Novel insights into phage biology of the pathogen *Clostridioides difficile* based on the active virome

Miriam A. Schüler, Rolf Daniel, Anja Poehlein*

Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University Göttingen, Germany

*Corresponding author: E-mail: apoehle3@gwdg.de

Abstract

The global pathogen *Clostridioides difficile* is a well-studied organism, and researchers work on unravelling its fundamental virulence mechanisms and biology. Prophages have been demonstrated to influence *C. difficile* toxin expression and contribute to the distribution of advantageous genes. All this underlines the importance of prophages in *C. difficile* virulence. Although several *C. difficile* prophages were sequenced and characterized, investigations on the entire active virome of a strain are still missing. Phages were mainly isolated after mitomycin C-induction, which does not resemble natural stressor for *C. difficile.* We examined active prophages from different *C. difficile* strains after cultivation in the absence of mitomycin C by sequencing and characterization of particle-protected DNA. Phage particles were collected after standard cultivation, or after cultivation in the presence of the secondary bile salt deoxycholate (DCA). DCA is a natural stressor for *C. difficile* and a potential prophage-inducing agent. We also investigated differences in prophage activity between clinical and non-clinical *C. difficile* strains. Our experiments demonstrated that spontaneous prophage release is common in *C. difficile*, and that DCA presence induces prophages. Fourteen different, active phages were identified by this experimental procedure. We could not identify a definitive connection between clinical background and phage activity. However, one phage exhibited distinctively higher activity upon DCA-induction in the clinical strain than in the corresponding non-clinical strain, although the phage is identical in both strains. We recorded that enveloped DNA mapped to genome regions with characteristics of mobile genetic elements other than prophages. This pointed to mechanisms of DNA mobility that are not well-studied in *C. difficile* so far. We also detected phage-mediated lateral transduction of bacterial DNA, which is the first described case in *C. difficile*. This study significantly contributes to our knowledge on prophage activity in *C. difficile* and revealed novel aspects on *C. difficile* (phage) biology.

Introduction

The pathogen *Clostridioides difficile* significantly contributes to nosocomial infections worldwide (1). A *C. difficile* infection mainly establishes after antibiotic treatment due to diverse resistances in *C. difficile* strains combined with the disturbed intestinal microflora (2). An intact microbiome usually provides resistance to *C. difficile* colonization and disease manifestation by producing different secondary bile salts such as lithocholate and deoxycholate (3). These compounds impede *C. difficile* spore germination and cell growth (4). Symptoms of an established *C. difficile* infection are caused by its toxins, predominantly toxin A and B, which are encoded in the pathogenicity locus (5). Symptom severity ranges from mild to severe manifestation, which also could include death of the infected individual (1). The personal health condition significantly affects resistance against a *C. difficile* infection, but the infecting strain is of relevance as well (6). Different *C. difficile* strains are linked to divergent virulence, and various aspects such as toxin production levels or secondary bile salt resistance were shown to correlate with disease severity (7,8). However, these studies also partially contradict in the concluded relevance of specific features on virulence. In addition to general virulence factors, mobile genetic elements (MGE) also contribute to *C. difficile* virulence (9–12). Those elements play a key role in fast adaptation to environmental conditions via horizontal gene transfer (HGT) (13). *C. difficile* genomes harbor various MGEs, including prophages (14). Prophages are widespread among the species *C. difficile*, and multiple prophages can exist within the same host (15). Primarily, prophages were assumed to affect *C. difficile* virulence solely by encoding advantageous traits such as antibiotic resistances, which was demonstrated by phage-mediated transduction of an erythromycin resistance (10).Phages can influence toxin production in *C. difficile* strains (11,12). These findings drew further attention to the influence of phages on *C. difficile* virulence. Phage research commonly works with prophage induction by introducing stressors such as UV radiation or mitomycin C. Meanwhile, studies confirmed spontaneous

61 prophage release from *C. difficile* isolates, and clinically relevant antibiotics were also
62 investigated for their phage-inducing effect (16,17). All experiments on *C. difficile* phages
63 however worked with cultivation conditions that do not represent the actual habitat. Some
64 components of the intestinal environment are stressful for *C. difficile*, such as the secondary
65 bile salts. One of the prominent secondary bile salts is deoxycholate (DCA), which can promote
66 biofilm formation in *C. difficile* (18). It was further demonstrated that DCA induces the bacterial
67 SOS response (19). Activation of the SOS response in turn induces prophages and lead to phage
68 particle production and release via host cell lysis (20). It is therefore likely that DCA induces
69 prophages as well, which would be a critical aspect in *C. difficile* biology and shed new light
70 on genetic transfer *in vivo*.
71 In this work, we examined prophage activity in different *C. difficile* strains under spontaneous
72 conditions and in the presence of DCA. The analyzed *C. difficile* strains were of non-clinical
73 and clinical origin and corresponded pairwise to each other based on their sequence type (ST).
74 Active prophage regions in these strains were identified by sequencing of particle-protected
75 DNA, analyzed for DCA-induced activity and for possible differences between clinical and
76 non-clinical strains that could contribute to virulence. The sequencing approach is more
77 sensitive than electron microscopy or plaque assays, the commonly used detection methods for
78 active *C. difficile* phages. We could therefore detect active prophages that might otherwise be
79 missed due to insufficient activity, but could contribute to HGT.
80

81 # Methods

82 **Strains and cultivation conditions**

83 The *C. difficile* strains used in this study were of clinical or non-clinical clinical background
84 (personal communication), with four pairs of clinical and non-clinical strains correpsonding in
85 ST, and one additional non-clinical strain (Table 1). Strains were routinely cultivated under

86 anaerobic conditions in supplemented Brain Heart Infusion Broth (BHIS; supplemented with
87 0.5% yeast extract, 0.05% L-cystein, 0.0001% Na-resazurin, purged with nitrogen) at 37°C.
88 Putative prophage regions of the strains were predicted with PHASTEST (21).
89
90 **Table 1. *C. difficile* strains used in this study.** Strain information on ST (Clade), profile of
91 toxins A. B, and CDT, clinical or non-clinical background, and GenBank accession of the
92 genome is listed.

Strain	ST (Clade)	Toxin profile	Background	GenBank accession
TS3_3	1 (2)	$A^+B^+CDT^+$	Non-clinical	CP134872
DSM 28196	1 (2)	$A^+B^+CDT^+$	Clinical	CP012320.1
B1_2	3 (1)	$A^+B^+CDT^-$	Non-clinical	CP132141-43
SC084-01-01	3 (1)	$A^+B^+CDT^-$	Clinical	CP132146-48
J2_1	8 (1)	$A^+B^+CDT^-$	Non-clinical	CP134690-1
SC083-01-01	8 (1)	$A^+B^+CDT^-$	Clinical	CP132144-45
MA_1	11 (5)	$A^+B^+CDT^+$	Non-clinical	CP132139-40
DSM 29747	11 (5)	$A^+B^+CDT^+$	Clinical	CP019864.1
MA_2	340 (C-III)	$A^-B^-CDT^-$	Non-clinical	CP129431-32

93
94 **DCA tolerance assessment**
95 The stress effect of DCA on the various strains was assessed via a minimum inhibitory
96 concentration (MIC) assay and relative growth determinations at different concentrations.
97 Cultivation was performed in cell culture plates (24 well for suspension cells, Sartorius AG,
98 Göttingen, Germany) in an anaerobic tent (Coy Laboratory, Grass Lake, USA). Overnight
99 cultures of *C. difficile* strains were cultivated as described above and used to inoculate two ml
100 medium to a final OD_{600} of 0.05, with either BHIS medium alone or supplemented with DCA
101 (sodium deoxycholate, Sigma-Aldrich Chemie Gmbh, Taufkirchen, Germany) in
102 concentrations ranging from 0.255 mM to 1.2 mM, which cover the physiological concentration
103 range in humans (22). Culture plates were incubated at 37°C for 22 hrs and kept anaerobic
104 during the OD_{600} measurement in a Synergy 2 Plate Reader (Biotek Agilent Technologies

105 Deutschland GmbH, Böblingen, Germany). Relative growth in the presence of DCA was
106 calculated in relation to the untreated cultures. Each experiment was performed in triplicates.
107 Relative growth was visualized with ggplot2 (v3.4.2) (23) in RStudio (v2022.06.0) (24) and
108 significance determined with the Tukey's 'Honest Significant Difference' method implemented
109 in the stats package (v4.2.0) (25).
110

111 **Prophage induction and phage particle isolation**

112 Phage particles were isolated from untreated and DCA-induced cultures. Two pre-warmed
113 flasks of 55 ml BHIS for each isolate were inoculated 1:100 from an overnight culture and
114 incubated at 37°C. Growth was monitored via OD_{600} measurements until an OD of ~0.6 (0.5-
115 0.7). One culture for each isolate was induced with 0.255 mM (0.01%) DCA final
116 concentration. The physiological concentration of DCA varies between individuals (22). We
117 therefore used this concentration that is within the physiological range and was also used in
118 various *C. difficile* studies regarding growth behavior or spore germination (7,26). The DCA
119 solution was freshly prepared under anaerobic conditions, with DCA suspended in distilled
120 water so that 100 µl inducing solution was required per 10 ml culture. The solution was
121 sterilized by filtration (Filtropur S 0.2 µm, Sarstedt AG & Co. KG, Nürnbrecht, Germany) and
122 anaerobically added to the induction cultures. The second culture was not induced for analysis
123 of spontaneous phage activity. Induced and non-induced cultures were further incubated until
124 22 hrs total incubation. The final OD_{600} of each culture was determined before isolating phage
125 particles.
126 For phage particle isolation, cells were pelleted via centrifugation at 4°C and 3,000 x *g* for
127 15 minutes. Remaining cells were removed by filtration of the supernatant with a 0.45 µm
128 Filtropur S filter (Sarstedt). Phage particles were pelleted via centrifugation at 8°C and
129 20,000 x *g* for 1 h. The pellet was suspended in 1 ml SM buffer (50 mM Tris-HCl, 100 mM
130 NaCl, 8 mM $MgSO_4 \cdot 7\ H_2O$, pH 7.4) supplemented with 0.5 mM $CaCl_2$ (supporting phage

131 stability and upcoming DNase treatment) and let soak overnight at 4°C. Particle suspension was
132 further supported the next day by shaking at 150 rpm (LT-V Lab-Shaker, Adolf Kühner AG,
133 Birsfelden, Germany) at room temperature for 2 hrs. Suspended samples were finally
134 transferred to 2 ml DNA LowBind micro tubes (Sarstedt) for following treatments using cut
135 filter tips to reduce possible shearing. Samples were stored at 4°C.
136
137 **Isolation of particle-protected DNA from phage samples**
138 Phage DNA was isolated using the MasterPure Gram Positive DNA Purification Kit (Epicentre,
139 Madison, WI, USA) with modifications. Prior to the phage DNA isolation, phage samples were
140 supplemented with 2 µl of 100 mg/ml lysozyme solution (lysozyme from chicken egg white
141 177,000 U/mg; Serva, Heidelberg, Germany) suspended in SM buffer to remove remaining host
142 cell debris, and with 50 µg/ml final concentration RNase A (Biozym Scientific GmbH, Hess.
143 Oldendorf, Germany) and 10 U Baseline-ZERO DNase (Biozym Scientific GmbH) to digest
144 host nucleic acids. Samples were incubated at 37°C for 6 hrs with gentle inversion every
145 30 minutes. Fragments of host nucleic acids resulting from the digestion were removed by
146 phage particle pelleting via centrifugation at 4°C and 20,000 x *g* for 1 hour. The recovered pellet
147 was suspended in 150 µl SM buffer. Suspension was supported by shaking at 150 rpm (LT-V
148 Lab-Shaker) at room temperature for 10 minutes and slight flicking. EDTA (0.5 M, pH 8.0)
149 was added to 10 mM final concentration for complete DNase inhibition.
150 Phage particles were lysed by adding 1% SDS (10% solution) and 2 µl Proteinase K (50 µg/µl;
151 Biozym Scientific GmbH), incubation at 56°C for 1.5 hrs and gentle inversion every
152 30 minutes. Subsequently, samples were completely cooled down on ice before addition of
153 130 µl MPC Protein Precipitation Reagent (pre-cooled to -20°C). After mixing by gentle
154 inversion, proteins were pelleted via centrifugation at 4°C and 10,000 x *g* for 10 minutes. The
155 DNA-containing supernatant was transferred to 1.5 ml DNA LowBind micro tubes (Sarstedt).
156 DNA was precipitated by addition of 0.3 M Na-acetate (3 M, pH 5.2), 10 mM $MgCl_2$ (2 M),

