%-threshold to the closest related phages CB457P1 and cteZU1, suggesting that phage PMBT24 represents a new phage species. However, a comparison with phage cteZU1 was difficult, because only the partial genome is available for this phage. Therefore, the genome of phage PMBT24 was compared to the complete genome of the next closely related phage CB457P1 at the amino acid level using EasyFig. Interestingly, despite high similarity between the structural proteins, i.e. major capsid protein, major tail protein and baseplate hub protein I and II, there is only low identity between the tape measure protein of phage PMBT24 and the corresponding protein(s) in the genome of phage CB457P1. The tail length of a phage is proportional to the length of the tape measure protein corresponding gene [44]. This could be an indication that the two phages differ in tail length.

Comparative analysis from VICTOR and VIRIDIC revealed that phage PMBT24 belongs to an unassigned genus together with phage cteZU1 (according to VIRIDIC) or 11 further closely related phages besides phage cteZU1 (according to VICTOR). However, the analysis revealed that at least two of these phages belong to the same species. As VIRIDIC builds on and improves the traditional BLASTN method used by Bacterial and Archaeal Viruses Subcommittee from ICTV to calculate and visualize virus intergenomic relatedness, these results should be more reliable [38]. In addition, a comparative analysis of the predicted terminase large subunit protein could confirm the hypothesis that phage PMBT24 and cteZU1 belong to the same unassigned genus, as the proteins showed 100% coverage and 99.84% aa identity (data not shown). All other terminase large subunit proteins used for the analysis showed no significant similarity. The terminase large subunit is considered the most universally conserved gene sequence in phages and therefore is often used to construct phylogeny [45,46].

Finally, the deep branch lengths compared to related phages observed in the ViPTree suggested that the same 13 phages including phage PMBT24 furthermore represent a new unassigned family near the *Herelleviridae* phage family. However, phylogenetic analysis of the phages was difficult because the most deposited genome sequences were partial genomes. To formally classify these phages, we propose to create the new bacteriophage family Kielviridae which includes the novel Kielvirus genus with PMBT24 as the type phage. Phage PMBT24 has the potential to be used as biotherapy tool for controlling infections with the opportunistic pathogen *E. bolteae*, although physiological parameters, e.g. latent period, burst size, pH and temperature stability as well as host range, still need to be investigated in a further study. However, some of these parameters are currently challenging to investigate, because of the need for culturing of the *E. bolteae* host bacterium under anaerobic conditions. Our study presents new data on phage genome sequence of the first described *E. bolteae* phage and increases current knowledge on *Enterocloster* phages, which may aid in future investigations on the phylogeny of these phages. To our knowledge, this is the first report of a virulent bacteriophage infecting *E. bolteae*.

