Comparative genome and phenotypic analysis of *Clostridioides difficile* strains

Dissertation
for the award of the degree
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen

within the doctoral program "Microbiology and Biochemistry"
of the Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences
(GGNB)

submitted by

Miriam Antonia Schüler

from Göttingen

Göttingen, 2023

THESIS ADVISORY COMMITTEE:

Prof. Dr. Rolf Daniel, Department of Genomic and Applied Microbiology
Institute of Microbiology and Genetics, Georg-August University Göttingen

PD Dr. Michael Hoppert, Department of General Microbiology
Institute of Microbiology and Genetics, Georg-August University Göttingen

Prof. Dr. Gerhard Braus, Department of Molecular Microbiology and Genetics
Institute of Microbiology and Genetics, Georg-August University Göttingen

MEMBERS OF THE EXAMINATION BOARD:

1st referee: **Prof. Dr. Rolf Daniel**, Department of Genomic and Applied Microbiology
Institute of Microbiology and Genetics, Georg-August University Göttingen

2nd referee: **PD Dr. Michael Hoppert**, Department of General Microbiology
Institute of Microbiology and Genetics, Georg-August University Göttingen

FURTHER MEMBERS OF THE EXAMINATION BOARD:

Prof. Dr. Gerhard Braus, Department of Molecular Microbiology and Genetics
Institute of Microbiology and Genetics, Georg-August University Göttingen

Prof. Dr. Kai Heimel, Department of Microbial Cell Biology
Institute of Microbiology and Genetics, Georg-August-University Göttingen

Prof. Dr. Stefanie Pöggeler, Department of Genetics of Eukaryotic Microorganisms
Institute of Microbiology and Genetics, Georg-August University Göttingen

Prof. Dr. Jan de Vries, Department of Applied Bioinformatics
Institute of Microbiology and Genetics, Georg-August University Göttingen

Date of oral examination: 14.11.2023

CONTENTS

1 SUMMARY

The anaerobic bacterium *Clostridioides difficile*, renamed from Clostridium *difficile*, substantially accounts for nosocomial, antibiotic-associated infections worldwide. The complex dynamics of various traits shape the bacterial virulence, such as toxin production and gene sequence variants, multidrug resistances, and sporulation. *C. difficile* is also known for its mobile genome that comprises diverse mobile genetic elements (MGE), thereby enabling the spread of advantageous genes. The pathogen is extensively investigated to enlighten its virulence and dissemination. Strain isolation is mainly done in clinical context, although *C. difficile* also inhabits human and mammalian intestines without disease manifestation or is present in environmental surroundings. Further, these isolations always employ antibiotic-based cultivation, which might lead to an isolation bias.

Within this thesis, non-clinical strains were isolated from environmental samples. Five strains in total were obtained from which two were isolated without antibiotic treatment. An effective protocol for antibiotic-free isolation could not be established but offers potential for further efforts. Isolation approaches were supported by the established detection PCR that targets the *C. difficile hpdBCA*-operon and detects even low abundances of *C. difficile* in environmental DNA. Specificity of the detection PCR was verified by next generation sequencing of detection PCR amplicons from environmental samples to ensure that amplified sequences only originated from *C. difficile* DNA. These environmental samples were also analyzed via 16S rRNA gene sequencing, in which *C. difficile* could not be detected, what clearly demonstrated their low abundance. Besides validation of environmental sources for promising *C. difficile* isolation, amplicons from the established detection PCR further allow phylogenetic estimation, which was determined by average nucleotide identity analyses of representative sequences. Analysis of PCR amplicons also supported the assumed isolation bias due to antibiotic treatment, as detected sequences from environmental samples did not accord with sequences from strains obtained by antibiotic-based protocols. This indicated an abundance shift of the present *C. difficile* strains and selective enrichment of the un-detected strains upon antibiotic treatment.

The genomes of the non-clinical strains isolated within this thesis were compared to clinical reference strains to identify clinical-related features with influence on strain virulence. Besides analysis of crucial virulence factors, the genome analyses addressed MGEs, and further included antibiotic resistances and accessory genes, which were all incorporated in pairwise

genome alignments. An overall trend of more ARGs, MGEs and accessory genes in the clinical strains was observed. The accessory genes mainly comprised specific functions such as transcription, which are correlated to higher strain virulence. This observation might further indicate the ability of the clinical strains to adapt more rapidly to environmental fluctuations. It is therefore advisable to increasingly implement non-clinical strains in comparative analyses to unravel *C. difficile* virulence progression. ARG, MGE and pan genome analyses were put into genomic context to identify conjunctions between specific elements. This revealed that accessory genes, which were identified to correlate with higher strain virulence, resided within MGEs. This connection highlights the relevance of MGEs in *C. difficile* virulence.

All non-clinical isolates and the clinical reference strains were subjected to phenotypic investigations on spontaneous and induced prophage activity. Active prophages were determined by isolation and sequencing of particle-protected DNA and mapping of sequencing reads onto the host genomes. Phages were either not induced, or induced with the secondary bile salt deoxycholate, which is a stressful component in the natural *C. difficile* habitat. Corresponding clinical and non-clinical strains showed no difference in prophage carriage or tolerance to deoxycholate, as determined by relative growth in the presence of varying concentrations. The sequencing data verified spontaneous prophage activity in all strains with one exception and revealed multiple active phages co-existing within the same host. The inducing effect of deoxycholate was also confirmed and observed for almost all phages, which underlines the relevance of natural conditions in investigations of *C. difficile* phage activity. Differences in increased phage activity upon induction could be observed between corresponding strains, but the data could not ensure a clinical relation. One corresponding strain pair however indicated a connection between the higher induction in the clinical strain and its higher reactivity determined in the comprehensive genome analyses. Fourteen novel phages in total were identified. Further, several regions with clear activity but missing phage identity were apparent in the sequencing read mapping. Different mechanisms might be responsible for the envelopment of these mobile DNA elements, such as phage-mediated transduction. One of these active DNA elements could be assigned to lateral transduction, which is the first description of this mechanism in *C. difficile*. These findings point to so far unraveled HGT strategies in *C. difficile* and encourage further investigations.

2 ABBREVIATIONS

ANI ..... Average Nucleotide Identity
ARG ..... Antibiotic Resistance Gene
ASV ..... Amplicon Sequence Variant
BLAST ..... Basic Local Alignment Search Tool
CDI ..... *Clostridioides Difficile* Infection
CDS ..... Coding Sequence
COG ..... Cluster of Orthologous Groups
CRISPR ..... Clustered Regularly Interspaced Short Palindromic Repeats
DCA ..... Deoxycholate
DNA ..... Deoxyribonucleic Acid
ECE ..... Extrachromosomal Element
GTA..... Gene-transfer Agent
HGT..... Horizontal Gene Transfer
ICE ..... Integrative and Conjugative Element
IME..... Integrative and Mobilizable Element
IS ..... Insertion Sequence
mgDNA ..... metagenomic DNA
MGE ..... Mobile Genetic Element
MLST ..... Multilocus Sequence Typing
NGS ..... Next Generation Sequencing
NPKM ..... Nucleotide activity Per Kilobase of exon model per Million mapped reads
OD ..... Optical Density
PCR ..... Polymerase Chain Reaction
rRNA ..... ribosomal Ribonucleic Acid
RT..... Ribotype
SPI ..... Spontaneous Prophage Induction
ST ..... Sequence Type

3 GENERAL INTRODUCTION

3.1 The pathogen *Clostridioides difficile*

The first description of *Clostridioides difficile*, back then termed *Bacillus difficilis*, arose from studies on the intestinal microflora of newborns in 1935 (Hall and O'Toole 1935). The Gram-positive, anaerobic, and spore-forming bacterium exhibited severe pathogenic properties in rabbits and guinea pigs, resulting in death in the majority of the study objects. Following, Hall and O'Toole proposed the pathogenic action of an exotoxin produced by *B. difficilis*. Only 40 years later, now termed *Clostridium difficile* was isolated from stool of patients suffering from pseudomembranous colitis after antibiotic treatment with Clindamycin (John G. Bartlett et al. 1978). Bartlett and colleagues also supported the assumption of an exotoxin being produced by the bacterium and responsible for the cytotoxic effect in the patients and hamster models. Following studies reinforced the connection between antibiotic treatment, mostly Clindamycin, and *C. difficile*-associated pseudomembranous colitis (Fekety et al. 1979; Humphrey et al. 1979; J G Bartlett et al. 1980), and identified two toxins, designated toxins A and B, as causative agent (Lyerly et al. 1982; Sullivan, Pellett, and Wilkins 1982). Later, *C. difficile* strains with a third, binary toxin CDT were discovered (Perelle et al. 1997).

In the early 2000s, a novel type of *C. difficile* strain with increased mortality emerged in association with the usage of fluoroquinolones antibiotics (Pépin et al. 2004; Loo et al. 2005; Muto et al. 2005). This strain, called NAP1/027/BI based on different molecular strain typing methods (Fatima and Aziz 2019), spread globally in the following years (Valiente, Cairns, and Wren 2014). Besides the emergence of hypervirulent strains, increasing antibiotic resistances in *C. difficile* intensified the global health threat (Spigaglia 2016). In 2016, Lawson and colleagues performed phenotypic, chemotaxonomic and phylogenetic characterizations of *Clostridium difficile* and proposed its reclassification to *Clostridioides difficile* (Lawson et al. 2016).

As *C. difficile* is a well-known globally pathogen, the majority of isolates were obtained from clinical sources in association with a *C. difficile* infection (CDI). However, isolates are also obtained from healthy, asymptomatic individuals (Ozaki et al. 2004), diverse animal species (Weese 2020), and various environmental surroundings, such as soil and water reservoirs (Janezic et al. 2016), compost (Dharmasena and Jiang 2018), or domestic and agricultural sites (Y. Zhou et al. 2021). Thousands of *C. difficile* strains have been isolated so far (17,118 GenBank entries of *C. difficile* genome assemblies on NCBI (Sayers et al. 2022),

on 23.09.2023), with only a small fraction originating from non-clinical sources. Regardless of the isolation source, *C. difficile* isolations are always performed with established protocols using selective cultivation methods with antibiotics (Levett 1985; Aspinall and Hutchinson 1992). These aspects lead to an isolation bias of *C. difficile* strains, which ultimately affects investigations and conclusions with relevance on the understanding of this pathogen and its virulence.

Several molecular techniques for *C. difficile* strain classification are established (Abad-Fau et al. 2023). Multilocus Sequence Typing (MLST) is based on allelic sequence variants of seven housekeeping genes (*adk*, *atpA*, *dxr*, *glyA*, *recA*, *sodA*, *tpi*), which are either extracted from whole-genome data, or obtained by amplification via PCR and amplicon sequencing (Griffiths et al. 2010). Combinations of the seven allelic sequences are assigned to sequence types (ST), forming monophyletic groups defined as clades 1 to 5 (Knight et al. 2021). Further, five cryptic clades C-I to C-V are known that include strains under discussion as novel *Clostridioides* species, which is based on genetic divergence determined via average nucleotide identity (ANI) analysis (Knight et al. 2021; Williamson et al. 2022). However, these strains can harbor *C. difficile* toxin homologues (Williamson et al. 2022) and can also cause CDI (Ramírez-Vargas et al. 2018). As another strain typing method, PCR ribotyping comprises the amplification of the intergenic spacer region between the 16S and 23S rRNA genes, which results in specific pattern of bands with different sizes due to multiple intergenic spacer regions with divergent lengths. This pattern is visualized by electrophoresis and can be assigned to an already known ribotype (RT) or otherwise defined as a novel RT. This technique is quite simple but bears the disadvantages of lacking discrimination between close relatives of the same RT and even between strains of different clades (Ducarmon et al. 2023).

Several studies approached detection of *C. difficile* strains, mostly from clinical samples to verify a CDI. These studies either focused on detection of toxin genes (Bélanger et al. 2003), or targeted widely distributed *C. difficile* genes but evaluated the PCR only against other clostridial bacteria and on samples with already enriched *C. difficile* (Lemee et al. 2004; Stone et al. 2016; van Rossen et al. 2021).

3.2 Underlying mechanism of a *C. difficile* infection

Advanced age and hospitalization are considerably involved in acquisition of a CDI, which usually manifests after antibiotic treatment, whereby a broad range of antibiotics can trigger CDI (Czepiel et al. 2019). Survival, colonization, and disease manifestation of *C. difficile* after

antibiotic treatment is enabled by several factors. First, *C. difficile* is commonly ingested as spores, which can survive outside of a host under unfavorable conditions such as oxygen exposure, and which are highly resistant to antimicrobial substances (Czepiel et al. 2019). Second, vegetative cells of *C. difficile* are also resistant to various antibiotics due to multidrug resistances (Spigaglia 2016). Third, the antibiotic treatment disturbs the normal, intestinal microflora, which provided protection from *C. difficile* outgrowth (Allegretti et al. 2016). The indigenous microbiota produces so-called secondary bile salts, the most prominent among them lithocholate and deoxycholate (further referred to as DCA), that impede vegetative growth of *C. difficile* (Sorg and Sonenshein 2008). These secondary bile salts are synthesized by microbial transformation of primary bile salts, which were produced by the liver to support absorption of fatty nutrients (Ridlon, Kang, and Hylemon 2006). Ingested *C. difficile* spores survive the gastric acid during passage of the stomach and reach the duodenum and jejunum, where primary bile salts such as taurocholate trigger *C. difficile* spore germination (Sorg and Sonenshein 2008). Germinated spores arrive at the cecum and colon, where secondary bile salts produced by an intact microbiota prevent *C. difficile* colonization (Sorg and Sonenshein 2008; Thanissery, Winston, and Theriot 2017). Protection against *C. difficile* colonization by secondary bile salts is repealed when the healthy microbiota is disrupted upon antibiotic treatment, allowing *C. difficile* to thrive (Allegretti et al. 2016). A schematic representation of the *C. difficile* life cycle is depicted in Fig. 1.

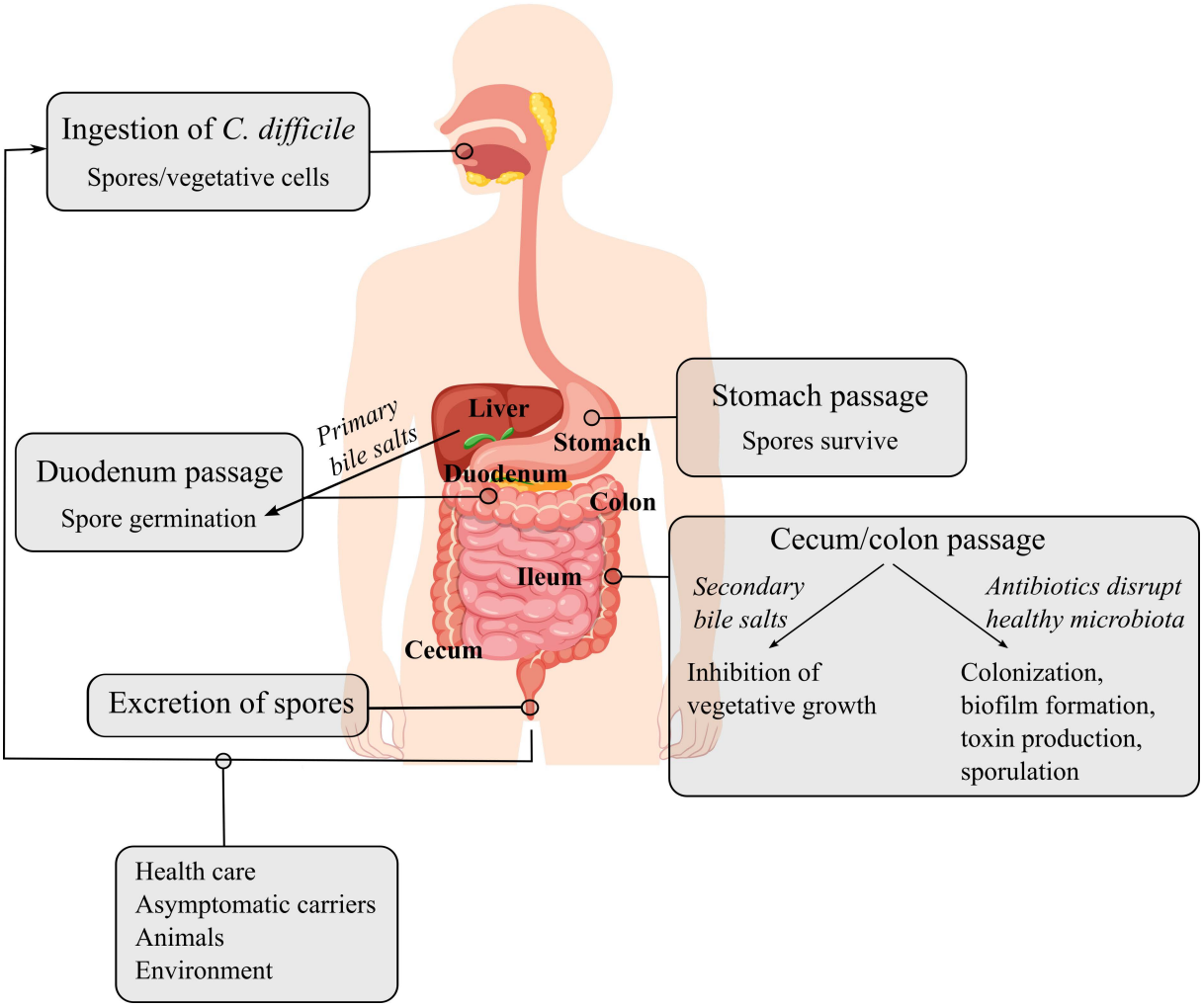


Figure 1. Schematic diagram of the *C. difficile* life cycle. All stages of the *C. difficile* life cycle that are relevant for disease establishment are depicted. Image of the human digestive system was taken from "Freepik.com" for modifications.

Carriage of *C. difficile* does not necessarily result in CDI with uniform symptoms and severity. Individuals can harbor *C. difficile* asymptomatically, or an established CDI varies in the clinical presentation between mild or moderate diarrhea and fatal colitis (Czepiel et al. 2019). This depends on the physiological state of the hosts (Czepiel et al. 2019), and on the *C. difficile* strains that can eminently vary in pathogenicity and virulence (Hunt and Ballard 2013). The major virulence factors of *C. difficile* are toxins A and B (also called TcdA and TcdB), which are encoded in the 19.6 kb big Pathogenicity Locus (PaLoc) that further harbors two regulatory genes *tcdR* and *tcdC*, and the holin-like protein encoding gene *tcdE* (Fig. 2) (Monot et al. 2015). TcdR and TcdC are up- and down-regulator, respectively, of TcdA and TcdB gene expression (Mani and Dupuy 2001; Matamouros, England, and Dupuy 2007), while TcdE is relevant for toxin secretion (Govind and Dupuy 2012). The toxic effects of TcdA and TcdB are synergistic and lead to the disruption of cell structural integrity of the intestinal mucosa (Chandrasekaran and Lacy 2017). However, certain *C. difficile* strains were discovered without

encoded TcdA (A^-B^+) that still caused severe CDI comparable to infections with strains possessing both toxins (A^+B^+) (Drudy, Fanning, and Kyne 2007). A modified TcdB protein that compensates the missing function of TcdA was identified as disease-causing agent (Chaves-Olarte et al. 1999). Besides toxins A and B, another toxin was identified in clinically relevant *C. difficile* strains, the binary toxin CDT comprising CdtA and CdtB (Gerding et al. 2014). These are encoded in the Cdt Locus (CdtLoc) in conjunction with the regulatory gene *cdtR* (Fig. 2). The relevance of CDT in *C. difficile* pathogenesis is still under research, but action of the binary toxin is associated with hypervirulent strains and severe disease outcome, especially in combination with toxins TcdA and TcdB ($A^+B^+CDT^+$) (Martínez-Meléndez et al. 2022). Non-toxigenic *C. difficile* strains without toxin genes or only truncated versions insufficient for disease manifestation are also known (Natarajan et al. 2013). They were investigated as potential protection from colonization by toxigenic *C. difficile* strains (Gerding, Sambol, and Johnson 2018), but were also shown to acquire the PaLoc from toxigenic strains and thereby gain the ability of toxin production (Brouwer et al. 2013).

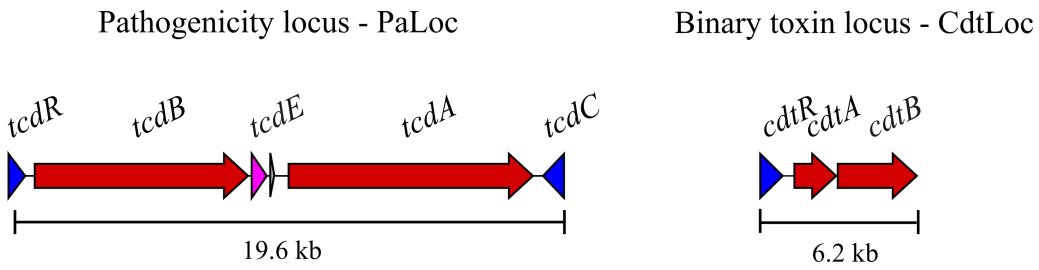


Figure 2. *C. difficile* toxin loci. The PaLoc (left) comprises the genes of toxins A and B (*tcdA* and *tcdB*), the regulatory genes *tcdR* and *tcdC*, and the holing-like gene *tcdE*. The CdtLoc (right) comprises the regulatory gene *cdtR* and the toxin genes *cdtA* and *cdtB*. Toxin genes are colored in red, regulatory genes in blue, the holin-like gene *tcdE* in purple.

Besides the disease-causing toxins as major virulence factors, several other features of a *C. difficile* strain determine its virulence. As already pointed out, the efficiency of sporulation in response to adverse environmental conditions is crucial for the survival of *C. difficile*, so that spores can survive until reaching favorable conditions in a new host (Czepiel et al. 2019). Once spores could germinate in an adequate environment, vegetative cells benefit from flagella-based motility for further host invasion and colonization (Baban et al. 2013). Motility promotes adherence and colonization, however also non-motile strains can cause severe disease symptoms (Baban et al. 2013), which emphasizes the interaction of various virulence factors in CDI. Surface proteins involved in adherence are crucial key players for both *C. difficile* colonization and biofilm formation (Buddle and Fagan 2023). They contribute to resistance against harmful substances, which is further supported by aggregation to bacterial biofilms

(Buddle and Fagan 2023). Bacterial cells associated in a biofilm can persist in the host by protection against various adverse influences, such as antibiotics or secondary bile salts (L. R. Frost, Cheng, and Unnikrishnan 2021). Consequently, biofilm formation facilitates infection recurrence after CDI therapy. All these virulence factors are often under investigation to elucidate the phenomenon of hypervirulent *C. difficile* strain. In the context of elucidating hypervirulence, toxin gene sequence variants were identified. While different variants of the toxin TcdB itself were associated with increased disease severity (Lanis et al. 2013), mutations in regulators for toxin gene expression, TcdC (Curry et al. 2007) and CdtR (Q. Dong et al. 2023), were identified as well.

3.3 The mobile genome of *C. difficile*

One prominent characteristic of *C. difficile* is its mobile genome. About 11% of the genome are estimated to be composed of mobile genetic elements (MGE) (Sebaihia et al. 2006; Mullany, Allan, and Roberts 2015). MGEs are DNA sequences that can move within one genome or be exchanged between different cells (L. S. Frost et al. 2005). They are fundamental agents of horizontal gene transfer (HGT), which enable the acquisition of advantageous genes that e.g. confer antibiotic resistance or encode virulence factors (de la Cruz and Davies 2000). Thus, MGEs are substantially involved in the worldwide increase of pathogen virulence and multidrug resistances, which are both critical health challenges nowadays (Bonneaud and Longdon 2020; EClinicalMedicine 2021) and observed in *C. difficile* progression (Hunt and Ballard 2013; Spigaglia 2016). *C. difficile* genomes harbor a variety of MGEs, such as plasmids, insertion sequences (IS), conjugative and mobilizable transposons, and prophages (Mullany, Allan, and Roberts 2015). These MGEs are involved in conjugation or transduction as the underlying mechanisms of HGT (L. S. Frost et al. 2005). Conjugation allows the exchange of conjugative plasmids or conjugative transposons (also termed integrative conjugative elements – in short ICE) via cell-cell contact (L. S. Frost et al. 2005). Interestingly, Brouwer and colleagues showed that the PaLoc can be mobilized via conjugative transfer as well (Brouwer et al. 2013). As a consequence, this action of HGT does not simply influence strain virulence but transforms a non-toxigenic *C. difficile* strain into the disease-causing pathogen. HGT via transduction is also mediated by lysogenic bacteriophages integrated into the bacterial chromosome as prophages (L. S. Frost et al. 2005).

3.4 Bacteriophages in general

Bacteriophages, in short termed phages, are bacterial viruses and estimated as the most abundant entity on earth, which are classified based on gene material or morphology (Dion,

Oechslin, and Moineau 2020). Phage genomes consist of ssDNA, dsDNA, ssRNA, or dsRNA. The phage morphology mainly differs in the DNA-containing capsid, whereby some phage families additionally possess a tail as DNA-injecting apparatus. Tailed and dsDNA phages of the order *Caudovirales* represent the majority of identified and described phages to date. They are morphologically classified by tail appearance and contractility into *Podoviridae* with a short non-contractile tail, *Myoviridae* with a long contractile tail, and *Siphoviridae* with a long non-contractile tail (Fig. 3).

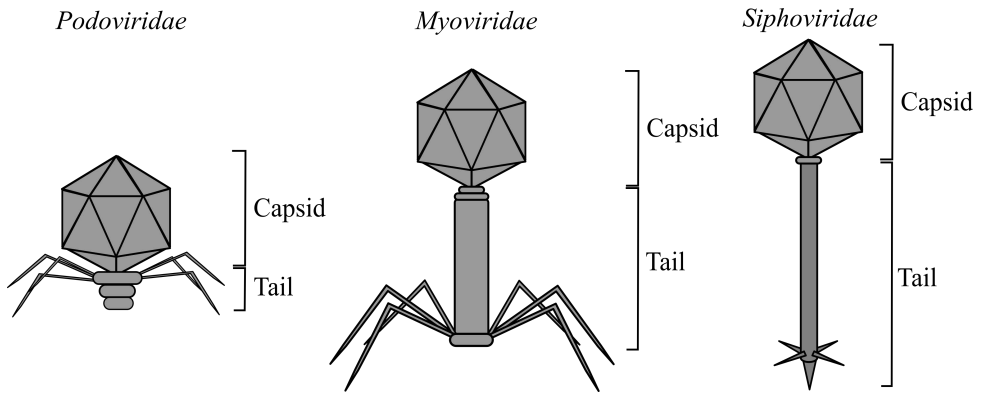


Figure 3. The three morphologies of tailed bacteriophages belonging to the order *Caudovirales*.

Phages follow either the lytic (virulent) or the lysogenic (temperate) lifestyle (Fig. 4) (Olszak et al. 2017). In both styles, phages recruit the bacterial metabolism for their genome replication and production of phage particles. Therefore, phage DNA is injected into a bacterial host cell. Phages are very specialized for host attachment, resulting in a considerable small host range (Weinbauer 2004). After phage particle assembly, phage-mediated host cell lysis releases the phage progeny. Phage-mediated cell lysis can be detected in solid bacterial cultivation by the appearance of clear zones in so-called plaque assay, which allows the isolation of single phages from separate plaques (Kropinski et al. 2009). While lytic phages follow this process straight-forward, lysogenic phages can integrate their gnome into the bacterial chromosome, which forms a prophage as intermediate state (Olszak et al. 2017). The prophage is replicated alongside the bacterial chromosome and remains in this dormant state until its induction, which is often triggered by stressful agents challenging the host, such as UV radiation, DNA damage, and antibiotics. However, spontaneous induction of prophages is a common mechanism, and although previously assumed to be deleterious for the host population, it is now considered as beneficial at a particular rate, depending on the host and its environment (Nanda, Thormann, and Frunzke 2015; Cortes, Krog, and Balázsi 2019). Upon induction, the prophage enters the lytic life cycle, including prophage excision from the bacterial chromosome (mediated by phage

integrase and excisionase) prior to genome replication and packaging of phage DNA into assembled phage particles (mediated by phage terminase).

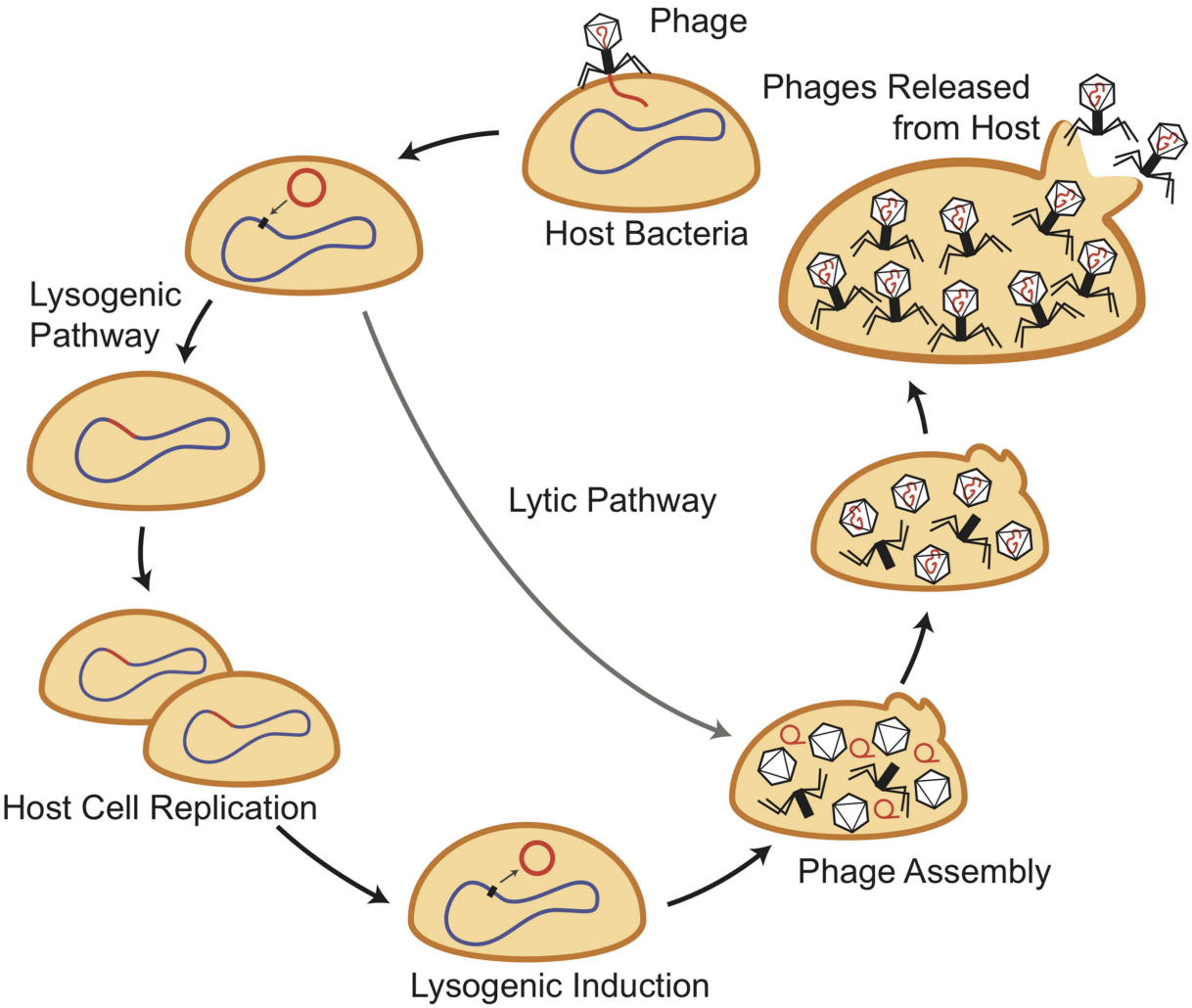


Figure 4. Life cycle of lysogenic and lytic bacteriophages. The phage genome is colored in red, while the bacterial is colored in blue. Illustration was taken from (Chiang, Penadés, and Chen 2019).

Prophages can impact their host in many different ways. Besides carrying genes advantageous for the bacterial host, like antibiotic resistance genes (ARG) or virulence factors, prophages can intervene in bacterial gene expression or biofilm formation (Fortier and Sekulovic 2013). They might also provide immunity of the host against closely related phages (Labrie, Samson, and Moineau 2010). A second form of bacterial immunity against invading phages is the CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR-associated proteins) system, a prokaryotic mechanism for digestion of foreign DNA (Koonin and Makarova 2009). This mechanism involves CRISPR sequences in the bacterial genome representing foreign, previously invading DNA, which guides the enzymatic digestion of this foreign DNA upon repeated invasion.

Phages are also important agents of HGT via transduction, in which DNA of the bacterial host cell is packaged into the phage capsid and transferred to another host upon phage infection. Three different transduction modes are known to date: generalized, specialized and lateral (Fig. 5) (Chiang, Penadés, and Chen 2019). All three mechanisms are performed by phages with the DNA-packaging strategy "headful", which comprises DNA packaging starting at a specific sequence and filling the capsid to its capacity. In generalized transduction, homologous packaging sites located anywhere in the bacterial chromosome are recognized by the phage terminase, so that bacterial instead of phage DNA is packaged into phage capsids. In specialized transduction, phage genome excision mistakenly occurs in the bacterial chromosome, and phage and bacterial DNA are jointly replicated and encapsidated. Both generalized and specialized transduction result from aberrant steps in the phage's life cycle, and are therefore infrequent among all produced phage particles, which can be identified by sequencing coverage (Kleiner et al. 2020). Lateral transduction was discovered quite recently (Chen et al. 2018). In contrast to generalized and specialized transduction, it is assumed to be no misstep but an elemental trait in the phage life cycle (Chiang, Penadés, and Chen 2019). This trait includes replication of the phage genome and DNA packaging into phage capsids prior to excision from the bacterial chromosome. As a consequence, adjacent host DNA is replicated and packaged alongside the phage genome at high frequencies, which allows the transfer of bacterial genes comprising even DNA sequences of hundreds of kilobases. This transduction frequency is likewise detectable by sequencing-based analysis (Kleiner et al. 2020).