157 and 0.8 volume isopropanol at room temperature. Samples were inverted 40 times and
158 incubated at room temperature for 10 minutes before centrifugation at 4°C and 15,000 x *g* for
159 30 minutes. The supernatant was removed carefully and the DNA pellet washed twice with
160 150 µl 70% ethanol (pre-cooled to -20°C) and centrifugation at 4°C and 15,000 x *g* for 5 to
161 10 minutes. The supernatant was removed and the sample briefly centrifuged again to collect
162 all residual ethanol for final removal. DNA pellets were air-dried under a sterile bench and
163 immediately suspended in 20 µl TE buffer. Complete DNA elution was supported by brief
164 storage at 4°C and slight flicking, before final storage at -20°C. DNA concentration was
165 assessed with the Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) using
166 the HS dsDNA assay kit.
167

168 **Phage DNA sequencing and sequencing read processing**

169 Phage DNA was subjected to Illumina sequencing for dsDNA by paired-end library preparation
170 with the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) as
171 recommended by the manufacturer. Libraries were sequenced using an Illumina MiSeq system
172 and MiSeq Reagent Kit version 3 (2 × 300 bp, 600 cycles) as recommended by the
173 manufacturer.
174 All following software was used in default mode. Sequencing raw reads were quality processed
175 with fastp (v0.23.4) (27) before removing the sequencing adapters with Trimmomatic (v0.39)
176 (28). Processed reads were mapped onto the corresponding host genome using bowtie2 (v2.5.0),
177 and the resulting SAM file was converted to a TDS file for bioinformatics analysis with the
178 TraV software (29).
179

180 **Data analysis of phage sequencing reads**

181 TDS files of processed reads for the same isolate were together inspected in TraV (29) for read
182 coverage, and reads were normalized by calculation of nucleotide activity per kilobase of exon

183 model per million mapped reads (NPKM). This results in a value for each CDS corresponding
184 to its read coverage in reference to the overall read amount. NPKM values were further
185 normalized to account for the differing growth behavior under the induction conditions by
186 transforming values to an OD_{600} of 2.0. In this way, NPKM values reflected phage abundance
187 under the different conditions at same cell density, which allows a better qualitative estimation
188 of phage activity. OD normalization and visualization of NPKMs values were done in Rstudio
189 (v2022.06.0) (24) using the packages tidyverse (v2.0.0) (30), ggforce (v0.4.1) (31), and ggplot2
190 (v3.4.2) (23). NPKM values were plotted against the host genome with regard to sequence start
191 of the corresponding CDS, and original and normalized NPKM values were plotted together to
192 visualize the effect of OD normalization. Phage regions predicted with PHASTEST (21) were
193 also implemented.
194
195 **Phage genome annotation and gene content analysis**
196 Active regions identified via sequence read mapping were extracted for anew genome
197 annotation. Sequence ranges were thereby adopted from PHASTEST (21) predictions or
198 selected based on read mapping in TraV (29), if the predicted prophage region did not cover
199 the entire mapped region. Annotation was customized for phage genomes using Pharokka
200 (v1.3.2) (32) in default mode with sequence re-orientation to the large terminase subunit. If the
201 large terminase subunit was not annotated automatically, it was determined via BLAST analysis
202 (33) and manually annotated. For specific genes and their encoded protein, additional analyses
203 with protein BLAST (33) and InterProScan (v5.63-95.0) (34) were performed.
204
205 **Phage genome-based classifications**
206 Genome-based classifications were done with different bioinformatic analysis tools. An
207 average nucleotide identity analysis (ANI) with pyani (v0.2.12) (35) and MUMmer3 alignment
208 (36) (ANIm) was used in default mode to compare the phages amongst each other. Assessment

209 of the DNA-packaging strategy was performed based on the study of Rashid et al. (37). The
210 large terminase subunit was aligned at protein sequence level with ClustalW and a Maximum-
211 Likelihood phylogenetic tree was constructed with the Whelan and Goldman (WAG)
212 substitution model and otherwise default parameter with the software MEGA (v11.0.13) (38).
213 Branches were collapsed in MEGA (38) if none of our phages clustered within and final
214 modifications for visualization were done in Inkscape (v0.48; https://inkscape.org/de/). A
215 nucleotide BLAST analysis (33) with default parameters was performed to check for similarity
216 to already known phage genomes and for the prevalence in genomes of other *C. difficile* strains.
217 BLAST results were ordered based on query coverage and hits with a query coverage below
218 10% were neglected, unless relevant matches with higher coverages were not obtained. For
219 assigning the phages to a morphological family of the order *Caudovirales*, phage genomes were
220 inspected for the presence of baseplate proteins and sequence length of the tail length tape
221 measure protein (39). If the tail length tape measure protein was not annotated, it was identified
222 via protein BLAST analysis (33) and manually curated in the genome.
223

224 **Data availability**

225 The sequencing raw data is deposited at the NIH short read archive (SRA) under accessions
226 listed in Table 2.

227 **Table 2. SRA accessions of sequencing raw data.**

Strain	Spontaneous	DCA-induced
TS3_3	SRR26060497	SRR26060498
DSM 28196	SRR26060430	SRR26060431
B1_2	SRR26060462	SRR26060463
SC084-01-01	SRR26064407	SRR26064408
J2_1	SRR26060485	SRR26060486
SC083-01-01	SRR26060487	SRR26060488
MA_1	SRR26060489	SRR26060490
MA_2	SRR26064320	SRR26064321

228

229 # Results and discussion

230 ## DCA-tolerance is linked to the genetic but not clinical strain
231 background

232 Before investigating the effect of the secondary bile salt DCA on prophage activity in *C.*
233 *difficile*, individual DCA tolerance of the various strains was assessed in form of a MIC assay
234 with relative growth determinations (Fig 1). DCA concentrations ranged from 0.255 mM to
235 1.2 mM, thereby comprising the physiological human concentration of DCA (22). All strains
236 already exhibited reduced growth at the lowest concentration, which further decreased with
237 increasing concentration. At all concentrations, strains of the same ST showed no significant
238 difference in DCA tolerance, which implied comparable stress levels. Similar stress levels in
239 turn might imply similar DCA-induced prophage activity. In contrast, ST-specific tolerance
240 differences were apparent across the DCA concentration range, with ST1 exhibiting highest
241 tolerance, followed by ST8, ST11, ST3, and lastly ST340, which was the most susceptible ST.
242 Consequently, DCA tolerance correlated with the ST but not with the clinical background.
243 Tolerance difference between the STs was most distinct at the lowest concentration, which was
244 also used in the prophage induction experiments. Determined MICs ranged from 1 mM (ST11
245 and ST340) to 1.2 mM (ST1, ST3, ST8).

246

247 **Fig 1. DCA tolerance at various concentrations.** Tolerance of the strains to DCA was
248 determined as relative growth compared to the reference culture (regarded as 100%) and MIC
249 assay at concentrations from 0.255 mM to 1.2 mM. Strains of the same ST are depicted next to
250 each other. No significant tolerance differences were observed within an ST at all
251 concentrations.

252

253 ## Sequencing-based assessments of prophage activity

254 Prophage prediction of all analyzed genomes exhibited various putative prophage regions, often
255 with multiple incomplete and intact predicted regions in one genome (Table S1). Active
256 prophages were determined by sequencing of particle-protected DNA. Sequencing reads were
257 mapped to the corresponding host genome. Normalized read coverage (NPKM values)
258 represented phage abundance as a relative measure indicative for phage activity (40). In the
259 following, the term phage activity describes the production and resulting abundance of DNA-
260 containing particles, and mentioned NPKM values refer to the OD-normalized data. Sequencing
261 of phage DNA libraries was successful for all samples except for strain DSM 29747, which was
262 the only strain without a predicted intact prophage genomic region (Table S1). This strain was
263 therefore missing in further analyses.

264 Mapping of phage DNA sequencing reads onto respective host genomes is depicted in Figs 2 -
265 6. Overall examination of the mapping results revealed distinct activity of at least one region in
266 all strains and under both induction conditions. This demonstrated spontaneous phage activity
267 in all strains, and simultaneous activity of multiple phages within the same host. Almost all
268 regions matched well with predicted and intact prophage regions. All these regions are
269 summarized in Table 3 for name assignment to facilitate following descriptions. As apparent in
270 Table 3, regions with positive prophage prediction accorded in size with typical genome sizes
271 of *C. difficile* phages (41), while those without were significantly smaller.

272

273 **Table 3. Overview of active regions in all strains.** All active regions identified based on
274 sequencing read coverage (Fig 2 - 6) were renamed according to a numbered scheme
275 (strain_phiX). Regions were validated for certain (✓) or uncertain (~) activity based on
276 coverage signal strength. Additionally, prophage prediction with PHASTEST (21) is included
277 by stating positive (✓) or negative (—) prediction, and region size in bp is stated as well.

Strain	Location	Region name	Activity certain	Phage predicted	Size (bp)
TS3_3	Chromosome	TS3_3_phi	✓	✓	55,976
DSM 28196	Chr. region 1	DSM28196_phi1	✓	✓	55,976
	Chr. region 2	DSM28196_phi2	✓	–	5,484
B1_2	Chr. region 1	B1_2_phi1	~	✓	54,334
	Chr. region 2	B1_2_phi2	✓	✓	57,163
	ECE1	B1_2_phi3	✓	✓	42,358
	ECE2	B1_2_phi4		–	7,624
SC084-01-01	Chromosome	SC084-01-01_phi1	~	✓	69,503
	ECE1	SC084-01-01_phi2	~	✓	47,363
	ECE2	SC084-01-01_phi3	✓	✓	130,799
J2_1	Chromosome	J2_1_phi1	✓	✓	55,958
	ECE	J2_1_phi2	✓	✓	46,261
SC083-01-01	Chr. region 1	SC083-01-01_phi1	✓	✓	56,419
	ECE	SC083-01-01_phi2	✓	✓	45,313
MA_1	ECE	MA_1_phi	✓	✓	33,670
MA_2	Chr. region 1	MA_2_phi1	✓	✓	42,327
	Chr. region 2	MA_2_phi2	✓	✓	46,234
	Chr. region 3	MA_2_phi3	✓	–	16,820
	ECE	MA_2_phi4	✓	–	10,144

278

279 Comparison of corresponding strains showed no apparent differences in carriage or location of
280 active phages. Correspondingly, a correlation to the clinical background of a strain was not
281 detected. As active prophages of corresponding strains resided at corresponding genome
282 positions, we performed an ANIm analysis on extracted sequences of all active regions to assess
283 their similarity amongst each other (Fig S1). This confirmed identical sequences of the
284 analogous phages TS3_3_phi/DSM28196_phi1 of ST1-strains, and high similarity of both
285 phages J2_1_phi1/SC083-01-01_phi1 and J2_1_phi2/SC083-01-01_phi2 among the ST8-
286 strains, while the other phages exhibited only little or no similarity to the others.

287

288 All regions (Table 3) were further closely inspected their activities under DCA-induction.
289 Overall NPKM-transformation to an OD of 2 revealed distinctly higher signals under DCA-
290 induction in most active regions and all strains. This verified the phage-inducing effect of DCA,
291 which varied apparently between the different regions even within the same host and thereby
292 implied phage-dependency.
293 The genomes of both ST1-strains carried an identical phage (Fig S1) (TS3_3_phi/
294 DSM28196_phi1) as determined by ANIm analysis (Fig S1) and similar genome position. The
295 analogous phages showed distinct spontaneous activity with approximate magnitude, but their
296 DCA-induced activity differed substantially. Phage DSM28196_phi1 (Fig 2B) exhibited ~3.5x
297 higher signal increase under DCA than TS3_3_phi in the non-clinical strain (Fig 2A). Since the
298 phages were identical, the differing reactions seemed to be host-related. The genome of strain
299 DSM 28196 possessed another active region DSM28196_phi2, which showed minor activity
300 under both conditions, and was phage-atypical by comprising only four genes and missing a
301 prophage prediction (Fig 2 B).