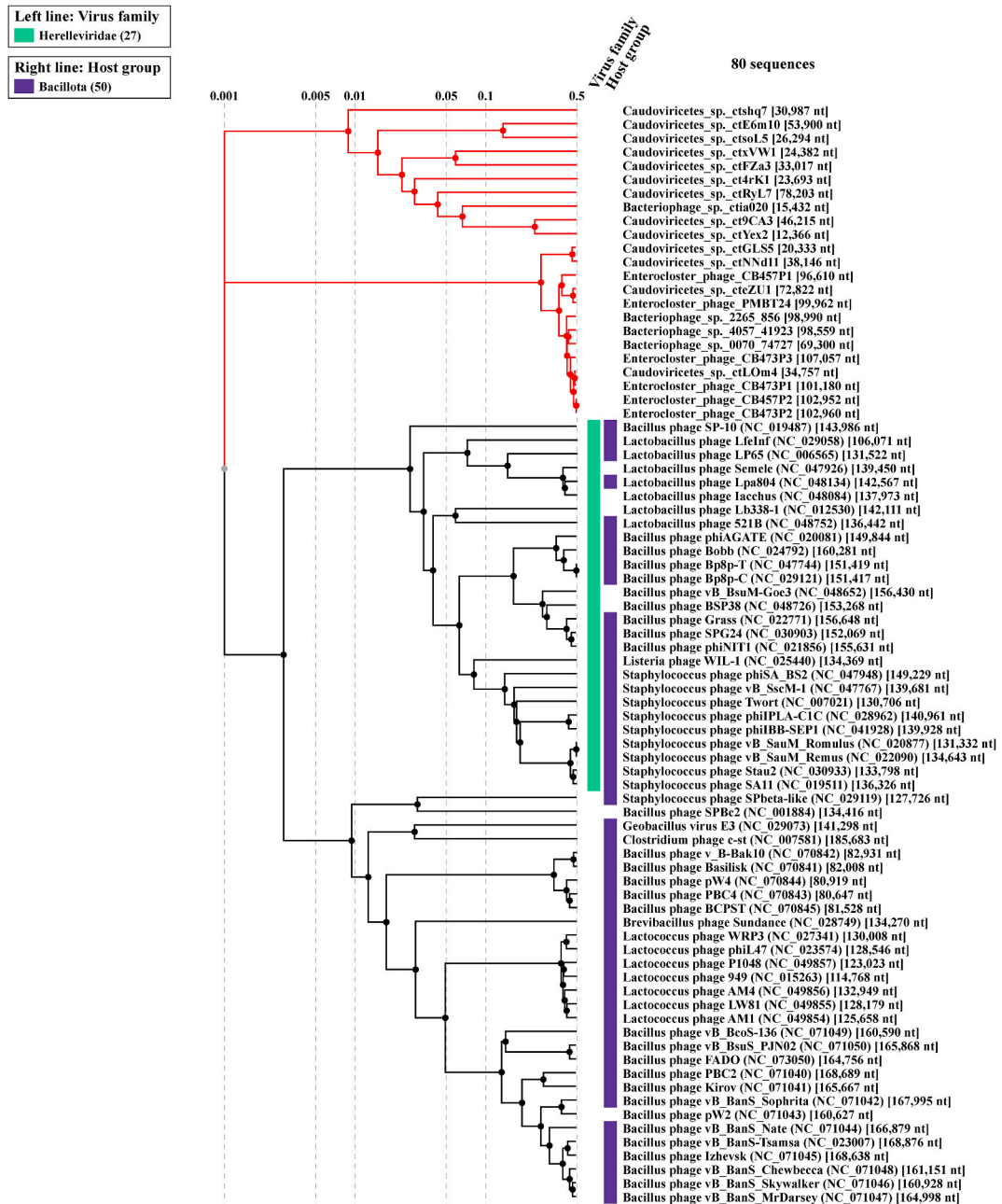


Fig. 5. ViPTree showing phylogenetic relationship of phage PMBT24 and its 22 closest related phages used as queries (marked with asterisks). The rectangular tree only shows related genomes (sequence similarities SG > 0.02).

Ethics statement

This study was approved by the Ethics Committee of the Faculty of Medicine, Kiel University, Germany with the ethics approval reference D420/20 on February 7, 2020. Written informed consent was obtained from all the participants.

Data availability statement

The complete genome sequence of phage PMBT24 generated in this project was deposited in the NCBI database under the GenBank accession number OQ326496.2. All raw reads from the host bacterium *E. bolteae* MBT-21 were deposited in the NCBI SRA with accession number PRJNA1078606.

CRediT authorship contribution statement

Sabrina Sprotte: Writing – original draft. **Erik Brinks:** Writing – review & editing. **Horst Neve:** Writing – review & editing. **Charles M.A.P. Franz:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We kindly thank Nadja Timm, Lena Westphal, Angela Back, Gesa Gehrke and Adrian Prager for technical assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e28813>.