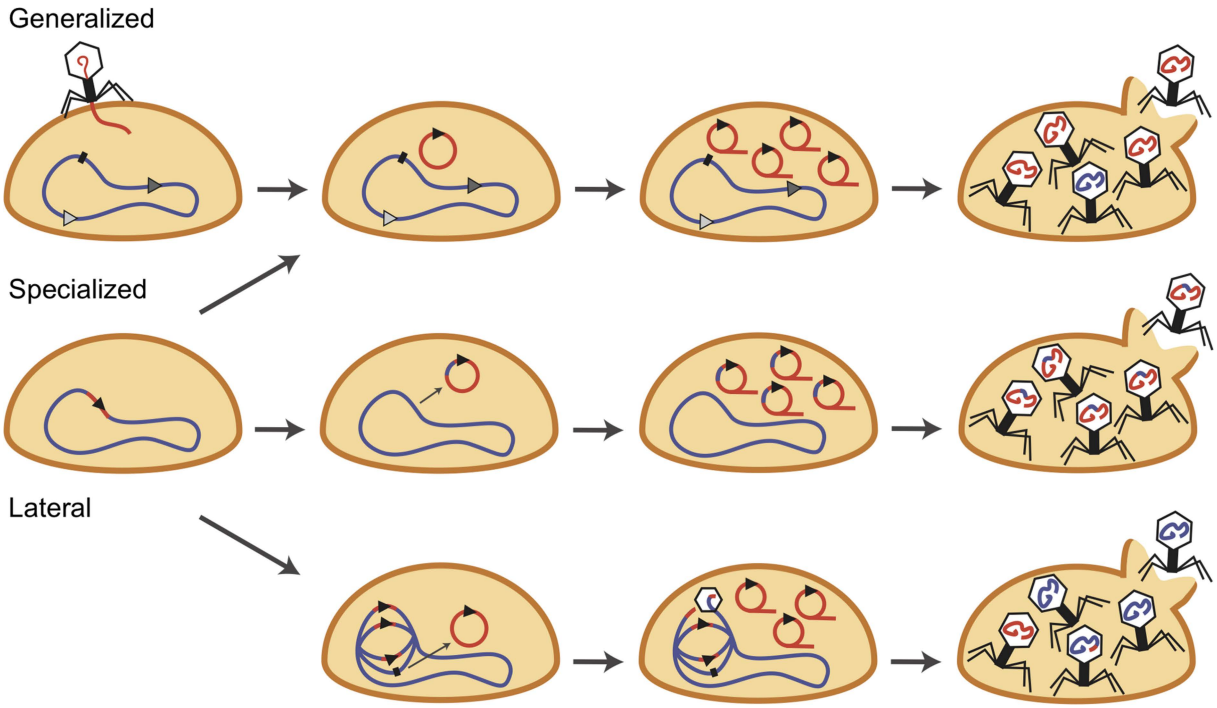


Figure 5. Phage-mediated generalized, specialized, and lateral transduction. The phage genome is depicted in red, the bacterial in blue. Black triangles represent phage packaging sites, while gray triangles represent homologous packaging sites in the bacterial chromosome. Illustration taken from (Chiang, Penadés, and Chen 2019).

3.5 Bacteriophages in *C. difficile*

To date, 43 genomes of *C. difficile* phages are publicly available (Table S1). All isolated phages are lysogenic and belong to the tailed phage-families of *Myoviridae* and *Siphoviridae* (Heuler, Fortier, and Sun 2021). In recent years, phage research in *C. difficile* has gained in importance. On the one hand, the increasing number of multidrug resistant strains aggravates successful CDI treatment, so that growing interest in alternatives resulted in the focus on phage therapy (Nale et al. 2016). On the other hand, different studies identified the effect of lysogenic phages on virulence-related aspects of *C. difficile*, such as influencing toxin gene expression (Goh, Chang, and Riley 2005; Govind et al. 2009), and transducing antibiotic resistance genes (Goh et al. 2013). Novel phages of *C. difficile* are usually isolated after induction with mitomycin C and less often with UV radiation (Heuler, Fortier, and Sun 2021) or quinolone antibiotics (Meessen-Pinard, Sekulovic, and Fortier 2012; Nale et al. 2012), all of which are no natural stressors that *C. difficile* would face during colonization and infection. Only quinolone antibiotics as common therapeutic treatment could actually interact with *C. difficile in vivo* (Pépin et al. 2005). However, spontaneous phage release has also been observed in *C. difficile* (Meessen-Pinard, Sekulovic, and Fortier 2012; Nale et al. 2012). Further, multiple prophages in the same host can be simultaneously active (Nale et al. 2012; Hargreaves et al. 2013).

Detection of active phages was often achieved by transmission electron microscopy examination only (Fortier and Moineau 2007; Shan et al. 2012; Hargreaves et al. 2013), which is a quite insensitive technique and results in overlooking of low abundant phages (Hertel et al. 2015). Furthermore, single phages were obtained from plaque assays and propagated in another bacterial host before phage particle isolation and phage DNA isolation for sequencing (Ramírez-Vargas, Goh, and Rodríguez 2018; Whittle et al. 2022). This approach however does not represent actual phage activity in the original host. Experimental setup for analysis of the entire active virome of *C. difficile* strains is so far not established.

3.6 Aim of the thesis

This thesis aimed to investigate genomic characteristics with focus on MGEs that can influence *C. difficile* virulence. A comparison between clinical and non-clinical strains was thereby intended since data on this kind of examination is still scarce.

First, non-clinical strains should be isolated from environmental samples, with focus on developing an isolation protocol without the selective antibiotic-based cultivation to omit a potential isolation bias. To support the isolation approach, a detection PCR should be established to evaluate *C. difficile* presence in environmental DNA, which allows identifying promising isolation sources. These non-clinical strains should be compared to clinical reference strains of the same ST to assess virulence-related differences by genome-based examinations, comprising the identification of diverse MGEs and detailed pan genome analyses. Lastly, phenotypic experiments on prophage activity in the diverse strains should provide insight into this particular MGE of *C. difficile*. Research on spontaneous prophage activity in *C. difficile* is quasi non-existent since studies mainly employ phage induction with UV radiation, mitomycin C or other antibiotics. These stressors do not represent a natural challenge for *C. difficile*, so that we also wanted to examine phage induction by a natural stressor, the secondary bile salt DCA. Phage activity should further be examined by next generation sequencing (NGS) of the entire strains' virome with subsequent mapping of sequencing reads on the host genome, which allows to identify the active prophage region as well as phage abundance based on read coverage. This is a new approach in *C. difficile* phage research, where active phages are mainly identified by electron microscopy and sequencing of single phages. Prophage activity should also be investigated regarding clinical-related characteristics with potential to influence strain virulence.

4 RESULTS

4.1 Direct and culture-independent detection of low-abundant *C. difficile* in environmental DNA via PCR

Miriam A. Schüler, Dominik Schneider, Anja Poehlein, Rolf Daniel[#]

Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, University of Göttingen, Göttingen, Germany

Running Head: *C. difficile*-specific detection PCR

#Address correspondence to Rolf Daniel, Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Georg-August University Göttingen, D-37077 Göttingen, Germany, Phone: +49-551-3923827, Fax: +49-551-3924619, Email: rdaniel@gwdg.de

Submitted to *Applied and Environmental Microbiology* (*American Society for Microbiology*) on 24 July 2023, ongoing correction of revision.

Published online on bioRxiv.org on 26 July 2023: https://doi.org/10.1101/2023.07.24.550392.

Author Contributions

Conceptualization: **MS**, AP, RD

Experiments: **MS**

Data analysis: **MS**, DS

Results interpretation: **MS**

Manuscript writing: **MS**

Manuscript review & editing: **MS**, DS, AP, RD

Abstract

Clostridioides difficile represents a major burden to public health. As a well-known nosocomial pathogen whose occurrence is highly associated with antibiotic treatment, most examined *C. difficile* strains originated from clinical specimen and were isolated under selective conditions employing antibiotics. This suggests a significant bias among analysed *C. difficile* strains, which impedes a holistic view on this pathogen. In order to support extensive isolation of *C. difficile* strains from environmental samples, we designed a detection PCR that targets the *hpdBCA*-operon and thereby identifies low abundances of *C. difficile* in environmental samples. Amplicon-based analyses of diverse environmental samples demonstrated that the designed PCR is highly specific for *C. difficile* and successfully detected *C. difficile* despite its absence in general 16S rRNA gene-based detection strategies. Further analyses revealed the potential of the *hpdBCA* detection PCR sequence for initial phylogenetic classification, which allows assessing *C. difficile* diversity in environmental samples via amplicon sequencing. Our findings furthermore showed that *C. difficile* strains isolated under antibiotic treatment from environmental samples were originally dominated by other strains according to detection PCR amplicon results. This provided evidence for selective cultivation of under-represented but antibiotic-resistant isolates. Thereby, we revealed a substantial bias in *C. difficile* isolation and research.

Importance

Clostridioides difficile is mainly responsible for hospital-acquired infections after antibiotic treatment with serious morbidity and mortality worldwide. Research on this pathogen and its virulence therefore focused on bacterial isolation from clinical specimen under antibiotic treatment, which implies a substantial bias in isolated strains. Comprehensive studies however require an unbiased strain collection, which is accomplished by isolation of *C. difficile* from diverse environmental samples and avoiding antibiotic-based enrichment strategies. Thus, isolation can significantly benefit from our *C. difficile*-specific detection PCR, which rapidly verifies *C. difficile* presence in environmental samples and further allows estimation of the *C. difficile* diversity by using NGS.

Introduction

Clostridioides difficile is a widely known pathogen, responsible for a majority of nosocomial infections worldwide with significant morbidity and mortality, especially after preceding antibiotic treatment (1). Symptom severity of a *C. difficile* infection not only depends on the individual physiological constitution but also on the infecting strain, as *C. difficile* virulence differs within the species (2, 3). Diverse characteristics of *C. difficile* are under research to elucidate its multifaceted pathogenicity (4). Thus, comprehensive investigations of the species *C. difficile* are essential to pinpoint the different characteristics of this pathogen. Most *C. difficile* strains isolated and examined to date originated from clinical specimens in the context of infections. However, this bacterium also resides in the intestinal microflora of healthy individuals and different animal species as well as environmental habitats like soil and water reservoirs (5). Consequently, there is a bias towards clinical-associated strains in *C. difficile* research (6, 7). Some researchers indeed already isolated *C. difficile* from the environment and thereby obtained multiple strains from one sample, including novel sequence types or even cryptic *C. difficile* strains that are classified as different genomospecies *Clostridioides* sp. by now (5, 6). However, further isolations from diverse habitats would contribute to a holistic view on this pathogen. Those broad-ranging isolation attempts would be promoted by initial assessment of environmental samples to harbor *C. difficile* before elaborate and time-consuming isolation experiments are performed. Common processes for *C. difficile* isolation last days up to few weeks and start with antibiotic-based cultivation to select for this presumably low-abundant species in complex environmental microbial communities (6). Selective cultivation is followed by sub-culturing and identification of isolates via e.g., MALDI-TOF, enzymatic immunoassays or 16S rRNA gene sequencing (5, 8, 9). The initial selective cultivation might eliminate environmental *C. difficile* strains that do not harbor the respective antibiotic resistance. For instance, a review from 2016 (10) summarized data from diverse studies on antibiotic resistances among clinical *C. difficile* isolates and showed that not all strains were resistant to the applied antibiotics, e.g. to cefoxitin that is used in *C. difficile* isolation attempts (11).

In this context, we designed a *C. difficile*-specific detection PCR for direct application on environmental DNA (metagenomic DNA, mgDNA), that produces a distinct amplicon upon *C. difficile* presence – even for an abundance below detection levels in general 16S rRNA gene amplicon-based community analysis (12). To our knowledge, this type of culture-independent detection directly performed on environmental DNA has not been established in *C. difficile*

research so far. Previous PCR approaches for *C. difficile* detection worked on single isolates and pre-enriched samples (13–15) or targeted the toxin genes, resulting in strains lacking these genes or possessing different gene sequences are not detected (6, 16). PCR specificity was assessed by next-generation-sequencing (NGS) of amplicons that were produced by detection PCR on diverse mgDNA samples to prove *C. difficile* origin. Further, PCR sensitivity was demonstrated by comparison to the amplicon-based 16S rRNA gene analysis. Besides establishing and demonstrating successful performance of our PCR on diverse environmental samples, we examined the detection sequence as potential phylogenetic determinant via average-nucleotide-identity analysis. The most common methods for phylogenetic classifications of *C. difficile* strains are ribotyping and multilocus sequence typing (MLST) (17, 18). The former one involves amplification of the 16S-23S rRNA gene intergenic spacer region, which results in a genotype-specific pattern. This pattern needs to be compared to already known ribotypes for identification, which is not feasible for all laboratories. Moreover, ribotypes can be polyphyletic and include various STs (19). The principle of MLST is based on sequence alleles of seven housekeeping genes, which can be determined via PCR and sequence analysis of all seven genes or whole-genome sequencing. Allele combinations are designated via the MLST database (http://pubmlst.org/) to sequence types (ST), which further group into monophyletic clades of C1 to C5 in the narrower sense of *C. difficile*, and additionally of cryptic clades C-I to C-V (6, 20). As such, MLST altogether exceeds ribotyping in phylogenetic classification of *C. difficile* strains. In contrast to MLST, phylogenetic classification with the here described detection PCR involves less effort and allows a superficial estimate of *C. difficile* strain diversity in environmental samples, which is so far not established in *C. difficile* research.

Materials and methods

C. difficile detection PCR target and primer design

To establish a *C. difficile*-specific detection PCR, literature research for species-specific genes or sequences were conducted and yielded amongst others the *hpdBCA* operon as potential target (21, 22). This operon encodes the *p*-hydroxyphenylacetate decarboxylase that catalyzes the production of *para*-cresol, an uncommon metabolic trait amongst bacteria (23).

The *hpdBCA*-nucleotide sequence was confirmed to be entirely present in all complete, unique *C. difficile* genomes deposited at the National Center for Biotechnology Information (NCBI (24); assessed 30 August 2022) using BLAST+ (v2.12.0) (25) with default parameters and the *hpdBCA*-sequence of *C. difficile* strain CD630 (NCBI accession CP010905.1, locus tags CDIF630_00272-00274) as query. A BLAST analysis (26) in megablast mode of the

CD630 operon was done as well to check for general and significant occurrence of this sequence.

The primer design focused on several criteria: (I) no binding to possible non-*C. difficile* sequences while (II) binding to all *C. difficile* strains and additionally including so-called cryptic ones. These cryptic *Clostridioides* are closest related to *C. difficile* (6, 7) and further show >99.87% 16S rRNA gene sequence similarity to *C. difficile*, which would even classify them as same species (27). (III) Yielding a PCR product of <500 bp length for amplicon analysis with NGS. (IV) Fulfilling standard criteria such as similar melting temperature, GC clamp, no secondary structure and no self-/cross-dimer formation (http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html).

(I) The *hpdBCA*-sequence of CD630 (3,913 bp) was used as reference in a blastn analysis excluding *C. difficile* (taxid 1496) to retrieve all similar non-*C. difficile* sequences. The alignment of these sequences to the CD630 reference was inspected using the NCBI Multiple Sequence Alignment Viewer (MSA Viewer, v1.22.0, https://www.ncbi.nlm.nih.gov/tools/msaviewer/) for promising primer regions that show least similarity between the CD630 and the non-*C. difficile* sequences, with special focus on the 3'-primer end as being the crucial part for amplification (28) (Fig. S1)**.** (II) Potential primer regions from step (I) were examined regarding conservation across *C. difficile*. Therefore, all aligned *hpdBCA* sequences from the previous megablast analysis against complete *C. difficile* genomes at chromosome level were downloaded and aligned in AliView v1.26 (29). (III) Corresponding forward/reverse primers for defined primer candidates were searched to amplify a 400-500 bp sequence while fulfilling criteria (I) and (II). Standard primer criteria such as GC clamp, similar melting temperature and no self- or primer-dimer formation (IV) were examined with the online tools of Thermo Fisher Scientific (Multiple Primer Analyzer, https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html) and Sigma-Aldrich (OligoEvaluator http://www.oligoevaluator.com/LoginServlet). Performance comparison of all potential primers resulted in the final pair of CDIF_cresol_3F (forward 5'-GAAAAGGTGGGTTCCATATTCAATATAATG) and CDIF_cresol_3R (reverse 5'-CCTTCTAATTGCTTTTGACTACTCATTAAACAC). Both primers aligned to all analyzed *C. difficile* strains with 100% coverage and predominantly 100% percent identity except for C5 strains with one mismatching nucleotide (forward 5'-position: 8, reverse 5'-position: 11; see

also Fig. S2), but still allowing efficient primer annealing. The cryptic *Clostridioides* strains showed also a coverage of 100% of the primers but percent identities lay between ~91% and ~96%, corresponding to one to four mismatching nucleotides.

Bioinformatic analyses of detection PCR primers and sequence

***C. difficile* specificity analysis.** The detection PCR primers were examined *in silico* for *C. difficile* specificity by blastn analysis with default settings excluding *C. difficile* (taxid 1496) (done on 15 September 2022). *C. difficile* specificity of the corresponding detection PCR product was likewise assessed by blastn analysis (done on 15 September 2022), using the sequence retrieved from CD630 via *in silico* PCR. The primer sequences were thereby omitted so that only the 401 bp-long sequence between the primers was used as BLAST query.

Detection PCR sequence as phylogenetic marker. The 401 bp-long detection sequence was further analyzed for its potential to phylogenetically classify *C. difficile* isolates. Phylogenetic examinations at whole genome and detection sequence level were performed using all complete *C. difficile* genomes available at NCBI by average-nucleotide-identity (ANI) calculations with PYANI v0.2.11 (30) in default settings and MUMmer3 (31) for sequence alignment (ANIm). ANIm results were visualized in Rstudio with the gplots-implemented tool heatmap.2. Phylogenetic assignment of the analyzed strains in form of MLST determination was performed with FastMLST v0.0.15 (32).

Alignment of unique detection sequences including the primers amongst the above analyzed complete *C. difficile* genomes corresponding to specific MLST types/clades were visualized in Jalview v2.11.2.0 (33) for comparison of MLST/clade-associated nucleotide deviations.

Isolation of mgDNA from environmental samples

Diverse environmental samples (n=46, Table 1, detailed information in Table S1) such as soil, compost or feces, of which the majority had been in contact to animals, were collected in sterile canonical falcon tubes and stored at 4 °C upon arrival in the laboratory. The mgDNA was extracted with the DNeasy PowerSoil Pro kit (Qiagen, Hilden, Germany) according to protocol. DNA concentration was determined with the Qubit 3.0 Fluorometer (Life Technologies Thermo Fisher Scientific, Waltham, MA, USA) using the BR dsDNA assay kit and DNA purity was assessed via NanoDrop ND-1000 (Peqlab VWR International GmbH, Darmstadt, Germany) measurement.

Table 1. Environmental samples used in this study with sample ID and description.

ID	Description	ID	Description
A1	Horse turd	**P2**	Horse turd (mother to J2)
B1	Sludge biogas fermenter	**P3**	Horse turd
D1	Muddy trickle	**PA**	Horse turd
Goe1	Compost – disposal company	**TS10a**	Wild boars wallow
Goe2	Mud dump	**TS10b**	Wild boars dung pile
J1	Meconium foal 1	**TS10c**	Wild boars dung
J2	Meconium foal 2	**TS1a**	Donkeys dung pile
J3	Horse turd foal 3	**TS1b**	Donkeys dung
J4	Horse turd foal 4	**TS2ab**	Red cattle dung pile + dung
J5	Horse turd foal 5	**TS3**	Mud from outdoor area farm
J6	Horse turd foal 6	**TS4ab**	German saddlebacks dung pile + dung
J7	Horse turd foal 7	**TS5a**	Exmoor ponies dung pile
K1	Wild boars dung	**TS5b**	Exmoor ponies dung
K2	Wild boars dung pile	**TS6**	Mud from swampy region (outside enclosure)
K3	Wild boars wallow	**TS6a**	Mud in fallow deer area
K4	Fallow deer dung	**TS7a**	Reindeer dung pile
K5	Fallow deer dung pile	**TS7b**	Reindeer dung
MA	Horse dung	**TS8a**	Woolly pigs wallow
MG1	Slurry	**TS8bc**	Woolly pigs dung pile + dung
MG2	Dung pile calf stable	**TS9a**	Pot-bellied pigs wallow
MG3	Mud from cow run	**TS9b**	Pot-bellied pigs dung pile
MG4	Sludge from biogas fermenter	**TS9c**	Pot-bellied pigs dung
P1	Horse turd	**TSA**	Muddy trickle

PCR protocols

Detection PCR on genomic DNA (g-PCR). Optimal detection PCR conditions were initially assessed on genomic DNA of in-house *C. difficile* strains belonging to different MLSTs/clades (ST1/C2, ST3/C1, ST8/C1, ST11/C5, ST340/C-III). PCR was performed with DreamTaq polymerase (Thermo Fisher Scientific) according to manufacturer's recommendations using 0.2 mM of each primer and 50 ng DNA template per 50 µl PCR reaction mixture. Established PCR cycling conditions comprised initial denaturation at 95 °C for 3 min, followed by 28 cycles of denaturation at 95 °C for 30 s, primer annealing at 61 °C for 30 s and extension at 72 °C for 40 s, and final extension at 72 °C for 5 min.

Detection PCR on metagenomic DNA (mg-PCR). The previously established g-PCR parameters were adapted for detection on mgDNA. Considering that mgDNA is a diverse

mixture of genetic material with presumably minor abundance of *C. difficile* DNA, we aimed to support primer annealing at a low abundant template. Adaptations included doubled DNA template amount of 100 ng per 50 µl PCR reaction mixture, increased primer concentration of 0.3 mM, and modified cycling conditions with 35 cycles comprising a specific time decrement program during primer annealing for 90 s that decreases by 2 s each cycle.

Analysis of bacterial community composition via 16S rRNA gene amplicon sequencing

For sensitivity assessment of the detection PCR compared to the common investigation of bacterial community composition, we conducted a 16S rRNA gene analysis of all 46 samples (12). Amplification of the V3 to V4 region of the 16S rRNA gene was performed in triplicates with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) as stated in the protocol using GC buffer, 5% DMSO, 0.2 mM of each primer S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21(12) with attached Illumina amplicon adapter overhang sequences (34) and 25 ng mgDNA template per 50 µl PCR reaction mixture. PCR triplicates were pooled equimolar, purified, and sequenced as described by Berkelmann et al. (35).

Illumina raw-reads were processed with an in-house pipeline as follows and thereby parallelized with GNU parallel 20190322 (36): raw-read quality filtering with fastp v0.23.2 (37) (extended default settings as in (38)), merging of paired-end reads with PEAR v0.9.11 (39), and clipping of primer sequences using cutadapt v3.2 (40). VSEARCH v2.12.0 (41) was used to perform size sorting and filtering (>300 bp), read dereplication, denoising with UNOISE3 (minimum abundance option: minsize 8) (42), chimera removal with UCHIME3 (43) *de novo* and reference-based against the SILVA SSU 138.1 NR database (44), and final read mapping to amplicon sequence variants (ASV) at 100% identity to create an abundance table. ASV taxonomies were assigned with BLAST 2.9.0+ (45) using the SILVA SSU 138.1 NR database with a 70% identity cut-off, and added to the abundance table with BIOM tools v1.0 (46). Final amplicon data was further processed in R by applying taxonomic boundaries according to sequence identities as defined by Yarza et al. (47), and removing spurious sequences by applying a 0.25% cutoff filter on each sample according to Reitmeier et al. (48) before visualization with the R package ggplot2 (49). Alpha diversity was calculated and visualized with the R packages ampvis2 (50) and ggplot2 (49) using “amp_alphadiv” and “amp_ordinate” functions.

Sequencing of detection PCR amplicons

Sequencing of PCR products from positive detections on diverse mgDNA samples was performed to verify *C. difficile* origin and, thus, assess PCR specificity and sensitivity. Sequencing was done with *C. difficile* isolates obtained from the analyzed environmental samples as indication and control for correct processing of amplicon data. For Illumina sequencing, the Illumina amplicon adapter overhang sequences (34) were attached to the 5'-end of the detection primers. Amplicon PCR was performed in triplicates with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) according to manufacturer's indications using GC buffer and no DMSO. The PCR parameters were used as determined in the above-mentioned mg-PCR protocol, including thermocycling conditions. All corresponding replicates were pooled equimolar before purification with the GeneRead Size Selection kit (Qiagen) as recommended by the manufacturer, including two successive purification rounds with repeated elution for highest amplicon recovery and elimination of primer-dimers. First and second purification round were eluted in 90 µl and 15 µl EB buffer, respectively. DNA concentration was determined in the Qubit 3.0 Fluorometer (Life Technologies) using the HS dsDNA assay kit. Amplicon quality was assessed employing a Bioanalyzer and the DNA 1000 kit as recommended by the manufacturer (Agilent Technologies, Waldbronn, Germany).

Raw reads were processed as described for 16S rRNA gene amplicon sequencing with few adaptations. The ASV size filter was set to ≥380 bp and chimeras were removed *de novo*. ASV taxonomies were assigned with BLAST against the NCBI BLAST database (as from 5 October 2022). MLST assignment of each ASV based on its BLAST identity (GenBank accession) was performed using the PubMLST database (51).

Final amplicon data was further processed in R. Based on data situation and according to Reitmeier et al. (48), an appropriate cutoff filter of 0.25% as employed in 16S-amplicon data processing was finally applied, and data was visualized with the R package ggplot2 (49).

Data records

The data was deposited at the National Center for Biotechnology Information and can be found under the BioProject accession numbers PRJNA991070, PRJNA991304, and PRJNA991297. These BioProjects contain the raw sequences of 16S rRNA gene amplicons (SRR25194157-SRR25194209) and of detection PCR amplicons (SRR25186671-SRR25186686 and SRR25132251-SRR25132254).

Results

Initial specificity assessment of *C. difficile* detection PCR

We initially assessed the *C. difficile* specificity of the detection PCR with bioinformatics examinations using BLAST. First, all unique and complete *C. difficile* genomes at chromosome level – 134 in total – were verified to contain the entire *hpdBCA*-sequence by BLAST+ alignment against the CD630 operon, which showed overall 100% query coverage and above 97% percent identity. A further BLAST megablast analysis of the CD630 operon was performed to check for significant non-*C. difficile* matches. Only significant matches against *C. difficile* entries and five genomes belonging to the cryptic *Clostridioides* strains were detected. The BLAST alignments against these *Clostridioides* sp. strains showed above 99% query coverage with over 92% percent identity to the CD630 operon.

All unique *C. difficile* genomes and the five *Clostridioides* sp. strains – 139 in total (Table S2) – were used for ANIm calculations.

The detection primers amplified a 464 bp-long part of the 3.9 kb *hpdBCA*-operon (position 2,525 to 2,988) that spanned across the entire *hpdC* gene and parts of genes *hpdB* and *hpdA*. The reverse primer covered *hpdC* and *hpdA*, which corresponded to an alignment gap in similar sequences in most non-*C. difficile* species, as determined by blastn *hpdBCA*-sequence alignment during primer design (Fig. S2). Consequently, the reverse primer was primarily responsible for *C. difficile* specificity. However, two strains – *Clostridium diolis* (CP043998.1) and *Pelosinus sp.* strain UFO1 (CP008852.1) – did not show this alignment gap but possessed sufficient sequence deviation to rule out effective binding of the reverse primer in theory (Fig. S2). Detection mg-PCR on genomic DNA of these strains proved to be true negative and, thus, supported PCR specificity. Further, mg/g-PCR was verified to be positive on a self-isolated cryptic C-III strain so that successful detection of at least C-III cryptic *Clostridioides* was verified.

Potential of the detection sequence *hpdBCA* as phylogenetic determinant

Heterogeneity of the detection PCR sequence and, thus, its potential for phylogenetic classification of *C. difficile* isolates, was examined by ANIm analysis (hereafter referred to as detANIm) in relation to a whole-genome ANIm (wgANIm) comprising all 139 genomes defined above (data of ANIm outputs in Table S3 and S4).

The wgANIm of the 139 NCBI strains (Fig. 1) exhibited clustering as already described for *C. difficile*, with a cluster-group comprising five distinct clusters and an overall ANI above 96% (species boundary) (7). Based on determined MLSTs (Table S2), each of the five clusters was assigned to clades C1 to C5. Intra- and inter-clade comparisons of mean, median, minimal and maximal ANIm values (Fig. 2) revealed that isolates within one clade shared ANI values above 99.24%. One isolate (ST963) could not be assigned to a clade with FastMLST (32), and did also not apparently cluster to a specific clade within the wgANIm. This was further visible by its inter-clade ANI values below 99% (Fig. 2), with highest similarity to C4. Subsequently, this isolate was treated as unclassified outlier. C1 was the biggest cluster, followed by C2, C5, C4 and finally C3. Additionally, C1 also showed highest heterogeneity among the five clades, indicated by a broader ANI range and apparent sub-clustering. C5 was least similar to C1 to C4 than these clades amongst themselves.

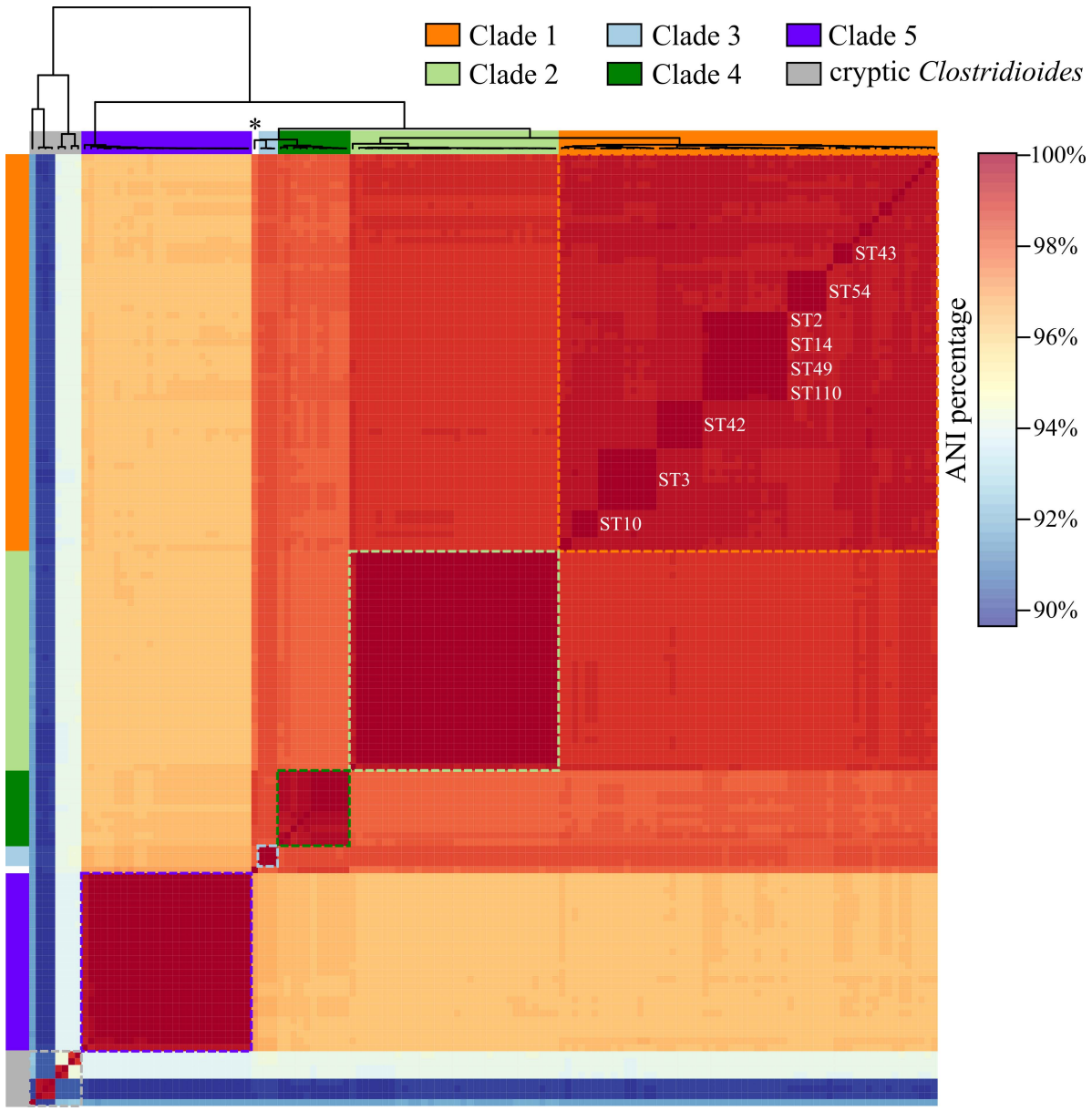


Figure 1. Heatmap of whole-genome ANIm analysis (wgANI). ANIm percentage values were visualized in heatmap.2 with corresponding dendrogram. Phylogenetic clades based on MLST determination are color-highlighted and MLST types are noted for C1 sub-clusters comprising more than three strains. The outlier strain ST963 is marked with * in the dendrogram.

In addition to clades C1 to C5, four further separate clusters were visible in the wgANIm, which represented the clades of cryptic *Clostridioides*. These clusters shared below 94.38% ANI to the five clades (Fig. 2) and comprised the five *Clostridioides* sp. (cryptic clades C-III and C-IV (6)) as well as three novel *C. difficile* strains (Cdiff_RT151_6, Cdiff_RT151_7, Cdiff_RT151_8), indicating that these strains belong to the cryptic *Clostridioides*. This was confirmed by further ANIm calculations with these strains against reference strains of cryptic clades C-I to C-V (6), which identified them as C-I and C-II strains, respectively.

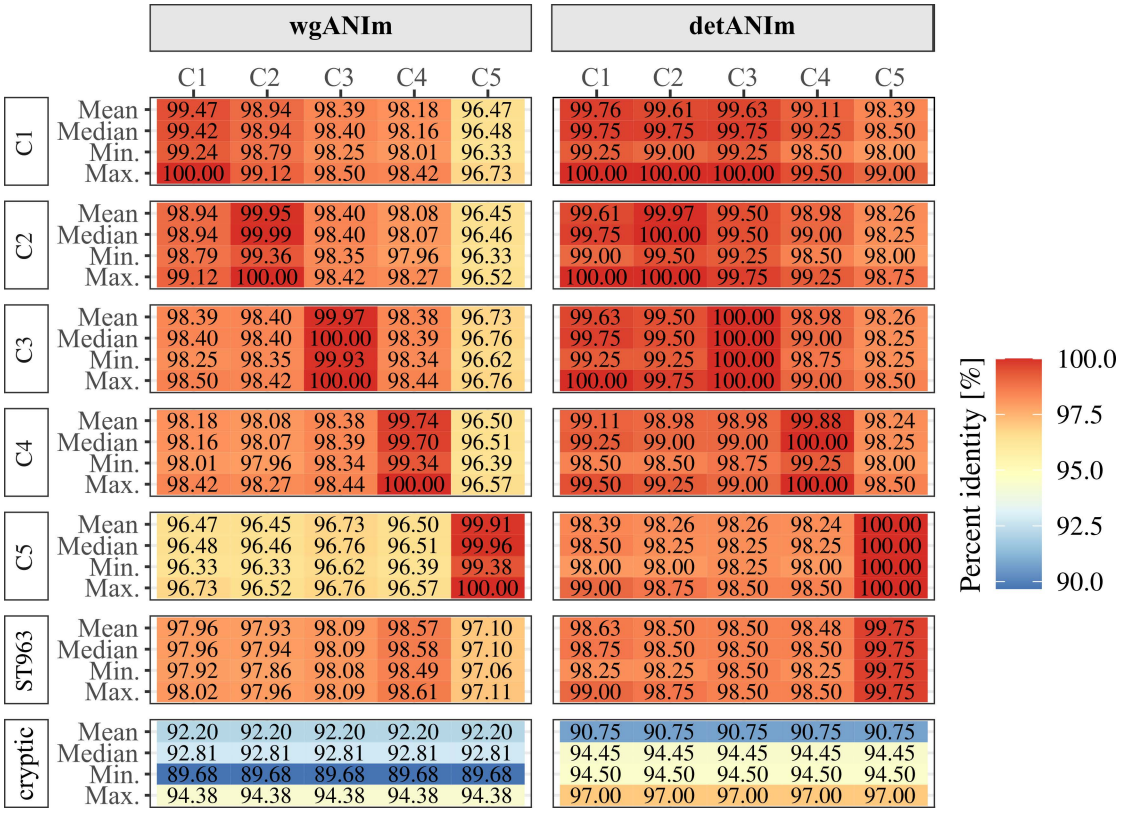


Figure 2. Clade-wise comparison of percentage identities for wgANIm and detANIm. Given are mean, median, minimal and maximal ANI percentages between two clades of C1-C5, outlier ST963 to each clade of C1-C5 or between all cryptic *Clostridioides sp.* to *C. difficile* (all five clades combined). Clade affiliation of an isolate was defined by MLST assignment using FastMLST (32).