302

303 **Fig 2. Coverage of phage sequencing reads of ST1-strains TS3_3 and DSM 28196.**
304 Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release
305 of strains (A) TS3_3 and (B) DSM 28196 were normalized to NPKM values (circles) with TraV
306 (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before
307 plotted against the chromosome (position in Mb) of the respective strain. Active regions were
308 magnified for better visualization. Prophage regions predicted by PHASTEST (21) were
309 highlighted in the background (intact = green, incomplete = pink).

310

311 The genomes of ST3-strains possessed several active regions (Fig 3), which were not similar to
312 each other (Fig S1). In the genome of the non-clinical strain B1_2 (Fig 3A), two of the four
313 active regions comprised the two ECEs. B1_2_phi4 on ECE2 exhibited the highest NPKM
314 values under DCA-induction. Remarkably, B1_2_phi4 is another phage-atypical but active
315 region without corresponding prophage prediction, as previously detected for DSM28196_phi2
316 (Fig 2B). B1_2_phi3 on ECE1 also showed increased activity under DCA, but both spontaneous
317 and induced activity were not particularly high. B1_2_phi2 on the chromosome showed
318 spontaneous activity and a substantial increase in DCA-induced activity. B1_2_phi1 on the
319 chromosome exhibited almost no signal under spontaneous conditions, which slightly increased
320 upon DCA-induction. This might indicate true DCA-induction of this phage without prior
321 spontaneous activity. In contrast to strain B1_2, the corresponding clinical strain SC084-01-01
322 possessed no prominently active region on the chromosome (Fig 3B). Activity could be
323 observed within the region SC084-01-01_phi1, but NPKM values were very low under both
324 conditions and signals did not cover the whole phage region. Abundancy of this phage was
325 probably insufficient to capture distinct activity by the sequencing approach. Similar activity
326 was observed for SC084-01-01_phi2 on ECE1 of SC084-01-01. SC084-01-01_phi3 on ECE2
327 was the only region in SC084-01-01 with prominent activity under spontaneous conditions, and
328 activity further increased apparently under DCA-induction.
329
330 **Fig 3. Coverage of phage sequencing reads of ST3-strains B1_2 and SC084-01-01.**
331 Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release
332 of strains (A) B1_2 and (B) SC084-01-01 were normalized to NPKM values (circles) with TraV
333 (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before
334 plotted against the chromosome (position in Mb) and ECE (position in kb) of the respective
335 strain. Active regions were magnified for better visualization. Prophage regions predicted by
336 PHASTEST (21) were highlighted (intact = green, incomplete = pink).

337

338 The genomes of both ST8-strains exhibited activity for their analogous chromosomal and
339 extrachromosomal regions (Fig 4), which were similar phages according to ANIm analysis (Fig
340 S1) and similar genomic location. J2_1_phi2 on the ECE of strain J2_1 showed distinct
341 spontaneous activity and a DCA-induced activity increase (Fig 4A), whereas SC083-01-
342 01_phi1 on the ECE of SC083-01-01 was only slightly active under spontaneous conditions,
343 and signal increase upon DCA-induction was only little. The activity of the chromosomal
344 phages differed apparently as well. J2_1_phi1 on the J2_1 chromosome was spontaneously
345 active and showed increased activity under DCA-induction (Fig 4A). The left part of the
346 prophage region started with minor activity, which drastically increased at the terminase genes.
347 Strikingly, the left part of the prophage region started with minor activity, which drastically
348 increased at the terminase genes. Further remarkable, sequencing reads mapped beyond the
349 predicted prophage region and spread upstream (~30 kb) and downstream (~135 kb). The
350 downstream region adjoined the phage activity with similar NPKM values that gradually
351 decreased over the entire section. The chromosomal phage SC083-01-01_phi1 of the
352 corresponding clinical ST8-strain SC083-01-01 did not exhibit these peculiarities (Fig 4B). It
353 was spontaneously active, and activity increased substantially under DCA treatment. This
354 increase was strikingly twice as high as observed for the analogous phage J2_1_phi1 (Fig 4A),
355 despite their similarity (Fig S1). Such dissimilar activity increase among analogous phages was
356 already observed in the ST1 strains (Fig 2). In both ST8- and ST1-pairs, a distinctly stronger
357 increase upon DCA-induction was connected to the clinical background of the strains. Since
358 DCA-stress levels did not significantly differ between corresponding clinical and non-clinical
359 strains (Fig 1) and thereby suggested similar induction levels, the question for the underlying
360 mechanism of these diverging activities in analogous phages arose. This prompted to an
361 undefined regulation of phage induction involved in the clinical strains.

362

363 **Fig 4. Coverage of phage sequencing reads of ST8-strains J2_1 and SC083-01-01.**
364 Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release
365 of strains (A) J2_1 and (B) SC083-01-01 were normalized to NPKM values (circles) with TraV
366 (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before
367 plotted against the chromosome (position in Mb) and ECE (position in kb) of the respective
368 strain. Active regions were magnified for better visualization. Prophage regions predicted by
369 PHASTEST (21) were highlighted (intact = green, incomplete = pink).

370

371 Strain MA_1 could not be compared to its corresponding clinical strain DSM 29747, but distinct
372 spontaneous activity was visible for MA_1_phi on the ECE, which increased under DCA
373 treatment (Fig 5).

374

375 **Fig 5. Coverage of phage sequencing reads of ST11-strain MA_1.** Sequencing reads of
376 phage particles from spontaneous (black) and DCA-induced (red) release of strain MA_1 were
377 normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally OD-
378 normalized to OD_{600} = 2 (line graphs), before plotted against the chromosome (position in Mb)
379 of MA_1. Active regions were magnified for better visualization. Prophage regions predicted
380 by PHASTEST (21) were highlighted (intact = green, incomplete = pink).

381

382 Strain MA_2 is to our knowledge the first cryptic *C. difficile* strain with detailed phage
383 examination. This strain possessed four active regions (Fig 6). MA_2_phi2 on the chromosome
384 and MA_2_phi4 on the ECE both showed prominent activity under spontaneous conditions and
385 a distinct activity increase upon DCA-induction. Interestingly, MA_2_phi4 was another active
386 region without phage prediction, as observed previously for DSM28196_phi2 (Fig 2B) and

387 B1_2_phi4 (Fig 3A). This was also true for MA_2_phi3 on the chromosomal, where activity
388 was however very low under spontaneous conditions. Activity of this region increased upon
389 DCA-induction. Chromosomal region MA_2_phi1 exhibited least activity in this genome even
390 under DCA treatment. Read mapping for MA_2_phi1 and MA_2_phi2 did not cover the entire
391 predicted prophage regions, but sections without read coverage contained only bacterial genes
392 and were therefore evidently mis-predicted.
393
394 **Fig 6. Coverage of phage sequencing reads of cryptic ST340-strain MA_2.** Sequencing
395 reads of phage particles from spontaneous (black) and DCA-induced (red) release of strain
396 MA_2 were normalized to NPKM values (circles) with TraV (29), and NPKM values were
397 additionally OD-normalized to OD_{600} = 2 (line graphs), before plotted against the chromosome
398 (position in Mb) and ECE (position in kb) of MA_2. Active regions were magnified for better
399 visualization. Prophage regions predicted by PHASTEST (21) were highlighted (intact = green,
400 incomplete = pink).
401
402 Almost all identified active prophage regions could be induced by the secondary bile salt DCA.
403 Mentionable, assessed phage activity after DCA treatment might be influenced by a direct effect
404 of DCA on the phage. A study on different bacteriophages in *Escherichia coli* investigated the
405 effect of bile salts on the host-phage interaction and observed varying survival rates of the
406 phages (42).
407 Moreover, almost all ECEs were detected as active phages. Although extrachromosomal
408 prophages have already been described in *C. difficile* strains (43,44), only a few of these were
409 isolated and characterized so far (41).
410
411

412 # Phage genome annotation and gene content analysis

413 ## Phage genomes harbor virulence-relevant genes

414 All active regions identified via sequence read mapping (Figs 2 – 6, Table 3) were inspected
415 after anew annotation with Pharokka (32) for genes that might increase virulence of the host
416 (genomic information in supplementary data file S1). All of them exhibited characteristic phage
417 genes in a modular organization according to the different encoded functions, as typically seen
418 in *C. difficile* phages (16,43,45). In some genomes, proteins characteristic for plasmids were
419 found, such as genes encoding partition proteins and replication initiation factors (46). Genes
420 encoding plasmid-related proteins besides phage-typical ones are common in a certain type of
421 MGE, so-called phage-plasmids (47). These phage-plasmid features were especially recorded
422 for extrachromosomal prophages, but were also present on the chromosomally integrated
423 prophage MA_2_phi1. Further frequently observed genes in the phage genomes encoded
424 proteins with potential involvement in cellular metabolism and growth, such as metallo-
425 proteases, kinases, a phosphatase, and most of all putative rhodanese-related sulfurtransferases.
426 These genes might be advantageous for the bacterial fitness, thereby indirectly contributing to
427 host virulence. Two phage genomes (B1_2_phi1 and SC084-01-01_phi1) carried genes
428 encoding hemolysin XhlA, an established virulence factor in other bacterial species (48),
429 capable of lysing mammalian erythrocytes (49). In the opportunistic pathogen *Mannheimia*
430 *haemolytica*, temperate phages were induced that encoded hemolysin XhlA and discussed to
431 contribute to bacterial pathogenicity and transfer of this virulence factor (50). Hemolysis in *C.*
432 *difficile* is not commonly known, but few studies demonstrated its hemolytic capability (51).
433 However, XhlA is also present in other prophage genomes as part of the lysis module (52).
434 Indeed, the gene encoding XhlA was found next to the lysis-relevant genes encoding holin and
435 endolysin in phages B1_2_phi1 and SC084-01-01_phi1. Thus, the actual role of hemolysin
436 XhlA in these phages and its potential impact on host virulence remains unclear. Phage SC084-

437 01-01_phi1 possessed a gene encoding an ABC-transporter, which might contribute to
438 antibiotic resistance of the host (53). Phage SC084-01-01_phi3 harbored a putative spore
439 protease, which could influence the bacterial sporulation or germination ability and, as a
440 consequence, alter bacterial fitness (43). Other genes conferring antibiotic resistances or
441 encoding known virulence factors were not identified in the active phage genomes. Six phages
442 (TS3_3_phi, DSM28196_phi1, B1_2_phi2, J2_1_phi1, SC083-01-01_phi, MA_2_phi2)
443 carried arrays of clustered regularly interspaced short palindromic repeats (CRISPR), which is
444 similar to other *C. difficile* phages with described CRISPRs (37,43,54). Temperate phages
445 carrying CRISPR arrays increase host immunity against other invading phages (55). The
446 corresponding host genomes were verified to encode Cas proteins required for CRISPR-Cas-
447 mediated phage immunity (56). CRISPRs in prophages represent horizontally transferrable
448 immunity against phages, which is specifically relevant in the context of phage therapy, an
449 alternative treatment approach for bacterial infections with growing importance in view of
450 increasing multidrug-resistances (57).