References

- [1] K.N. Haas, J.L. Blanchard, Reclassification of the *Clostridium clostridioforme* and *Clostridium sphenoides* clades as *Enterocloster* gen. nov. and *Lacrimispora* gen. nov., including reclassification of 15 taxa, *Int. J. Syst. Evol. Microbiol.* 70 (2020) 23–34, <https://doi.org/10.1099/ijsem.0.003698>.
- [2] Y. Song, C. Liu, D.R. Molitoris, T.J. Tomzynski, P.A. Lawson, M.D. Collins, S.M. Finegold, *Clostridium bolteae* sp. nov., isolated from human sources, *Syst. Appl. Microbiol.* 26 (2003) 84–89, <https://doi.org/10.1078/072320203322337353>.
- [3] P. Dehoux, J.C. Marvaud, A. Abouelleil, A.M. Earl, T. Lambert, C. Dauga, Comparative genomics of *Clostridium bolteae* and *Clostridium clostridioforme* reveals species-specific genomic properties and numerous putative antibiotic resistance determinants, *BMC Genom.* 17 (2016) 819, <https://doi.org/10.1186/s12864-016-3152-x>.
- [4] B. Pequegnat, M. Sagermann, M. Valliani, M. Toh, H. Chow, E. Allen-Vercoe, M.A. Monteiro, A vaccine and diagnostic target for *Clostridium bolteae*, an autism-associated bacterium, *Vaccine* 31 (2013) 2787–2790, <https://doi.org/10.1016/j.vaccine.2013.04.018>.
- [5] O.M. Williams, J. Brazier, V. Peraino, E.J.C. Goldstein, A review of three cases of *Clostridium aldenense* bacteremia, *Anaerobe* 16 (2010) 475–477, <https://doi.org/10.1016/j.anaerobe.2010.08.004>.
- [6] S.M. Finegold, D. Molitoris, Y. Song, C. Liu, M.-L. Vaisanen, E. Bolte, M. McTeague, R. Sandler, H. Wexler, E.M. Marlowe, et al., Gastrointestinal microflora studies in late-onset autism, *Clin. Infect. Dis.* 35 (2002) S6–S16, <https://doi.org/10.1086/341914>.
- [7] M. Zuppi, H.L. Hendrickson, J.M. O'Sullivan, T. Vatanen, Phages in the gut ecosystem, *Front. Cell. Infect. Microbiol.* 11 (2021) 822562, <https://doi.org/10.3389/fcimb.2021.822562>.
- [8] T.S. Rasmussen, A.K. Koefoed, R.R. Jakobsen, L. Deng, J.L. Castro-Mejía, A. Brunse, H. Neve, F.K. Vogensen, D.S. Nielsen, Bacteriophage-mediated manipulation of the gut microbiome - promises and presents limitations, *FEMS Microbiol. Rev.* 44 (2020) 507–521, <https://doi.org/10.1093/femsre/fuaa020>.
- [9] J.Y. Nale, T.A. Redgwell, A. Millard, M.R.J. Clokie, Efficacy of an optimised bacteriophage cocktail to clear *Clostridium difficile* in a batch fermentation model, *Antibiotics (Basel)* 7 (2018), <https://doi.org/10.3390/antibiotics7010013>.
- [10] L.A. Draper, F.J. Ryan, M.K. Smith, J. Jalanka, E. Mattila, P.A. Arkkila, R.P. Ross, R. Satokari, C. Hill, Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation, *Microbiome* 6 (2018) 220, <https://doi.org/10.1186/s40168-018-0598-x>.