The detANIm analysis revealed apparent clustering of specific sequence variants (Fig. 3), which resembled the wgANIm clustering. Alike the wgANIm, a large group of several clusters contained all *C. difficile* strains, while smaller separate clusters comprised the five *Clostridioides* sp. and the three presumable cryptic *Clostridioides* strains. These separate clusters shared below 97% ANI to *C. difficile* strains, while *C. difficile* detection sequences amongst C1 to C5 strains were at least 98% similar and often 100% identical within clusters (Fig. 2).

For detailed phylogenetic analysis of the detection sequence, previous MLST assignments of the genomes included in ANIm calculations were transferred to the corresponding detection sequences in the detANIm. This revealed a slightly different clustering or cluster positioning of the five clades. C1 sequences showed again highest heterogeneity (Fig. 2 and 3) but split into sub-clusters that partially shared 100% ANI and comprised various MLSTs. Further, sequences of two C1 strains (DSM 29632 and Z31) did not cluster in their correct clade or were even more similar to C3 sequences. Alike the wgANIm, detection sequences of C2 strains were closest related to those of C1 strains, with one outlying C2 strain (Cd1) even clustering within C1 sequences. This strain already attracted attention in the

wgANIm by sharing ANI values to the other C2 strains below 99.40%, whereas all other C2 strains showed higher sequence similarity of at least 99.94%. All detection sequences of C3, C4 and C5 isolates clustered according to their phylogenetic clade. The detection sequence of wgANIm-outlier strain ST963 with undefined MLST/clade and highest wgANIm similarity to C4 now clustered with C5. However, while all sequences within C5 were 100% identical, strain ST963 was distinguishable and only shared 99.75% ANI to all C5 isolates (Fig. 2).

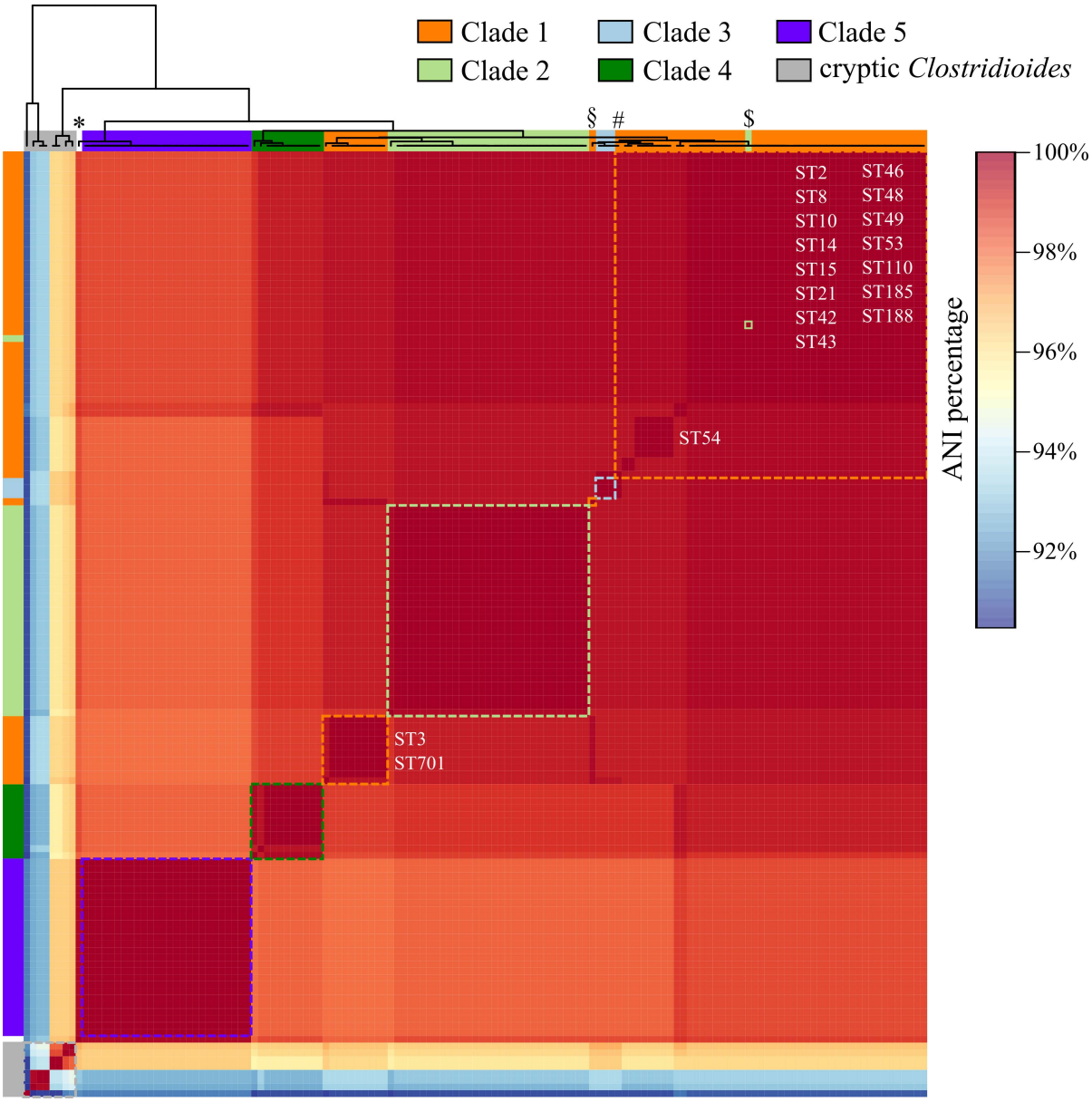


Figure 3. Heatmap of detection sequence ANIm analysis (detANIm). ANIm percentage values were visualized in heatmap.2 with corresponding dendrogram. Phylogenetic clades based on MLST determination are color-highlighted and MLST types are noted for C1 sub-clusters comprising more than three strains. Outliers are marked as follows in the dendrogram: strain ST963 with *, C1 outliers DSM 29632 and Z31 with § and #, C2 outlier Cd1 with $.

Alignment of all 139 detection sequences revealed the unique nucleotide differences that represent distinct ANIm-clusters and MLSTs/clades (Fig. S3). All detection sequences used for ANIm analyses are listed in the supplementary data file S1, with GenBank accession numbers of the corresponding *C. difficile* genomes as well as MLST/clade information as determined with FastMLST (32) or otherwise deduced from ANIm results. This file can be used for future ANI calculations or employed as BLAST+ database to examine one's own detection sequences of interest.

Further apparent in Fig. S3, the primer regions are conserved amongst diverse *C. difficile* strains except for the already mentioned mismatch in the regions of C5 strains.

Amplicon-based examination of detection PCR sensitivity and specificity.

Detection PCR sensitivity was evaluated in comparison to the common 16S rRNA gene analysis. None of the 46 environmental samples contained the *C. difficile*-16S rRNA gene ASV in the final amplicon data after removing spurious sequences with the 0.25% cutoff filter (48), which eliminated the few present *C. difficile*-ASV reads (maximal 0.029% per sample) (data file S2). Consequently, this analysis suggested the lack of *C. difficile*. However, this data based on the standardized 16S rRNA gene PCR with 25 cycles whereas the detection PCR involves 35. Thus, PCR sensitivities were not directly comparable. To rule out that *C. difficile* absence in 16S rRNA gene-amplicon data resulted from insufficient PCR cycling, 16S rRNA gene PCR was repeated with 35 cycles using specific environmental samples that possessed the *C. difficile*-ASV in the unfiltered data or from which we had isolated *C. difficile*. In this final amplicon data, only one out of eight analyzed samples contained reads corresponding to the *C. difficile*-ASV at all, and the abundance of 0.0134% lay again below the 0.25% filter cutoff (data file S3). These findings supported that general 16S rRNA gene sequencing-based analyses suggest *C. difficile* absence in the environmental samples.

Further analyses of the 16S rRNA gene-amplicon data focused on community composition and diversity (Fig. S4-6). Bacterial abundance at genus was level diverse within and between the environmental samples (Fig. S4 and S5a) and similar composition at genus level was rather linked to sample type (Fig. S4 A) but not to detection PCR outcome (Fig. S4 B). Alpha diversity was assessed as ASV richness (Fig. S6). Most environmental samples showed a richness between 2,000 and 6,000 ASVs. Thereby, no correlation between ASV richness of a sample and the corresponding detection PCR result was apparent.

Detection PCR specificity was already supported experimentally by true negative results using genomic DNA of bacteria that were suspected to be false positive (*C. diolis* and *Pelosinus sp.* strain UFO1). For further confirming PCR specificity, NGS of detection PCR products was performed to examine the origin of amplicon sequences. First, the original detection on mgDNA was positive for 16/46 environmental samples with varying PCR product yield as seen by band strength on agarose gel, which might be an indicator for *C. difficile* abundance. The detection PCR for NGS purpose was performed using the 16 samples yielding positive results. Multiple reactions per PCR replicate were necessary for samples where low PCR product yield was observed in the original detection PCR. For assessing adequate processing of detection amplicon data, NGS was performed using control amplicons from four different *C. difficile* isolates that were isolated from three of the environmental samples, which should only give rise to one sequence. Indeed, final amplicon data for these isolates contained only one ASV each with 100% relative abundance. These ASVs matched the expected detection sequence for each isolate as determined by *in silico* PCR. The final amplicon data of all environmental samples contained only ASVs with *C. difficile* identity, proving that all detection PCR products originated from *C. difficile* DNA present in the mgDNA samples (Fig. 4).

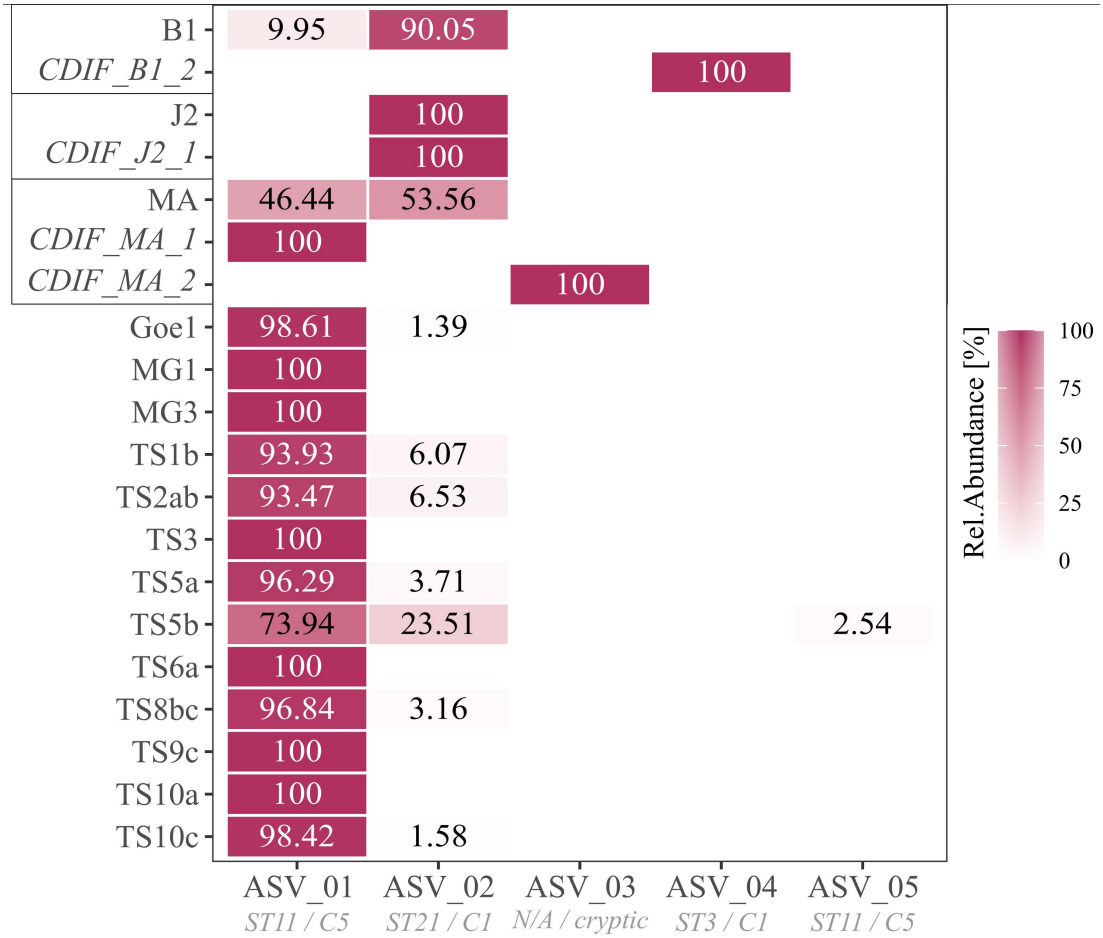


Figure 4. Heatmap of detection-ASV distribution and their relative abundances. *C. difficile* isolates are in italic and depicted next to their original environmental sample. MLST assignment (ST / clade) of each ASV as described in Table 2 is additionally specified below each ASV. Percentage values of the relative abundances are noted for values >0%.

The environmental samples comprised five different ASVs representing four sequence types within three phylogenetic clades, amongst others a cryptic clade (Fig. 4, Table 2). From the 16 mgDNA samples, seven samples contained one, eight samples two, and sample TS5b comprised even three *C. difficile*-affiliated ASVs. The most prevalent ASV_01, classified as ST11/C5 sequence, was present in all mgDNA samples except of sample J2. The second most abundant ASV_02 with determined ST21/C1 identity was present in ten mgDNA samples with an abundance ranging from 1.39% to 90.05%. These two ASVs were present in control PCRs with DNA from *C. difficile* isolates MA_1 (ST11/C5) and J2_1 (ST8/C1), coinciding with ASV presence in their original environmental sample MA and J2. Contrary, the cryptic isolate MA_2 (ST340/C-III) correctly exhibited ASV_03, which was however not observed in the corresponding environmental sample MA used for isolation. Environmental sample B1 also comprised the two most abundant ASV_01 and ASV_02. Remarkably, *C. difficile* strain B1_2 that was isolated from sample B1 did not possess any of these ASVs but the ST3/C1 sequence ASV_04. It should be noted here that both isolates MA_2 and B1_2, which were not observed

in the detection amplicons of their respective environmental sample, were isolated with antibiotic-based cultivation according to Dharmasena & Jiang (8).

The other environmental samples (Goe1 – TS10c in Fig. 4) were dominated by ASV_01 with at least 93.47% relative abundance, often accompanied by the second ASV_02 with maximal 6.53% abundance. The only exception was the afore-mentioned sample TS5b, in which ASV_01 comprised 73.94%, ASV_02 23.51% and the third ASV_05 2.54%. ASV_05 was as ASV_02 identified as ST11/C5 sequence. However, although matching best to C5 sequences, the similarity of ASV_05 to the C5-BLAST reference of 98.504% lay below our previously determined intra-clade detANI value (100%) of C5 strains (Fig. 2).

Table 2. BLAST results from amplicon data processing of the five ASVs shown in Fig. 4. Listed are BLAST hits with species and strain (GenBank accession), percentage identity, query coverage and E-value. Additionally, MLST/clade assignment for each strain is given as determined via PubMLST.org.

ASV	BLAST hit	Identity (%)	Query coverage (%)	E-value	MLST	Clade
01	*C. difficile* strain TW11-RT078 CP035499.1	100	100	0.0	ST11	C5
02	*C. difficile* strain 020709 CP028529.1	100	100	0.0	ST21	C1
03	*C. difficile* strain ST632 CP103804.1	98.005	100	0.0	ST632	cryptic[a]
04	*C. difficile* strain DSM 29745 CP019857.1	100	100	0.0	ST3	C1
05	*C. difficile* strain TW11-RT078 CP035499.1	98.504	100	0.0	ST11	C5

[a]no specific cryptic clade could be assigned to CP103804.1, highest similarity was found to C-III with a percentage identity of 95% - 96% based on ANIm analysis against reference strains for cryptic clades C-I to C-V selected from (6).

Discussion

We established a PCR that specifically detects the global pathogen *C. difficile* in environmental DNA, even in low abundance. Thus, environmental samples can be evaluated for *C. difficile* presence, which in turn can support *C. difficile* isolation from various environments and thereby contributes to holistic investigations of this significant pathogen. In addition to detection purposes, the detection PCR allows to initially classify the phylogenetic affiliation of *C. difficile* isolates or to assess *C. difficile* diversity present environmental samples.

The sensitivity of our detection PCR was validated in comparison to the commonly used amplicon-based 16S rRNA gene analysis. This method did not show *C. difficile* in various environmental DNA samples, meaning the abundance of *C. difficile* 16S rRNA gene amplicons was below the filter cutoff during amplicon processing or not present at all. Even increased PCR cycling in 16S rRNA gene-amplicon production as adjustment to the higher cycle amount within the detection PCR did not reveal *C. difficile* presence. Contrary, our detection PCR was positive for one third of the environmental samples. As such, the PCR was proven to outperform the 16S rRNA gene approach in detection of low-abundant *C. difficile*, even without prior enrichment as done in other studies (15).

Detection PCR specificity was evaluated by NGS of detection amplicons from diverse environmental samples. The bacterial community of these samples was divergent as shown by 16S rRNA gene analysis. Sequencing of detection amplicons obtained from those diverse environmental samples verified high *C. difficile* specificity of the detection PCR, as only *C. difficile*-ASVs were found. With this proof of specificity, the detection PCR can be used beyond the detection purpose in environmental DNA and quickly verify newly isolated strains as *C. difficile* in a simple colony PCR. This identification procedure is a fast and easy alternative to common methods such as MALDI-TOF, 16S rRNA gene sequencing or enzymatic assays (5, 8, 9). We also examined the detection sequence itself as potential phylogenetic determinant via ANIm analyses with regard to a whole-genome ANIm. ANIm clusters in combination with MLST assignment showed that clade sizes and MLST type distribution among our genome dataset matched the holistic, taxonomic investigation by Knight et al. (7), supporting that the dataset adequately represented the species *C. difficile*. ANIm clusters observed at whole-genome level were mostly identified in the detANIm, which showed that the detection sequence reflects phylogenetic information to some extent. While the detection sequence allowed correct clade assignment for almost all 139 analyzed strains except of the four described outliers, it could not distinguish between closely related STs. Thus, a complete phylogenetic classification

with distinct MLST assignment requires detailed MLST analysis. Nevertheless, the detection PCR not only allows identification of *C. difficile* isolates with a positive PCR result but also enables an initial phylogenetic classification of the isolate by sequencing the recovered PCR product.

Further, we investigated the NGS detection amplicons from environmental DNA samples and predominantly found sequences that match *C. difficile* ST11/C5 and ST21/C1 strains. Prevalence of C5 strains is linked to various animal species such as horses, pigs and cattle (52). These observations coincide with our results of prevalent C5 strains in the analyzed environmental samples, of which the majority were in previous contact with the afore-mentioned animals. Prevalent sequences with C1-association in the environmental samples in turn reflect the global occurrence of C1 strains in both animals and humans, and are in this context also of interest concerning transmission of potentially toxigenic strains (20, 53, 54). Noteworthy, in the light of our detANIm analysis, the sequence of ST21/C1 is not unique for this sequence type but shares 100% similarity to sequences of further C1 sequence types. This was demonstrated by the environmental sample J2 and the corresponding isolate J2_1 for which the detected ASV_02 was assigned to ST21/C1, albeit strain J2_1 belonging to ST8/C1. Consequently, detection ASVs with determined ST21/C1 identity might belong to another C1-ST sequence as well and even the presence of multiple C1 strains in the environmental sample can be considered.

We also performed the detection amplicon sequencing on *C. difficile* strains that originated from the environmental samples and checked for ASV accordance between the isolates and their corresponding environmental sample. The two analyzed strains MA_2 and B1_2 and their ASVs did not accord with ASV presence in the corresponding environmental data, which instead contained ASVs of two different *C. difficile* strains. This implied that the isolated strains were insufficient abundant in the environmental sample to be identified by the detection PCR, contrary to those strains that showed up in the detection amplicons. In that regard, the successful isolation of these strains despite their underrepresentation indicated a selective cultivation of these underrepresented strains, in contrast to the other more abundant strains. Such selective cultivation can be performed for instance with antibiotic treatment. Indeed, we obtained these *C. difficile* isolates with antibiotic-based cultivation using moxalactam norfloxacin, an established supplement for selective *C. difficile* isolation (8, 55). In conclusion, these results suggest an abundance shift of *C. difficile* strains in environmental samples by applying selective antibiotic cultivation procedures with environmental samples as

starting material, which supports our hypothesis to miss some strains during antibiotic-based isolation approaches. Thus, for assessing *C. difficile* in environmental samples, a combination of both strategies is advisable.

Acknowledgements

We thank the laboratory of Prof. Dr. Adams at the Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA for providing the genomic DNA of *Pelosinus sp.* strain UFO1. We also thank Melanie Heinemann for technical assistance.

References

1. Balsells E, Shi T, Leese C, Lyell I, Burrows J, Wiuff C, Campbell H, Kyaw MH, Nair H. 2019. Global burden of *Clostridium difficile* infections: a systematic review and meta-analysis. J Glob Health 9:010407.
2. Czepiel J, Dróżdż M, Pituch H, Kuijper EJ, Perucki W, Mielimonka A, Goldman S, Wultańska D, Garlicki A, Biesiada G. 2019. *Clostridium difficile* infection: review. Eur J Clin Microbiol Infect Dis 38:1211–1221.
3. Awad MM, Johanesen PA, Carter GP, Rose E, Lyras D. 2014. *Clostridium difficile* virulence factors: Insights into an anaerobic spore-forming pathogen. Gut Microbes 5:579–593.
4. Lewis BB, Carter RA, Ling L, Leiner I, Taur Y, Kamboj M, Dubberke ER, Xavier J, Pamer EG. 2017. Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to *Clostridium difficile* Virulence. MBio 8:e00885-17.
5. Janezic S, Potocnik M, Zidaric V, Rupnik M. 2016. Highly Divergent *Clostridium difficile* Strains Isolated from the Environment. PLoS One 11:e0167101.
6. Williamson CHD, Stone NE, Nunnally AE, Roe CC, Vazquez AJ, Lucero SA, Hornstra H, Wagner DM, Keim P, Rupnik M, Janezic S, Sahl JW. 2022. Identification of novel, cryptic *Clostridioides* species isolates from environmental samples collected from diverse geographical locations. Microb Genomics 8:000742.
7. Knight DR, Imwattana K, Kullin B, Guerrero-Araya E, Paredes-Sabja D, Didelot X, Dingle KE, Eyre DW, Rodríguez C, Riley T V. 2021. Major genetic discontinuity and novel toxigenic species in *Clostridioides difficile* taxonomy. Elife 10:e64325.
8. Dharmasena M, Jiang X. 2018. Improving culture media for the isolation of *Clostridium difficile* from compost. Anaerobe 51:1–7.
9. Cheng J-W, Xiao M, Kudinha T, Xu Z-P, Sun L-Y, Hou X, Zhang L, Fan X, Kong F, Xu Y-C. 2015. The Role of Glutamate Dehydrogenase (GDH) Testing Assay in the

Diagnosis of *Clostridium difficile* Infections: A High Sensitive Screening Test and an Essential Step in the Proposed Laboratory Diagnosis Workflow for Developing Countries like China. PLoS One 10:e0144604.

10. Spigaglia P. 2016. Recent advances in the understanding of antibiotic resistance in *Clostridium difficile* infection. Ther Adv Infect Dis 3:23–42.
11. Tyrrell KL, Citron DM, Leoncio ES, Merriam CV, Goldstein EJC. 2013. Evaluation of Cycloserine-Cefoxitin Fructose Agar (CCFA), CCFA with Horse Blood and Taurocholate, and Cycloserine-Cefoxitin Mannitol Broth with Taurocholate and Lysozyme for Recovery of *Clostridium difficile* Isolates from Fecal Samples. J Clin Microbiol 51:3094–3096.
12. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1.
13. Lemee L, Dhalluin A, Testelin S, Mattrat M-A, Maillard K, Lemeland J-F, Pons J-L. 2004. Multiplex PCR Targeting *tpi* (Triose Phosphate Isomerase), *tcdA* (Toxin A), and *tcdB* (Toxin B) Genes for Toxigenic Culture of *Clostridium difficile*. J Clin Microbiol 42:5710–5714.
14. van Rossen TM, van Prehn J, Koek A, Jonges M, van Houdt R, van Mansfeld R, Kuijper EJ, Vandenbroucke-Grauls CMJE, Budding AE. 2021. Simultaneous detection and ribotyping of *Clostridioides difficile*, and toxin gene detection directly on fecal samples. Antimicrob Resist Infect Control 10:23.
15. Stone NE, Sidak-Loftis LC, Sahl JW, Vazquez AJ, Wiggins KB, Gillece JD, Hicks ND, Schupp JM, Busch JD, Keim P, Wagner DM. 2016. More than 50% of *Clostridium difficile* Isolates from Pet Dogs in Flagstaff, USA, Carry Toxigenic Genotypes. PLoS One 11:e0164504.
16. Zeidan AR, Strey K, Vargas MN, Reveles KR. 2021. Prevalence of *Clostridioides difficile* strains found in Texas soil, p. 24. *In* Journal of Clinical and Translational Science.
17. Bidet P, Barbut F, Lalande V, Burghoffer B, Petit J-C. 1999. Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. FEMS Microbiol Lett 175:261–266.
18. Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, Golubchik T, Harding RM, Jeffery KJM, Jolley KA, Kirton R, Peto TE, Rees G, Stoesser N, Vaughan A, Walker AS, Young BC, Wilcox M, Dingle KE. 2010. Multilocus

Sequence Typing of *Clostridium difficile*. J Clin Microbiol 48:770–778.

19. Ducarmon QR, van der Bruggen T, Harmanus C, Sanders IMJG, Daenen LGM, Fluit AC, Vossen RHAM, Kloet SL, Kuijper EJ, Smits WK. 2023. *Clostridioides difficile* infection with isolates of cryptic clade C-II: a genomic analysis of polymerase chain reaction ribotype 151. Clin Microbiol Infect 29:538.e1-538.e6.
20. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley T V. 2015. Diversity and Evolution in the Genome of *Clostridium difficile*. Clin Microbiol Rev 28:721–741.
21. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, Holden MT, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabbinowitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet 38:779–786.
22. Harrison MA, Kaur H, Wren BW, Dawson LF. 2021. Production of p-cresol by Decarboxylation of p-HPA by All Five Lineages of *Clostridioides difficile* Provides a Growth Advantage. Front Cell Infect Microbiol 11:757599.
23. Passmore IJ, Letertre MPM, Preston MD, Bianconi I, Harrison MA, Nasher F, Kaur H, Hong HA, Baines SD, Cutting SM, Swann JR, Wren BW, Dawson LF. 2018. *Para*-cresol production by *Clostridium difficile* affects microbial diversity and membrane integrity of Gram-negative bacteria. PLOS Pathog 14:e1007191.
24. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res 50:D20–D26.
25. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10:1–9.
26. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5–W9.
27. Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M, Sodergren E, Weinstock GM. 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun 10:5029.

28. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134.
29. Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278.
30. Pritchard L. 2014. PYANI script. https://github.com/widdowquinn/pyani.
31. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12.
32. Guerrero-Araya E, Muñoz M, Rodríguez C, Paredes-Sabja D. 2021. FastMLST: A Multi-core Tool for Multilocus Sequence Typing of Draft Genome Assemblies. Bioinform Biol Insights 15:117793222110592.
33. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191.
34. Schneider D, Thürmer A, Gollnow K, Lugert R, Gunka K, Groß U, Daniel R. 2017. Gut bacterial communities of diarrheic patients with indications of *Clostridioides difficile* infection. Sci Data 4:170152.
35. Berkelmann D, Schneider D, Engelhaupt M, Heinemann M, Christel S, Wijayanti M, Meryandini A, Daniel R. 2018. How Rainforest Conversion to Agricultural Systems in Sumatra (Indonesia) Affects Active Soil Bacterial Communities. Front Microbiol 9:2381.
36. Tange O. 2018. GNU Parallel 2018Isbn 9781387509881. http://ole.tange.dk.
37. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.
38. Friedrich I, Klassen A, Neubauer H, Schneider D, Hertel R, Daniel R. 2021. Living in a Puddle of Mud: Isolation and Characterization of Two Novel *Caulobacteraceae* Strains *Brevundimonas pondensis* sp. nov. and *Brevundimonas goettingensis* sp. nov. Appl Microbiol 1:38–59.
39. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620.
40. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12.
41. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584.

42. Nearing JT, Douglas GM, Comeau AM, Langille MGI. 2018. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ 6:e5364.
43. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200.
44. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596.
45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.
46. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, Caporaso JG. 2012. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. Gigascience 1:7.
47. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12:635–645.
48. Reitmeier S, Hitch TCA, Treichel N, Fikas N, Hausmann B, Ramer-Tait AE, Neuhaus K, Berry D, Haller D, Lagkouvardos I, Clavel T. 2021. Handling of spurious sequences affects the outcome of high-throughput 16S rRNA gene amplicon profiling. ISME Commun 1:31.
49. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org.
50. Andersen KS, Kirkegaard RH, Karst SM, Albertsen M. 2018. ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. bioRxiv 10–11.
51. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 3:124.
52. Weese JS. 2020. *Clostridium* (*Clostridioides*) *difficile* in animals. J Vet Diagnostic Investig 32:213–221.
53. Janezic S, Zidaric V, Pardon B, Indra A, Kokotovic B, Blanco JL, Seyboldt C, Diaz CR, Poxton IR, Perreten V, Drigo I, Jiraskova A, Ocepek M, Weese JS, Songer JG, Wilcox MH, Rupnik M. 2014. International *Clostridium difficile* animal strain

collection and large diversity of animal associated strains. BMC Microbiol 14:173.

54. Rodriguez Diaz C, Seyboldt C, Rupnik M. 2018. Non-human *C. difficile* Reservoirs and Sources: Animals, Food, Environment, p. 227–243. *In* Advances in Experimental Medicine and Biology.

55. Aspinall ST, Hutchinson DN. 1992. New selective medium for isolating *Clostridium difficile* from faeces. J Clin Pathol 45:812–814.

Supplementaries

Following supplementary material can be found on the enclosed CD/Supplements/Chapter_4.1:

Figure S1. MSA Viewer output of blastn alignment of CD630-*hpdBCA* excluding *C. difficile*. The CD630 *hpdBCA* sequence was analyzed with blastn excluding *C. difficile* to retrieve all significant non-*C. difficile* hits, and alignment was examined with the MSA Viewer to identify potential detection PCR regions. (JPEG)

Figure S3. Visualization of nucleotide differences in representative detection sequences of specific *C. difficile* MLST/clade genomes. Depicted are detection sequences (including primer sequences, purple box) that represent distinct detANIm clusters of MLSTs/clades. MLST/clade affiliation together with GenBank accession of the representative genomes are indicated. (TIF)

Figure S5b. Taxonomy legend to Fig. S5a. Taxonomic assignment comprises bacterial phylum, family, and genus. (SVG)

Table S1. Detail environmental samples.

Table S2. Analyzed strains & MLST information. Information of complete, unique C. difficile genomes available on NCBI on 30 August 2022.

Table S3. Output of wgANIm. Output of whole-genome ANIm (wgANIm, Fig. 1); used for calculations of clade-wise wgANIm comparisons (Fig. 2).

Table S4. Output of detANIm. Output of detection sequence ANIm (detANIm, Fig. 3); used for calculations of clade-wise detANIm comparisons (Fig. 2).

Data file S1. Multifasta of the 139 detection sequences used in detANIm analysis, with corresponding GenBank accession and MLST information.

Data file S2. Raw 16S rRNA gene sequencing ASV table.

Data file S3. Raw 16S rRNA gene sequencing ASV table of amplicons ran with 35 PCR cycles.

Following supplementary material is included below:

Figure S2, S4, S5a, SS6

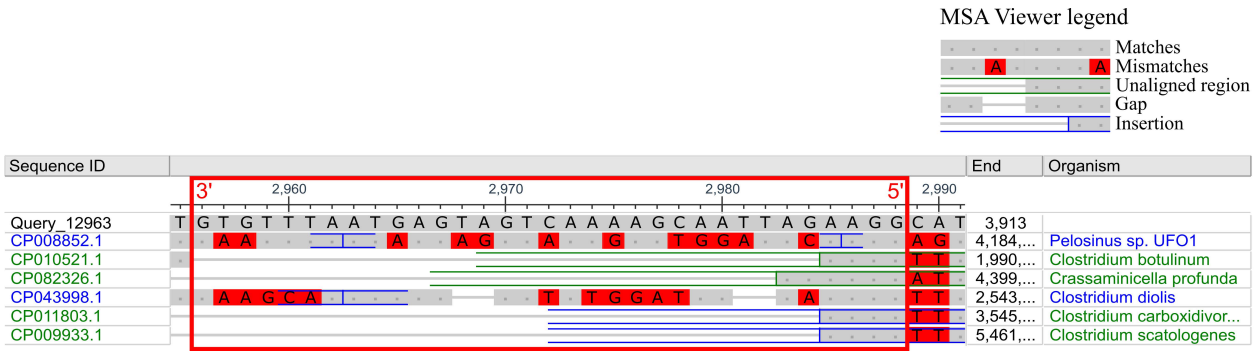


Figure S2. MSA Viewer alignment of the CDIF_cresol_R3 primer region against exemplary non-*C. difficile* strains. The CD630 *hpdBCA* sequence was analyzed with blastn excluding *C. difficile* and examined with the MSA Viewer (Fig. S1). *hpdBCA* region 2,956 – 2,988 bp (red box, 3'- and 5'-region indicated) was chosen as reverse primer target due to the alignment gap in most hits (exemplary genomes in green). *Pelosinus* sp. strain UFO1 and *Clostridium diolis* (in blue) did not have the alignment gap but considerable sequence deviation.