451 Active non-phage elements likely belong to so far undescribed HGT 452 mechanisms in *C. difficile*

453 The regions without corresponding prophage prediction (Table 3) did not possess a phage-
454 typical genome accordingly. Therefore, further gene analysis of these non-phage elements were
455 performed based on the original genome annotation with Prokka (58) (genomic information in
456 supplementary data file S1). No proteins involved in capsid or tail production, DNA packaging
457 or host lysis were present. The lack of structural proteins was striking, since these DNA
458 elements were enveloped according to the DNA isolation procedure. This indicated the
459 involvement of unrelated particles. Even additional analysis of hypothetical proteins with
460 InterProScan (34) and BLASTp (33) could not identify further functions. All other proteins
461 were assigned to functions with DNA-binding activity, like helicases, integrases, relaxases, and

462 transcriptional regulators. These genes are typical for phage genomes but also for MGEs like
463 transposons as Integrative and Conjugative or Mobilizable Elements (ICE / IME) (59).
464 Screening for these MGEs with ICEscreen (60) validated DSM28196_phi2 as complete IME,
465 while MA_2_phi3 was detected as invalid ICE. Interestingly, the peculiar upstream-region of
466 J2_1_phi1 (Fig 4A) was also identified as an ICE, although being incomplete. These integrative
467 MGEs do not encode proteins for production of particles that carry the respective mobile
468 sequence (59). Interestingly, transposons were found to hitchhike co-residing phages in several
469 bacteria such as *Staphylococcus aureus* (61), *Vibrio cholerae* (62), and *Enterococcus faecalis*
470 (63), enabling the phage-mediated transduction of virulence-relevant genes. This type of
471 transduction was demonstrated once in *C. difficile* with the transfer of a conjugative transposon
472 carrying an antibiotic resistance gene (10). Transduction can be either generalized, specialized
473 or lateral (64). They all imply the "headful" DNA-packaging, in which the terminase starts
474 DNA packaging at a bacterial homologue to the phage packaging site until the capsid is full,
475 which consequently implies that transduced DNA is at least of similar phage genome size (64).
476 The transduction mechanisms differ in transduced DNA and frequency (65). Specialized and
477 lateral transduction involve host DNA adjacent to the prophage, while random host DNA is
478 packaged in generalized transduction (65). Generalized and specialized transduction are
479 processes of erroneous DNA-packaging, which results in low transduction frequency detectable
480 by sequencing read coverage (64,65). In contrast, lateral transduction results in high sequencing
481 read coverage comparable to actual phage activity, as this mechanism is assumed as natural
482 phage trait instead of mistaken processes (64,65). This trait comprises phage genome replication
483 and simultaneous DNA-packaging before excision from the chromosome, whereby a
484 substantial amount of adjacent host DNA is packaged as well (66,67). All these characteristics
485 of lateral transduction accorded with the observed mapped region downstream of phage
486 J2_1_phi1 (Fig 4A), which indicated involvement of this DNA segment in phage-mediated
487 lateral transduction.

488 The downstream region of J2_1_phi1 (Fig 4A) did not comprise characteristic genes for MGEs.
489 Instead, several genes encoded proteins with potential relevance for strain virulence, such as
490 genes for ABC-transporters, a multidrug efflux system ATP-binding protein, stress-related
491 proteins, proteins involved in resistance to vancomycin and daunorubicin, and the putative
492 virulence factor BrkB. Therefore, mobilization and transfer of this region is critical regarding
493 the spread of antibiotic resistances or virulence-related genes. The drastic NPKM difference
494 within the genome phage J2_1_phi1 (Fig 4A) might also result from the process of lateral
495 transduction, as inaccurate excision of the phage genome after *in situ* replication leads to phage
496 genome truncation.
497 Since no evidence for lateral transduction was observed for phage J2_1_phi1's analogue
498 SC083-01-01_phi1 (Fig 4B) despite their high similarity (Fig S1), the question about
499 underlying differences arose. Direct phage genome comparison revealed diverging excisionases
500 and integrases as well as four additional amino acids in the large terminase protein of
501 J2_1_phi1. All these proteins perform activities destining for lateral transduction (64).
502 The extrachromosomal non-phage elements were significantly smaller than the co-existing
503 phages (Table 3), which is in contrast with the headful packaging mechanism required in
504 transduction. This indicate a form of DNA-protecting particle other than phages, such as gene
505 transfer agents (GTA). These phage-like particles carry DNA between 4 to 14 kb (68), which
506 is similar to the sizes of the detected non-phage elements (Table 3). However, GTAs package
507 bacterial DNA randomly (68), which does not fit to the detected distinct activity of specific
508 regions, making GTAs unlikely as mode of action. Extracellular vesicles are known in various
509 bacteria and described to carry and transfer genetic content, e.g. plasmids, between cells (69–
510 71). This type of alternative HGT is not well characterized so far, but was demonstrated to allow
511 interspecies gene transfer (71), which underlines the importance of vesicle-mediated DNA
512 exchange. Noteworthy, vesicle-driven HGT in *C. difficile* has not been described so far.
513

514 # Classification of the active phages

515 ## Terminase-based determination of the phage DNA-packaging strategy

516 Assessment of the phage-DNA packaging mechanisms was performed to validate the above
517 hypothesized transduction events. The large terminase subunit was analyzed via protein
518 sequence alignment and phylogenetic tree construction referring to Rashid et al. (37). This
519 assigned the phages to different phage DNA-packaging mechanisms (Fig 7). All our phages
520 were assigned to clusters comprising other *C. difficile* phages, which predominantly represented
521 the P22-like headful packaging mechanism, followed by the 3'-extended COS ends and an
522 unknown strategy. Consequently, most of the phages were predicted to utilize the headful
523 packaging mechanism and would, thus, be indeed capable of transducing host DNA. The
524 mechanism "P22-like headful" originates from the packaging strategy employed by phage P22
525 of *Salmonella enterica*. Phage P22 was originally described to perform generalized transduction
526 (72), but recent evidence demonstrated also specialized as well as lateral transduction activity
527 (67). These terminase analysis results supported the assumption of lateral transduction of the
528 phage J2_1_phi1 downstream region, which is by that the first described case in *C. difficile*.

529

530 **Fig 7. Maximum-likelihood phylogenetic tree of large terminase protein sequences.** The
531 large terminase of the active phages (highlighted in bold and italic) were aligned on protein
532 level to the reference sequences from Rashid et al. (37). Branches of the different DNA-
533 packaging strategies were colored according to Rashid et al. (37), and branches were collapsed
534 for better visualization if none of our phages was included.

535

536 ## Nucleotide BLAST analyses assess phage prevalence and novelty

537 A nucleotide BLAST analysis (33) was performed on all active regions in Table 3 to check for
538 similar phages and elements, and to assess prevalence among *C. difficile* strains. The results are

539 available in supplementary data file S2 and summarized in Table 4. Most of the phages matched
540 against *C. difficile* phages with query coverages between 4% and 90% and percent identities
541 between 86.26% and 99.86%. This confirmed that our phages are indeed similar to known
542 phages but still represent novel types, which underlines the contribution of this work to the
543 general knowledge on *C. difficile* phages. Further, chromosomal phages also matched against a
544 multitude of *C. difficile* chromosomes, while the extrachromosomal phages often corresponded
545 to *C. difficile* plasmids and few chromosomes. This demonstrated the prevalence of the
546 identified phages in other *C. difficile* strains. The non-phage elements B1_2_phi4 and
547 MA_2_phi4 yielded no significant BLAST hit against a phage, but matched against *C. difficile*
548 plasmids. Both matching plasmids belong to classes of *C. difficile* plasmids with similar
549 organization that are present in diverse *C. difficile* strains (73,74). All these plasmids might
550 therefore likewise be inducible and particle-protected, which implies a different mechanism of
551 HGT than currently assumed.
552

553 **Table 4. Nucleotide BLAST results of the identified active regions.** Results were
554 summarized regarding the first phage-related *C. difficile* BLAST hit by stating its description,
555 query coverage, percent identity, and number of pre- and succeeding chromosome/plasmid
556 entries. The few hits of metagenome-assembled phages were not listed. If available, information
557 on the phage family of the respective hit is indicated (M – *Myoviridae*, S – *Siphoviridae*).

	First best *C. difficile*-phage BLAST hit				
Phage	**Entry**	**Query cover %**	**Percent identity %**	***C. difficile* matches preceding / succeeding**	**Family**
TS3_3_phi	CD2301	36	91.64	98 chromosomes / 32 chromosomes/assemblies 5 phages	M (75)
DSM28196_phi1	phiC2	36	91.64	98 chromosomes / 32 chromosomes/assemblies 5 phages	M (76)
DSM28196_phi2	–	< 79	< 92.44	Only chromosomes	–
B1_2_phi1	CDMH1	60	88.33	35 chromosomes / 102 chromosomes 13 phages	M (77)
B1_2_phi2	phiC2	43	97.80	80 chromosomes / 55 chromosomes 3 phages	M (76)
B1_2_phi3	phiCD111	84	95.62	0 / 3 chromosomes 9 plasmids 16 phages	S (78)
B1_2_phi4	plasmid pJMR5-4[a]	9	86.26	0 / 2 chromosomes 1 plasmid	–
SC084-01-01_phi1	phiCD418	54	92.42	29 chromosomes / 137 chromosomes 8 phages	M (75)
SC084-01-01_phi2	HGP05	45	93.09	20 plasmids, 2 chromosomes / 145 chromosomes 14 phages	–
SC084-01-01_phi3	phiCD211	89	99.86	0 / 10 plasmids 6 phages	S (43)
J2_1_phi1	phiC2	37	92.19	112 chromosomes / 27 chromosomes 3 phages	M (76)
J2_1_phi2	HGP05	46	91.63	20 plasmids, 2 chromosomes / 93 chromosomes 8 phages	–
SC083-01-01_phi1	phiC2	37	92.19	98 chromosomes / 43 chromosomes 4 phages	M (76)

558

559

560 **Table 4 continued.**

	First best *C. difficile*-phage BLAST hit				
Phage	**Entry**	**Query cover %**	**Percent identity %**	***C. difficile* matches preceding / succeeding**	**Family**
SC083-01-01_phi2	HGP05	47	91.64	20 plasmids, 2 chromosomes / 140 chromosomes 8 phages	–
MA_1_phi	phiCD506	90	99.60	2 plasmids / 172 chromosomes 27 plasmids 20 phages	M (78)
MA_2_phi1	phiCD24-1	4	88.74	17 chromosomes or genome assemblies / 0	S (79)
MA_2_phi2	phiCDKH01	74	94.21	1 chromosome / 89 chromosomes 1 phage	S (80)
MA_2_phi3	–	< 44	< 97.72	Only chromosomes	–
MA_2_phi4	plasmid pCD-WTSI1[a]	76	91.95	0 / 25 plasmids	–

561 [a]since ECEs B1_2_phi4 and MA_2_phi4 identified to be no phage, BLAST results were
562 checked for *C. difficile* plasmids instead.

563

564 ## Genome-based phage assignment to *Myoviridae* and *Siphoviridae*

565 We lastly classified our phages morphologically. All known *C. difficile* phages so far belong to
566 the *Caudovirales* family of *Myoviridae* or *Siphoviridae*, which distinguish by tail appearance
567 (41). Genome inspection for the presence of baseplate proteins characteristic for *Myoviridae*
568 and the length of the tail length tape measure protein as indicator for *Siphoviridae* could assign
569 eleven phages to *Myoviridae* and four phages to *Siphoviridae* (Table S2).