- [11] S. Sprotte, T.S. Rasmussen, G.-S. Cho, E. Brinks, R. Lametsch, H. Neve, F.K. Vogensen, D.S. Nielsen, C.M.A.P. Franz, Morphological and genetic characterization of *Eggerthella lenta* bacteriophage PMBT5, *Viruses* 14 (2022), <https://doi.org/10.3390/v14081598>.
- [12] J. Sambrook, D.W. Russell, *Molecular Cloning: A Laboratory Manual*, third ed., Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2001. ISBN 0879695773.
- [13] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Pribelski, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477, <https://doi.org/10.1089/cmb.2012.0021>.
- [14] T. Brettin, J.J. Davis, T. Disz, R.A. Edwards, S. Gerdes, G.J. Olsen, R. Olson, R. Overbeek, B. Parrello, G.D. Pusch, et al., RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes, *Sci. Rep.* 5 (2015) 8365, <https://doi.org/10.1038/srep08365>.
- [15] A.R. Wattam, J.J. Davis, R. Assaf, S. Boisvert, T. Brettin, C. Bun, N. Conrad, E.M. Dietrich, T. Disz, J.L. Gabbard, et al., Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center, *Nucleic Acids Res.* 45 (2017) D535–D542, <https://doi.org/10.1093/nar/gkw1017>.
- [16] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410, [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- [17] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A.N. Lupas, V. Alva, A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core, *J. Mol. Biol.* 430 (2018) 2237–2243, <https://doi.org/10.1016/j.jmb.2017.12.007>.
- [18] M. Sampaio, M. Rocha, H. Oliveira, O. Dias, Predicting promoters in phage genomes using PhagePromoter, *Bioinformatics* 35 (2019) 5301–5302, <https://doi.org/10.1093/bioinformatics/btz580>.
- [19] M. Naville, A. Ghuillot-Gaudeffroy, A. Marchais, D. Gautheret, ARNold: a web tool for the prediction of Rho-independent transcription terminators, *RNA Biol.* 8 (2011) 11–13, <https://doi.org/10.4161/rna.8.1.13346>.
- [20] J.R. Garneau, F. Depardieu, L.-C. Fortier, D. Bikard, M. Monot, PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data, *Sci. Rep.* 7 (2017) 8292, <https://doi.org/10.1038/s41598-017-07910-5>.
- [21] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.* 25 (1997) 955–964, <https://doi.org/10.1093/nar/25.5.955>.
- [22] P.P. Chan, B.Y. Lin, A.J. Mak, T.M. Lowe, tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes, *Nucleic Acids Res.* 49 (2021) 9077–9096, <https://doi.org/10.1093/nar/gkab688>.
- [23] K. McNair, B.A. Bailey, R.A. Edwards, PHACTS, a computational approach to classifying the lifestyle of phages, *Bioinformatics* 28 (2012) 614–618, <https://doi.org/10.1093/bioinformatics/bts014>.
- [24] V. Bortolaia, R.S. Kaas, E. Ruppe, M.C. Roberts, S. Schwarz, V. Cattori, A. Philippon, R.L. Allesoe, A.R. Rebelo, A.F. Florensa, et al., ResFinder 4.0 for predictions of phenotypes from genotypes, *J. Antimicrob. Chemother.* 75 (2020) 3491–3500, <https://doi.org/10.1093/jac/dkaa345>.