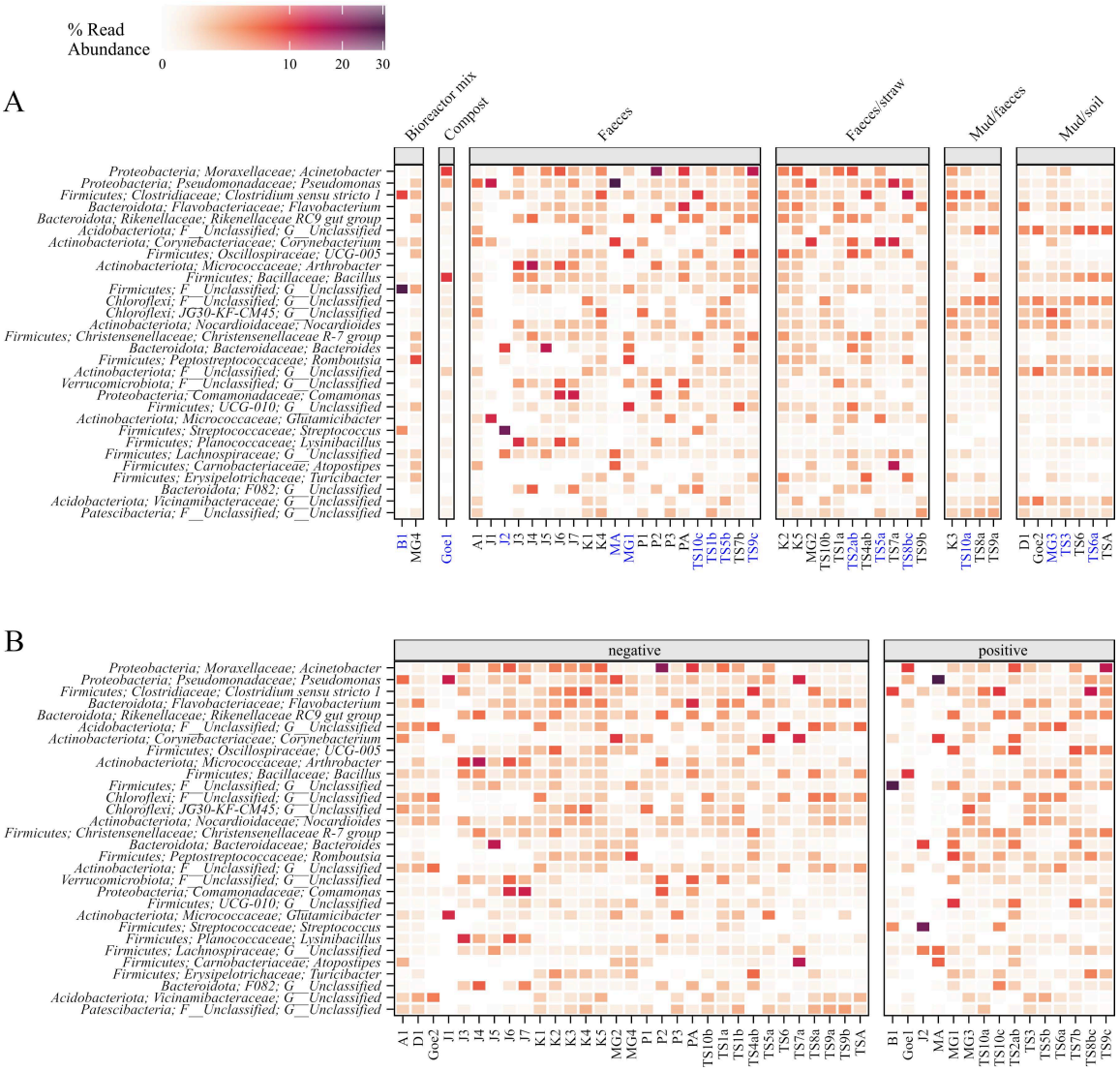


Figure S4. Top 30 genera in all environmental samples based on 16S rRNA gene sequencing analysis. Relative abundance of the 30 most prevalent genera (given as phylum; family; genus) within the entire dataset is depicted as heatmap for the 46 environmental samples. Samples are thereby grouped according their (A) sample type with detection-positive samples (blue, see below) or (B) detection PCR result.

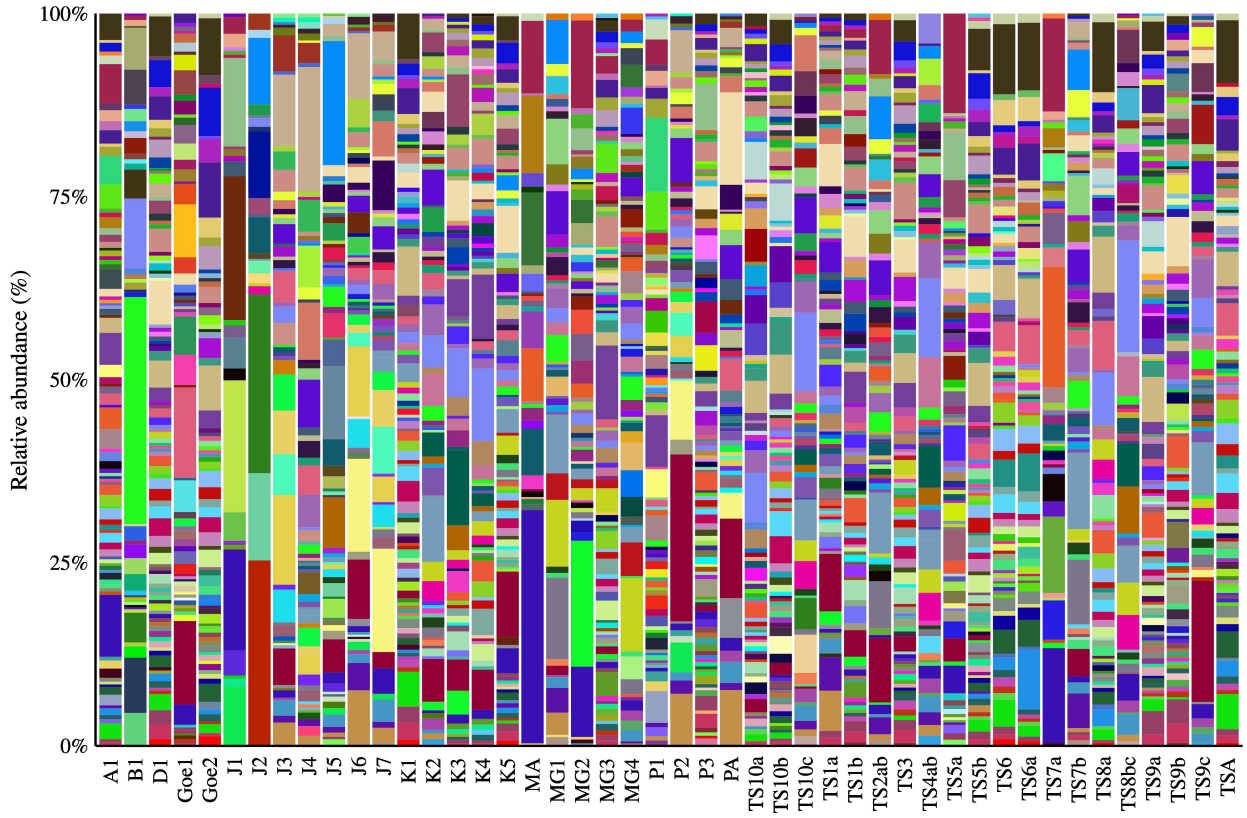


Figure S5a. Bacterial community composition of all environmental samples. Relative abundance at genus level based on 16S rRNA gene ASVs in final amplicon data is depicted as bar charts, the corresponding legend is displayed separately as Fig. S5b.

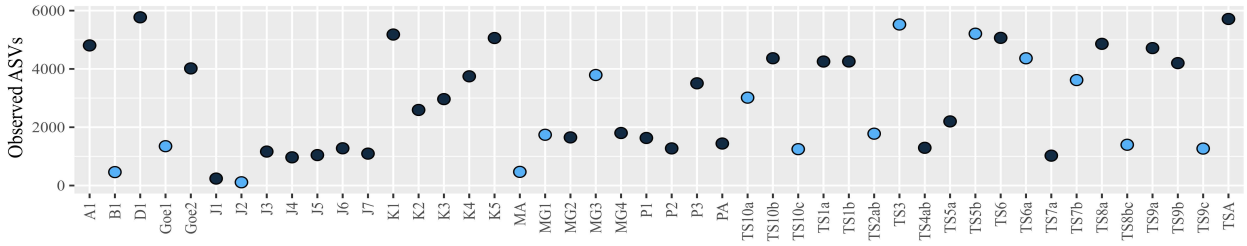


Figure S6. Alpha diversity as ASV richness of all environmental samples. Diversity is defined as richness in observed 16S rRNA gene ASVs. Samples are color-coded according to their detection PCR result (blue = positive).

4.2 Complete genome sequence of a *Clostridioides difficile* cryptic C-III strain isolated from horse feces

Miriam Antonia Schüler, Rolf Daniel, Anja Poehlein*

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

Running title: Genome of a *Clostridioides difficile* cryptic C-III strain
Keywords: *Clostridioides difficile,* cryptic C-III strain, horse feces, non-toxigenic strain

*Correspondence to
Anja Poehlein, apoehle3@gwdg.de.

Revised manuscript re-submitted to *Microbiology Resource Announcements* (*American Society for Microbiology*) on 22 September 2023, accepted on 30 September 2023

Author Contributions

Conceptualization: **MS**, AP, RD

Experiments: **MS**

Data analysis: **MS**

Manuscript writing: **MS**, AP, RD

Abstract

We provide the complete genome of a non-toxigenic *Clostridioides difficile* strain isolated from horse faeces. The strain represents a sub-cluster in the cryptic clade C-III. The genome consists of one chromosome (4,144,784 bp) and one plasmid (10,144 bp) and encodes 3,798 putative genes.

Announcement

The worldwide pathogen *Clostridioides difficile* covers at least five phylogenetic clades plus the cryptic clades C-I to C-V. We provide the complete genome of the C-III *C. difficile* strain MA_2, which was isolated from horse feces based on Dharmasena and Jiang (1). All enrichment and cultivation steps were performed under anaerobic conditions with incubation at 37°C. Twenty-five grams of sample were suspended in 100 ml PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4), solid particles filtered out, and cells washed repeatedly by centrifugation (4,000 x *g*, 10 minutes) before cultivation in 20 ml BHIB-YE-CYS-MN-T medium (1) for seven days. One milliliter of enriched culture was heat-shocked (60°C, 25 minutes), cells pelleted, suspended in 100 µl PBS, plated on CDA-CYS-H-MN-T plates (1) and cultivated for two days. Single colonies were cultivated in BHIS medium (Brain Heart Infusion Broth, 0.5% yeast extract, 0.05% L-cystein), and 2 µl were used as input for 16S rRNA gene PCR, using Phusion High-Fidelity polymerase (Thermo Fisher Scientific, Waltham, MA, USA) with GC buffer and PCR components according to protocol and 0.2 µM of primers (08f: 5′-AGAGTTTGATCCTGGC-3′, 1504r: 5′-TACCTTGTTACGACTT-3′). PCR cycling comprised: 98°C for 6 minutes, 28 cycles of 10 seconds at 98°C, 15 seconds at 55°C, and 40 seconds at 72°C, followed by 5 minutes at 72°C. Purified PCR products were Sanger sequenced (Microsynth Seqlab GmbH, Göttingen, Germany). Best sequence match was *C. difficile* strain ZZV14-6009 16S rRNA gene (KX792125.1; query coverage 99%, percent identity 99.93%) according to BLASTn (2).

The previous culture was inoculated in BHIS medium and cultivated as stated above before using two milliliters for DNA isolation with the MasterPure Gram Positive DNA Purification kit as recommended by the manufacturer (Epicentre, Madison, WI, USA). Illumina Nextera XT DNA libraries were sequenced on a MiSeq instrument (v3 chemistry; 2 × 300 bp, 600 cycles) as recommended by the manufacturer (Illumina, San Diego, CA, USA). For Nanopore sequencing, genomic DNA was isolated as described before from four milliliters of another fresh overnight culture cultivated as stated above. Libraries were prepared without size selection of DNA using the ligation sequencing kit 1D (SQK-LSK109) and the native barcode

expansion kit (EXP-NBD104) according to the manufacturer (Oxford Nanopore Technologies, Oxford, UK), sequencing was performed for 72 h using a SpotON flow cell Mk I (R9.4.0) and MinKNOW software v19.12.5, with integrated Guppy v3.2.10 for base calling (fast mode) and demultiplexing. Short reads were processed with fastp v0.23.3 (3) and trimmed with Trimmomatic v0.39 (4). Long reads were trimmed using Porechop v0.2.4 (https://github.com/rrwick/Porechop), filtered with Filtlong v0.2.1 (https://github.com/rrwick/Filtlong) and assembled with Flye v2.9.2 (5). The assembly was polished with short reads using BWA v0.7.17 (r1188) (6) and Polypolish v0.5.0 (7), resulting in one complete chromosome (4,144,784 bp) and plasmid (10,144 bp). Circlator v1.5.5 was used for start position adjustment (8). Prokka v1.14.5 (9) was used for annotation (selenoproteins were curated manually), PubMLST (10) for sequence type assessment, and toxin genes were detected with VFDB (11). All software was used with default settings unless otherwise stated. Sequencing statistics and genome features are listed in Table 1.

An average nucleotide identity analysis (ANI) with pyani v0.2.12 (12) using MUMmer3 (13) (ANIm) was performed for clade assignment. Strain MA_2 clustered with representative strains of cryptic clades C-I to C-V (14, 15) (Fig. 1).

Table 1. Sequencing statistics and genome features of strain MA_2.

Sequencing statistics		**Genome features**	
No. of short reads	3,026,482	Length chromosome (bp)	4,144,784
No. of long reads	18,911	Length plasmid (bp)	10,144
Long-read N_{50} (bp)	29,017	GC content (%)	28.68
Coverage short reads	166	No. of tRNAs	92
Coverage long reads	356	No. of tmRNA	1
		No. of rRNAs	35
		No. of CRISPRs	5
		No. of CDSs	3,798
		No. of curated selenoproteins	3
		Sequence type	340
		Toxin gene profile	A^- B^- CDT^-

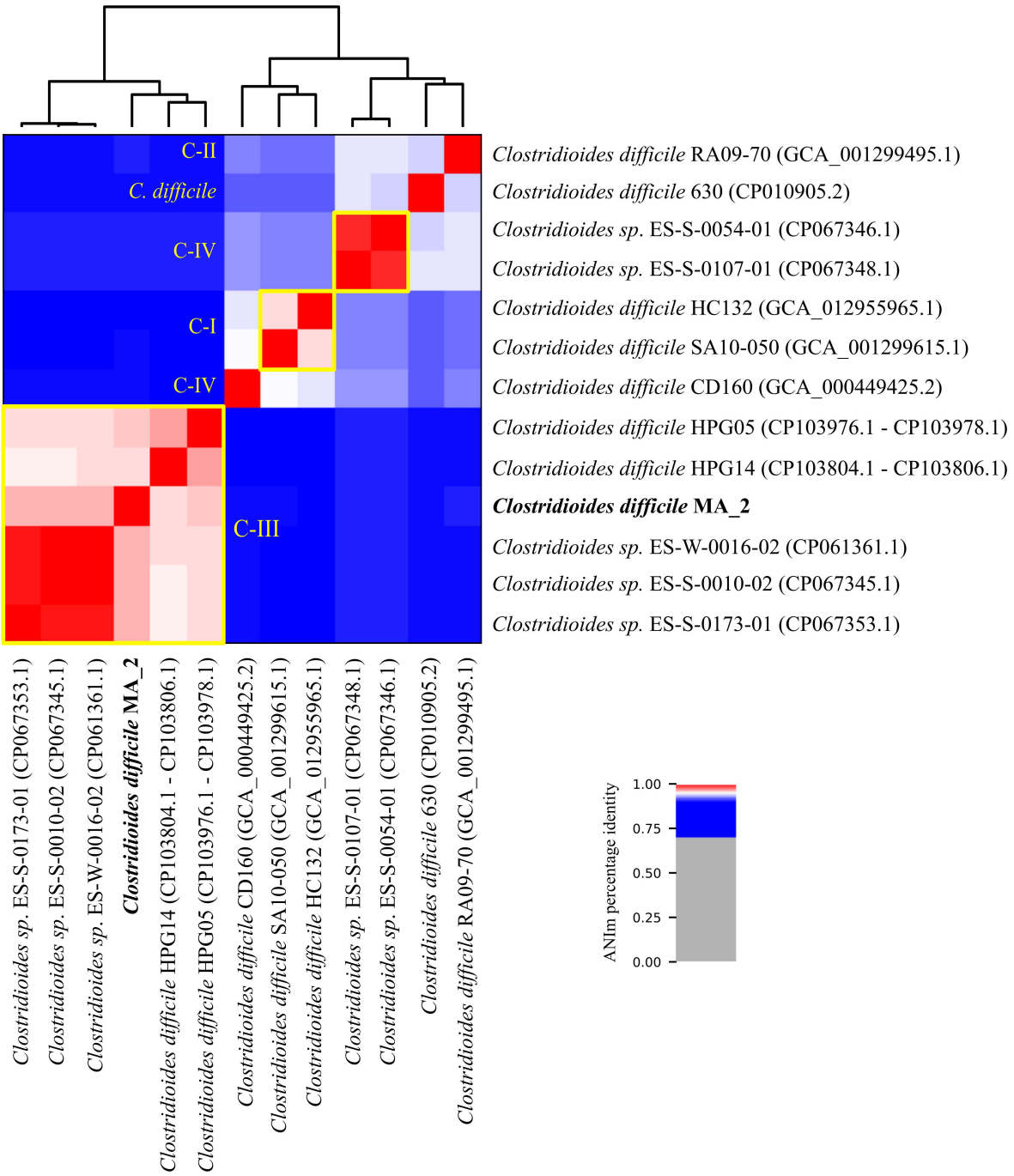


Fig. 1. ANIm analysis of strain *C. difficile* MA_2. Representative genomes of cryptic clades C-I to C-V (14, 15) and of *C. difficile* strain 630 representing clades C1 to C5 were included in ANIm calculations against *C. difficile* MA_2 (bold). GenBank accession numbers are provided in parenthesis, clade affiliation is stated in yellow.

Data availability statement

The complete genome sequence is deposited under GenBank accession numbers CP129431.1 (chromosome) and CP129432.1 (plasmid), the 16S rRNA gene sequence is deposited under OR144343.1. Corresponding raw reads are accessible in the NCBI Sequence Read Archive

(SRA) under accession numbers SRR24958190 (Nanopore reads) and SRR24958186 (Illumina reads).

Acknowledgments

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. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Dharmasena M, Jiang X. 2018. Improving culture media for the isolation of *Clostridium difficile* from compost. Anaerobe 51:1–7.
2. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5–W9.
3. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.
4. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
5. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546.
6. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.
7. Wick RR, Holt KE. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLOS Comput Biol 18:e1009802.
8. Hunt M, Silva N De, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294.
9. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.
10. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 3:124.

11. Liu B, Zheng D, Zhou S, Chen L, Yang J. 2022. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res 50:D912–D917.

12. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 8:12–24.

13. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12.

14. Williamson CHD, Stone NE, Nunnally AE, Roe CC, Vazquez AJ, Lucero SA, Hornstra H, Wagner DM, Keim P, Rupnik M, Janezic S, Sahl JW. 2022. Identification of novel, cryptic *Clostridioides* species isolates from environmental samples collected from diverse geographical locations. Microb Genomics 8:000742.

15. Shivaperumal N, Hain-Saunders NMR, Chang BJ, Riley T V., Knight DR. 2023. Complete Genome Sequences of Evolutionary Clade C-III Strains of *Clostridioides* (*Clostridium*) *difficile* Isolated from the Environment in Western Australia. Microbiol Resour Announc 12:e0023923.

4.3 Comparative genome analyses of clinical and non-clinical *Clostridioides difficile* strains

Miriam A. Schüler[1], Thomas Riedel[2], Jörg Overmann[2,3], Rolf Daniel[1], Anja Poehlein[1,*]

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

[2] Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, 38124 Braunschweig, Germany

[3] Microbiology, Technische Universität Braunschweig, Spielmannstraße 7, 38106 Braunschweig, Germany

*Corresponding author

E-mail: apoehle3@gwdg.de

Prepared for submission at *PLOS ONE;* by now submitted to *PLOS ONE* on 04 January 2024

Author Contributions

Conceptualization: **MS**, AP, RD

Experiments – sequencing: **MS**, TR, JO

Experiments – bioinformatics: **MS**

Data analysis: **MS**

Results interpretation: **MS**

Manuscript writing: **MS**

Manuscript review & editing: **MS**, JO, AP, RD

Abstract

The pathogenic bacterium *Clostridioides difficile* is a worldwide health burden with increasing morbidity, mortality and antibiotic resistances. Therefore, extensive research efforts are made to unravel its virulence and dissemination. One crucial aspect for *C. difficile* is its mobilome, which for instance allows the spread of antibiotic resistance genes (ARG) or influence strain virulence. As a nosocomial pathogen, the majority of strains analyzed originated from clinical environments and infected individuals. Nevertheless, *C. difficile* can also be present in human intestines without disease development or occur in diverse environmental habitats such as puddle water and soil, from which several strains could already be isolated. We therefore performed comprehensive genome comparisons of closely related clinical and non-clinical strains to identify the effects of the clinical background. Analyses included the prediction of virulence factors, ARGs, mobile genetic elements (MGEs), and detailed examinations of the pan genome. Clinical-related trends were thereby observed. While no significant differences were identified in fundamental *C. difficile* virulence factors, the clinical strains carried more ARGs and MGEs, and possessed a larger accessory genome. Detailed inspection of accessory genes revealed higher abundance of genes with unknown function, transcription-associated, or recombination-related activity. Accessory genes of these functions were already highlighted in other studies in association with higher strain virulence. This specific trend might allow the strains to react more efficiently on changing environmental conditions in the human host such as emerging stress factors, and potentially increase strain survival, colonization, and strain virulence. These findings indicated an adaptation of the strains to the clinical environment. Further, implementation of the analysis results in pairwise genome comparisons revealed that the majority of these accessory genes were encoded on predicted MGEs, shedding further light on the mobile genome of *C. difficile*. We therefore encourage the inclusion of non-clinical strains in comparative analyses.

Introduction

The bacterium *Clostridioides difficile* (formerly *Clostridium difficile*) is a globally widespread pathogen that constitutes a major cause of nosocomial and antibiotic-associated infections, with disease severity ranging from mild diarrhea to pseudomembranous colitis, eventually leading to death (1). A *C. difficile* infection is mainly elicited after antibiotic treatment and increasing antibiotic resistances in this species impede successful treatment of an infection (2). *C. difficile* is extensively studied, especially in the context of increasing multi-drug resistances, but also concerning its virulence heterogeneity. *C. difficile* strains can extremely vary in the induced

symptoms, and even non-toxigenic strains without disease-causing toxins exist (3). Research on *C. difficile* virulence already pointed towards the importance of mobile genetic elements (MGE). About 11% of a *C. difficile* genome is composed of MGEs, including plasmids, bacteriophages, IS elements, and conjugative and mobilizable transposons (4,5). Plasmids can contribute to virulence by carrying toxin genes or promoting antibiotic resistances (6), and also bacteriophages can influence *C. difficile* virulence (7–11). MGEs are especially crucial for horizontal gene transfer that allows fast adaptation to environmental conditions, e.g. spreading genes conferring antibiotic resistances between different strains or even species (12). In addition, the pathogenicity locus of *C. difficile*, which encodes the *C. difficile*-typical toxin genes, exhibits a mobile character and can transfer to a previously non-toxigenic strain (13). Although the toxin genes represent the major virulence factors of *C. difficile*, their contribution to overall virulence is still under debate, and other aspects such as tolerance to secondary bile acids or specific accessory genes rather correlated with disease severity (14).

As a prominent pathogen with increasing morbidity and mortality, most of the analyzed *C. difficile* strains originate from clinical specimen of infected individuals. However, *C. difficile* was also found in asymptomatic, healthy individuals, and is also a natural inhabitant of various animal species and environmental reservoirs (15–17). Although several *C. difficile* strains were isolated from diverse environmental sources in recent years, genome-based comparisons always comprised strains associated with infection ("clinical") or only worked on draft genomes (18–21). Comprehensive genomic analyses specifically comparing clinical and non-clinical strains have not been conducted, yet.

In this study, we performed genomic analyses between *C. difficile* strains of non-clinical background from environmental samples and closely related, clinical reference strains isolated from infected humans. We focused on MGEs and potentially linked genes encoding antibiotic resistances (ARG) or virulence factors and conducted a pan genome analysis. All these analyses were put into the genomic context through direct genome comparisons of corresponding clinical and non-clinical strains. We detected genomic differences that were linked to clinical background and might reflect increased physiological adaptation ability.

Materials and methods

Non-clinical *C. difficile* strains were isolated from horse feces, biogas fermenter sludge and mud. The environmental samples were collected between November 2019 and July 2020 with sterile canonical falcon tubes and were stored at 4°C upon arrival in the laboratory. Different antibiotic-free and antibiotic-based isolation approaches were employed. As a result, strains

J2_1 and TS3_3 were isolated without antibiotics, whereas MA_1 and B1_2 originated from isolation approaches with moxalactam norfloxacin (CDMN, Oxoid Deutschland GmbH, Wesel, Germany). Details of isolation protocols are described in supplementary text file S1. In general, environmental samples were dissolved in anoxic PBS (pH 7.4) (22) and pasteurized before inoculating the enrichment media. Grown enrichment cultures were plated on solid media with 1.5% agar and colonies examined for identity via 16S rRNA gene Sanger sequencing using colony PCR with Phusion High-Fidelity polymerase (Thermo Fisher Scientific, Waltham, MA, USA) and primers 08f (5′-AGAGTTTGATCCTGGC-3′) and 1504r (5′-TACCTTGTTACGACTT-3′), following the recommendations of the manufacturer. PCR products were purified with the QIAquick PCR Purification kit (Qiagen, Hilden, Germany) as recommended by the manufacturer and subjected to Sanger sequencing by Microsynth Seqlab GmbH (Göttingen, Germany).

Clinical reference strains of sequence types (ST)/ribotypes (RT) corresponding to the four non-clinical strains were kindly provided by the Institute of Medical Microbiology, Göttingen, Germany. Strains DSM 28196, DSM 29747, SC083-01-01 and SC084-01-01 had been isolated from infected humans as described in Riedel et al (23).

Isolates were routinely cultivated at 37°C under anoxic conditions in supplemented Brain Heart Infusion Broth (BHIS; supplemented with 0.5% yeast extract, 0.05% L-cystein, 0.0001% Na-resazurin, purged with nitrogen).

DNA extraction

Genomic DNA was extracted from overnight cultures using the MasterPure Gram Positive DNA Purification kit as recommended by the manufacturer (Epicentre, Madison, WI, USA). DNA quality was assessed on a NanoDrop ND-1000 (Peqlab Biotechnologie GmbH, Erlangen, Germany), and DNA concentration was measured using the Qubit 3.0 Fluorometer (Thermo Fisher Scientific) with the BR dsDNA assay kit.

Ribotyping of *C. difficile* isolates

Isolated strains were phylogenetically examined via ribotyping based on Bidet et al (24). Amplification of the 16S-23S rRNA intergenic spacer region was conducted with the Dreamtaq polymerase (Thermo Fisher Scientific) using reagents as recommended by the manufacturer with 0.2 mM of each primer and 50 ng template DNA per 50 µl PCR reaction. PCR cycling comprised initial denaturation at 95°C for 3 min, followed by 30 cycles of 95°C for 1 min, 56°C for 30 s, and 72°C for 1 min. Final elongation was performed at 72°C for 5 min. PCR products

were separated on a 2% agarose gel ran at 5 V/cm with subsequent staining using ethidium bromide and visualization with the AlphaImager HP (Alpha Innotech Corp., San Leandro, USA) and AlphaView Software (v3.5.0). For RT assignment, observed band patterns were compared to already known RTs.

Genome sequencing, assembly, and annotation

For whole-genome sequencing of the non-clinical isolates, genomic DNA was subjected to short-read and long-read sequencing using Illumina and Oxford Nanopore technology, respectively. Illumina sequencing libraries were prepared with the Nextera XT DNA sample preparation kit and sequenced using a MiSeq instrument and reagent kit v3 (2 × 300 bp, 600 cycles) as recommended by the manufacturer (Illumina, San Diego, CA, USA). For Nanopore sequencing, genomic DNA without specific size selection was processed using the ligation sequencing kit 1D (SQK-LSK109) and the native barcode expansion kit (EXP-NBD104) according to the manufacturer's specifications (Oxford Nanopore Technologies, Oxford, UK). Nanopore sequencing was performed with the MinION system using a SpotON flow cell Mk I (R9.4.0) for 72 h. All following software was used with default settings unless otherwise stated. The MinKNOW software (v19.12.5) with implemented Guppy (v3.2.10) was used in fast mode for demultiplexing and base calling. Nanopore reads were first trimmed using Porechop (v0.2.4) (https://github.com/rrwick/Porechop) and filtered with Filtlong (v0.2.1) (https://github.com/rrwick/Filtlong), following assembly with Flye (v2.9.2) (25). Illumina reads were processed with fastp (v0.23.3) (26) and trimmed using Trimmomatic (v0.39) (27). The long-read assembly was polished with the processed short reads using softwares BWA (v0.7.17, r1188) (28) and Polypolish (v0.5.0) (29). Circularization of the assemblies were verified with Bandage v0.8.1 (30) and assemblies rotated to *dnaA* as start position with Circlator (v1.5.5) (31). The assembled genome sequences were annotated with Prokka (v1.14.5) (32). Selenoproteins were curated manually.

Genome sequencing and assembly of the clinical reference strains was done by Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany and the Institute of Microbiology at the Technische Universität Braunschweig, Germany. High molecular weight DNA was prepared using the Qiagen Genomic Tip/100 G kit (Qiagen, Hilden, Germany). SMRTbell template libraries were prepared according to the instructions from Pacific Biosciences, Menlo Park, CA, USA, following the Procedure & Checklist - 20 kb Template Preparation Using BluePippin Size-Selection System. Briefly, for preparation of 15kb libraries 5µg genomic DNA were end-repaired and ligated overnight to

hairpin adapters applying components from the DNA/Polymerase Binding Kit P6 from Pacific Biosciences, Menlo Park, CA, USA. Reactions were carried out according to the manufacturer´s instructions. BluePippin Size-Selection was performed according to the manufacturer´s instructions (Sage Science, Beverly, MA, USA). Conditions for annealing of sequencing primers and binding of polymerase to purified SMRTbell template were assessed with the Calculator in RS Remote, PacificBiosciences, Menlo Park, CA, USA. SMRT sequencing was carried out on the PacBio RSII (PacificBiosciences, Menlo Park, CA, USA) taking 240-minutes movies. Long read genome assembly was performed with the "RS_HGAP_Assembly.3" protocol included in SMRTPortal version 2.3.0 using default parameters. Chromosomal contigs and plasmids were circularized, particularly artificial redundancies at the ends of the contigs were removed and adjusted to *dnaA*. Identification of redundancies and the replication genes has been done based on BLAST, circularization and rotation to the replication genes has been performed by genomecirculator.jar tool (https://github.com/boykebunk/genomefinish). Error-correction was performed by a mapping of Illumina short reads onto finished genome using Burrows-Wheeler Alignment bwa 0.6.2 in paired-end (sampe) mode using default settings (28) with subsequent variant and consensus calling using VarScan 2.3.6 (33).

Genomic analyses

In general, plots were created with RStudio (v2022.06.0) (34) using the package ggplot2 (v3.4.2) (35), and final modifications were done with Inkscape (v0.48; https://inkscape.org/de/). MLST assignment of the non-clinical strains was done using PubMLST (36). Genome qualities were assessed with CheckM2 (v1.0.2) (37) before performing genome analyses.

The program antiSMASH (v7.0.0) (38) was used for predicting secondary metabolite biosynthetic gene clusters. Putative ARGs were identified with RGI (v6.0.2), CARD (v3.2.7) (39), and AMRFinderPlus (v3.11.14) (40) employing the NCBI Bacterial Antimicrobial Resistance Reference Gene Database (v2023-07-13.2).

Screening for virulence factors was performed by BLAST+ blastp analysis (v2.12.0) (41) using the *C. difficile*-associated protein sequences present in the full dataset (retrieved on 14.07.2023) from the virulence factor database (VFDB (42)) as query against the whole-genome protein sequences of the analyzed strains. The *spo0A* sporulation gene from *C. difficile* strain CD630 was additionally included in the analysis (CP010905.2, CDIF630_01363). Protein sequences of each virulence factor between corresponding genomes were additionally compared with blastp (41) to check for sequence deviations. Presence/absence of ARGs and protein sequence query coverage of virulence factors was visualized as heatmaps.

After initial assessment of toxin gene presence with the aforementioned VFDB analysis, the corresponding toxin-operons and adjacent genes were inspected for nucleotide sequence similarity and genomic location by sequence alignment with clinker (v1.32) (43), including reference sequences from *C. difficile* strain CD630 (CP010905.2, CDIF630_00771-00782) and R20291 (CP029423.1, CDIF27147_02765-02770), respectively.

Analysis of MGEs. Genomes were analyzed with PlasmidFinder (v2.1) (44) for plasmid family identification. Insertion sequences (IS) were identified with ISEScan (v1.7.2.3) (45). Genomic islands (GI) were predicted with various tools, including PHASTEST in deep mode (on 25.6.23) (46) for prophage prediction, IslandViewer 4 (accessed on 4.7.23) (47) and ICEscreen (v1.2.0) (48). The numbers of identified MGEs and their types were visualized as heatmaps.

Pan genome analysis. The core/pan genome including all genomes was calculated with Roary (v3.13.0) (49), and a Venn diagram visualizing the results was created with Inkscape (v0.48; https://inkscape.org/de/). Venn diagrams showing the shared and unique genes for each pair of non-clinical and corresponding clinical strain as estimated by Roary were visualized with ggplot2. The unique genes were assigned to functional clusters of orthologous groups of proteins (COG) with eggNOG-mapper (v2.1.9) (50). Relative abundance of unique genes of a specified COG was determined for each genome relative to its total number of coding sequences (CDS) and visualized as bar charts. Further, the differences between clinical and non-clinical strain of these relative COG-gene abundances were calculated by subtracting the relative values of the non-clinical from the clinical strain. These difference values were also plotted.