570

571 # Conclusion

572 We aimed to investigate prophage activity in different clinical and non-clinical *C. difficile*
573 strains and unravelling potential relationships between phage activity and clinical background
574 of the strain. Our investigations did not find specific connections to the clinical background,

575 although we observed stronger DCA-related activity with clinical background for phages that
576 were similar between clinical and non-clinical strains. We further revealed several interesting
577 findings with relevance for future *C. difficile* phage research. We identified and characterized
578 several active prophages in various *C. difficile* strains with a sequencing-based approach. This
579 sensitive approach allowed detecting multiple co-existing prophages with diverse activity. Most
580 of these phages were distinctly active without specific induction, but they showed increased
581 activity when induced with the secondary bile salt DCA. This proved that spontaneous activity
582 is common in *C. difficile* prophages, and that the natural stressor DCA triggers prophage
583 induction. These findings are crucial for investigating *C. difficile* biology, since phages
584 evidently affect *C. difficile* fitness and virulence by influencing toxin production or
585 participating in the exchange of clinically relevant genes. We also found such genes with
586 potential connection to virulence in some phage genomes. In this context, research on actual *in*
587 *vivo* phage mobility should increasingly resemble *C. difficile*'s natural habitat. The sequencing
588 approach further revealed active regions without phage identity. Based on genomic
589 examinations, these regions were identified as another form of MGE, in most cases possibly
590 integrative elements. These elements apparently participated in a strategy of mobilization that
591 involves some kind of DNA envelopment, which pointed to phage particles or bacterial
592 vesicles. This phenomenon was observed in several of the analyzed strains, which indicated
593 that this type of DNA mobilization might be a frequent mechanism in *C. difficile* and should be
594 further investigated. We also observed transduced bacterial DNA that was likely the result of
595 lateral transduction employed by one phage, which has not been described so far in *C. difficile*
596 and opens up new perspective on *C. difficile* phage research.
597
598
599

Acknowledgements

We thank Melanie Heinemann for technical assistance. This work was funded by the Federal State of Lower Saxony, Niedersächsisches Vorab CDiff and CDInfect projects (VWZN2889/3215/3266). We acknowledge support by the Open Access Publication Funds of the Göttingen University. This work was partly supported by the Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences at the Georg-August-Universität Göttingen. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Balsells E, Shi T, Leese C, Lyell I, Burrows J, Wiuff C, et al. Global burden of *Clostridium difficile* infections: a systematic review and meta-analysis. J Glob Health [Internet]. 2019 Jun;9(1):010407. Available from: http://jogh.org/documents/issue201901/jogh-09-010407.pdf
2. Spigaglia P. Recent advances in the understanding of antibiotic resistance in *Clostridium difficile* infection. Ther Adv Infect Dis [Internet]. 2016 Feb 23;3(1):23–42. Available from: http://journals.sagepub.com/doi/10.1177/2049936115622891
3. Sorg JA, Sonenshein AL. Bile Salts and Glycine as Cogerminants for *Clostridium difficile* Spores. J Bacteriol [Internet]. 2008 Apr;190(7):2505–12. Available from: https://journals.asm.org/doi/10.1128/JB.01765-07
4. Thanissery R, Winston JA, Theriot CM. Inhibition of spore germination, growth, and toxin activity of clinically relevant *C. difficile* strains by gut microbiota derived secondary bile acids. Anaerobe [Internet]. 2017 Jun;45:86–100. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1075996417300471
5. Chandrasekaran R, Lacy DB. The role of toxins in *Clostridium difficile* infection.

625 FEMS Microbiol Rev [Internet]. 2017 Nov 1;41(6):723–50. Available from:
626 https://academic.oup.com/femsre/article/41/6/723/4557158
627 6. Czepiel J, Dróżdż M, Pituch H, Kuijper EJ, Perucki W, Mielimonka A, et al.
628 *Clostridium difficile* infection: review. Eur J Clin Microbiol Infect Dis [Internet]. 2019
629 Jul 3;38(7):1211–21. Available from: https://journals.lww.com/00013542-201807000-
630 00001
631 7. Lewis BB, Carter RA, Ling L, Leiner I, Taur Y, Kamboj M, et al. Pathogenicity Locus,
632 Core Genome, and Accessory Gene Contributions to *Clostridium difficile* Virulence.
633 Ballard JD, editor. MBio [Internet]. 2017 Sep 6;8(4):e00885-17. Available from:
634 https://journals.asm.org/doi/10.1128/mBio.00885-17
635 8. Alonso CD, Kelly CP, Garey KW, Gonzales-Luna AJ, Williams D, Daugherty K, et al.
636 Ultrasensitive and Quantitative Toxin Measurement Correlates With Baseline Severity,
637 Severe Outcomes, and Recurrence Among Hospitalized Patients With *Clostridioides*
638 *difficile* Infection. Clin Infect Dis [Internet]. 2022 Jul 6;74(12):2142–9. Available from:
639 https://academic.oup.com/cid/article/74/12/2142/6372425
640 9. Smits WK, Roseboom AM, Corver J. Plasmids of *Clostridioides difficile*. Curr Opin
641 Microbiol [Internet]. 2022 Feb;65:87–94. Available from:
642 https://linkinghub.elsevier.com/retrieve/pii/S1369527421001521
643 10. Goh S, Hussain H, Chang BJ, Emmett W, Riley T V., Mullany P. Phage ϕC2 Mediates
644 Transduction of Tn *6215* , Encoding Erythromycin Resistance, between *Clostridium*
645 *difficile* Strains. Onderdonk AB, editor. MBio [Internet]. 2013 Dec 31;4(6):1–7.
646 Available from: https://journals.asm.org/doi/10.1128/mBio.00840-13
647 11. Sekulovic O, Meessen-Pinard M, Fortier L-C. Prophage-Stimulated Toxin Production
648 in *Clostridium difficile* NAP1/027 Lysogens. J Bacteriol [Internet]. 2011
649 Jun;193(11):2726–34. Available from: https://journals.asm.org/doi/10.1128/JB.00787-
650 10

651 12. Govind R, Vediyappan G, Rolfe RD, Dupuy B, Fralick JA. Bacteriophage-Mediated
652 Toxin Gene Regulation in *Clostridium difficile*. J Virol [Internet]. 2009
653 Dec;83(23):12037–45. Available from:
654 https://journals.asm.org/doi/10.1128/JVI.01256-09
655 13. de la Cruz F, Davies J. Horizontal gene transfer and the origin of species: lessons from
656 bacteria. Trends Microbiol [Internet]. 2000 Mar;8(3):128–33. Available from:
657 https://linkinghub.elsevier.com/retrieve/pii/S0966842X00017030
658 14. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The
659 multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic
660 genome. Nat Genet [Internet]. 2006 Jul 25;38(7):779–86. Available from:
661 https://www.nature.com/articles/ng1830
662 15. Fortier L-C. Bacteriophages Contribute to Shaping *Clostridioides (Clostridium)*
663 *difficile* Species. Front Microbiol [Internet]. 2018 Aug 31;9:2033. Available from:
664 https://www.frontiersin.org/article/10.3389/fmicb.2018.02033/full
665 16. Meessen-Pinard M, Sekulovic O, Fortier L-C. Evidence of *In Vivo* Prophage Induction
666 during *Clostridium difficile* Infection. Appl Environ Microbiol [Internet]. 2012
667 Nov;78(21):7662–70. Available from:
668 https://journals.asm.org/doi/10.1128/AEM.02275-12
669 17. Nale JY, Shan J, Hickenbotham PT, Fawley WN, Wilcox MH, Clokie MRJ. Diverse
670 Temperate Bacteriophage Carriage in *Clostridium difficile* 027 Strains. Popoff MR,
671 editor. PLoS One [Internet]. 2012 May 18;7(5):e37263. Available from:
672 https://dx.plos.org/10.1371/journal.pone.0037263
673 18. Dubois T, Tremblay YDN, Hamiot A, Martin-Verstraete I, Deschamps J, Monot M, et
674 al. A microbiota-generated bile salt induces biofilm formation in *Clostridium difficile*.
675 npj Biofilms Microbiomes [Internet]. 2019 May 9;5(1):14. Available from:
676 https://www.nature.com/articles/s41522-019-0087-4

677 19. Kandell RL, Bernstein C. Bile salt/acid induction of DNA damage in bacterial and
678 mammalian cells: Implications for colon cancer. Nutr Cancer [Internet]. 1991
679 Jan;16(3–4):227–38. Available from:
680 http://www.tandfonline.com/doi/abs/10.1080/01635589109514161
681 20. Hu J, Ye H, Wang S, Wang J, Han D. Prophage Activation in the Intestine: Insights
682 Into Functions and Possible Applications. Front Microbiol [Internet]. 2021 Dec
683 13;12:785634. Available from:
684 https://www.frontiersin.org/articles/10.3389/fmicb.2021.785634/full
685 21. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than
686 PHASTER, better than PHAST. Nucleic Acids Res [Internet]. 2023 Jul
687 5;51(W1):W443–50. Available from:
688 https://academic.oup.com/nar/article/51/W1/W443/7167344
689 22. Hamilton JP, Xie G, Raufman J-P, Hogan S, Griffin TL, Packard CA, et al. Human
690 cecal bile acids: concentration and spectrum. Am J Physiol Liver Physiol [Internet].
691 2007 Jul;293(1):G256–63. Available from:
692 https://www.physiology.org/doi/10.1152/ajpgi.00027.2007
693 23. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag
694 New York; 2016. 1–260 p. Available from: https://ggplot2.tidyverse.org
695 24. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. 2020.
696 Available from: http://www.rstudio.com/
697 25. Team RC. R: A Language and Environment for Statistical Computing [Internet].
698 Vienna, Austria: R Foundation for Statistical Computing; 2013. Available from:
699 http://www.r-project.org/
700 26. Theriot CM, Bowman AA, Young VB. Antibiotic-Induced Alterations of the Gut
701 Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium difficile*
702 Spore Germination and Outgrowth in the Large Intestine. Ellermeier CD, editor.

703 mSphere [Internet]. 2016 Feb 25;1(1):e00045-15. Available from:
704 https://journals.asm.org/doi/10.1128/mSphere.00045-15
705 27. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
706 Bioinformatics [Internet]. 2018 Sep 1;34(17):i884–90. Available from:
707 https://academic.oup.com/bioinformatics/article/34/17/i884/5093234
708 28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
709 sequence data. Bioinformatics [Internet]. 2014 Aug 1;30(15):2114–20. Available from:
710 https://academic.oup.com/bioinformatics/article/30/15/2114/2390096
711 29. Dietrich S, Wiegand S, Liesegang H. TraV: A Genome Context Sensitive
712 Transcriptome Browser. Scaria V, editor. PLoS One [Internet]. 2014 Apr
713 7;9(4):e93677. Available from: https://dx.plos.org/10.1371/journal.pone.0093677
714 30. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome
715 to the Tidyverse. J Open Source Softw [Internet]. 2019 Nov 21;4(43):1686. Available
716 from: https://joss.theoj.org/papers/10.21105/joss.01686
717 31. Pedersen TL. ggforce: Accelerating "ggplot2" Title [Internet]. 2022. Available from:
718 https://github.com/thomasp85/ggforce
719 32. Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald P-J, Vreugde S. Pharokka: a fast
720 scalable bacteriophage annotation tool. Marschall T, editor. Bioinformatics [Internet].
721 2023 Jan 1;39(1):btac776. Available from:
722 https://academic.oup.com/bioinformatics/article/doi/10.1093/bioinformatics/btac776/68
723 58464
724 33. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI
725 BLAST: a better web interface. Nucleic Acids Res [Internet]. 2008 May 19;36(Web
726 Server issue):W5–9. Available from: https://academic.oup.com/nar/article-
727 lookup/doi/10.1093/nar/gkn201
728 34. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5:

729 genome-scale protein function classification. Bioinformatics [Internet]. 2014 May
730 1;30(9):1236–40. Available from:
731 https://academic.oup.com/bioinformatics/article/30/9/1236/237988
732 35. Pritchard L. PYANI script [Internet]. 2014. Available from:
733 https://github.com/widdowquinn/pyani
734 36. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al.
735 Versatile and open software for comparing large genomes. Genome Biol [Internet].
736 2004;5(2):R12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14759262
737 37. Rashid S, Barylski J, Hargreaves K, Millard A, Vinner G, Clokie M. Two Novel
738 Myoviruses from the North of Iraq Reveal Insights into *Clostridium difficile* Phage
739 Diversity and Biology. Viruses [Internet]. 2016 Nov 16;8(11):310. Available from:
740 http://www.mdpi.com/1999-4915/8/11/310
741 38. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis
742 Version 11. Battistuzzi FU, editor. Mol Biol Evol [Internet]. 2021 Jun 25;38(7):3022–
743 7. Available from: https://academic.oup.com/mbe/article/38/7/3022/6248099
744 39. Zinke M, Schröder GF, Lange A. Major tail proteins of bacteriophages of the order
745 Caudovirales. J Biol Chem [Internet]. 2022 Jan;298(1):101472. Available from:
746 https://linkinghub.elsevier.com/retrieve/pii/S0021925821012813
747 40. Hertel R, Rodríguez DP, Hollensteiner J, Dietrich S, Leimbach A, Hoppert M, et al.
748 Genome-Based Identification of Active Prophage Regions by Next Generation
749 Sequencing in *Bacillus licheniformis* DSM13. Schuch R, editor. PLoS One [Internet].
750 2015 Mar 26;10(3):e0120759. Available from:
751 https://dx.plos.org/10.1371/journal.pone.0120759
752 41. Heuler J, Fortier L-C, Sun X. *Clostridioides difficile* phage biology and application.
753 FEMS Microbiol Rev [Internet]. 2021 Sep 8;45(5):fuab012. Available from:
754 https://academic.oup.com/femsre/article/doi/10.1093/femsre/fuab012/6134755