- [25] E. Zankari, R. Allesøe, K.G. Joensen, L.M. Cavaco, O. Lund, F.M. Aarestrup, PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens, *J. Antimicrob. Chemother.* 72 (2017) 2764–2768, <https://doi.org/10.1093/jac/dkx217>.
- [26] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, *BMC Bioinf.* 10 (2009) 421, <https://doi.org/10.1186/1471-2105-10-421>.
- [27] A.F. Florensa, R.S. Kaas, P.T.L.C. Clausen, D. Aytan-Aktug, F.M. Aarestrup, ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes, *Microb. Genom.* 8 (2022), <https://doi.org/10.1099/mgen.0.000748>.
- [28] A.M. Malberg Tetzschner, J.R. Johnson, B.D. Johnston, O. Lund, F. Scheutz, *In Silico* genotyping of *Escherichia coli* isolates for extraintestinal virulence genes by use of whole-genome sequencing data, *J. Clin. Microbiol.* 58 (2020), <https://doi.org/10.1128/JCM.01269-20>.
- [29] K.G. Joensen, F. Scheutz, O. Lund, H. Hasman, R.S. Kaas, E.M. Nielsen, F.M. Aarestrup, Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*, *J. Clin. Microbiol.* 52 (2014) 1501–1510, <https://doi.org/10.1128/JCM.03617-13>.
- [30] D. Turner, A.M. Kropinski, E.M. Adriaenssens, A roadmap for genome-based phage taxonomy, *Viruses* 13 (2021), <https://doi.org/10.3390/v13030506>.
- [31] M.J. Sullivan, N.K. Petty, S.A. Beatson, Easyfig: a genome comparison visualizer, *Bioinformatics* 27 (2011) 1009–1010, <https://doi.org/10.1093/bioinformatics/btr039>.
- [32] J.P. Meier-Kolthoff, A.F. Auch, H.-P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, *BMC Bioinf.* 14 (2013) 60, <https://doi.org/10.1186/1471-2105-14-60>.
- [33] J.P. Meier-Kolthoff, M. Göker, VICTOR: genome-based phylogeny and classification of prokaryotic viruses, *Bioinformatics* 33 (2017) 3396–3404, <https://doi.org/10.1093/bioinformatics/btx440>.
- [34] J. Farris, Estimating phylogenetic trees from distance matrices, *Am. Nat.* (1972) 645–668.
- [35] G. Yu, Using ggtree to visualize data on tree-like structures, *Curr. Protoc. Bioinformatics* 69 (2020) e96, <https://doi.org/10.1002/cpbi.96>.
- [36] J.P. Meier-Kolthoff, R.L. Hahnke, J. Petersen, C. Scheuner, V. Michael, A. Fiebig, C. Rohde, M. Rohde, B. Fartmann, L.A. Goodwin, et al., Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy, *Stand. Genomic Sci.* 9 (2014) 2, <https://doi.org/10.1186/1944-3277-9-2>.
- [37] M. Göker, G. García-Blázquez, H. Voglmayr, M.T. Tellería, M.P. Martín, Molecular taxonomy of phytopathogenic fungi: a case study in *Peronospora*, *PLoS One* 4 (2009) e6319, <https://doi.org/10.1371/journal.pone.0006319>.
- [38] C. Moraru, A. Varsani, A.M. Kropinski, VIRIDIC-A novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses, *Viruses* 12 (2020), <https://doi.org/10.3390/v12111268>.
- [39] Y. Nishimura, T. Yoshida, M. Kuronishi, H. Uehara, H. Ogata, S. Goto, ViPTree: the viral proteomic tree server, *Bioinformatics* 33 (2017) 2379–2380, <https://doi.org/10.1093/bioinformatics/btx157>.
- [40] M.J. Tisza, C.B. Buck, A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases, *Proc. Natl. Acad. Sci. U. S. A.* 118 (2021), <https://doi.org/10.1073/pnas.2023202118>.
- [41] L. Hanemaaijer, P. Kelleher, H. Neve, C.M.A.P. Franz, P.P. de Waal, N.N.M.E. van Peij, D. van Sinderen, J. Mahony, Biodiversity of phages infecting the dairy bacterium *Streptococcus thermophilus*, *Microorganisms* 9 (2021), <https://doi.org/10.3390/microorganisms9091822>.
- [42] J. Murphy, F. Bottacini, J. Mahony, P. Kelleher, H. Neve, A. Zomer, A. Nauta, D. van Sinderen, Comparative genomics and functional analysis of the 936 group of lactococcal *Siphoviridae* phages, *Sci. Rep.* 6 (2016) 21345, <https://doi.org/10.1038/srep21345>.
- [43] A. Kupczok, H. Neve, K.D. Huang, M.P. Hoepfner, K.J. Heller, C.M.A.P. Franz, T. Dagan, Rates of mutation and recombination in *Siphoviridae* phage genome evolution over three decades, *Mol. Biol. Evol.* 35 (2018) 1147–1159, <https://doi.org/10.1093/molbev/msy027>.
- [44] M. Belcaid, A. Bergeron, G. Poisson, The evolution of the tape measure protein: units, duplications and losses, *BMC Bioinf.* 12 (Suppl 9) (2011) S10, <https://doi.org/10.1186/1471-2105-12-S9-S10>.
- [45] R.K. Lokareddy, C.-F.D. Hou, S.G. Doll, F. Li, R.E. Gillilan, F. Forti, D.S. Horner, F. Briani, G. Cingolani, Terminase subunits from the *Pseudomonas*-phage E217, *J. Mol. Biol.* 434 (2022) 167799, <https://doi.org/10.1016/j.jmb.2022.167799>.
- [46] J.K. Akhwale, M. Rohde, C. Rohde, B. Bunk, C. Spröer, H.I. Boga, H.-P. Klenk, J. Wittmann, Isolation, characterization and analysis of bacteriophages from the haloalkaline lake Elmenteita, Kenya, *PLoS One* 14 (2019) e0215734, <https://doi.org/10.1371/journal.pone.0215734>.