Pairwise genome alignment and comparison. For direct genome comparison, genomes of non-clinical and corresponding clinical reference strains were first aligned with Mauve (v20150226) (51) and inspected for significant sequence deviations detected as alignment gaps over multiple CDSs. These CDSs were inspected for their predicted function and compared to the previous pan genome analysis. Further, Proksee (specifically: CGView Builder v1.1.2 + Features v1.0.0) (52) was used for visualization of each genome complemented with its previously predicted ARGs, MGEs and unique genes. Additionally, pairwise genome alignments of clinical and non-clinical strains were performed with MUMmer (v3.23) (53) (options: -maxmatch; -l 100; -b), with each genome used as query. The resulting alignment positions at the reference sequence were also implemented in the genome visualization. Proksee depictions of corresponding genomes were combined and modified using Inkscape (v0.48; https://inkscape.org/de/) for direct genome comparison.

Results and discussion

General genome characteristics

The four non-clinical *C. difficile* strains belong to ST1/RT027 (*C. difficile* TS3_3), ST3/RT001/072 (*C. difficile* B1_2), ST8/RT002 (*C. difficile* J2_1), and ST11/RT078 (*C. difficile* MA_1). These STs/RTs are known for their high clinical relevance and/or prevalence and prominent representatives of the phylogenetic clades (Table 1) (54,55). Clinical strains DSM 28196 (ST1/RT027), SC084-01-01 (ST3/RT001/072), SC083-01-01 (ST8/RT002), and DSM 29747 (ST11/RT078) were used in genome-based investigations covering analyses of MGEs as well as core and accessory genes. Throughout this work, the mentioning of corresponding strains refers to clinical and non-clinical strains of the same ST.

All eight genomes were initially evaluated for quality. The CheckM2 analysis thoroughly verified uniform genome completeness (99.86% - 99.99%) and purity (0.1% - 0.78% contamination).

General genomic features of the analyzed strains are listed in Table 1. For the sake of simplicity, the analyzed strains/genomes will be designated in the following with their ST and clinical background (clinical = med, non-clinical = env) instead of their actual strain name. Most of the genomes comprised extrachromosomal elements (ECE). The ST3 genomes exhibited even two co-occurring ECEs. However, ECE carriage was not necessarily linked to the ST. Additionally, ECE size varied between the genomes of corresponding clinical and non-clinical strains, indicating their divergence. The clinical strains exhibited larger total genome size (including ECEs), and correspondingly more CDSs than non-clinical strains. No differences between clinical and non-clinical strains were recorded with respect to the number of rRNA and tRNA genes.

The screening for putative gene clusters encoding biosynthesis of secondary metabolites did not show differences between corresponding strains. All strains possessed regions predicted to encode cyclic-lactone-autoinducer, non-ribosomal peptide synthetase, or ranthipeptide. Since the capacity for secondary metabolite production did not differ between the strains, they were not considered in further analyses.

Table 1. General genomic features of the analyzed strains.

	Non-clinical *C. difficile* strains			
Strain name	TS3_3	B1_2	J2_1	MA_1
ST / clade	1 / 2	3 / 1	8 / 1	11 / 5
Designation	ST1-env	ST3-env	ST8-env	ST11-env
RT	027	001/072	002	078
GenBank accession	CP134872	CP132141 CP132142 CP132143	CP134690 CP134691	CP132139 CP132140
Chromosome size (bp)	4,116,134	4,194,230	4,081,925	3,970,170
ECEs size (bp)	-	42,358 / 7,624	46,261	33,670
No. of CDS	3,630	3,763	3,652	3,545
No. of curated selenoproteins	4	4	4	4
No. of rRNA	35	35	35	35
No. of tRNA	90	90	90	90

	Clinical *C. difficile* strains			
Strain name	DSM 28196	SC084-01-01[a]	SC083-01-01[a]	DSM 29747
ST / clade	1 / 2	3 / 1	8 / 1	11 / 5
Designation	ST1-med	ST3-med	ST8-med	ST11-med
RT	027	001/072	002	078
GenBank accession	CP012320.1	CP132146 CP132147 CP132148	CP132144 CP132145	CP019864.1
Chromosome size (bp)	4,205,365	4,184,644	4,122,919	4,071,596
ECEs size (bp)	-	47,363 / 130,799	45,313	-
No. of CDS	3,707	3,950	3,704	3,556
No. of curated selenoproteins	4	4	4	4
No. of rRNA	35	34	35	35
No. of tRNA	91	92	90	90

[a]Internal strain ID as defined by the Institute of Medical Microbiology, Göttingen, Germany.

In silico examination of virulence factors for genomic assessment of virulence potential

Genomic examinations of the strains for the presence of fundamental virulence factors of *C. difficile* (listed in VFDB (42)) were performed to assess the virulence potential of the corresponding strains. Most of the characterized *C. difficile* virulence factors were present in the genomes of all strains (Fig 1). The main virulence factors in *C. difficile* pathogenicity, the disease-causing toxin genes *tcdA* and *tcdB* encoded by the pathogenicity locus (PaLoc), were identified in all strains (Fig 1). Closer investigations of the PaLoc in all genomes including the regulatory genes *tcdC* and *tcdR*, along with *tcdE* (56–58) confirmed its consistent genomic location between the same genes (*cdu1 and cdd1*) like in the reference genome of *C. difficile* strain CD630 (59) (Fig 2A-B). DNA alignment of the PaLoc-operons demonstrated that genes were 100% identical between corresponding clinical and non-clinical strains and even between strains of different STs (Fig 2A). All genes except *tcdE* shared at least 80% nucleotide sequence similarity to all other aligned genomes (Fig 2B). In addition to the PaLoc, another toxin-

harboring locus (CdtLoc) is known in certain *C. difficile* strains of clades 2, 3 and 5 (60), which harbors the binary toxin CDT encoded by the genes *cdtA* and *cdtB* (61). The entire binary toxin genes were identified in both clinical and non-clinical strains of ST1 and ST11, while only 13% of the gene sequences were present in the other strains (Fig 1). The CdtLoc was also found at a consistent genomic location between the same genes (CDIF27147_02765 and *trpS)* in accordance with the reference genome of *C. difficile* strain R20291 (62) (Fig 2C-D). Moreover, the regulatory gene *cdtR* was observed in all genomes. The *cdtR* gene sequences were identical in most of the genomes (Fig 2C). Strains lacking genes *cdtA* and *cdtB* harbored the same five small CDS instead (Fig 2C-D). In summary, toxin gene analyses confirmed uniform presence and location of toxin genes among corresponding clinical and non-clinical strains as well as identical toxin gene sequences. Sequence variants of toxin genes or the corresponding regulatory genes were demonstrated to influence strain virulence (19,63). Based on our sequence comparisons, the corresponding clinical and non-clinical strains exhibit the same genomic virulence potential.

In addition to the toxin genes, other proteins are relevant for *C. difficile* pathogenicity, including genes involved in cell adherence, sporulation or motility, that together determine colonization efficiency (64). Adherence proteins are involved in biofilm formation, which in turn affects resistance to harmful substances such as antibiotics (65). Most of the known adherence-associated virulence factors were identified, and only few of the corresponding genes were only partially present or missing in particular strains (Fig 1). The patterns of virulence factor involved in adherence was largely consistent between corresponding strains except for ST1-CD0873 and ST11-CwpV. In some cases, including the aforementioned exceptions, protein sequences of adherence virulence factors deviated between corresponding strains by only 0.10% to 1.59%. In most cases, one amino acid was deviating from the reference sequence. These differences occurred in strains of both clinical/environmental background so that no link to clinical background could be recorded. Few sequence deviations consisted of a missing stretch of several amino acids in protein CbpA. This protein is described with a modular architecture comprising different repeat types and repetitions (66). Both missing amino acid stretches in CbpA occurred in the non-clinical strains, but affected different regions of the protein. The effect of all these sequence variants in the analyzed adherence proteins on strain virulence so far remains unclear. Different studies addressed divergent protein sequences of the adherence virulence factors CwpV, CbpA, or Cwp66 (66–68). However, the former studies focused on the modular architecture of the protein, instead of single amino acid deviations. In contrast, the latter studies investigated the effect of complete gene deletion or disruption, which

significantly altered cell adhesion, but also stress tolerance and antibiotic resistance in the case of Cwp66. Thus, determination of the adherence virulence factors functionalities with regard to colonization and the accompanying virulence of the clinical and non-clinical strains would require experimental investigations. Since differences between corresponding strains included only few of numerous proteins involved in adherence and mainly consisted of single amino acid exchanges, the data basis did not allow deducing resulting significant virulence alterations. Further, no clinical-related pattern in the various differences was observed.

The protein sequences of the proteases Zmp1 and Cwp84, which are involved in exoenzymatic reactions, were entirely present in all strains. The Zmp1 sequences were all identical, while protein Cwp84 of both ST8 strains covered 98% of the reference protein sequence with 100% sequence similarity, whereas the Cwp84 protein sequences of ST3 strains aligned with 100% query coverage but differed amongst each other by 0.12%. A gene encoding the sporulation protein Spo0A was present in all strains with 100% identity. Another important aspect in *C. difficile* virulence is its flagellum-based motility. The flagellar operon was represented by 41 CDSs in the *C. difficile*-associated protein sequences dataset of VFDB. Here, the fraction of the 41 flagellar CDSs present was used instead of protein sequence coverage to compare the strains. Clinical and non-clinical counterparts showed identical coverages of predominantly 90% of the flagellar CDSs, whereas strains of ST11 only possessed 24% (Fig 1). These two strains were already observed to lack a flagellum (data not shown). Sequence comparisons between corresponding genomes revealed complete congruence for almost all strains and CDSs. Solely ST3 strains deviated in two sequences (VFG043312 - MinD/ParA, VFG043317 - FliM/FliN) by maximal 0.8%. Thus, only a few differences in the analyzed virulence factors were observed in corresponding clinical and non-clinical strains. Potentially, a similar virulence would therefore be expected for all strains, independent of their environmental or clinical origin.

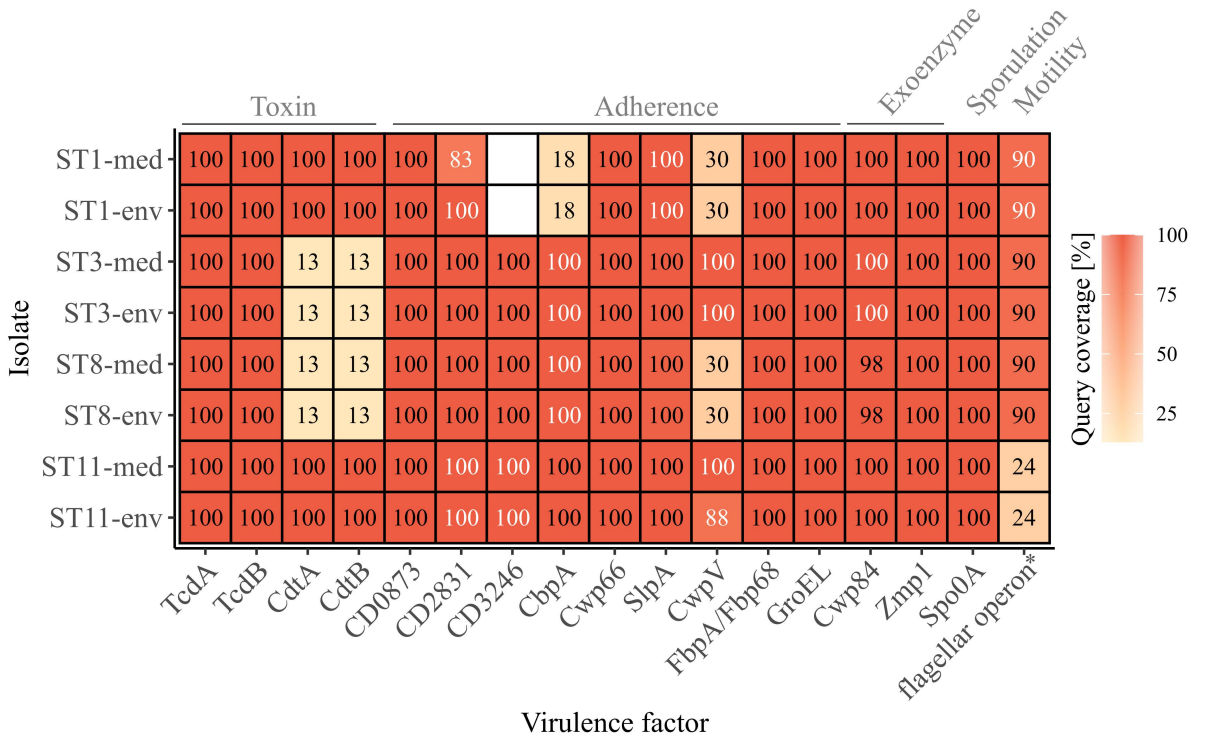


Fig 1. *C. difficile*-associated virulence factors in the analyzed strains. Presence of the examined virulence factors is indicated as the protein sequence query coverage to the reference VFDB dataset; by color and noted coverage value. White coverage values highlight differing sequences between clinical and non-clinical strain. Virulence factors are labelled with their gene names as obtained from the VFDB dataset, and their related functions are stated on top. *flagellar operon comprising 41 CDSs obtained from the VFDB dataset, and its query coverage calculated as the relative number of present CDSs of the total 41.

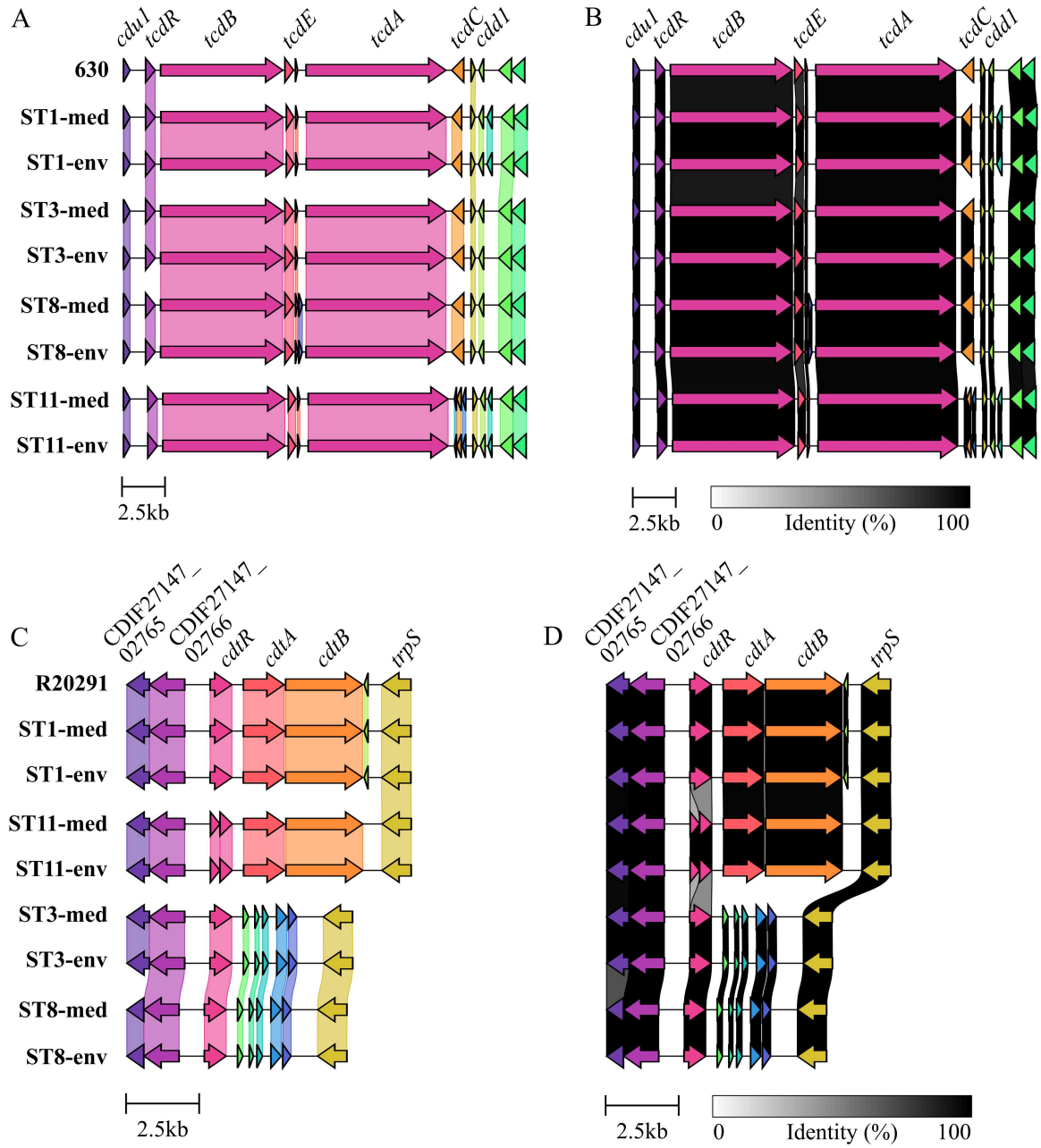


Fig 2. Gene cluster comparisons of toxin loci. Genes within and next to the toxin-encoding loci were compared on nucleotide sequence level between all analyzed strains for (A-B) the PaLoc with CD630 as reference, and (C-D) the CdtLoc with R20291 as reference sequence. (A) and (C) depict 100% sequence identity, while (B) and (D) represent identities above 80%.

Core/pan genome analysis

Overall core/pan genome analysis with all eight genomes resulted in 2,735 groups of core genes and varying numbers of unique genes between 12 and 351 (Fig 3A). A high number of unique genes was not associated with the clinical background of the strains. However, pairwise comparison of corresponding clinical and non-clinical strains showed that the clinical ones always possessed more accessory genes. Focusing on this result, pairwise pan genome analyses of clinical and non-clinical strain pairs were performed, and the relative proportions of unique

genes per genome in relation to the total number of CDSs calculated to take account of the different genome sizes (Fig 3B). These analyses confirmed the previous observation of more unique genes forming the accessory genomes of the clinical strains. Differences ranged from 0.21% between ST11 strains to 4.36% between ST3 strains.

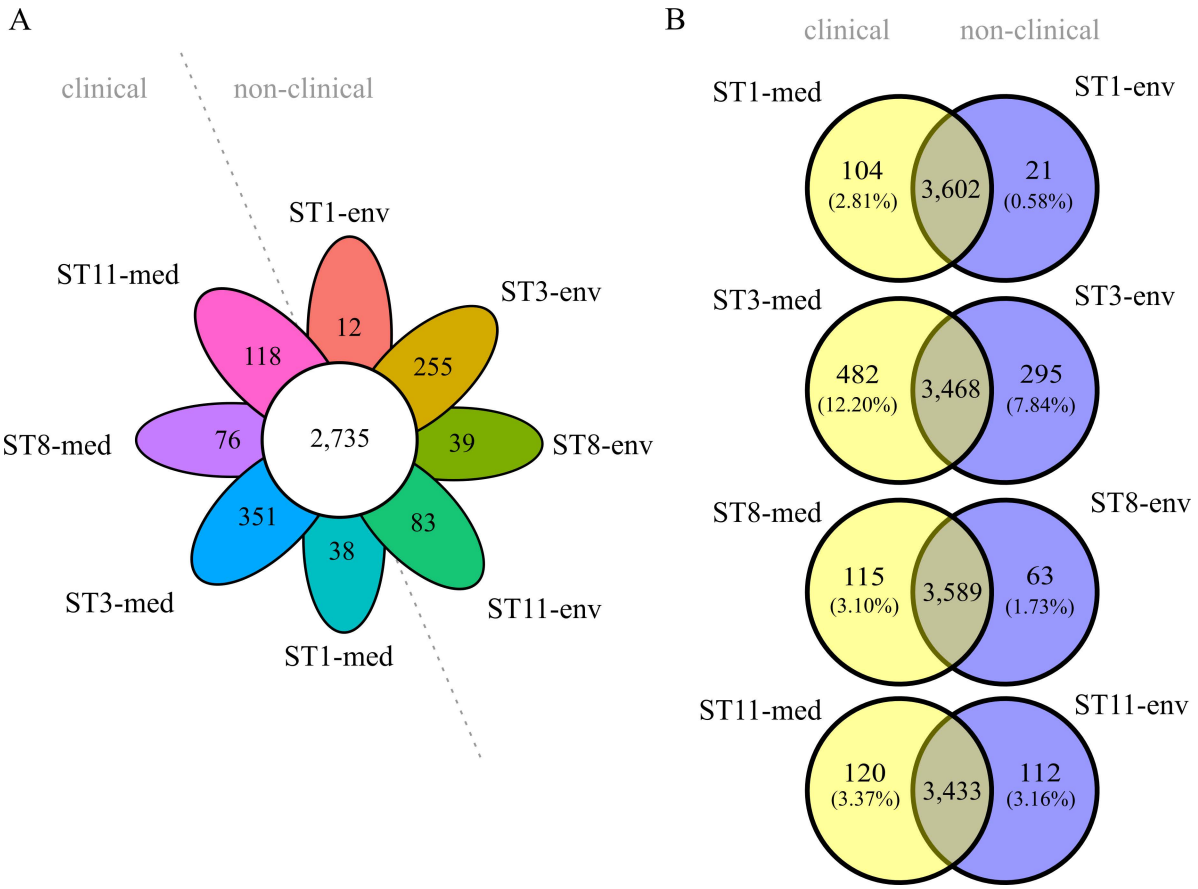


Fig 3. Core and accessory genome sizes of the analyzed *C. difficile* strains. Venn diagrams depicting the shared and unique genes among (A) all eight strains, or (B) pairwise between ST-corresponding clinical and non-clinical strain. The relative proportions of unique genes with regard to the total number of CDS per genome are indicated in parentheses below each absolute number of unique genes.

To elucidate genomic differences between clinical and non-clinical strains, unique genes were functionally classified into COGs with eggNOG-mapper (50). Thereby, not all input genes could be classified, and several classified genes were not assigned to a specific COG. The number of un-classified or un-assigned genes varied between the genomes, but almost all these genes were annotated as hypothetical proteins. For better comparison, the relative abundance of assigned COGs in each accessory genome was calculated as described above (Fig 4A). This illustrated the previously determined differences in relative unique gene carriage within clinical and non-clinical strain pairs. Besides the noticeable numbers of unclassified and unassigned genes (hypothetical proteins), further bars representing the COG categories S ("Function unknown"), K ("Transcription"), L ("Replication, recombination and repair") and also M ("Cell

wall/membrane/envelope biogenesis") were prominent to varying degrees in all strains, and even seemed to be more abundant in the clinical strains. To further examine this, the differences in unique gene/COG proportions between corresponding clinical and non-clinical strains were determined (Fig 4B). This allowed identifying abundance trends of specific unique genes-COGs. The COG category S was the most abundant category and dominated in the clinical strains. Prokka-annotated functions (32) of the genes assigned to COG category S were diverse and included for example phage-related proteins. Inspection of these genes for further potential virulence factors revealed the two genes encoding haemolysin XhlA and the virulence-associated protein E. Haemolysin XhlA and virulence-associated protein E are not associated with *C. difficile* virulence according to the data in VFDB (42), but involved in the virulence of other bacteria such as *Clostridium chauvoei* (69) and *Streptococcus suis* serotype 2 (70). Nevertheless, though *C. difficile* is not established as hemolytic pathogen, some evidence of hemolysis was recorded (71). Influence of haemolysin XhlA and virulence-associated protein E on *C. difficile* virulence however has not been examined, yet, so that their virulence potential remains unknown. Both of these genes were present in ST3-env, while ST3-med and ST8-med possessed the haemolysin gene and ST11-med the virulence-associated protein E only. Thus, a direct relation to clinical background was not recorded.

The next highest differences were visible for COG categories L and K in particular, both being more abundant in three of the four clinical strains. Therefore, the accessory genomes of the clinical strains encoded more genes of unknown function, recombination- or transcription-associated activity than the genomes of non-clinical strains. Interestingly, Lewis et al. identified accessory genes of these functions to correlate with higher strain virulence (14). For example, the genes *rep* and *recF* encoding DNA helicase and DNA recombinase, and the gene *iap* encoding the lysozyme-like family protein were found in ST1 strains of high virulence, whereas they were absent in low-virulence ST1 strains. This gene presence/absence pattern was likewise true for our ST1 strains with respect to clinical background instead of *in vivo*-measured disease severity, implying a higher virulence potential of the clinical strain based on the accessory genome. Another study also identified accessory transcriptional regulators in hypervirulent strains of RT027 in contrast to the less virulent predecessor strain and pointed to the significance of these genomic accessories on strain virulence (73). However, strain virulence has to be validated *in vivo* to support these assumptions. The increased proportion of transcriptional accessory genes compared to total CDS content in the clinical strains might reflect adaptation to higher environmental variability, likewise the increased genome size/CDS carriage (Table 1) (72). Noteworthy, Sebaihia et al. already pointed to the high amount of

transcriptional regulators in *C. difficile* strain 630 and associated them with its potential ability to adapt to a rapidly changing environment (4). Consequently, the clinical strains could be appraised to physiologically react faster to changing environmental conditions such as emerging stress factors, which in turn affects survival as well as colonization and disease manifestation.

Although COG category M seemed noticeably abundant in the individual proportions (Fig 4A), it did neither exhibit a specific trend nor remarkable abundance in the proportional differences (Fig 4B). Contrary, category U ("Intracellular trafficking, secretion, and vesicular transport") showed a noticeable trend towards genomes of clinical strains. Examination of the U-unique genes revealed that they were effectively restricted to clinical strains of ST1, ST3, and ST11, and thereby encoded only proteins involved in conjugal transfer of DNA, such as type IV secretory system components and relaxases/mobilization nuclease domain proteins. This is interesting in the context of horizontal gene transfer for fast adaptation to changing environmental conditions such as the presence of antibiotics (12). Moreover, Brouwer et al. demonstrated the conjugative transfer of the PaLoc from a toxigenic to a non-toxigenic *C. difficile* strain, which turned the non-toxigenic strain into a toxin-producing one (13). Bacterial conjugation is mediated by cell-to-cell contact, which is eminently present in bacterial biofilms. Biofilm production and conjugation activity are therefore intertwined. This was already demonstrated in other bacteria such as *Escherichia coli* (74) and *Bacillus subtilis* (75), where biofilm formation was shown to significantly impact conjugation efficiency. Taking this and the occurrence of conjugal accessory genes in clinical strains into account, the above identified sequence deviations in some adherence virulence factors could also be related to conjugal activity in the context of biofilm formation. This potential association has not been addressed in *C. difficile* so far but would be worth considering in investigations on its virulence.

The only tendency to the non-clinical strains showed unique genes of the COG category P ("Inorganic ion transport and metabolism"), which was represented by only one gene per each non-clinical accessory genome encoding a cobalt transport protein, an ABC transporter transmembrane region, or a sodium sulphate symporter.

Conclusively, the comprehensive pan genome analysis in clinical and non-clinical strain comparisons established associations between clinical background and higher abundance of hypothetical proteins or proteins of unknown functions, and of genes linked to increased potential of conjugal and transcriptional activity. This trend is possibly linked to higher

virulence, and in general can contribute to rapid physiological and evolutionary adaptation, which implies elevated virulence.

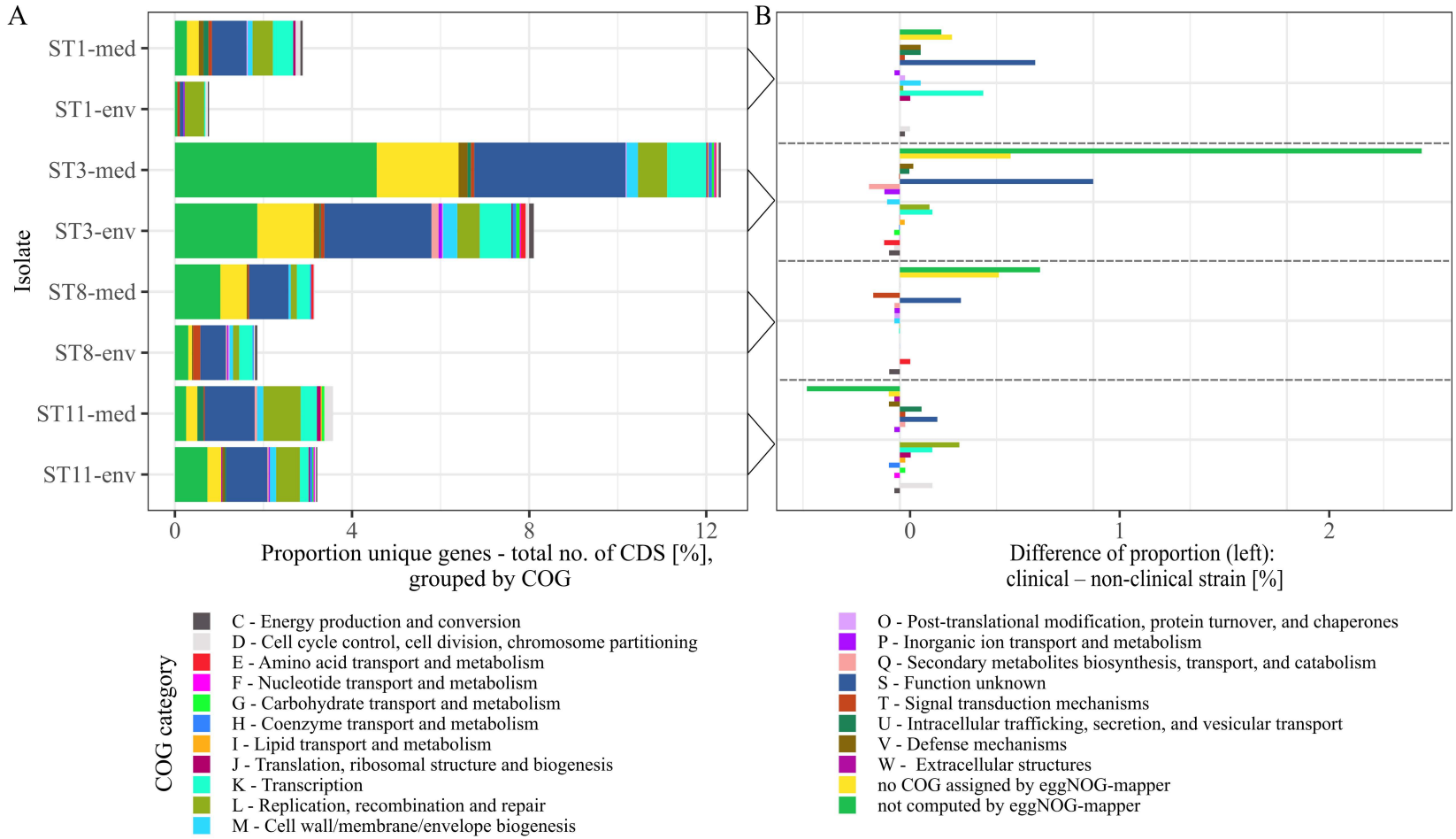


Fig 4. Relative abundance of COGs assigned to the unique genes from pairwise pan genome analyses. (A) The relative proportions of unique genes to total number of CDSs per genome in pairwise comparisons was transferred to their assigned COGs, which are designated with COG category name and function. (B) Individual COG proportions of unique genes of non-clinical strains were subtracted from the corresponding clinical strain to see if specific COGs are more frequent among accessory genes of a certain clinical background

Prediction of ARGs

Antibiotic resistances are another crucial factor in *C. difficile* virulence, as they often allow *C. difficile* colonization and infection manifestation (2). Further, ARGs can be linked to MGEs and contribute to the accessory genome (4), which might partially explain the previously determined higher number of unique genes in the genomes of the clinical strains (Fig 3). We inspected all eight genomes for putative ARGs (including AR-conferring mutations) (Fig 5). Corresponding strains exhibited similar ARG patterns. Few genes were only identified in one strain, or in multiple strains of the same ST or clinical background. For example, genes conferring resistance against macrolides (*erm*(B)), rifampin (*rpo*B^{R505K}) or tetracyclines (*tet*(40), *tet*(M)) were only detected in certain clinical strains, while a gene contributing to streptothricin (*sat4*) or aminoglycoside resistance (*ant*(6)-Ia, *aph*(3')-IIIa) was only identified in one non-clinical strain. Thus, a specific resistance was not linked to all clinical strains. Interestingly, ARGs *erm*(B) and *tet*(M) that were restricted to clinical strains are known to be associated to MGEs and thereby transferrable between strains (76,77). Regarding the ST, ST1 strains encoded most ARGs. Overall, clinical strains encoded one to four more ARGs than non-clinical strains.

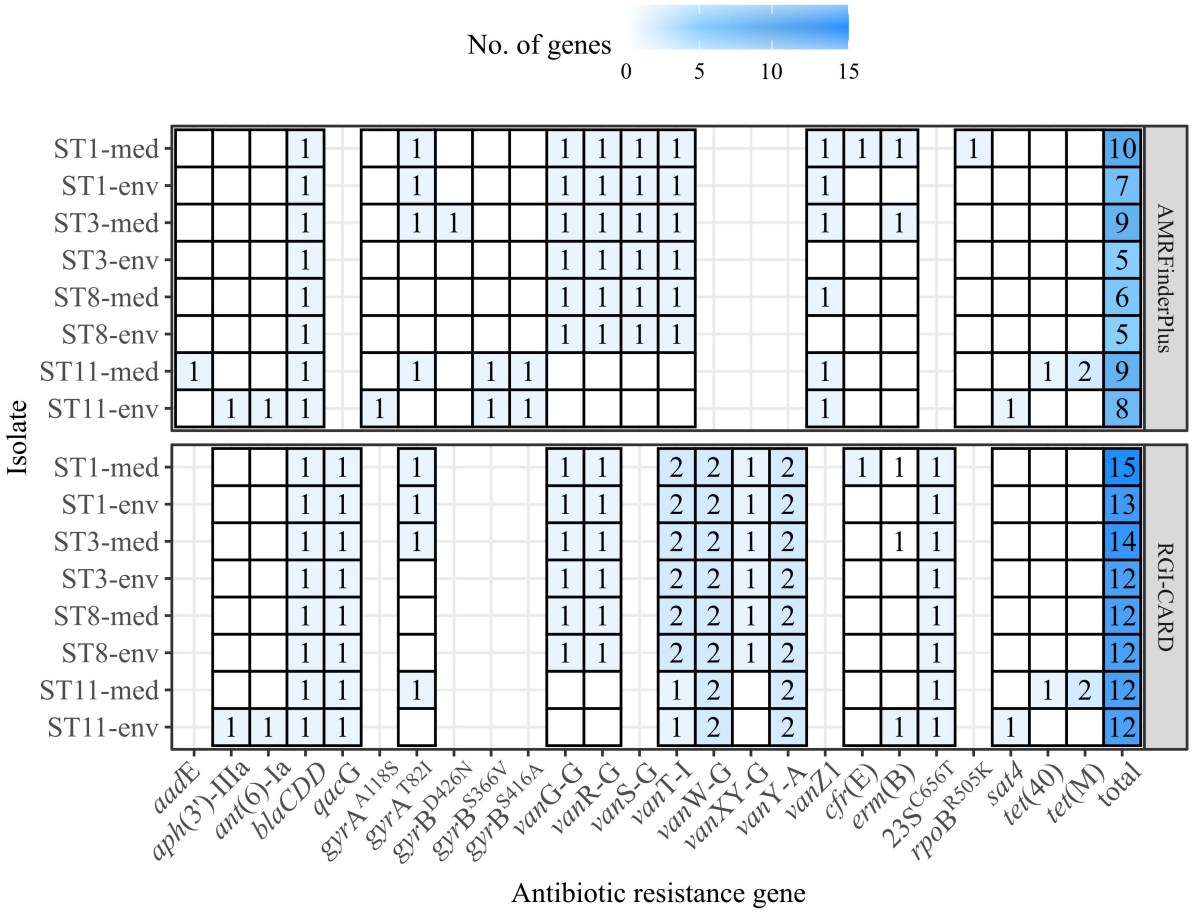


Fig 5. Predicted ARGs in the analyzed *C. difficile* strains. The number of identified ARGs and AR-conferring mutations as predicted with RGI-CARD (39) and AMRFinderPlus (40) were indicated by color and respective value, with white/no value meaning gene absence. Heatmap-tiles are missing for genes that were not part of the analysis tools. The total number of predicted ARGs for each program is additionally stated.