755 42. Scanlan PD, Bischofberger AM, Hall AR. Modification of *Escherichia coli* –
756 bacteriophage interactions by surfactants and antibiotics *in vitro*. Smalla K, editor.
757 FEMS Microbiol Ecol [Internet]. 2017 Jan;93(1):fiw211. Available from:
758 https://academic.oup.com/femsec/article-lookup/doi/10.1093/femsec/fiw211
759 43. Garneau JR, Sekulovic O, Dupuy B, Soutourina O, Monot M, Fortier L-C. High
760 Prevalence and Genetic Diversity of Large phiCD211 (phiCDIF1296T)-Like
761 Prophages in *Clostridioides difficile*. Dozois CM, editor. Appl Environ Microbiol
762 [Internet]. 2018 Feb;84(3):e02164-17. Available from:
763 https://journals.asm.org/doi/10.1128/AEM.02164-17
764 44. Ramírez-Vargas G, Goh S, Rodríguez C. The Novel Phages phiCD5763 and
765 phiCD2955 Represent Two Groups of Big Plasmidial Siphoviridae Phages of
766 *Clostridium difficile*. Front Microbiol [Internet]. 2018 Jan 22;9:26. Available from:
767 http://journal.frontiersin.org/article/10.3389/fmicb.2018.00026/full
768 45. Govind R, Fralick JA, Rolfe RD. Genomic organization and molecular characterization
769 of *Clostridium difficile* bacteriophage ΦCD119. J Bacteriol. 2006;188(7):2568–77.
770 46. Filutowicz M. Plasmids, Bacterial. In: Encyclopedia of Microbiology [Internet].
771 Elsevier; 2009. p. 644–65. Available from:
772 https://linkinghub.elsevier.com/retrieve/pii/B9780123739445000146
773 47. Pfeifer E, Moura de Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous
774 distinctive groups of phage–plasmids with conserved phage and variable plasmid gene
775 repertoires. Nucleic Acids Res [Internet]. 2021 Mar 18;49(5):2655–73. Available from:
776 https://academic.oup.com/nar/article/49/5/2655/6137301
777 48. Thomas P, Abdel-Glil MY, Eichhorn I, Semmler T, Werckenthin C, Baumbach C, et
778 al. Genome Sequence Analysis of *Clostridium chauvoei* Strains of European Origin
779 and Evaluation of Typing Options for Outbreak Investigations. Front Microbiol
780 [Internet]. 2021 Sep 29;12:.732106. Available from:

781 https://www.frontiersin.org/articles/10.3389/fmicb.2021.732106/full
782 49. Cowles KN, Goodrich-Blair H. Expression and activity of a *Xenorhabdus nematophila*
783 haemolysin required for full virulence towards *Manduca sexta* insects. Cell Microbiol
784 [Internet]. 2004 Sep 8;7(2):209–19. Available from:
785 https://onlinelibrary.wiley.com/doi/10.1111/j.1462-5822.2004.00448.x
786 50. Niu YD, Cook SR, Wang J, Klima CL, Hsu Y, Kropinski AM, et al. Comparative
787 analysis of multiple inducible phages from *Mannheimia haemolytica*. BMC Microbiol
788 [Internet]. 2015 Dec 30;15(1):175. Available from: http://dx.doi.org/10.1186/s12866-
789 015-0494-5
790 51. Alkudmani ZSB. The identification and characterization of novel haemolysin genes
791 from *Clostridium difficile* [Internet]. University College London; 2018. Available from:
792 https://discovery.ucl.ac.uk/id/eprint/10078163/7/Alkudmani_00_Thesis_edited.pdf
793 52. Krogh S, Jørgensen ST, Devine KM. Lysis Genes of the *Bacillus subtilis* Defective
794 Prophage PBSX. J Bacteriol [Internet]. 1998 Apr 15;180(8):2110–7. Available from:
795 https://journals.asm.org/doi/10.1128/JB.180.8.2110-2117.1998
796 53. Orelle C, Mathieu K, Jault J-M. Multidrug ABC transporters in bacteria. Res Microbiol
797 [Internet]. 2019 Nov;170(8):381–91. Available from:
798 https://linkinghub.elsevier.com/retrieve/pii/S0923250819300580
799 54. Hargreaves KR, Flores CO, Lawley TD, Clokie MRJ. Abundant and Diverse Clustered
800 Regularly Interspaced Short Palindromic Repeat Spacers in *Clostridium difficile* Strains
801 and Prophages Target Multiple Phage Types within This Pathogen. Paul J, Azam F,
802 editors. MBio [Internet]. 2014 Oct 31;5(5):e01045-13. Available from:
803 https://journals.asm.org/doi/10.1128/mBio.01045-13
804 55. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al.
805 CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science (80-).
806 2007;315(March):1709–12.

807 56. Koonin E V, Makarova KS. CRISPR-Cas: an adaptive immunity system in
808 prokaryotes. F1000 Biol Rep [Internet]. 2009 Dec 9;1(95). Available from:
809 https://facultyopinions.com/prime/reports/b/1/95/
810 57. Gutiérrez B, Domingo-Calap P. Phage Therapy in Gastrointestinal Diseases.
811 Microorganisms [Internet]. 2020 Sep 16;8(9):1420. Available from:
812 https://www.mdpi.com/2076-2607/8/9/1420
813 58. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics [Internet].
814 2014 Jul 15;30(14):2068–9. Available from:
815 https://academic.oup.com/bioinformatics/article/30/14/2068/2390517
816 59. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable
817 genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev [Internet].
818 2014 Jul;38(4):720–60. Available from: https://academic.oup.com/femsre/article-
819 lookup/doi/10.1111/1574-6976.12058
820 60. Lao J, Lacroix T, Guédon G, Coluzzi C, Payot S, Leblond-Bourget N, et al. ICEscreen:
821 a tool to detect Firmicute ICEs and IMEs, isolated or enclosed in composite structures.
822 NAR Genomics Bioinforma [Internet]. 2022 Oct 6;4(4):lqac079. Available from:
823 https://academic.oup.com/nargab/article/doi/10.1093/nargab/lqac079/6765278
824 61. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP. The gene for toxic shock
825 toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*.
826 Mol Microbiol [Internet]. 1998 Jul;29(2):527–43. Available from:
827 https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.1998.00947.x
828 62. Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own
829 CRISPR/Cas adaptive response to evade host innate immunity. Nature [Internet]. 2013
830 Feb 27;494(7438):489–91. Available from:
831 https://www.nature.com/articles/nature11927
832 63. Duerkop BA, Clements C V., Rollins D, Rodrigues JLM, Hooper L V. A composite

833 bacteriophage alters colonization by an intestinal commensal bacterium. Proc Natl
834 Acad Sci [Internet]. 2012 Oct 23;109(43):17621–6. Available from:
835 https://pnas.org/doi/full/10.1073/pnas.1206136109
836 64. Borodovich T, Shkoporov AN, Ross RP, Hill C. Phage-mediated horizontal gene
837 transfer and its implications for the human gut microbiome. Gastroenterol Rep
838 [Internet]. 2022 Jan 25;10:goac012. Available from:
839 https://academic.oup.com/gastro/article/doi/10.1093/gastro/goac012/6567682
840 65. Kleiner M, Bushnell B, Sanderson KE, Hooper L V., Duerkop BA. Transductomics:
841 sequencing-based detection and analysis of transduced DNA in pure cultures and
842 microbial communities. Microbiome [Internet]. 2020 Dec 15;8(1):158. Available from:
843 https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-020-00935-5
844 66. Chiang YN, Penadés JR, Chen J. Genetic transduction by phages and chromosomal
845 islands: The new and noncanonical. Kline KA, editor. PLOS Pathog [Internet]. 2019
846 Aug 8;15(8):e1007878. Available from:
847 https://dx.plos.org/10.1371/journal.ppat.1007878
848 67. Fillol-Salom A, Bacigalupe R, Humphrey S, Chiang YN, Chen J, Penadés JR. Lateral
849 transduction is inherent to the life cycle of the archetypical *Salmonella* phage P22. Nat
850 Commun [Internet]. 2021 Nov 8;12(1):6510. Available from:
851 https://www.nature.com/articles/s41467-021-26520-4
852 68. Lang AS, Zhaxybayeva O, Beatty JT. Gene transfer agents: phage-like elements of
853 genetic exchange. Nat Rev Microbiol [Internet]. 2012 Jul 11;10(7):472–82. Available
854 from: https://www.nature.com/articles/nrmicro2802
855 69. Tran F, Boedicker JQ. Genetic cargo and bacterial species set the rate of vesicle-
856 mediated horizontal gene transfer. Sci Rep [Internet]. 2017 Aug 18;7(1):8813.
857 Available from: https://www.nature.com/articles/s41598-017-07447-7
858 70. Bitto NJ, Chapman R, Pidot S, Costin A, Lo C, Choi J, et al. Bacterial membrane

859 vesicles transport their DNA cargo into host cells. Sci Rep [Internet]. 2017 Aug
860 1;7(1):7072. Available from: https://www.nature.com/articles/s41598-017-07288-4
861 71. Fulsundar S, Harms K, Flaten GE, Johnsen PJ, Chopade BA, Nielsen KM. Gene
862 Transfer Potential of Outer Membrane Vesicles of *Acinetobacter baylyi* and Effects of
863 Stress on Vesiculation. Kivisaar M, editor. Appl Environ Microbiol [Internet]. 2014
864 Jun;80(11):3469–83. Available from:
865 https://journals.asm.org/doi/10.1128/AEM.04248-13
866 72. Ebel-Tsipis J, Fox MS, Botstein D. Generalized transduction by bacteriophage P22 in
867 *Salmonella typhimurium*. J Mol Biol [Internet]. 1972 Nov;71(2):449–69. Available
868 from: https://linkinghub.elsevier.com/retrieve/pii/0022283672903622
869 73. Smits WK, Weese JS, Roberts AP, Harmanus C, Hornung B. A helicase-containing
870 module defines a family of pCD630-like plasmids in *Clostridium difficile*. Anaerobe
871 [Internet]. 2018 Feb;49:78–84. Available from:
872 https://linkinghub.elsevier.com/retrieve/pii/S1075996417302287
873 74. Roseboom AM, Ducarmon QR, Hornung BVH, Harmanus C, Crobach MJT, Kuijper
874 EJ, et al. Carriage of three plasmids in a single human clinical isolate of *Clostridioides*
875 *difficile*. Plasmid [Internet]. 2023 Jan;125:102669. Available from:
876 https://linkinghub.elsevier.com/retrieve/pii/S0147619X22000531
877 75. Whittle MJ, Bilverstone TW, van Esveld RJ, Lücke AC, Lister MM, Kuehne SA, et al.
878 A Novel Bacteriophage with Broad Host Range against *Clostridioides difficile*
879 Ribotype 078 Supports SlpA as the Likely Phage Receptor. Auchtung JM, editor.
880 Microbiol Spectr [Internet]. 2022 Feb 23;10(1):e0229521. Available from:
881 https://journals.asm.org/doi/10.1128/spectrum.02295-21
882 76. Goh S, Ong PF, Song KP, Riley T V, Chang BJ. The complete genome sequence of
883 *Clostridium difficile* phage ϕC2 and comparisons to ϕCD119 and inducible prophages
884 of CD630. Microbiology [Internet]. 2007 Mar 1;153(3):676–85. Available from:

885 https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.2006/0024
886 36-0
887 77. Hargreaves KR, Kropinski AM, Clokie MRJ. What Does the Talking?: Quorum
888 Sensing Signalling Genes Discovered in a Bacteriophage Genome. Kaufmann GF,
889 editor. PLoS One [Internet]. 2014 Jan 24;9(1):e85131. Available from:
890 https://dx.plos.org/10.1371/journal.pone.0085131
891 78. Sekulovic O, Garneau JR, Néron A, Fortier L-C. Characterization of Temperate Phages
892 Infecting *Clostridium difficile* Isolates of Human and Animal Origins. Griffiths MW,
893 editor. Appl Environ Microbiol [Internet]. 2014 Apr 15;80(8):2555–63. Available
894 from: https://journals.asm.org/doi/10.1128/AEM.00237-14
895 79. Fortier L-C, Moineau S. Morphological and Genetic Diversity of Temperate Phages in
896 *Clostridium difficile*. Appl Environ Microbiol [Internet]. 2007 Nov 15;73(22):7358–66.
897 Available from: https://journals.asm.org/doi/10.1128/AEM.00582-07
898 80. Hinc K, Kabała M, Iwanicki A, Martirosian G, Negri A, Obuchowski M. Complete
899 genome sequence of the newly discovered temperate *Clostridioides difficile*
900 bacteriophage phiCDKH01 of the family *Siphoviridae*. Arch Virol [Internet]. 2021 Aug
901 20;166(8):2305–10. Available from: https://link.springer.com/10.1007/s00705-021-
902 05092-0
903

904 # Supporting information

905 **Table S1. Detailed prophage prediction results.**

906 **Fig S1. ANIm analysis of the active regions as identified by sequencing read coverage.**

907 **Data file S1. Genomic information of the active regions.** Genomic information of all phage
908 genomes and non-phage elements: position on host genome, CDS or repeat-region (type) with
909 corresponding start/stop (starting from 1 bp), strand (+/-), locus tag (phage genome specific;

910 host genome locus tags for non-phage elements), function as determined by Pharokka (32), and
911 gene product as determined by Pharokka (32) for phage genomes or Prokka (58) for non-phage
912 elements.

913 **Data file S2. Nucleotide BLAST analysis results of all active regions.** First best *C. difficile*-
914 phage BLAST hits listed in Table 4 are highlighted in bold with yellow background.

915 **Table S2. Genome-based prediction of morphological family.**

916

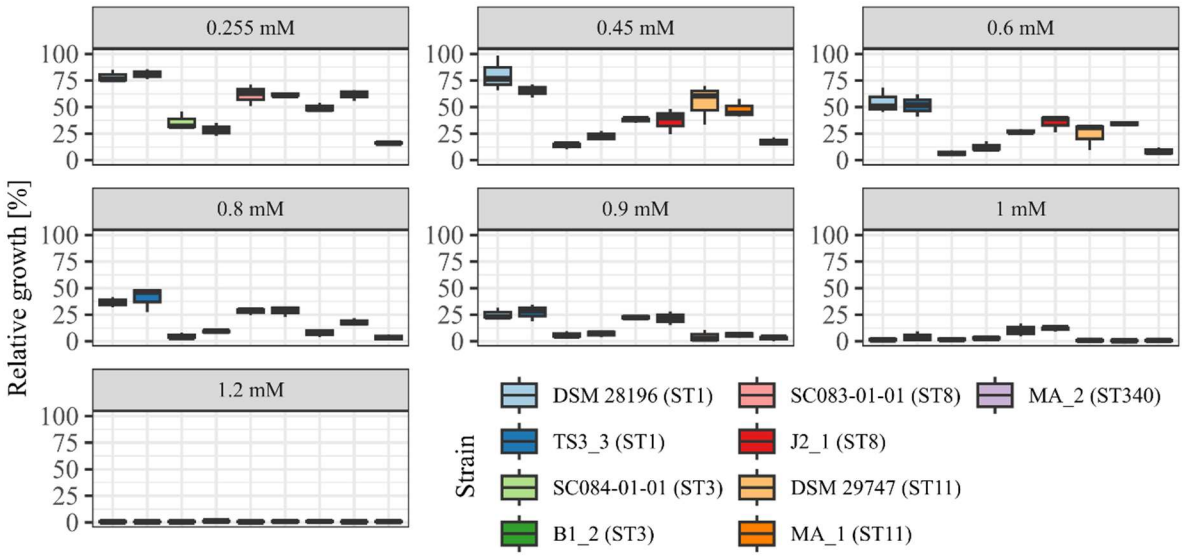


Fig 1. DCA tolerance at various concentrations. Tolerance of the strains to DCA was determined as relative growth compared to the reference culture (regarded as 100%) and MIC assay at concentrations from 0.255 mM to 1.2 mM. Strains of the same ST are depicted next to each other. No significant tolerance differences were observed within an ST at all concentrations.

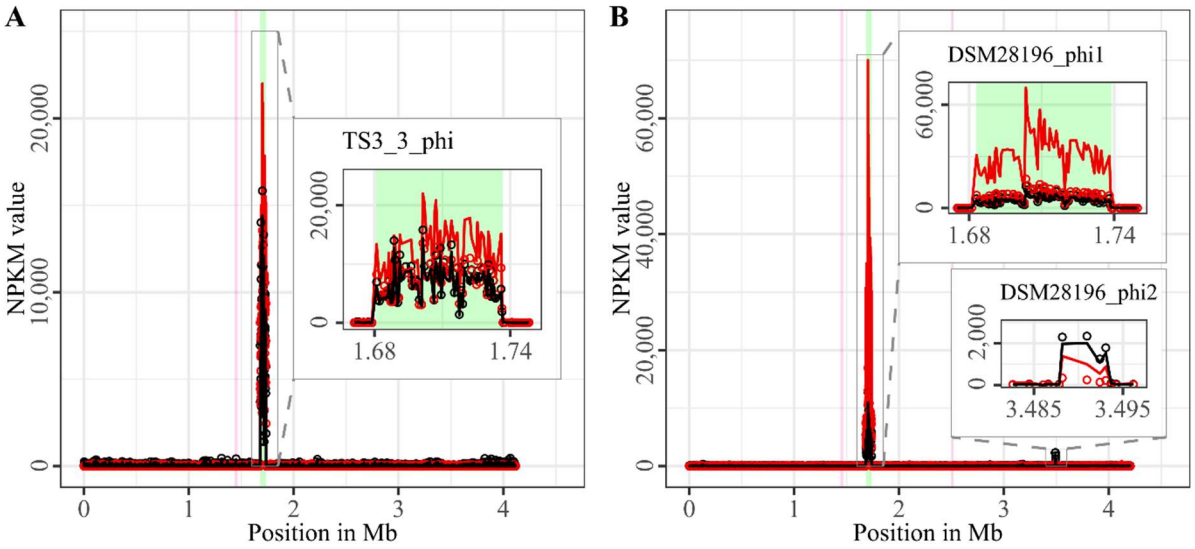
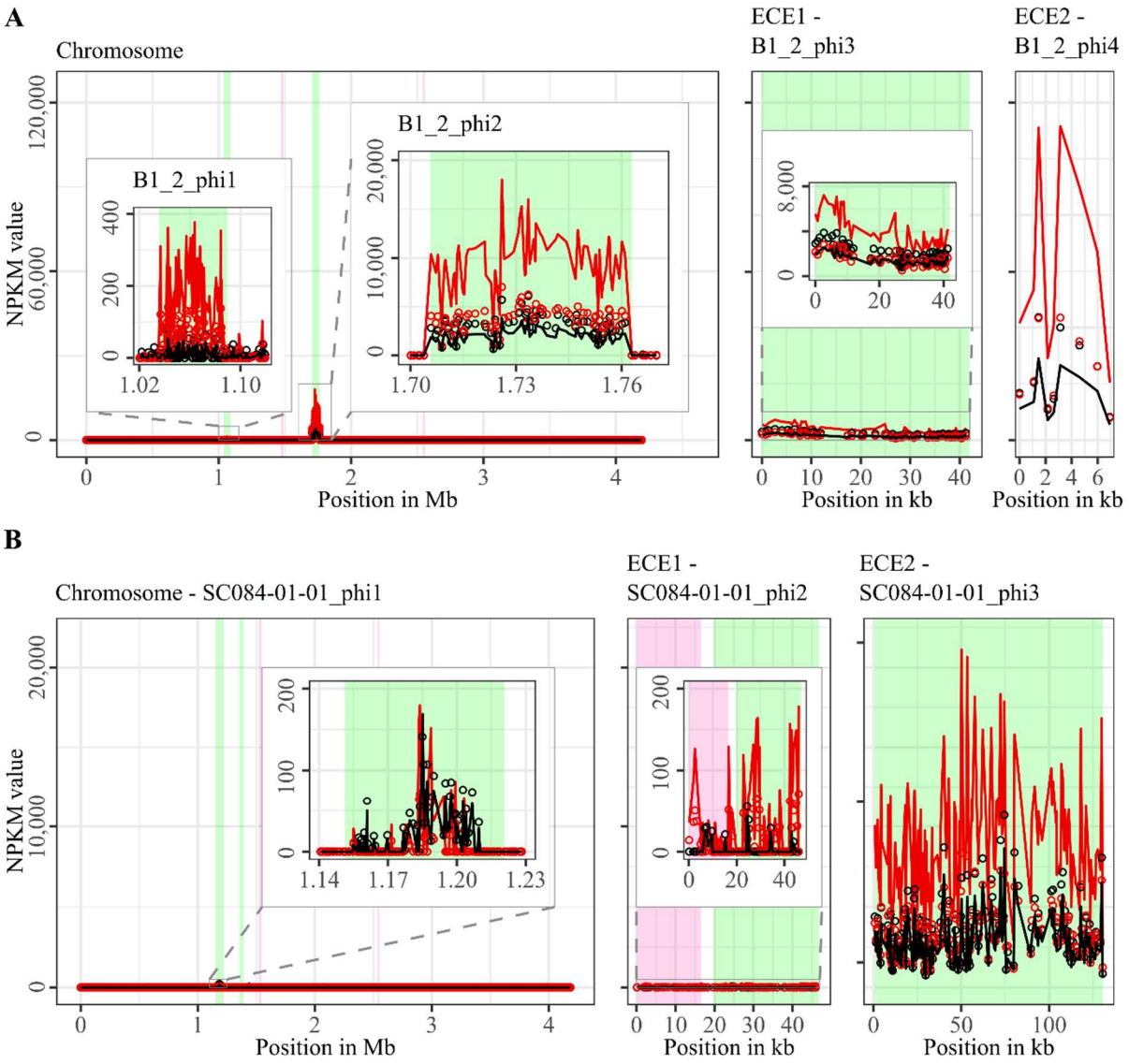


Fig 2. Coverage of phage sequencing reads of ST1-strains TS3_3 and DSM 28196. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strains (A) TS3_3 and (B) DSM 28196 were normalized to NPKM values (circles) with TraV

928 (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before
929 plotted against the chromosome (position in Mb) of the respective strain. Active regions were
930 magnified for better visualization. Prophage regions predicted by PHASTEST (21) were
931 highlighted in the background (intact = green, incomplete = pink).



932

933 **Fig 3. Coverage of phage sequencing reads of ST3-strains B1_2 and SC084-01-01.**
934 Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release
935 of strains (A) B1_2 and (B) SC084-01-01 were normalized to NPKM values (circles) with TraV
936 (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before
937 plotted against the chromosome (position in Mb) and ECE (position in kb) of the respective

strain. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (21) were highlighted (intact = green, incomplete = pink).