Prediction of MGEs

All analyzed *C. difficile* genomes were investigated for MGEs. A plasmid family was solely present in ST11-med, representing the type repUS43 twice. Consequently, none of the observed ECEs was classified as plasmid, indicating another extrachromosomal type such as cryptic plasmids or prophages (78,79). Thus, the genomes of all strains were analyzed for putative prophage regions. Contrary to the plasmid analysis, all ECEs besides the 7.6 kb element of ST3-env represented putatively intact prophages spanning the entire ECEs. Up to four incomplete or intact prophage regions were predicted per genome, except for ST11-med, which carried only one putative, intact prophage (Fig 6). Thereby, corresponding clinical and non-clinical strains showed comparable prophage carriage.

Integrative elements were detected in all strains except ST8-med (Fig 6). Two different types of elements, Integrative and Conjugative Elements (ICEs) and Integrative Mobilizable Elements (IMEs) were identified in both clinical and non-clinical strains. The two superfamilies

Tn*916* and Tn*5252* represented ICEs, both observed as complete or incomplete modules without the ability of integration ("conjugation module"). The complete Tn*916* elements dominated, while incomplete Tn*916* and in-/complete Tn*5252* modules were likewise minor abundant. The ICE Tn*916* was already described in *C. difficile* and associated with antibiotic resistances, predominantly with *tet*(M) followed by *erm*(B) (76,80,81), although the latter one is rather linked to other MGEs (77). Consequently, this ICE is of interest in the context of spreading antibiotic resistances. Tn*5252* modules were restricted to clinical strains, while the incomplete Tn*916* modules were only found in non-clinical strains. Moreover, clinical ST1 and ST11 strains possessed three times more ICE modules than their non-clinical counterpart. Altogether, the ICE predictions could explain the presence of accessory genes with conjugal function (COG U) in the clinical strains described above, as they possessed elements that were missing in the corresponding non-clinical strains. Looking on IMEs, none were identified in ST8-strains, while all other isolates possessed IME modules of the families MOB_Q, MOB_T, and MOB_V. MOB_T elements dominated with similar occurrence in clinical and non-clinical strains, whereas each of the other two IME modules occurred once, but only in clinical strains. Overall, IMEs were less common than ICEs. Similarly to ICEs, IMEs are potential carrier for ARGs such as *tet*(M) and, thus, are also involved in distribution of antibiotic resistances (82).

Thirteen and predominantly 15 GIs were detected on each chromosome (Fig 6). These numbers exceeded the occurrences of the previously examined GI prophages and integrative elements and indicated the presence of other GI types. Depending on the ST, clinical strains possessed the same number of predicted GIs or more than the corresponding non-clinical strain. In addition to various GIs, we also examined IS elements and detected eight different IS families (Fig 6). The prevailing family was IS*200*/IS*605* with an incidence of six to 36 elements in the genomes of both ST11 strains. Elements belonging to families IS*21* and IS*256* occurred twice per genome on average, while family IS*3* was present several times ST1 strain genomes and once in all other genomes. These IS families were identified to be potentially linked to ARGs and, consequently, might further contribute to ARG spread amongst *C. difficile* strains (83). All other IS families were identified once per genome or not at all. Elements of type IS*1595* and IS*L3* were only found in ST11 or ST3-med strain genomes, respectively. Altogether, corresponding clinical and non-clinical strains exhibited similar IS patterns. Thereby, the total number of IS elements ranged between ten and 43 per genome, with clinical strains mostly possessing more IS elements (between one and seven more) than non-clinical strains. Further, general IS abundance was correlated to ST strain genomes, with ST8-strain genomes possessing the lowest and ST11-strain genomes the highest number of IS elements. Thus, regarding all

above described MGEs, an overall trend of higher MGE carriage in clinical than in non-clinical strains was recorded.

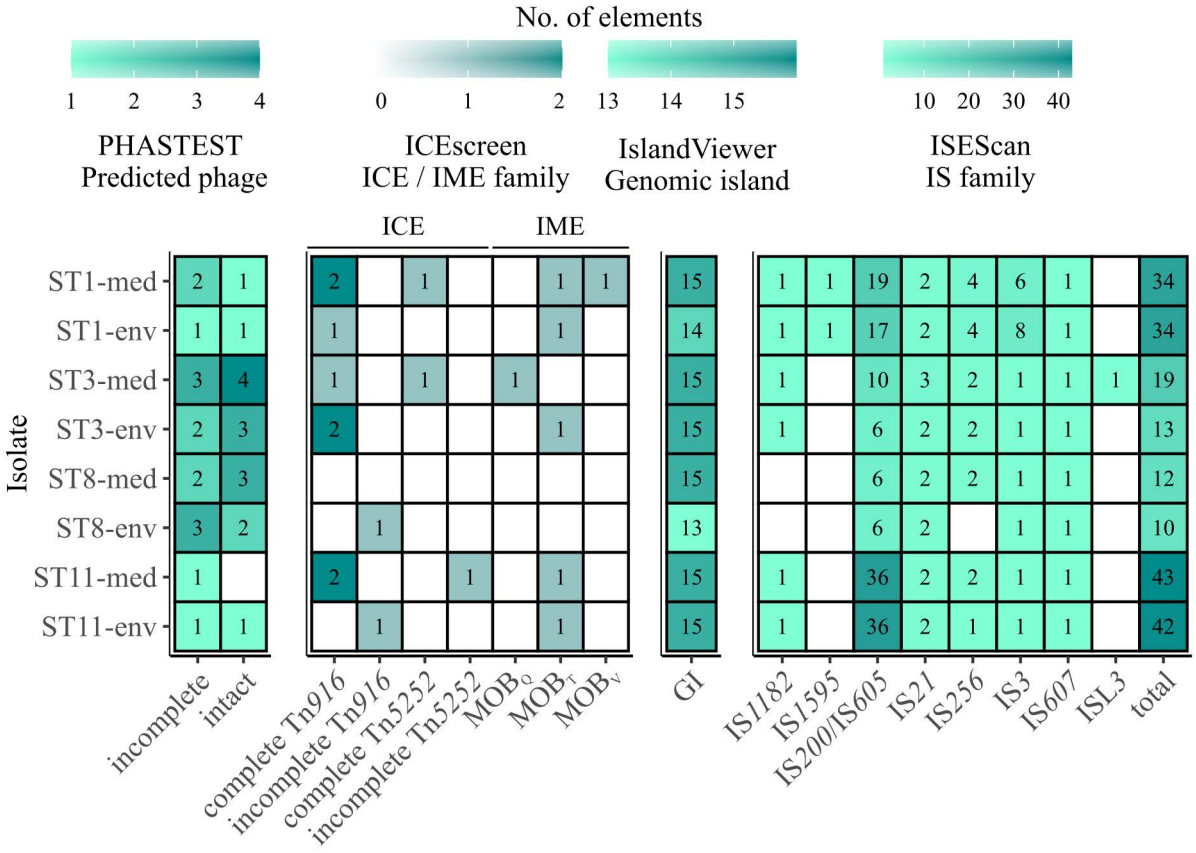


Fig 6. Predicted MGEs in the analyzed strains. The number of the analyzed MGEs prophages, integrative elements, GIs, and IS elements were indicated by color and the respective number, with white/no value meaning no prediction. Prophages are described as incomplete or intact as predicted by PHASTEST (46). Integrative elements are categorized by the assigned superfamily and grouped into ICE and IME. Presence of IS elements is described in the context of the identified families, and additionally given as total number of IS elements.

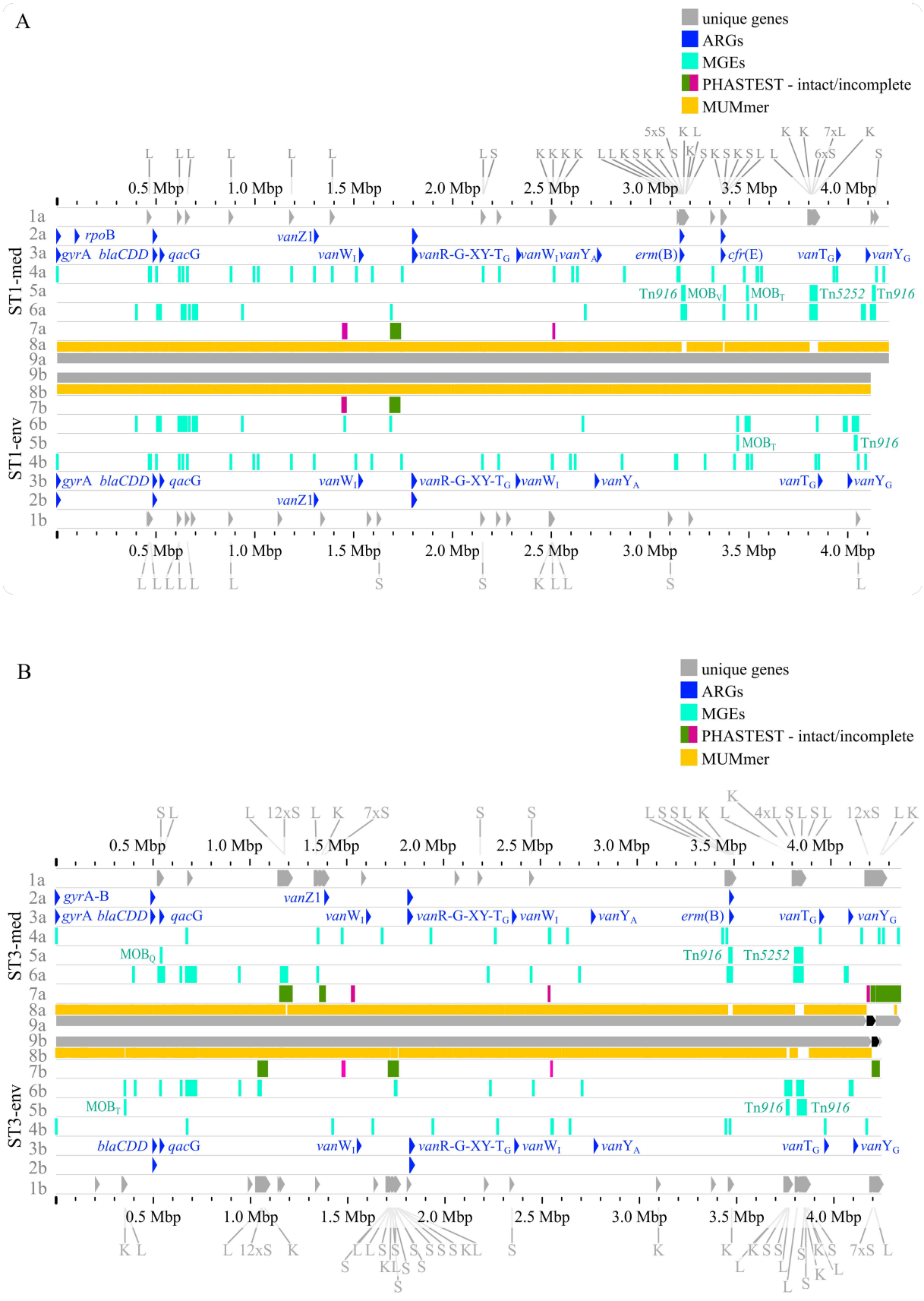
Pairwise genome comparisons with implementation of preceding analyses

The previous analyses of accessory genes, ARGs, and MGEs showed differences between clinical and non-clinical strains. All these results together with MUMmer alignments (53) were combined and put into genomic context in aligned, pairwise, linear genome analyses (Fig 7). This type of representation revealed connections between the various analyzed elements based on co-occurrence. First, the MUMmer alignments demonstrated once again higher abundance of unique genes in the clinical strains, as they exhibited more alignment gaps that corresponded to expansive regions missing in the corresponding non-clinical strain. Consequently, these regions accorded with the majority of the mapped unique genes occurring in clusters. Mapping of MGEs and ARGs initially illustrated that those elements/genes identified in corresponding strains likewise resided at the same genomic positions. MGEs and ARGs that were only present in one of the two compared strains, mostly the clinical strain, often occurred together. The

genome of strain ST1-med (Fig 7A) possessed the ARG *erm*(B) that resided within ICE Tn916. This conjunction was observed in ST3-med genome as well (Fig 7B) and supported the already mentioned connection between Tn*916* elements and ARG *erm*(B) (80,81). ARG *tet*(M) in the ST11-med genome exhibited the presumed connection with Tn*916* (Fig 7D) (76). ST11-med further possessed ARGs *tet*(40) and *aad*E, which occurred within an incomplete Tn*5252* element (Fig 7D). In contrast, the non-clinical ST11 strain carried three ARGs (*ant*(6)-Ia, *sat4, aph*(3')-IIIa) close to each other and to an incomplete Tn*916* element, thus located outside of this predicted ICE region (Fig 7D). However, the GI prediction identified a larger mobile region than determined for the incomplete Tn*916* that included the three ARGs. The region was confirmed by an alignment gap and a cluster of unique genes with similar sizes. The close location of these ARGs conferring resistance to aminoglycosides and streptothricins was already observed in genomes of ST13 and ST49 *C. difficile* strains (RT014, clade 1), which derived from porcine, but without a connection to a MGE (21). ST11-env was isolated from horse feces, which indicates a broader distribution of this ARG-cluster beyond strains of RT014 and pigs as hosts. Another ARG, *qac*G of the ST3-med genome (Fig 7B), was located within a predicted GI and next to a complete IME of family MOB_Q. All further *qag*G genes and other ARGs were not observed within MGEs. However, some of these ARGs were located in close proximity to all types of predicted MGEs, such as the remaining *qac*G genes (Fig 7A-D), *cfr*(E) in the ST1-med genomes (Fig 7A), or *van*T-G in ST1 and ST3 strain genomes (Fig 7A and B). Connections between these ARGs and MGEs are speculative but might still be interesting for further investigations regarding dissemination of antibiotic resistances. For instance, ARG *cfr*(E) was also found within an undescribed MGE in genomes of RT027 *C. difficile* strains from Mexico (84).

The predicted prophage regions, integrative elements, and GIs occurring in only one of the corresponding strains matched with alignment gaps and coincided with the unique gene clusters. Correspondingly, this accounted for the majority of the genomic differences between corresponding strains. The unique genes in these clusters with COG assignments to categories S, K, and L could be associated to MGEs. This is especially interesting for the COG category S of "Unknown function", which could now be prompted to be involved in the function of the respective MGE or encoding accessory functions potentially relevant for virulence. Accessory genes of COG category S with virulence potential were already identified during pan genome analysis, which encoded haemolysin XhlA and the virulence-associated protein E. Within this genome comparison analyses, these potential virulence factors were now associated to MGEs and, thus, might be transferrable between cells, which sheds another light on these potential

virulence factors. Many of the unique genes that were only annotated as hypothetical proteins belonged to the MGE-associated clusters, which indicates involvement in MGE-related functions.


A
unique genes
ARGs
MGEs
PHASTEST - intact/incomplete
MUMmer
ST1-med
ST1-env
0.5 Mbp
1.0 Mbp
1.5 Mbp
2.0 Mbp
2.5 Mbp
3.0 Mbp
3.5 Mbp
4.0 Mbp
rpoB
vanZ1
gyrA blaCDD qacG
vanW_I
vanR-G-XY-T_G
vanW_I vanY_A
erm(B)
efr(E)
vanT_G
vanY_G
Tn916
MOB_V
MOB_T
Tn5252
B
unique genes
ARGs
MGEs
PHASTEST - intact/incomplete
MUMmer
ST3-med
ST3-env
gyrA-B
vanZ1
gyrA blaCDD qacG
vanW_I
vanR-G-XY-T_G
vanW_I
vanY_A
erm(B)
vanT_G
vanY_G
MOB_Q
Tn916
Tn5252
MOB_T
blaCDD qacG

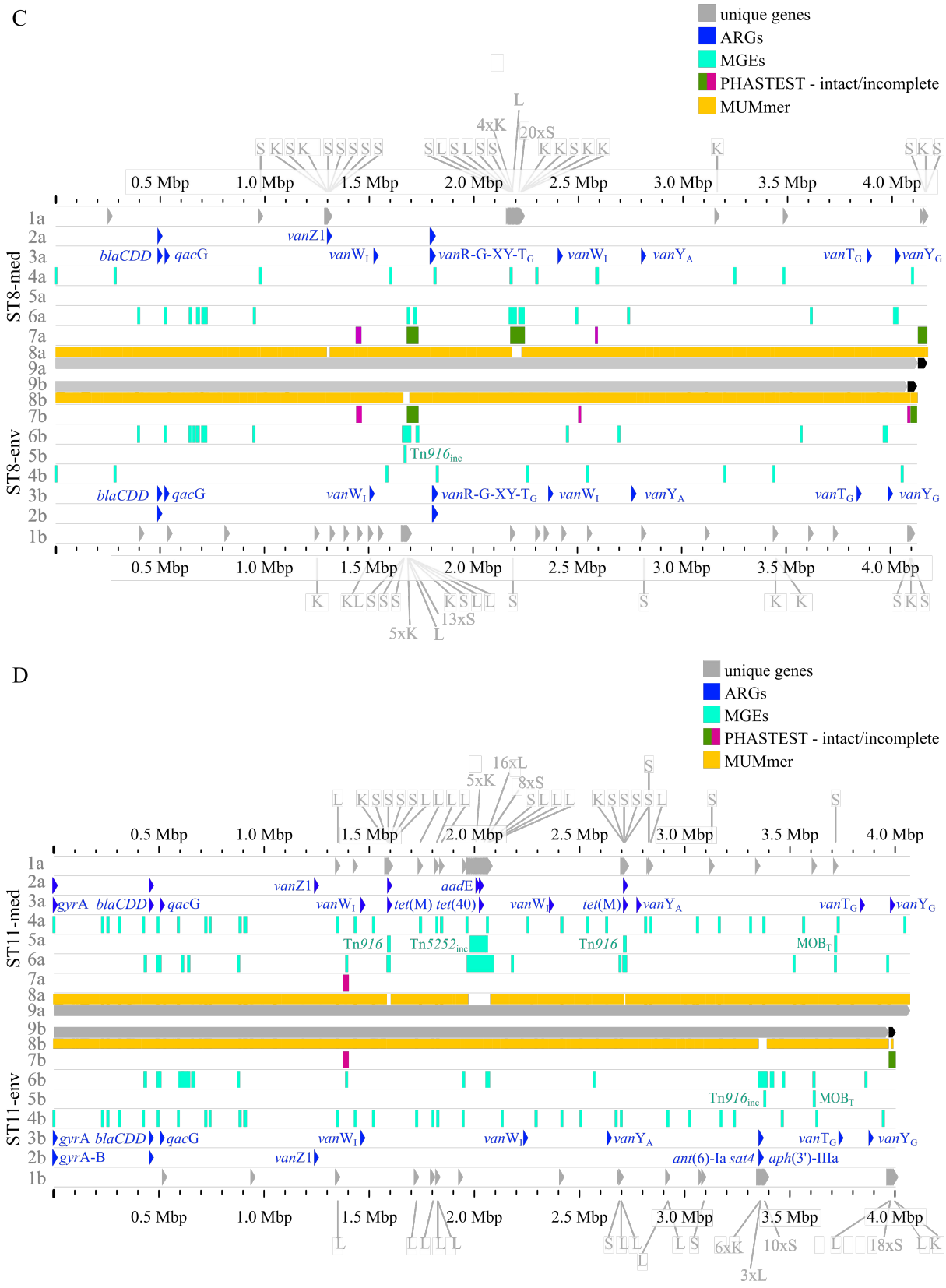


Fig 7. Pairwise genome comparisons complemented with the predicted ARGs, MGEs, and accessory genes. Genome comparisons of (A) ST1-strains, (B) ST3-strains, (C) ST8-strains, and (D) ST11-strains are depicted with different tracks for each visualized feature in the clinical strain at the top (track letter a) and non-clinical strain at the bottom (track letter b). The tracks represent: 1 – unique genes with genes assigned to COG S, K, L highlighted, and multiple genes of the same COG grouped together if necessary for better visibility, 2 – AMRFinderPlus (40) predicted ARGs, 3 – RGI+CARD (39) predicted ARGs, 4 – predicted IS elements, 5 – predicted integrative

elements labelled with assigned superfamily, 6 – predicted GIs, 7 – prophage prediction with completeness color-coded according to PHASTEST (46), 8 – MUMmer alignment (53), 9 – replicons.

Conclusion

The comprehensive genome analyses and comparisons of corresponding clinical and non-clinical *C. difficile* strains revealed genomic patterns associated with clinical background. No distinct differences in virulence factors known to be crucial for *C. difficile* virulence, such as toxins, and proteins involved in adherence and motility were detected. Thus, corresponding strains possessed the same fundamental virulence equipment, which suggested same virulence regardless of the clinical/non-clinical background. Pan genome analysis revealed that clinical strains possessed a larger accessory genome. Assignment of the unique genes to functional clusters demonstrated the trend in clinical strains with more unique genes previously annotated as hypothetical proteins or functionally assigned to COG categories S, K, L, and U. Such trend of those accessory genes/functions is linked to higher virulence and enables the strain to rapidly respond to changing environmental conditions such as emerging stress, which supports bacterial survival, colonization, and disease manifestation. Further analyses predicted various ARGs and MGEs. No particular ARG or MGE was specifically linked to clinical background, but the overall trend of more ARGs and MGEs in clinical strains. Results from pan genome, ARG, and MGE analyses revealed conjunctions between specific ARGs and MGEs. The genome comparisons further demonstrated that genomic differences between clinical and non-clinical strains mainly originated from MGEs. This also included the majority of unique genes with higher abundance in clinical strains that were assigned to COG categories with connection to increased virulence and faster physiological reactivity. Consequently, these trends suggested adaptations of the clinical strains by gene acquisition that might manifest in higher strain virulence. This should be further investigated to elucidate *C. difficile* virulence and progression, especially in the context of clinical and non-clinical strain comparison. Therefore, future investigations are advised to incorporate non-clinical strains in comparative analyses for a comprehensive understanding of *C. difficile* virulence development. These findings also highlight the importance of MGEs for *C. difficile* since they seem to be involved not only in the dissemination of ARGs or virulence factors but also impact virulence in another way. We further advise to examine genomic analyses in whole-genome context to reveal conjunctions between the various elements.

Acknowledgements

We thank Uwe Groß, Wolfgang Bohne, Julian Schwanbeck, and Ines Oehmig for providing the clinical reference strains. We also thank Dr. Cathrin Spröer (DSMZ) for support and Melanie Heinemann as well as Nicole Heyer (DSMZ) 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. 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. Czepiel J, Dróżdż M, Pituch H, Kuijper EJ, Perucki W, Mielimonka A, et al. *Clostridium difficile* infection: review. Eur J Clin Microbiol Infect Dis [Internet]. 2019 Jul 3;38(7):1211–21. Available from: https://journals.lww.com/00013542-201807000-00001
4. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet [Internet]. 2006 Jul 25;38(7):779–86. Available from: https://www.nature.com/articles/ng1830
5. Mullany P, Allan E, Roberts AP. Mobile genetic elements in *Clostridium difficile* and their role in genome function. Res Microbiol [Internet]. 2015 May;166(4):361–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S092325081400254X
6. Smits WK, Roseboom AM, Corver J. Plasmids of *Clostridioides difficile*. Curr Opin Microbiol [Internet]. 2022 Feb;65:87–94. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1369527421001521
7. Goh S, Hussain H, Chang BJ, Emmett W, Riley T V., Mullany P. Phage ϕC2 Mediates Transduction of Tn *6215*, Encoding Erythromycin Resistance, between *Clostridium difficile* Strains. Onderdonk AB, editor. MBio [Internet]. 2013 Dec 31;4(6):1–7.

Available from: https://journals.asm.org/doi/10.1128/mBio.00840-13

8. Sekulovic O, Meessen-Pinard M, Fortier L-C. Prophage-Stimulated Toxin Production in *Clostridium difficile* NAP1/027 Lysogens. J Bacteriol [Internet]. 2011 Jun;193(11):2726–34. Available from: https://journals.asm.org/doi/10.1128/JB.00787-10
9. Govind R, Vediyappan G, Rolfe RD, Dupuy B, Fralick JA. Bacteriophage-Mediated Toxin Gene Regulation in *Clostridium difficile*. J Virol [Internet]. 2009 Dec;83(23):12037–45. Available from: https://journals.asm.org/doi/10.1128/JVI.01256-09
10. Riedel T, Wittmann J, Bunk B, Schober I, Spröer C, Gronow S, et al. A *Clostridioides difficile* bacteriophage genome encodes functional binary toxin-associated genes. J Biotechnol [Internet]. 2017 May;250:23–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168165617300809
11. Mehner-Breitfeld D, Rathmann C, Riedel T, Just I, Gerhard R, Overmann J, et al. Evidence for an adaptation of a phage-derived holin/endolysin system to toxin transport in Clostridioides difficile. Front Microbiol. 2018;9:2446.
12. de la Cruz F, Davies J. Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol [Internet]. 2000 Mar;8(3):128–33. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0966842X00017030
13. Brouwer MSM, Roberts AP, Hussain H, Williams RJ, Allan E, Mullany P. Horizontal gene transfer converts non-toxigenic *Clostridium difficile* strains into toxin producers. Nat Commun [Internet]. 2013 Oct 17;4(1):2601. Available from: https://www.nature.com/articles/ncomms3601
14. Lewis BB, Carter RA, Ling L, Leiner I, Taur Y, Kamboj M, et al. Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to *Clostridium difficile* Virulence. Ballard JD, editor. MBio [Internet]. 2017 Sep 6;8(4):e00885-17. Available from: https://journals.asm.org/doi/10.1128/mBio.00885-17
15. Weese JS. *Clostridium* (*Clostridioides*) *difficile* in animals. J Vet Diagnostic Investig [Internet]. 2020 Mar 6;32(2):213–21. Available from: http://journals.sagepub.com/doi/10.1177/1040638719899081
16. Janezic S, Potocnik M, Zidaric V, Rupnik M. Highly Divergent *Clostridium difficile* Strains Isolated from the Environment. Paredes-Sabja D, editor. PLoS One [Internet]. 2016 Nov 23;11(11):e0167101. Available from: https://dx.plos.org/10.1371/journal.pone.0167101

17. Ozaki E, Kato H, Kita H, Karasawa T, Maegawa T, Koino Y, et al. *Clostridium difficile* colonization in healthy adults: transient colonization and correlation with enterococcal colonization. J Med Microbiol [Internet]. 2004 Feb 1;53(2):167–72. Available from: https://www.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.05376-0
18. Zhou Y, Zhou W, Xiao T, Chen Y, Lv T, Wang Y, et al. Comparative genomic and transmission analysis of *Clostridioides difficile* between environmental, animal, and clinical sources in China. Emerg Microbes Infect [Internet]. 2021 Jan 1;10(1):2244–55. Available from: https://www.tandfonline.com/doi/full/10.1080/22221751.2021.2005453
19. Dong Q, Lin H, Allen M-M, Garneau JR, Sia JK, Smith RC, et al. Virulence and genomic diversity among clinical isolates of ST1 (BI/NAP1/027) *Clostridioides difficile*. Cell Rep [Internet]. 2023 Aug;42(8):112861. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2211124723008720
20. Xu X, Bian Q, Luo Y, Song X, Lin S, Chen H, et al. Comparative Whole Genome Sequence Analysis and Biological Features of *Clostridioides difficile* Sequence Type 2‡. Front Microbiol [Internet]. 2021 Jul 5;12:651520. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.651520/full
21. Knight DR, Squire MM, Collins DA, Riley T V. Genome Analysis of *Clostridium difficile* PCR Ribotype 014 Lineage in Australian Pigs and Humans Reveals a Diverse Genetic Repertoire and Signatures of Long-Range Interspecies Transmission. Front Microbiol [Internet]. 2017 Jan 11;7:2138. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2016.02138/full
22. Phosphate-buffered saline (PBS). Cold Spring Harb Protoc [Internet]. 2006 Jun;2006(1):pdb.rec8247. Available from: http://www.cshprotocols.org/lookup/doi/10.1101/pdb.rec8247
23. Riedel T, Wetzel D, Hofmann JD, Plorin SPEO, Dannheim H, Berges M, et al. High metabolic versatility of different toxigenic and non-toxigenic *Clostridioides difficile* isolates. Int J Med Microbiol [Internet]. 2017 Sep;307(6):311–20. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1438422117301029
24. Bidet P, Barbut F, Lalande V, Burghoffer B, Petit J-C. Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. FEMS Microbiol Lett [Internet]. 1999 Jun 15;175(2):261–6. Available from: https://academic.oup.com/femsle/article-lookup/doi/10.1111/j.1574-

6968.1999.tb13629.x

25. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol [Internet]. 2019 May 1;37(5):540–6. Available from: https://www.nature.com/articles/s41587-019-0072-8
26. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta [Internet]. 2023 May 8;2(2):e107. Available from: https://onlinelibrary.wiley.com/doi/10.1002/imt2.107
27. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics [Internet]. 2014 Aug 1;30(15):2114–20. Available from: https://academic.oup.com/bioinformatics/article/30/15/2114/2390096
28. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics [Internet]. 2010 Mar 1;26(5):589–95. Available from: https://academic.oup.com/bioinformatics/article/26/5/589/211735
29. Wick RR, Holt KE. Polypolish: Short-read polishing of long-read bacterial genome assemblies. Schneidman-Duhovny D, editor. PLOS Comput Biol [Internet]. 2022 Jan 24;18(1):e1009802. Available from: http://dx.doi.org/10.1371/journal.pcbi.1009802
30. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics [Internet]. 2015 Oct 15;31(20):3350–2. Available from: https://academic.oup.com/bioinformatics/article/31/20/3350/196114
31. Hunt M, Silva N De, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol [Internet]. 2015 Dec 29;16(1):294. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0849-0
32. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics [Internet]. 2014 Jul 15;30(14):2068–9. Available from: https://academic.oup.com/bioinformatics/article/30/14/2068/2390517
33. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res [Internet]. 2012 Mar;22(3):568–76. Available from: http://genome.cshlp.org/lookup/doi/10.1101/gr.129684.111
34. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. 2020. Available from: http://www.rstudio.com/
35. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. 1–260 p. Available from: https://ggplot2.tidyverse.org

36. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res [Internet]. 2018 Sep 24;3:124. Available from: https://wellcomeopenresearch.org/articles/3-124/v1
37. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods [Internet]. 2023 Aug 27;20(8):1203–12. Available from: https://www.nature.com/articles/s41592-023-01940-w
38. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res [Internet]. 2023 Jul 5;51(W1):W46–50. Available from: https://academic.oup.com/nar/article/51/W1/W46/7151336
39. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res [Internet]. 2023 Jan 6;51(D1):D690–9. Available from: https://academic.oup.com/nar/article/51/D1/D690/6764414
40. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep [Internet]. 2021 Jun 16;11(1):12728. Available from: https://www.nature.com/articles/s41598-021-91456-0
41. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. BMC Bioinformatics. 2009;10:1–9.
42. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res [Internet]. 2022 Jan 7;50(D1):D912–7. Available from: https://academic.oup.com/nar/article/50/D1/D912/6446532
43. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. Robinson P, editor. Bioinformatics [Internet]. 2021 Aug 25;37(16):2473–5. Available from: https://academic.oup.com/bioinformatics/article/37/16/2473/6103786
44. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. *In Silico* Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrob Agents Chemother [Internet]. 2014

Jul;58(7):3895–903. Available from: https://journals.asm.org/doi/10.1128/AAC.02412-14

45. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Hancock J, editor. Bioinformatics [Internet]. 2017 Nov 1;33(21):3340–7. Available from: https://academic.oup.com/bioinformatics/article/33/21/3340/3930124
46. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST : faster than PHASTER, better than PHAST. Nucleic Acids Res. 2023;gkad382:1–8.
47. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res [Internet]. 2017 Jul 3;45(W1):W30–5. Available from: https://academic.oup.com/nar/article/45/W1/W30/3787837
48. Lao J, Lacroix T, Guédon G, Coluzzi C, Payot S, Leblond-Bourget N, et al. ICEscreen: a tool to detect Firmicute ICEs and IMEs, isolated or enclosed in composite structures. NAR Genomics Bioinforma [Internet]. 2022 Oct 6;4(4):lqac079. Available from: https://academic.oup.com/nargab/article/doi/10.1093/nargab/lqac079/6765278
49. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics [Internet]. 2015 Nov 15;31(22):3691–3. Available from: https://academic.oup.com/bioinformatics/article/31/22/3691/240757
50. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Tamura K, editor. Mol Biol Evol [Internet]. 2021 Dec 9;38(12):5825–9. Available from: https://academic.oup.com/mbe/article/38/12/5825/6379734
51. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res [Internet]. 2004 Jul;14(7):1394–403. Available from: http://genome.cshlp.org/lookup/doi/10.1101/gr.2289704
52. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C, et al. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Res [Internet]. 2023 Jul 5;51(W1):W484–92. Available from: https://academic.oup.com/nar/article/51/W1/W484/7151341
53. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al.