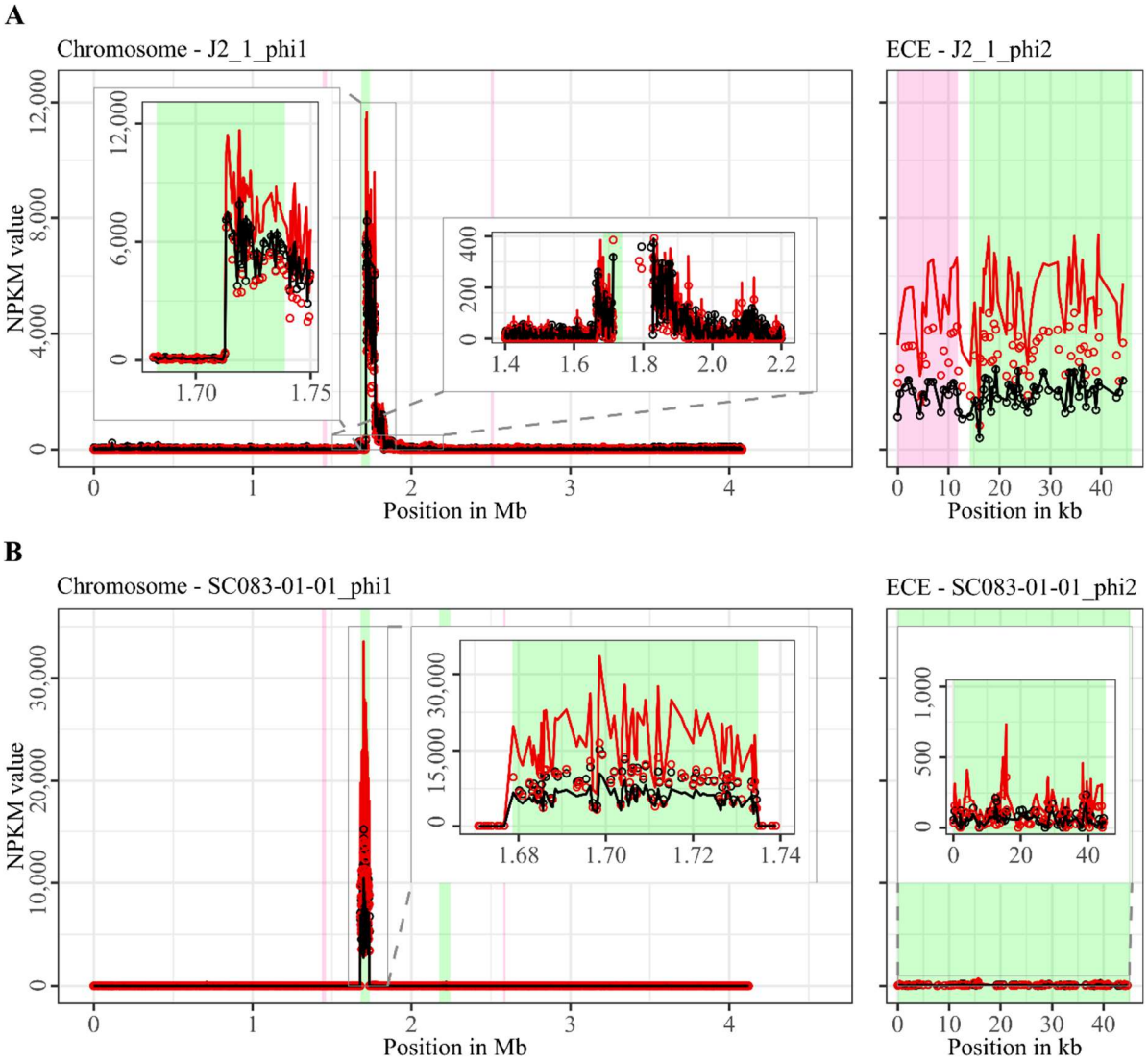


Fig 4. Coverage of phage sequencing reads of ST8-strains J2_1 and SC083-01-01. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strains (A) J2_1 and (B) SC083-01-01 were normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before plotted against the chromosome (position in Mb) and ECE (position in kb) of the respective strain. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (21) were highlighted (intact = green, incomplete = pink).

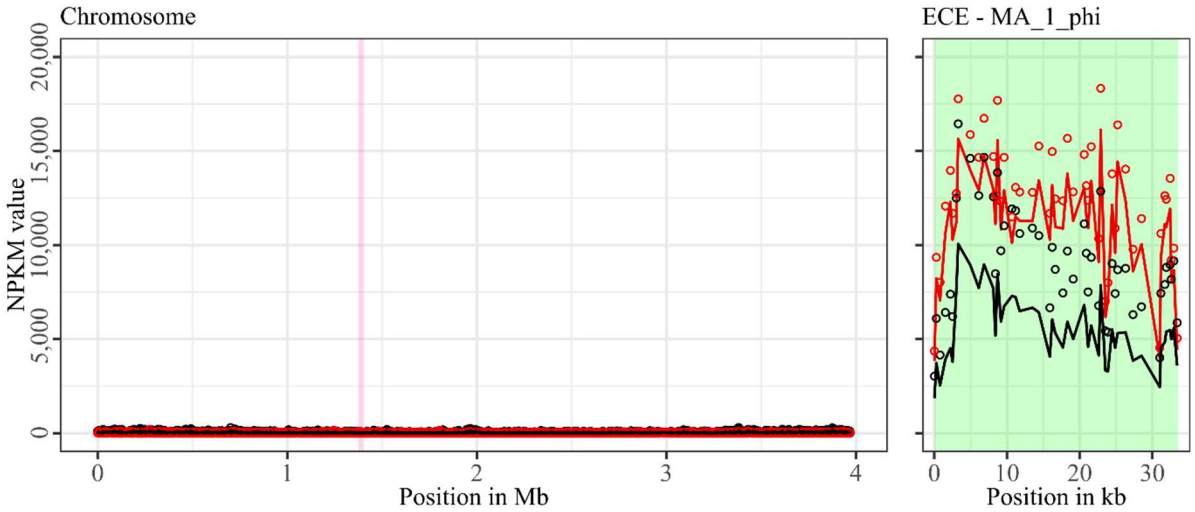


Fig 5. Coverage of phage sequencing reads of ST11-strain MA_1. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strain MA_1 were normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before plotted against the chromosome (position in Mb) of MA_1. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (21) were highlighted (intact = green, incomplete = pink).

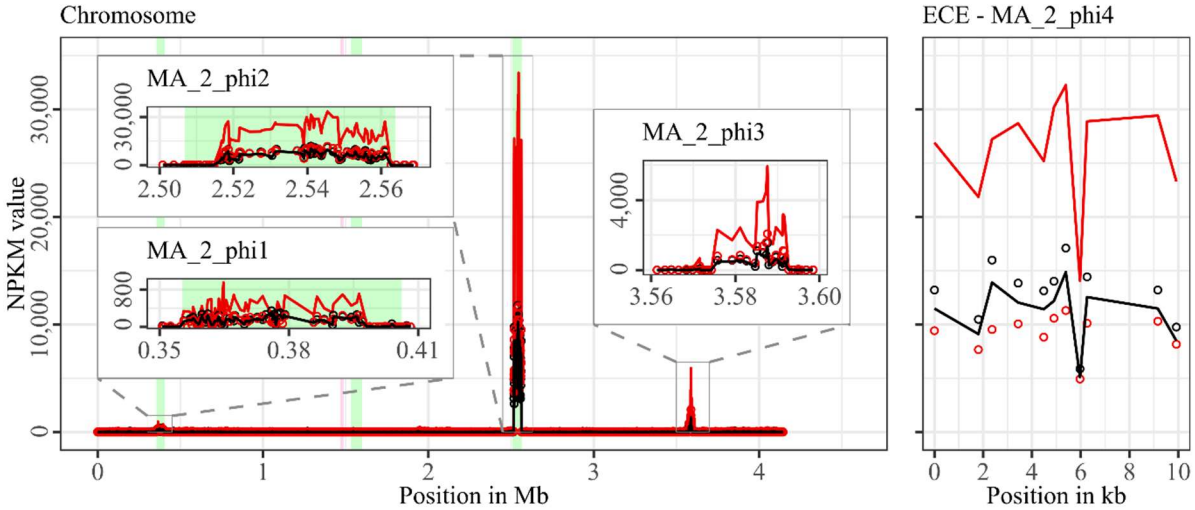


Fig 6. Coverage of phage sequencing reads of cryptic ST340-strain MA_2. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strain MA_2 were normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), before plotted against the chromosome

(position in Mb) and ECE (position in kb) of MA_2. Active regions were magnified for better visualization. Prophage regions predicted by PHASTEST (21) were highlighted (intact = green, incomplete = pink).

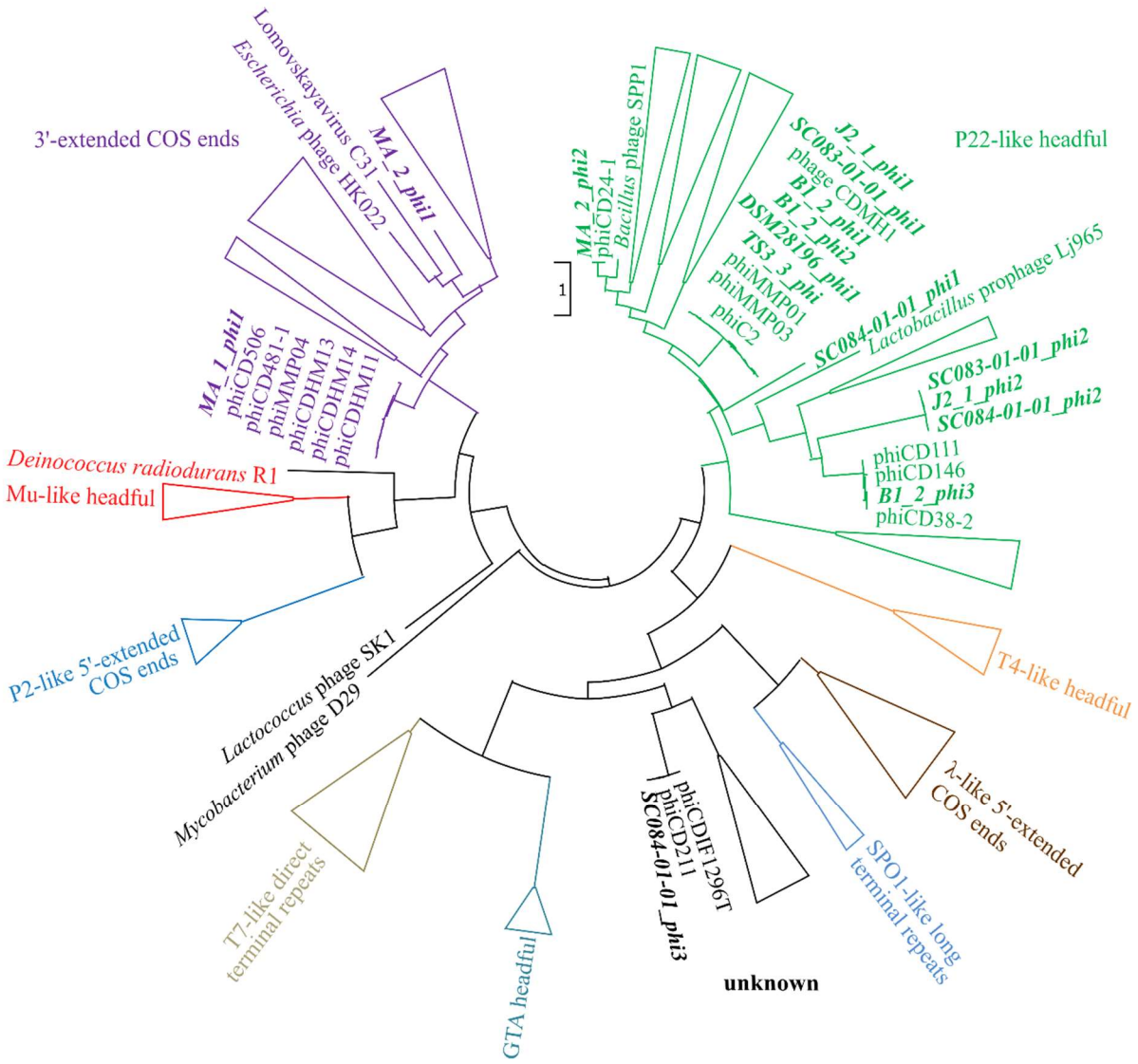


Fig 7. Maximum-likelihood phylogenetic tree of large terminase protein sequences. The large terminase of the active phages (highlighted in bold and italic) were aligned on protein level to the reference sequences from Rashid et al. (37). Branches of the different DNA-packaging strategies were colored according to Rashid et al. (37), and branches were collapsed for better visualization if none of our phages was included.