Versatile and open software for comparing large genomes. Genome Biol [Internet]. 2004;5(2):R12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14759262

54. Knight DR, Imwattana K, Kullin B, Guerrero-Araya E, Paredes-Sabja D, Didelot X, et al. Major genetic discontinuity and novel toxigenic species in *Clostridioides difficile* taxonomy. Elife [Internet]. 2021 Jun 11;10:e64325. Available from: https://elifesciences.org/articles/64325
55. Walker AS, Eyre DW, Wyllie DH, Dingle KE, Griffiths D, Shine B, et al. Relationship Between Bacterial Strain Type, Host Biomarkers, and Mortality in *Clostridium difficile* Infection. Clin Infect Dis [Internet]. 2013 Jun 1;56(11):1589–600. Available from: https://academic.oup.com/cid/article/56/11/1589/304397
56. Mani N, Dupuy B. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. Proc Natl Acad Sci [Internet]. 2001 May 8;98(10):5844–9. Available from: https://pnas.org/doi/full/10.1073/pnas.101126598
57. Govind R, Dupuy B. Secretion of *Clostridium difficile* Toxins A and B Requires the Holin-like Protein TcdE. Cheung A, editor. PLoS Pathog [Internet]. 2012 Jun 7;8(6):e1002727. Available from: https://dx.plos.org/10.1371/journal.ppat.1002727
58. Matamouros S, England P, Dupuy B. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. Mol Microbiol [Internet]. 2007 May 30;64(5):1274–88. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.2007.05739.x
59. Monot M, Eckert C, Lemire A, Hamiot A, Dubois T, Tessier C, et al. *Clostridium difficile*: New Insights into the Evolution of the Pathogenicity Locus. Sci Rep [Internet]. 2015 Oct 8;5(1):15023. Available from: http://dx.doi.org/10.1038/srep15023
60. Martínez-Meléndez A, Cruz-López F, Morfin-Otero R, Maldonado-Garza HJ, Garza-González E. An Update on *Clostridioides difficile* Binary Toxin. Toxins (Basel) [Internet]. 2022 Apr 27;14(5):305. Available from: https://www.mdpi.com/2072-6651/14/5/305
61. Gerding DN, Johnson S, Rupnik M, Aktories K. *Clostridium difficile* binary toxin CDT. Gut Microbes [Internet]. 2014 Jan 31;5(1):15–27. Available from: http://www.tandfonline.com/doi/abs/10.4161/gmic.26854
62. Wang S, Heuler J, Wickramage I, Sun X. Genomic and Phenotypic Characterization of the Nontoxigenic *Clostridioides difficile* Strain CCUG37785 and Demonstration of Its Therapeutic Potential for the Prevention of *C. difficile* Infection. Uzal F, editor. Microbiol Spectr [Internet]. 2022 Apr 27;10(2). Available from: https://journals.asm.org/doi/10.1128/spectrum.01788-21

63. Lanis JM, Heinlen LD, James JA, Ballard JD. *Clostridium difficile* 027/BI/NAP1 Encodes a Hypertoxic and Antigenically Variable Form of TcdB. Bradley KA, editor. PLoS Pathog [Internet]. 2013 Aug 1;9(8):e1003523. Available from: https://dx.plos.org/10.1371/journal.ppat.1003523
64. Awad MM, Johanesen PA, Carter GP, Rose E, Lyras D. *Clostridium difficile* virulence factors: Insights into an anaerobic spore-forming pathogen. Gut Microbes [Internet]. 2014 Sep 3;5(5):579–93. Available from: http://www.tandfonline.com/doi/full/10.4161/19490976.2014.969632
65. Dapa T, Unnikrishnan M. Biofilm formation by *Clostridium difficile*. Gut Microbes. 2013;4(5):397–402.
66. Tulli L, Marchi S, Petracca R, Shaw HA, Fairweather NF, Scarselli M, et al. CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen. Cell Microbiol [Internet]. 2013 May;15(10):1674–87. Available from: https://onlinelibrary.wiley.com/doi/10.1111/cmi.12139
67. Reynolds CB, Emerson JE, de la Riva L, Fagan RP, Fairweather NF. The *Clostridium difficile* Cell Wall Protein CwpV is Antigenically Variable between Strains, but Exhibits Conserved Aggregation-Promoting Function. Cheung A, editor. PLoS Pathog [Internet]. 2011 Apr 21;7(4):e1002024. Available from: https://dx.plos.org/10.1371/journal.ppat.1002024
68. Zhou Q, Rao F, Chen Z, Cheng Y, Zhang Q, Zhang J, et al. The *cwp66* Gene Affects Cell Adhesion, Stress Tolerance, and Antibiotic Resistance in *Clostridioides difficile*. Khursigara CM, editor. Microbiol Spectr [Internet]. 2022 Apr 27;10(2):e0270421. Available from: https://journals.asm.org/doi/10.1128/spectrum.02704-21
69. Thomas P, Abdel-Glil MY, Eichhorn I, Semmler T, Werckenthin C, Baumbach C, et al. Genome Sequence Analysis of *Clostridium chauvoei* Strains of European Origin and Evaluation of Typing Options for Outbreak Investigations. Front Microbiol [Internet]. 2021 Sep 29;12:.732106. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.732106/full
70. Ji X, Sun Y, Liu J, Zhu L, Guo X, Lang X, et al. A novel virulence-associated protein, vapE, in Streptococcus suis serotype 2. Mol Med Rep [Internet]. 2016 Mar;13(3):2871–7. Available from: https://www.spandidos-publications.com/10.3892/mmr.2016.4818
71. Alkudmani ZSB. The identification and characterization of novel haemolysin genes from *Clostridium difficile* [Internet]. University College London; 2018. Available from:

https://discovery.ucl.ac.uk/id/eprint/10078163/7/Alkudmani_00_Thesis_edited.pdf

72. Parter M, Kashtan N, Alon U. Environmental variability and modularity of bacterial metabolic networks. BMC Evol Biol [Internet]. 2007 Dec 23;7:169. Available from: https://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-7-169
73. Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, et al. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol [Internet]. 2009;10(9):R102. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-9-r102
74. Ghigo J-M. Natural conjugative plasmids induce bacterial biofilm development. Nature [Internet]. 2001 Jul 26;412(6845):442–5. Available from: https://www.nature.com/articles/35086581
75. Lécuyer F, Bourassa J-S, Gélinas M, Charron-Lamoureux V, Burrus V, Beauregard PB. Biofilm Formation Drives Transfer of the Conjugative Element ICE *Bs1* in *Bacillus subtilis*. Ellermeier CD, editor. mSphere [Internet]. 2018 Oct 31;3(5):e00473. Available from: https://journals.asm.org/doi/10.1128/mSphere.00473-18
76. Dong D, Chen X, Jiang C, Zhang L, Cai G, Han L, et al. Genetic analysis of Tn916-like elements conferring tetracycline resistance in clinical isolates of *Clostridium difficile*. Int J Antimicrob Agents [Internet]. 2014 Jan;43(1):73–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0924857913003348
77. Kartalidis P, Skoulakis A, Tsilipounidaki K, Florou Z, Petinaki E, Fthenakis GC. *Clostridioides difficile* as a Dynamic Vehicle for the Dissemination of Antimicrobial-Resistance Determinants: Review and In Silico Analysis. Microorganisms [Internet]. 2021 Jun 25;9(7):1383. Available from: https://www.mdpi.com/2076-2607/9/7/1383
78. Amy J, Bulach D, Knight D, Riley T, Johanesen P, Lyras D. Identification of large cryptic plasmids in *Clostridioides (Clostridium) difficile*. Plasmid [Internet]. 2018 Mar;96–97:25–38. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0147619X18300040
79. Ramírez-Vargas G, Goh S, Rodríguez C. The Novel Phages phiCD5763 and phiCD2955 Represent Two Groups of Big Plasmidial *Siphoviridae* Phages of *Clostridium difficile*. Front Microbiol [Internet]. 2018 Jan 22;9:26. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2018.00026/full
80. Spigaglia P, Barbanti F, Mastrantonio P. Detection of a Genetic Linkage Between Genes Coding for Resistance to Tetracycline and Erythromycin in *Clostridium difficile*.

Microb Drug Resist [Internet]. 2007 Jun;13(2):90–5. Available from: http://www.liebertpub.com/doi/10.1089/mdr.2007.723

81. Spigaglia P, Carucci V, Barbanti F, Mastrantonio P. ErmB Determinants and Tn *916* -Like Elements in Clinical Isolates of *Clostridium difficile*. Antimicrob Agents Chemother [Internet]. 2005 Jun;49(6):2550–3. Available from: https://journals.asm.org/doi/10.1128/AAC.49.6.2550-2553.2005
82. López de Egea G, González-Díaz A, Guédon G, Lao J, Berbel D, Casabella A, et al. A New Integrative and Mobilizable Element Is a Major Contributor to Tetracycline Resistance in *Streptococcus dysgalactiae* subsp. *equisimilis*. Antibiotics [Internet]. 2023 Mar 15;12(3):579. Available from: https://www.mdpi.com/2079-6382/12/3/579
83. Razavi M, Kristiansson E, Flach C-F, Larsson DGJ. The Association between Insertion Sequences and Antibiotic Resistance Genes. LaPara TM, editor. mSphere [Internet]. 2020 Oct 28;5(5). Available from: https://journals.asm.org/doi/10.1128/mSphere.00418-20
84. Stojković V, Ulate MF, Hidalgo-Villeda F, Aguilar E, Monge-Cascante C, Pizarro-Guajardo M, et al. *cfr*(B), *cfr*(C), and a New *cfr*-Like Gene, *cfr*(E), in *Clostridium difficile* Strains Recovered across Latin America. Antimicrob Agents Chemother [Internet]. 2019 Dec 20;64(1):e01074-19. Available from: https://journals.asm.org/doi/10.1128/AAC.01074-19

Supplementaries

Supplementary file S1. **Isolation protocols of the non-clinical *C. difficile* strains.**

BHIS - supplemented Brain Heart Infusion Broth (BHIS; supplemented with 0.5% yeast extract, 0.05% L-cystein, 0.0001% Na-resazurin, purged with nitrogen)

***C. difficile* strain MA_1 – antibiotic-based isolation.** Isolation was based on the study of Dharmasena and Jiang (1). 25 g of environmental sample were suspended in 100 ml PBS, and solid particles removed with a filter paper. Cells were repeatedly washed via centrifugation before inoculated into the enrichment medium BHIB-YE-CYS-MN-T medium (Brain Heart Infusion Broth, 0.5% yeast extract, 0.05% *L*-cystein, 0.1% sodium taurocholate, moxalactam-norfloxacin (CDMN, Oxoid Deutschland GmbH, Wesel, Germany) and following the described enrichment procedure.

***C. difficile* strain J2_1 – antibiotic-free isolation.** Isolation was performed under strict anaerobic conditions. The environmental sample was suspended in 30 ml PBS via slight shaking and soaking at room temperature for five hours, and 1 ml was transferred to a reaction tube without withdrawing solid particles. The sample was pasteurized at 80°C for 10 minutes, and 100 µl undiluted or 1:2 diluted aliquots were plated on solid BHIS 1.5% agar supplemented with 0.1% taurocholic acid to promote *C. difficile* spore germination (2), some plates additionally contained the antibiotic moxalactam norfloxacin (CDMN, Oxoid). Plates were incubated at 37°C until colonies were visible, and strain J2_1 was isolated from an antibiotic-free plate.

***C. difficile* strain B1_2 – antibiotic-based isolation.** Isolation was performed under strict anaerobic conditions. The environmental sample was treated as described above for the isolation of *C. difficile* strain J2_1 until plating of sample aliquots. 100 µl aliquots of 1:2, 1:10, and 1:100 dilutions were plated on plates and incubated as described for strain J2_1. Strain B1_2 was isolated from an antibiotic-containing plate.

***C. difficile* strain TS3_3 – antibiotic-free isolation.** Isolation was performed under strict anaerobic conditions, and enrichment was done in a minimum amino acid-defined medium (3,4) to metabolically select for *C. difficile*. The medium based on the defined medium of Yamakawa et al., 1996 (4) with 1x amino acid composition and omitting Glucose, and was additionally supplemented with Na_2SeO_3, $Na_2MoO_4 \cdot 2\ H_2O$, and $Na_2WO_4 \cdot 2\ H_2O$ at the final concentration of 0.1 mg/L and Na-resazurin 0.0001% as anaerobic indicator. The medium was purged with N_2/CO_2 (80%/20%) and further reduced if necessary by dropwise addition of a Na_2S solution until the resazurin indicator turned colorless. The pH was adjusted to 7.2 if necessary. The medium was finally sterilized with a Filtropur S 0.2 µm filter (SARSTEDT AG & Co. KG, Nürnbrecht, Germany) while kept anaerobic. A scoop of the environmental sample was dissolved in 30 ml PBS via slight shaking and soaking at room temperature for five hours, 2 ml aliquots were briefly centrifuged to pellet solid particles and 1.5 ml of the supernatant transferred to new tubes for pasteurization at 80°C for 10 minutes. Subsequently, spores were briefly pelleted and incubated at 37°C for 45 minutes in an equal volume of PBS supplemented with 0.1% taurocholic acid and 1.3 mM glycine to promote *C. difficile* spore germination (2). This approach was used to inoculate 250 ml of enrichment medium, and incubated at 37°C for three days. 100 µl aliquots and 2 ml aliquots concentrated to 100 µl were plated on standard solid BHIS 1.5% agar and incubated at 37°C until colonies were visible.

References

1. Dharmasena M, Jiang X. Improving culture media for the isolation of *Clostridium difficile* from compost. Anaerobe [Internet]. 2018 Jun;51:1–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1075996418300374
2. 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
3. Karasawa T, Ikoma S, Yamakawa K, Nakamura S. A defined growth medium for *Clostridium difficile*. Microbiology [Internet]. 1995 Feb 1;141(Pt 2):371–5. Available from: https://www.microbiologyresearch.org/content/journal/micro/10.1099/13500872-141-2-371
4. Yamakawa K, Karasawa T, Ikoma S, Nakamura S. Enhancement of Clostridium difficile toxin production in biotin-limited conditions. J Med Microbiol [Internet]. 1996 Feb 1;44(2):111–4. Available from: https://www.microbiologyresearch.org/content/journal/jmm/10.1099/00222615-44-2-111

4.4 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

Published online on bioRxiv.org on 27 September 2023:
https://doi.org/10.1101/2023.09.27.559748.

Author Contributions

Conceptualization: **MS**, AP, RD

Experiments: **MS**

Data analysis: **MS**

Results interpretation: **MS**

Manuscript writing: **MS**

Manuscript review & editing: **MS**, AP, RD

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 prophage release from *C. difficile* isolates, and clinically relevant antibiotics were also investigated for their phage-inducing effect (16,17). All experiments on *C. difficile* phages however worked with cultivation conditions that do not represent the actual habitat. Some components of the intestinal environment are stressful for *C. difficile*, such as the secondary bile salts. One of the prominent secondary bile salts is deoxycholate (DCA), which can promote biofilm formation in *C. difficile* (18). It was further demonstrated that DCA induces the bacterial SOS response (19). Activation of the SOS response in turn induces prophages and lead to phage particle production and release via host cell lysis (20). It is therefore likely that DCA induces prophages as well, which would be a critical aspect in *C. difficile* biology and shed new light on genetic transfer *in vivo*.

In this work, we examined prophage activity in different *C. difficile* strains under spontaneous conditions and in the presence of DCA. The analyzed *C. difficile* strains were of non-clinical and clinical origin and corresponded pairwise to each other based on their sequence type (ST). Active prophage regions in these strains were identified by sequencing of particle-protected DNA, analyzed for DCA-induced activity and for possible differences between clinical and non-clinical strains that could contribute to virulence. The sequencing approach is more sensitive than electron microscopy or plaque assays, the commonly used detection

methods for active *C. difficile* phages. We could therefore detect active prophages that might otherwise be missed due to insufficient activity, but could contribute to HGT.

Methods

Strains and cultivation conditions

The *C. difficile* strains used in this study were of clinical or non-clinical clinical background (personal communication), with four pairs of clinical and non-clinical strains correpsonding in ST, and one additional non-clinical strain (Table 1). Strains were routinely cultivated under anaerobic conditions in supplemented Brain Heart Infusion Broth (BHIS; supplemented with 0.5% yeast extract, 0.05% L-cystein, 0.0001% Na-resazurin, purged with nitrogen) at 37°C. Putative prophage regions of the strains were predicted with PHASTEST (21).

Table 1. *C. difficile* strains used in this study. Strain information on ST (Clade), profile of toxins A. B, and CDT, clinical or non-clinical background, and GenBank accession of the 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

DCA tolerance assessment

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

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

Prophage induction and phage particle isolation

Phage particles were isolated from untreated and DCA-induced cultures. Two pre-warmed flasks of 55 ml BHIS for each isolate were inoculated 1:100 from an overnight culture and incubated at 37°C. Growth was monitored via OD_{600} measurements until an OD of ~0.6 (0.5-0.7). One culture for each isolate was induced with 0.255 mM (0.01%) DCA final concentration. The physiological concentration of DCA varies between individuals (22). We therefore used this concentration that is within the physiological range and was also used in various *C. difficile* studies regarding growth behavior or spore germination (7,26). The DCA solution was freshly prepared under anaerobic conditions, with DCA suspended in distilled water so that 100 µl inducing solution was required per 10 ml culture. The solution was sterilized by filtration (Filtropur S 0.2 µm, Sarstedt AG & Co. KG, Nürnbrecht, Germany) and anaerobically added to the induction cultures. The second culture was not induced for analysis of spontaneous phage activity. Induced and non-induced cultures were further incubated until 22 hrs total incubation. The final OD_{600} of each culture was determined before isolating phage particles.

For phage particle isolation, cells were pelleted via centrifugation at 4°C and 3,000 x *g* for 15 minutes. Remaining cells were removed by filtration of the supernatant with a 0.45 µm Filtropur S filter (Sarstedt). Phage particles were pelleted via centrifugation at 8°C and 20,000 x *g* for 1 h. The pellet was suspended in 1 ml SM buffer (50 mM Tris-HCl, 100 mM NaCl, 8 mM $MgSO_4 \cdot 7\ H_2O$, pH 7.4) supplemented with 0.5 mM $CaCl_2$ (supporting phage stability and upcoming DNase treatment) and let soak overnight at 4°C. Particle suspension was further supported the next day by shaking at 150 rpm (LT-V Lab-Shaker, Adolf Kühner AG, Birsfelden, Germany) at room temperature for 2 hrs. Suspended samples were finally transferred to 2 ml DNA LowBind micro tubes (Sarstedt) for following treatments using cut filter tips to reduce possible shearing. Samples were stored at 4°C.

Isolation of particle-protected DNA from phage samples

Phage DNA was isolated using the MasterPure Gram Positive DNA Purification Kit (Epicentre, Madison, WI, USA) with modifications. Prior to the phage DNA isolation, phage samples were supplemented with 2 µl of 100 mg/ml lysozyme solution (lysozyme from chicken egg white 177,000 U/mg; Serva, Heidelberg, Germany) suspended in SM buffer to remove remaining host cell debris, and with 50 µg/ml final concentration RNase A (Biozym Scientific GmbH, Hess. Oldendorf, Germany) and 10 U Baseline-ZERO DNase (Biozym Scientific GmbH) to digest host nucleic acids. Samples were incubated at 37°C for 6 hrs with gentle inversion every 30 minutes. Fragments of host nucleic acids resulting from the digestion were removed by phage particle pelleting via centrifugation at 4°C and 20,000 x *g* for 1 hour. The recovered pellet was suspended in 150 µl SM buffer. Suspension was supported by shaking at 150 rpm (LT-V Lab-Shaker) at room temperature for 10 minutes and slight flicking. EDTA (0.5 M, pH 8.0) was added to 10 mM final concentration for complete DNase inhibition.

Phage particles were lysed by adding 1% SDS (10% solution) and 2 µl Proteinase K (50 µg/µl; Biozym Scientific GmbH), incubation at 56°C for 1.5 hrs and gentle inversion every 30 minutes. Subsequently, samples were completely cooled down on ice before addition of 130 µl MPC Protein Precipitation Reagent (pre-cooled to -20°C). After mixing by gentle inversion, proteins were pelleted via centrifugation at 4°C and 10,000 x *g* for 10 minutes. The DNA-containing supernatant was transferred to 1.5 ml DNA LowBind micro tubes (Sarstedt). DNA was precipitated by addition of 0.3 M Na-acetate (3 M, pH 5.2), 10 mM $MgCl_2$ (2 M), and 0.8 volume isopropanol at room temperature. Samples were inverted 40 times and incubated at room temperature for 10 minutes before centrifugation at 4°C and 15,000 x *g* for 30 minutes. The supernatant was removed carefully and the DNA pellet washed twice with 150 µl 70% ethanol (pre-cooled to -20°C) and centrifugation at 4°C and 15,000 x *g* for 5 to 10 minutes. The supernatant was removed and the sample briefly centrifuged again to collect all residual ethanol for final removal. DNA pellets were air-dried under a sterile bench and immediately suspended in 20 µl TE buffer. Complete DNA elution was supported by brief storage at 4°C and slight flicking, before final storage at -20°C. DNA concentration was assessed with the Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) using the HS dsDNA assay kit.

Phage DNA sequencing and sequencing read processing

Phage DNA was subjected to Illumina sequencing for dsDNA by paired-end library preparation with the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) as recommended by the manufacturer. Libraries were sequenced using an Illumina MiSeq system and MiSeq Reagent Kit version 3 (2 × 300 bp, 600 cycles) as recommended by the manufacturer.

All following software was used in default mode. Sequencing raw reads were quality processed with fastp (v0.23.4) (27) before removing the sequencing adapters with Trimmomatic (v0.39) (28). Processed reads were mapped onto the corresponding host genome using bowtie2 (v2.5.0), and the resulting SAM file was converted to a TDS file for bioinformatics analysis with the TraV software (29).

Data analysis of phage sequencing reads

TDS files of processed reads for the same isolate were together inspected in TraV (29) for read coverage, and reads were normalized by calculation of nucleotide activity per kilobase of exon model per million mapped reads (NPKM). This results in a value for each CDS corresponding to its read coverage in reference to the overall read amount. NPKM values were further normalized to account for the differing growth behavior under the induction conditions by transforming values to an OD_{600} of 2.0. In this way, NPKM values reflected phage abundance under the different conditions at same cell density, which allows a better qualitative estimation of phage activity. OD normalization and visualization of NPKMs values were done in Rstudio (v2022.06.0) (24) using the packages tidyverse (v2.0.0) (30), ggforce (v0.4.1) (31), and ggplot2 (v3.4.2) (23). NPKM values were plotted against the host genome with regard to sequence start of the corresponding CDS, and original and normalized NPKM values were plotted together to visualize the effect of OD normalization. Phage regions predicted with PHASTEST (21) were also implemented.

Phage genome annotation and gene content analysis

Active regions identified via sequence read mapping were extracted for anew genome annotation. Sequence ranges were thereby adopted from PHASTEST (21) predictions or selected based on read mapping in TraV (29), if the predicted prophage region did not cover the entire mapped region. Annotation was customized for phage genomes using Pharokka (v1.3.2) (32) in default mode with sequence re-orientation to the large terminase subunit. If the large terminase subunit was not annotated automatically, it was determined via BLAST analysis

(33) and manually annotated. For specific genes and their encoded protein, additional analyses with protein BLAST (33) and InterProScan (v5.63-95.0) (34) were performed.

Phage genome-based classifications

Genome-based classifications were done with different bioinformatic analysis tools. An average nucleotide identity analysis (ANI) with pyani (v0.2.12) (35) and MUMmer3 alignment (36) (ANIm) was used in default mode to compare the phages amongst each other. Assessment of the DNA-packaging strategy was performed based on the study of Rashid et al. (37). The large terminase subunit was aligned at protein sequence level with ClustalW and a Maximum-Likelihood phylogenetic tree was constructed with the Whelan and Goldman (WAG) substitution model and otherwise default parameter with the software MEGA (v11.0.13) (38). Branches were collapsed in MEGA (38) if none of our phages clustered within and final modifications for visualization were done in Inkscape (v0.48; https://inkscape.org/de/). A nucleotide BLAST analysis (33) with default parameters was performed to check for similarity to already known phage genomes and for the prevalence in genomes of other *C. difficile* strains. BLAST results were ordered based on query coverage and hits with a query coverage below 10% were neglected unless relevant matches with higher coverages were not obtained. For assigning the phages to a morphological family of the order *Caudovirales*, phage genomes were inspected for the presence of baseplate proteins and sequence length of the tail length tape measure protein (39). If the tail length tape measure protein was not annotated, it was identified via protein BLAST analysis (33) and manually curated in the genome.

Data availability

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

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

Results and discussion

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

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

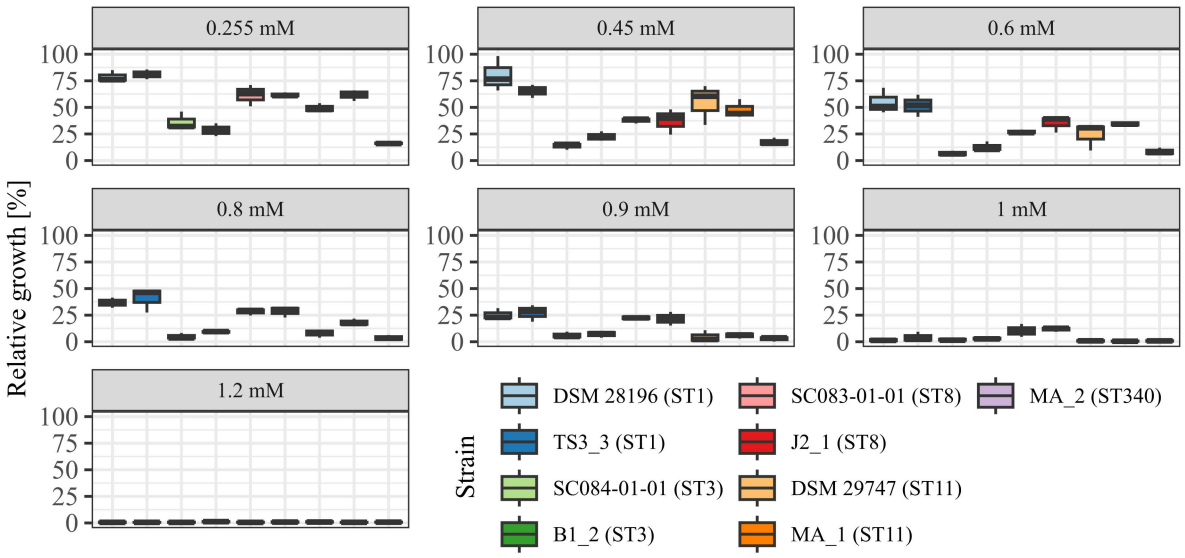


Fig 1. DCA tolerance at various concentrations. Tolerance of the strains to DCA was determined as relative growth compared to the reference culture (regarded as 100%) and MIC assay at concentrations from 0.255 mM to 1.2 mM. Strains of the same ST are depicted next to each other. No significant tolerance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

Sequencing-based assessments of prophage activity

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

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

Table 3. Overview of active regions in all strains. All active regions identified based on sequencing read coverage (Fig 2 - 6) were renamed according to a numbered scheme (strain_phiX). Regions were validated for certain (✓) or uncertain (~) activity based on coverage signal strength. Additionally, prophage prediction with PHASTEST (21) is included 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

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

All regions (Table 3) were further closely inspected regarding their activities under DCA-induction. Overall NPKM-transformation to an OD of 2 revealed distinctly higher signals under DCA-induction in most active regions and all strains. This verified the phage-inducing effect of DCA, which varied apparently between the different regions even within the same host and thereby implied phage-dependency.

The genomes of both ST1-strains carried an identical phage (Fig S1) (TS3_3_phi/ DSM28196_phi1) as determined by ANIm analysis (Fig S1) and similar genome position. The analogous phages showed distinct spontaneous activity with approximate magnitude, but their DCA-induced activity differed substantially. Phage DSM28196_phi1 (Fig 2B) exhibited ~3.5x higher signal increase under DCA than TS3_3_phi in the non-clinical strain (Fig 2A). Since the phages were identical, the differing reactions seemed to be host related. The genome of strain DSM 28196 possessed another active region DSM28196_phi2, which showed minor activity under both conditions, and was phage-atypical by comprising only four genes and missing a prophage prediction (Fig 2 B).

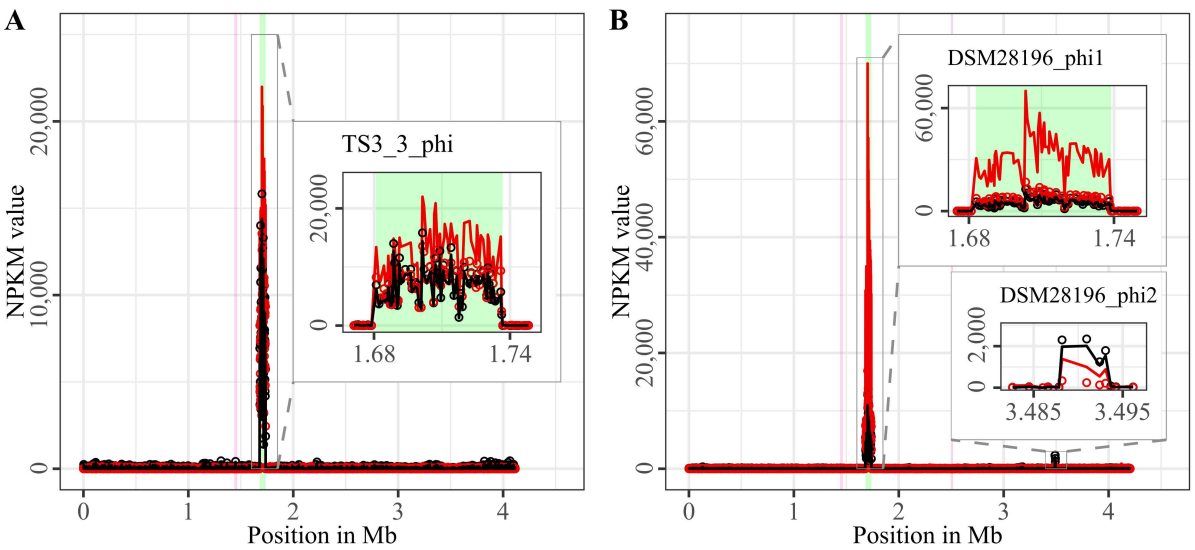


Fig 2. Coverage of phage sequencing reads of ST1-strains TS3_3 and DSM 28196. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strains (A) TS3_3 and (B) DSM 28196 were normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally OD-normalized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

The genomes of ST3-strains possessed several active regions (Fig 3), which were not similar to each other (Fig S1). In the genome of the non-clinical strain B1_2 (Fig 3A), two of the four active regions comprised the two ECEs. B1_2_phi4 on ECE2 exhibited the highest NPKM values under DCA-induction. Remarkably, B1_2_phi4 is another phage-atypical but active

region without corresponding prophage prediction, as previously detected for DSM28196_phi2 (Fig 2B). B1_2_phi3 on ECE1 also showed increased activity under DCA, but both spontaneous and induced activity were not particularly high. B1_2_phi2 on the chromosome showed spontaneous activity and a substantial increase in DCA-induced activity. B1_2_phi1 on the chromosome exhibited almost no signal under spontaneous conditions, which slightly increased upon DCA-induction. This might indicate true DCA-induction of this phage without prior spontaneous activity. In contrast to strain B1_2, the corresponding clinical strain SC084-01-01 possessed no prominently active region on the chromosome (Fig 3B). Activity could be observed within the region SC084-01-01_phi1, but NPKM values were very low under both conditions and signals did not cover the whole phage region. Abundancy of this phage was probably insufficient to capture distinct activity by the sequencing approach. Similar activity was observed for SC084-01-01_phi2 on ECE1 of SC084-01-01. SC084-01-01_phi3 on ECE2 was the only region in SC084-01-01 with prominent activity under spontaneous conditions, and activity further increased apparently under DCA-induction.

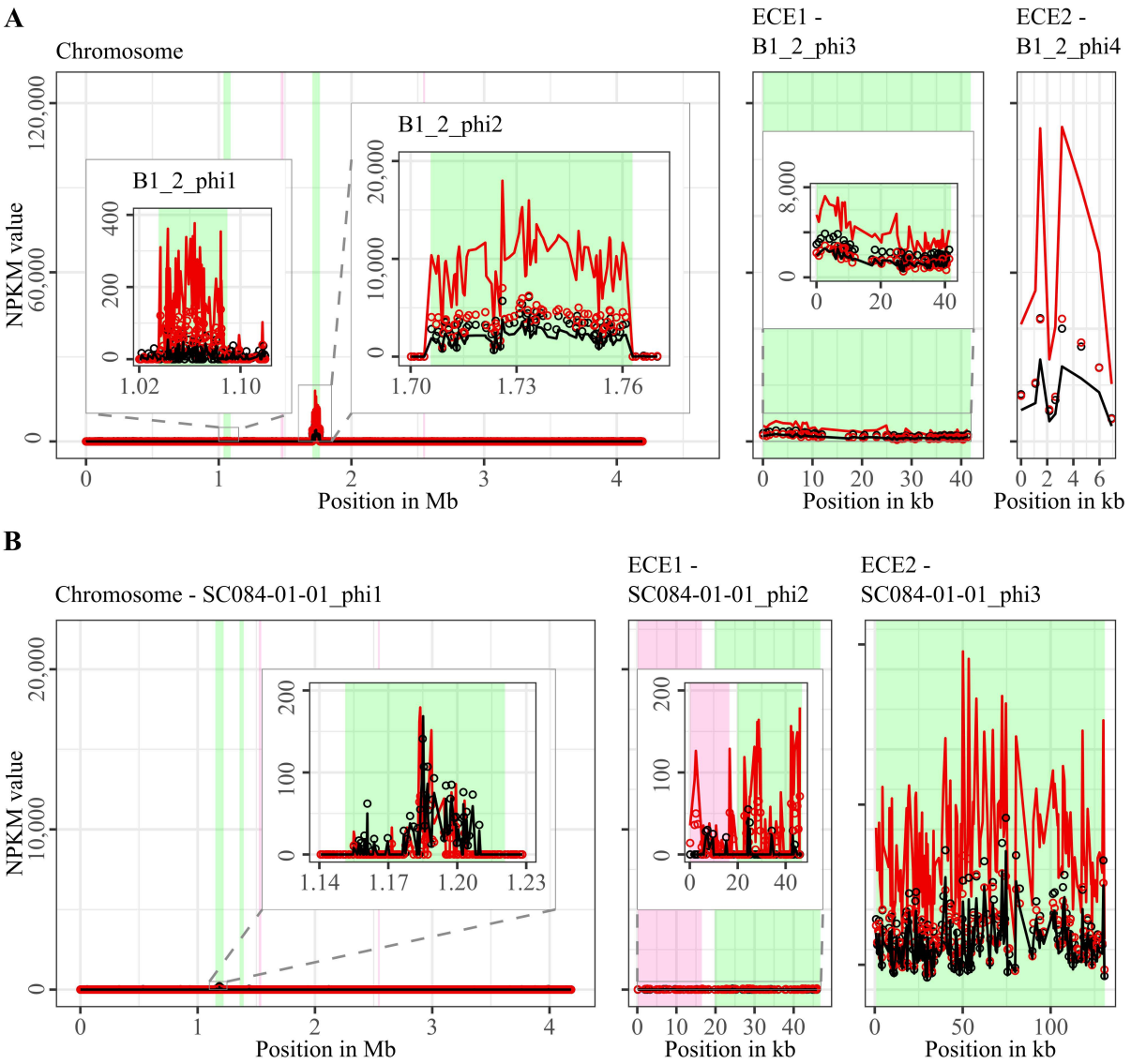


Fig 3. Coverage of phage sequencing reads of ST3-strains B1_2 and SC084-01-01. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strains (A) B1_2 and (B) SC084-01-01 were normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally 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).

The genomes of both ST8-strains exhibited activity for their analogous chromosomal and extrachromosomal regions (Fig 4), which were similar phages according to ANIm analysis (Fig S1) and similar genomic location. J2_1_phi2 on the ECE of strain J2_1 showed distinct spontaneous activity and a DCA-induced activity increase (Fig 4A), whereas SC083-01-01_phi1 on the ECE of SC083-01-01 was only slightly active under spontaneous conditions, and signal increase upon DCA-induction was only little. The activity of the chromosomal phages differed apparently as well. J2_1_phi1 on the J2_1 chromosome was spontaneously active and showed increased activity under DCA-induction (Fig 4A). The left part of the prophage region started with minor activity, which drastically increased at the terminase genes.

Strikingly, the left part of the prophage region started with minor activity, which drastically increased at the terminase genes. Further remarkable, sequencing reads mapped beyond the predicted prophage region and spread upstream (~30 kb) and downstream (~135 kb). The downstream region adjoined the phage activity with similar NPKM values that gradually decreased over the entire section. The chromosomal phage SC083-01-01_phi1 of the corresponding clinical ST8-strain SC083-01-01 did not exhibit these peculiarities (Fig 4B). It was spontaneously active, and activity increased substantially under DCA treatment. This increase was strikingly twice as high as observed for the analogous phage J2_1_phi1 (Fig 4A), despite their similarity (Fig S1). Such dissimilar activity increase among analogous phages was already observed in the ST1 strains (Fig 2). In both ST8- and ST1-pairs, a distinctly stronger increase upon DCA-induction was connected to the clinical background of the strains. Since DCA-stress levels did not significantly differ between corresponding clinical and non-clinical strains (Fig 1) and thereby suggested similar induction levels, the question for the underlying mechanism of these diverging activities in analogous phages arose. This prompted to an undefined regulation of phage induction involved in the clinical strains.

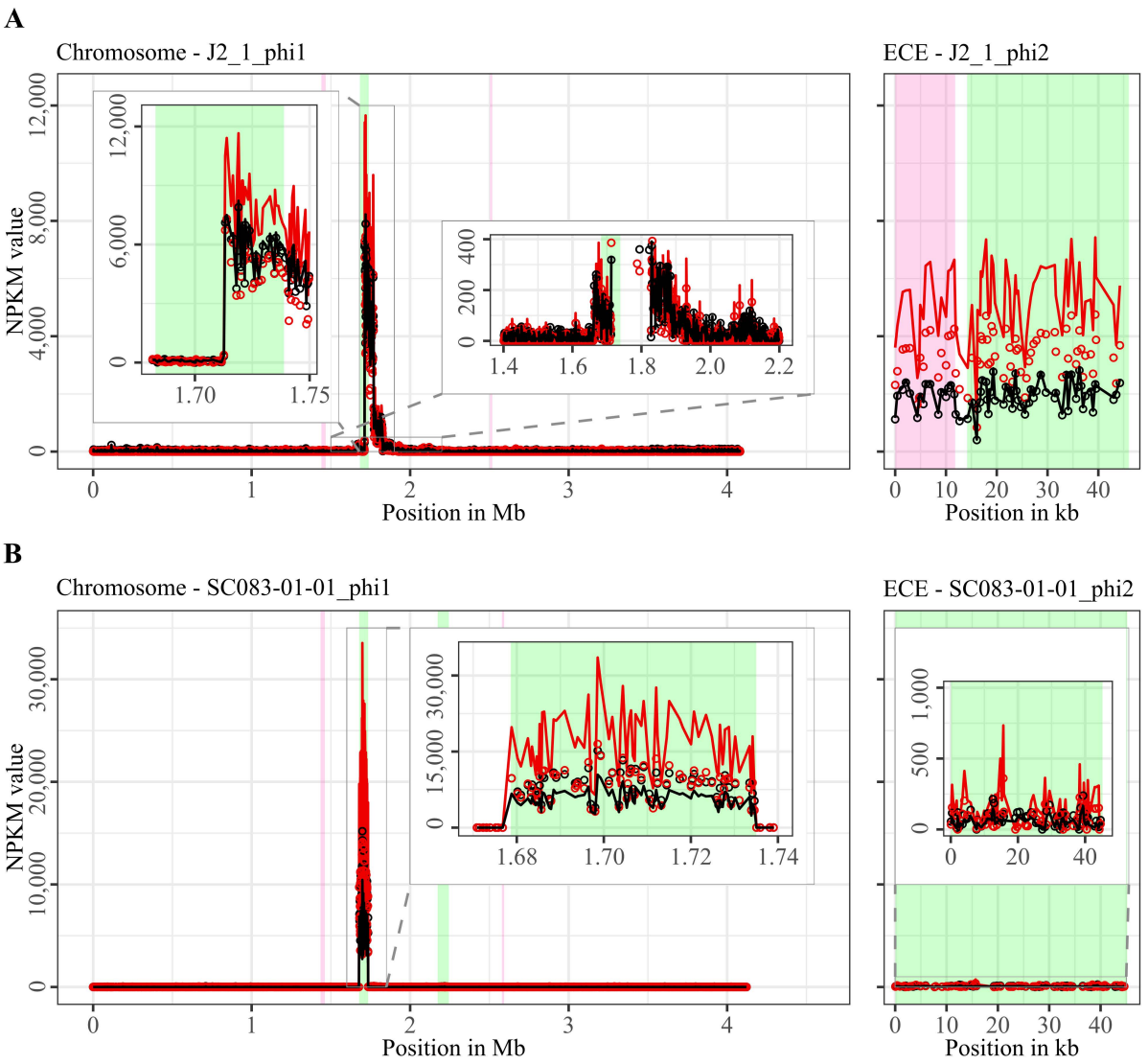


Fig 4. Coverage of phage sequencing reads of ST8-strains J2_1 and SC083-01-01. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strains (A) J2_1 and (B) SC083-01-01 were normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally 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).

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

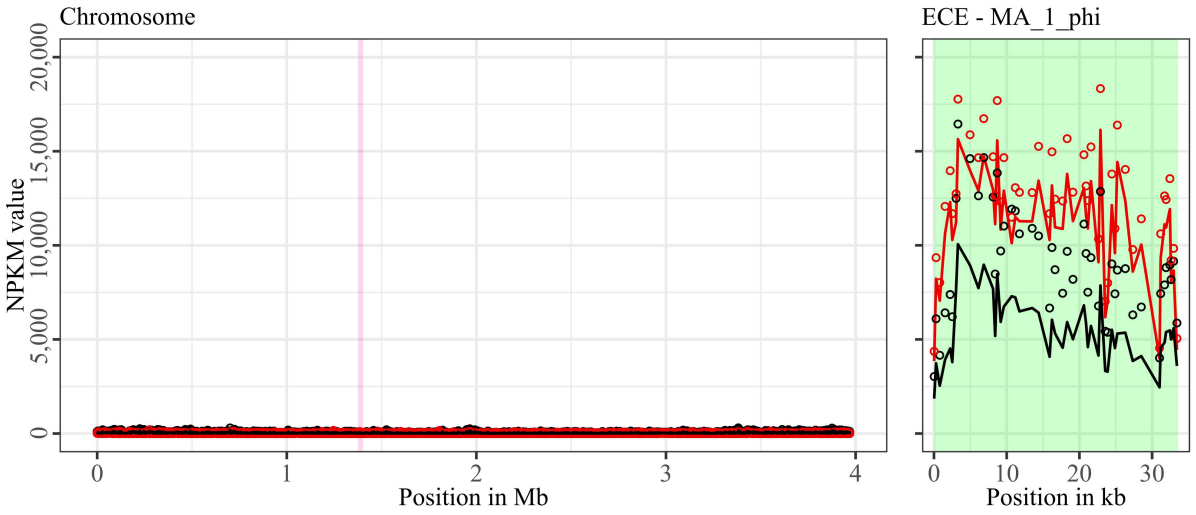


Fig 5. Coverage of phage sequencing reads of ST11-strain MA_1. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strain MA_1 were normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line graphs), 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).

Strain MA_2 is to our knowledge the first cryptic *C. difficile* strain with detailed phage examination. This strain possessed four active regions (Fig 6). MA_2_phi2 on the chromosome and MA_2_phi4 on the ECE both showed prominent activity under spontaneous conditions and a distinct activity increase upon DCA-induction. Interestingly, MA_2_phi4 was another active region without phage prediction, as observed previously for DSM28196_phi2 (Fig 2B) and B1_2_phi4 (Fig 3A). This was also true for MA_2_phi3 on the chromosomal, where activity was however very low under spontaneous conditions. Activity of this region increased upon DCA-induction. Chromosomal region MA_2_phi1 exhibited least activity in this genome even under DCA treatment. Read mapping for MA_2_phi1 and MA_2_phi2 did not cover the entire predicted prophage regions, but sections without read coverage contained only bacterial genes and were therefore evidently mis-predicted.

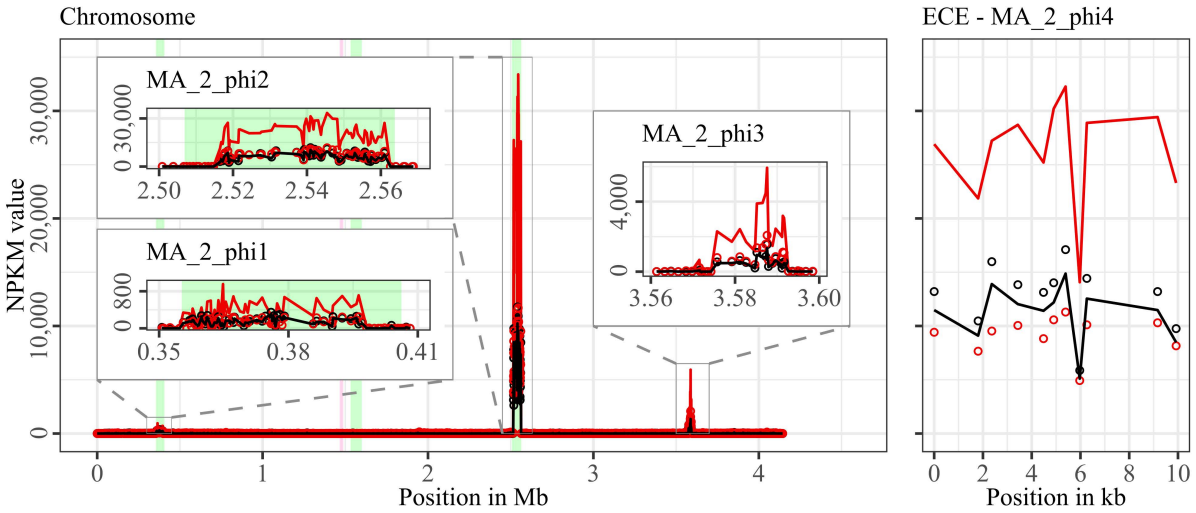


Fig 6. Coverage of phage sequencing reads of cryptic ST340-strain MA_2. Sequencing reads of phage particles from spontaneous (black) and DCA-induced (red) release of strain MA_2 were normalized to NPKM values (circles) with TraV (29), and NPKM values were additionally OD-normalized to OD_{600} = 2 (line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

Almost all identified active prophage regions could be induced by the secondary bile salt DCA. Mentionable, assessed phage activity after DCA treatment might be influenced by a direct effect of DCA on the phage. A study on different bacteriophages in *Escherichia coli* investigated the effect of bile salts on the host-phage interaction and observed varying survival rates of the phages (42).

Moreover, almost all ECEs were detected as active phages. Although extrachromosomal prophages have already been described in *C. difficile* strains (43,44), only a few of these were isolated and characterized so far (41).

Phage genome annotation and gene content analysis

Phage genomes harbor virulence-relevant genes. All active regions identified via sequence read mapping (Figs 2 – 6, Table 3) were inspected after anew annotation with Pharokka (32) for genes that might increase virulence of the host (genomic information in supplementary data file S1). All of them exhibited characteristic phage genes in a modular organization according to the different encoded functions, as typically seen in *C. difficile* phages (16,43,45). In some genomes, proteins characteristic for plasmids were found, such as genes encoding partition proteins and replication initiation factors (46). Genes encoding plasmid-related proteins besides phage-typical ones are common in a certain type of MGE, so-called phage-plasmids (47). These phage-plasmid features were especially recorded for extrachromosomal prophages but were

also present on the chromosomally integrated prophage MA_2_phi1. Further frequently observed genes in the phage genomes encoded proteins with potential involvement in cellular metabolism and growth, such as metallo-proteases, kinases, a phosphatase, and most of all putative rhodanese-related sulfurtransferases. These genes might be advantageous for the bacterial fitness, thereby indirectly contributing to host virulence. Two phage genomes (B1_2_phi1 and SC084-01-01_phi1) carried genes encoding hemolysin XhlA, an established virulence factor in other bacterial species (48), capable of lysing mammalian erythrocytes (49). In the opportunistic pathogen *Mannheimia haemolytica*, temperate phages were induced that encoded hemolysin XhlA and discussed to contribute to bacterial pathogenicity and transfer of this virulence factor (50). Hemolysis in *C. difficile* is not commonly known, but few studies demonstrated its hemolytic capability (51). However, XhlA is also present in other prophage genomes as part of the lysis module (52). Indeed, the gene encoding XhlA was found next to the lysis-relevant genes encoding holin and endolysin in phages B1_2_phi1 and SC084-01-01_phi1. Thus, the actual role of hemolysin XhlA in these phages and its potential impact on host virulence remains unclear. Phage SC084-01-01_phi1 possessed a gene encoding an ABC-transporter, which might contribute to antibiotic resistance of the host (53). Phage SC084-01-01_phi3 harbored a putative spore protease, which could influence the bacterial sporulation or germination ability and, as a consequence, alter bacterial fitness (43). Other genes conferring antibiotic resistances or encoding known virulence factors were not identified in the active phage genomes. Six phages (TS3_3_phi, DSM28196_phi1, B1_2_phi2, J2_1_phi1, SC083-01-01_phi, MA_2_phi2) carried arrays of clustered regularly interspaced short palindromic repeats (CRISPR), which is similar to other *C. difficile* phages with described CRISPRs (37,43,54). Temperate phages carrying CRISPR arrays increase host immunity against other invading phages (55). The corresponding host genomes were verified to encode Cas proteins required for CRISPR-Cas-mediated phage immunity (56). CRISPRs in prophages represent horizontally transferrable immunity against phages, which is specifically relevant in the context of phage therapy, an alternative treatment approach for bacterial infections with growing importance in view of increasing multidrug-resistances (57).

Active non-phage elements likely belong to so far undescribed HGT mechanisms in *C. difficile*. The regions without corresponding prophage prediction (Table 3) did not possess a phage-typical genome accordingly. Therefore, further gene analysis of these non-phage elements were performed based on the original genome annotation with Prokka (58) (genomic information in supplementary data file S1). No proteins involved in capsid or tail production, DNA packaging or host lysis were present. The lack of structural proteins was striking since

these DNA elements were enveloped according to the DNA isolation procedure. This indicated the involvement of unrelated particles. Even additional analysis of hypothetical proteins with InterProScan (34) and BLASTp (33) could not identify further functions. All other proteins were assigned to functions with DNA-binding activity, like helicases, integrases, relaxases, and transcriptional regulators. These genes are typical for phage genomes but also for MGEs like transposons as Integrative and Conjugative or Mobilizable Elements (ICE / IME) (59). Screening for these MGEs with ICEscreen (60) validated DSM28196_phi2 as complete IME, while MA_2_phi3 was detected as invalid ICE. Interestingly, the peculiar upstream-region of J2_1_phi1 (Fig 4A) was also identified as an ICE, although being incomplete. These integrative MGEs do not encode proteins for production of particles that carry the respective mobile sequence (59). Interestingly, transposons were found to hitchhike co-residing phages in several bacteria such as *Staphylococcus aureus* (61), *Vibrio cholerae* (62), and *Enterococcus faecalis* (63), enabling the phage-mediated transduction of virulence-relevant genes. This type of transduction was demonstrated once in *C. difficile* with the transfer of a conjugative transposon carrying an antibiotic resistance gene (10). Transduction can be either generalized, specialized or lateral (64). They all imply the "headful" DNA-packaging, in which the terminase starts DNA packaging at a bacterial homologue to the phage packaging site until the capsid is full, which consequently implies that transduced DNA is at least of similar phage genome size (64). The transduction mechanisms differ in transduced DNA and frequency (65). Specialized and lateral transduction involve host DNA adjacent to the prophage, while random host DNA is packaged in generalized transduction (65). Generalized and specialized transduction are processes of erroneous DNA-packaging, which results in low transduction frequency detectable by sequencing read coverage (64,65). In contrast, lateral transduction results in high sequencing read coverage comparable to actual phage activity, as this mechanism is assumed as natural phage trait instead of mistaken processes (64,65). This trait comprises phage genome replication and simultaneous DNA-packaging before excision from the chromosome, whereby a substantial amount of adjacent host DNA is packaged as well (66,67). All these characteristics of lateral transduction accorded with the observed mapped region downstream of phage J2_1_phi1 (Fig 4A), which indicated involvement of this DNA segment in phage-mediated lateral transduction.

The downstream region of J2_1_phi1 (Fig 4A) did not comprise characteristic genes for MGEs. Instead, several genes encoded proteins with potential relevance for strain virulence, such as genes for ABC-transporters, a multidrug efflux system ATP-binding protein, stress-related proteins, proteins involved in resistance to vancomycin and daunorubicin, and the putative

virulence factor BrkB. Therefore, mobilization and transfer of this region is critical regarding the spread of antibiotic resistances or virulence-related genes. The drastic NPKM difference within the genome phage J2_1_phi1 (Fig 4A) might also result from the process of lateral transduction, as inaccurate excision of the phage genome after *in situ* replication leads to phage genome truncation.

Since no evidence for lateral transduction was observed for phage J2_1_phi1's analogue SC083-01-01_phi1 (Fig 4B) despite their high similarity (Fig S1), the question about underlying differences arose. Direct phage genome comparison revealed diverging excisionases and integrases as well as four additional amino acids in the large terminase protein of J2_1_phi1. All these proteins perform activities destining for lateral transduction (64).

The extrachromosomal non-phage elements were significantly smaller than the co-existing phages (Table 3), which is in contrast with the headful packaging mechanism required in transduction. This indicated a form of DNA-protecting particle other than phages, such as gene transfer agents (GTA). These phage-like particles carry DNA between 4 to 14 kb (68), which is similar to the sizes of the detected non-phage elements (Table 3). However, GTAs package bacterial DNA randomly (68), which does not fit to the detected distinct activity of specific regions, making GTAs unlikely as mode of action. Extracellular vesicles are known in various bacteria and described to carry and transfer genetic content, e.g. plasmids, between cells (69–71). This type of alternative HGT is not well characterized so far, but was demonstrated to allow interspecies gene transfer (71), which underlines the importance of vesicle-mediated DNA exchange. Noteworthy, vesicle-driven HGT in *C. difficile* has not been described so far.

Classification of the active phages

Terminase-based determination of the phage DNA-packaging strategy. Assessment of the phage-DNA packaging mechanisms was performed to validate the above hypothesized transduction events. The large terminase subunit was analyzed via protein sequence alignment and phylogenetic tree construction referring to Rashid et al. (37). This assigned the phages to different phage DNA-packaging mechanisms (Fig 7). All our phages were assigned to clusters comprising other *C. difficile* phages, which predominantly represented the P22-like headful packaging mechanism, followed by the 3'-extended COS ends and an unknown strategy. Consequently, most of the phages were predicted to utilize the headful packaging mechanism and would, thus, be indeed capable of transducing host DNA. The mechanism "P22-like headful" originates from the packaging strategy employed by phage P22 of *Salmonella enterica*. Phage P22 was originally described to perform generalized transduction (72), but

recent evidence demonstrated also specialized as well as lateral transduction activity (67). These terminase analysis results supported the assumption of lateral transduction of the phage J2_1_phi1 downstream region, which is by that the first described case in *C. difficile*.

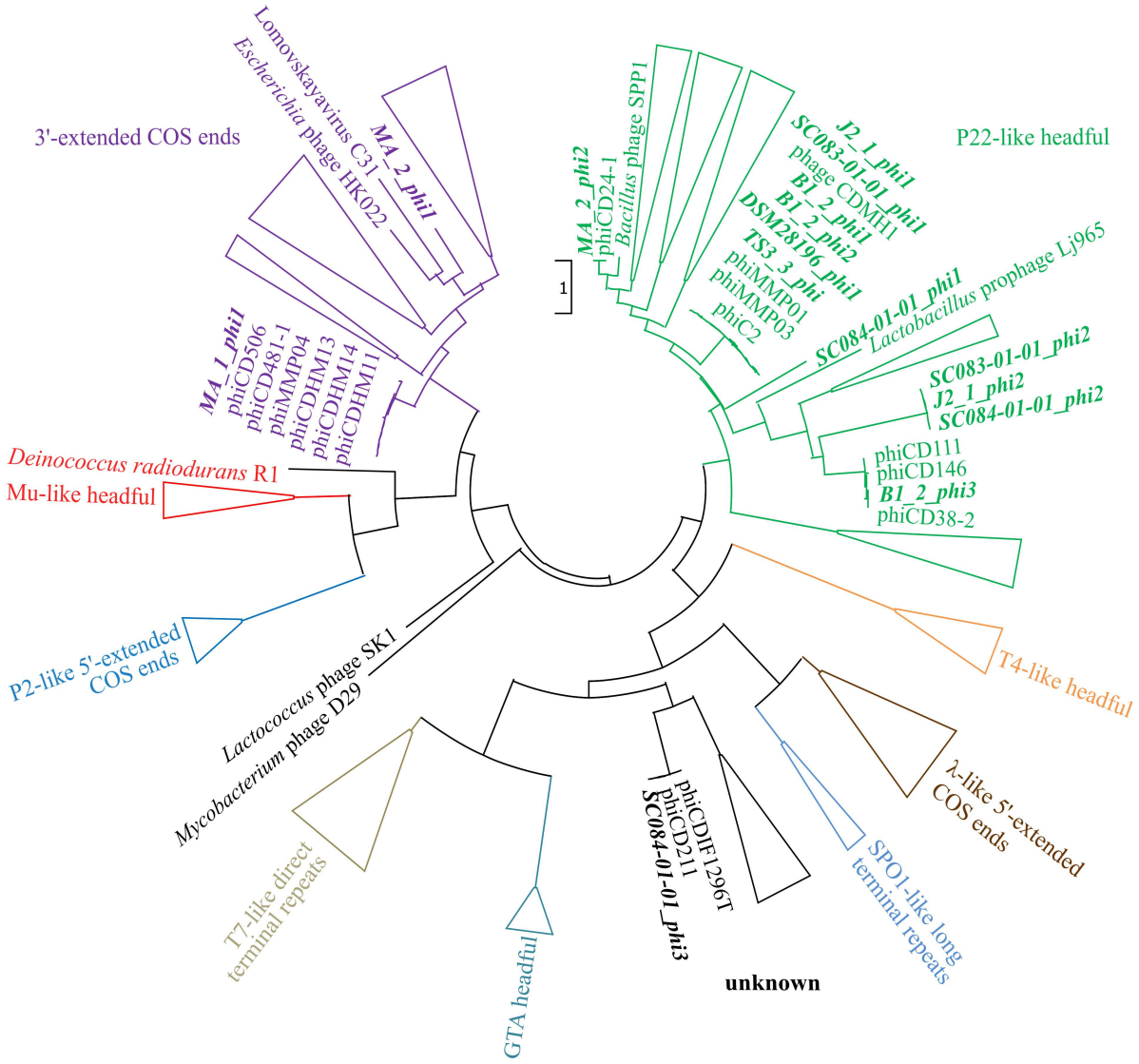


Fig 7. Maximum-likelihood phylogenetic tree of large terminase protein sequences. The large terminase of the active phages (highlighted in bold and italic) were aligned on protein level to the reference sequences from Rashid et al. (37). Branches of the different DNA-packaging strategies were colored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

Nucleotide BLAST analyses assess phage prevalence and novelty. A nucleotide BLAST analysis (33) was performed on all active regions in Table 3 to check for similar phages and elements, and to assess prevalence among *C. difficile* strains. The results are available in supplementary data file S2 and summarized in Table 4. Most of the phages matched against *C. difficile* phages with query coverages between 4% and 90% and percent identities between 86.26% and 99.86%. This confirmed that our phages are indeed similar to known phages but

still represent novel types, which underlines the contribution of this work to the general knowledge on *C. difficile* phages. Further, chromosomal phages also matched against a multitude of *C. difficile* chromosomes, while the extrachromosomal phages often corresponded to *C. difficile* plasmids and few chromosomes. This demonstrated the prevalence of the identified phages in other *C. difficile* strains. The non-phage elements B1_2_phi4 and MA_2_phi4 yielded no significant BLAST hit against a phage but matched against *C. difficile* plasmids. Both matching plasmids belong to classes of *C. difficile* plasmids with similar organization that are present in diverse *C. difficile* strains (73,74). All these plasmids might therefore likewise be inducible and particle-protected, which implies a different mechanism of HGT than currently assumed.

Table 4. Nucleotide BLAST results of the identified active regions. Results were summarized regarding the first phage-related *C. difficile* BLAST hit by stating its description, query coverage, percent identity, and number of pre- and succeeding chromosome/plasmid entries. The few hits of metagenome-assembled phages were not listed. If available, information 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	–

Table 4 continued.

Phage	Entry	Query cover %	Percent identity %	*C. difficile* matches preceding / succeeding	Family
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)
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	–

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

Genome-based phage assignment to *Myoviridae* and *Siphoviridae*. We lastly classified our phages morphologically. All known *C. difficile* phages so far belong to the *Caudovirales* family of *Myoviridae* or *Siphoviridae*, which distinguish by tail appearance (41). Genome inspection for the presence of baseplate proteins characteristic for *Myoviridae* and the length of the tail

length tape measure protein as indicator for *Siphoviridae* could assign eleven phages to *Myoviridae* and four phages to *Siphoviridae* (Table S2).

Conclusion

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

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. 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. FEMS Microbiol Rev [Internet]. 2017 Nov 1;41(6):723–50. Available from: https://academic.oup.com/femsre/article/41/6/723/4557158
6. Czepiel J, Dróżdż M, Pituch H, Kuijper EJ, Perucki W, Mielimonka A, et al. *Clostridium difficile* infection: review. Eur J Clin Microbiol Infect Dis [Internet]. 2019 Jul 3;38(7):1211–21. Available from: https://journals.lww.com/00013542-201807000-00001
7. Lewis BB, Carter RA, Ling L, Leiner I, Taur Y, Kamboj M, et al. Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to *Clostridium difficile* Virulence. Ballard JD, editor. MBio [Internet]. 2017 Sep 6;8(4):e00885-17. Available from: https://journals.asm.org/doi/10.1128/mBio.00885-17
8. Alonso CD, Kelly CP, Garey KW, Gonzales-Luna AJ, Williams D, Daugherty K, et al. Ultrasensitive and Quantitative Toxin Measurement Correlates With Baseline Severity, Severe Outcomes, and Recurrence Among Hospitalized Patients With *Clostridioides difficile* Infection. Clin Infect Dis [Internet]. 2022 Jul 6;74(12):2142–9. Available from: https://academic.oup.com/cid/article/74/12/2142/6372425

9. Smits WK, Roseboom AM, Corver J. Plasmids of *Clostridioides difficile*. Curr Opin Microbiol [Internet]. 2022 Feb;65:87–94. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1369527421001521
10. Goh S, Hussain H, Chang BJ, Emmett W, Riley T V., Mullany P. Phage ϕC2 Mediates Transduction of Tn *6215*, Encoding Erythromycin Resistance, between *Clostridium difficile* Strains. Onderdonk AB, editor. MBio [Internet]. 2013 Dec 31;4(6):1–7. Available from: https://journals.asm.org/doi/10.1128/mBio.00840-13
11. Sekulovic O, Meessen-Pinard M, Fortier L-C. Prophage-Stimulated Toxin Production in *Clostridium difficile* NAP1/027 Lysogens. J Bacteriol [Internet]. 2011 Jun;193(11):2726–34. Available from: https://journals.asm.org/doi/10.1128/JB.00787-10
12. Govind R, Vediyappan G, Rolfe RD, Dupuy B, Fralick JA. Bacteriophage-Mediated Toxin Gene Regulation in *Clostridium difficile*. J Virol [Internet]. 2009 Dec;83(23):12037–45. Available from: https://journals.asm.org/doi/10.1128/JVI.01256-09
13. de la Cruz F, Davies J. Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol [Internet]. 2000 Mar;8(3):128–33. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0966842X00017030
14. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet [Internet]. 2006 Jul 25;38(7):779–86. Available from: https://www.nature.com/articles/ng1830
15. Fortier L-C. Bacteriophages Contribute to Shaping *Clostridioides (Clostridium) difficile* Species. Front Microbiol [Internet]. 2018 Aug 31;9:2033. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2018.02033/full
16. Meessen-Pinard M, Sekulovic O, Fortier L-C. Evidence of *In Vivo* Prophage Induction during *Clostridium difficile* Infection. Appl Environ Microbiol [Internet]. 2012 Nov;78(21):7662–70. Available from: https://journals.asm.org/doi/10.1128/AEM.02275-12
17. Nale JY, Shan J, Hickenbotham PT, Fawley WN, Wilcox MH, Clokie MRJ. Diverse Temperate Bacteriophage Carriage in *Clostridium difficile* 027 Strains. Popoff MR, editor. PLoS One [Internet]. 2012 May 18;7(5):e37263. Available from: https://dx.plos.org/10.1371/journal.pone.0037263
18. Dubois T, Tremblay YDN, Hamiot A, Martin-Verstraete I, Deschamps J, Monot M, et

al. A microbiota-generated bile salt induces biofilm formation in *Clostridium difficile*. npj Biofilms Microbiomes [Internet]. 2019 May 9;5(1):14. Available from: https://www.nature.com/articles/s41522-019-0087-4

19. Kandell RL, Bernstein C. Bile salt/acid induction of DNA damage in bacterial and mammalian cells: Implications for colon cancer. Nutr Cancer [Internet]. 1991 Jan;16(3–4):227–38. Available from: http://www.tandfonline.com/doi/abs/10.1080/01635589109514161
20. Hu J, Ye H, Wang S, Wang J, Han D. Prophage Activation in the Intestine: Insights Into Functions and Possible Applications. Front Microbiol [Internet]. 2021 Dec 13;12:785634. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.785634/full
21. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res [Internet]. 2023 Jul 5;51(W1):W443–50. Available from: https://academic.oup.com/nar/article/51/W1/W443/7167344
22. Hamilton JP, Xie G, Raufman J-P, Hogan S, Griffin TL, Packard CA, et al. Human cecal bile acids: concentration and spectrum. Am J Physiol Liver Physiol [Internet]. 2007 Jul;293(1):G256–63. Available from: https://www.physiology.org/doi/10.1152/ajpgi.00027.2007
23. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. 1–260 p. Available from: https://ggplot2.tidyverse.org
24. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. 2020. Available from: http://www.rstudio.com/
25. Team RC. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. Available from: http://www.r-project.org/
26. Theriot CM, Bowman AA, Young VB. Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium difficile* Spore Germination and Outgrowth in the Large Intestine. Ellermeier CD, editor. mSphere [Internet]. 2016 Feb 25;1(1):e00045-15. Available from: https://journals.asm.org/doi/10.1128/mSphere.00045-15
27. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics [Internet]. 2018 Sep 1;34(17):i884–90. Available from: https://academic.oup.com/bioinformatics/article/34/17/i884/5093234

28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics [Internet]. 2014 Aug 1;30(15):2114–20. Available from: https://academic.oup.com/bioinformatics/article/30/15/2114/2390096
29. Dietrich S, Wiegand S, Liesegang H. TraV: A Genome Context Sensitive Transcriptome Browser. Scaria V, editor. PLoS One [Internet]. 2014 Apr 7;9(4):e93677. Available from: https://dx.plos.org/10.1371/journal.pone.0093677
30. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw [Internet]. 2019 Nov 21;4(43):1686. Available from: https://joss.theoj.org/papers/10.21105/joss.01686
31. Pedersen TL. ggforce: Accelerating "ggplot2" Title [Internet]. 2022. Available from: https://github.com/thomasp85/ggforce
32. Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald P-J, Vreugde S. Pharokka: a fast scalable bacteriophage annotation tool. Marschall T, editor. Bioinformatics [Internet]. 2023 Jan 1;39(1):btac776. Available from: https://academic.oup.com/bioinformatics/article/doi/10.1093/bioinformatics/btac776/6858464
33. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res [Internet]. 2008 May 19;36(Web Server issue):W5–9. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn201
34. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics [Internet]. 2014 May 1;30(9):1236–40. Available from: https://academic.oup.com/bioinformatics/article/30/9/1236/237988
35. Pritchard L. PYANI script [Internet]. 2014. Available from: https://github.com/widdowquinn/pyani
36. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol [Internet]. 2004;5(2):R12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14759262
37. Rashid S, Barylski J, Hargreaves K, Millard A, Vinner G, Clokie M. Two Novel Myoviruses from the North of Iraq Reveal Insights into *Clostridium difficile* Phage Diversity and Biology. Viruses [Internet]. 2016 Nov 16;8(11):310. Available from: http://www.mdpi.com/1999-4915/8/11/310
38. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis

Version 11. Battistuzzi FU, editor. Mol Biol Evol [Internet]. 2021 Jun 25;38(7):3022–7. Available from: https://academic.oup.com/mbe/article/38/7/3022/6248099

39. Zinke M, Schröder GF, Lange A. Major tail proteins of bacteriophages of the order Caudovirales. J Biol Chem [Internet]. 2022 Jan;298(1):101472. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0021925821012813
40. Hertel R, Rodríguez DP, Hollensteiner J, Dietrich S, Leimbach A, Hoppert M, et al. Genome-Based Identification of Active Prophage Regions by Next Generation Sequencing in *Bacillus licheniformis* DSM13. Schuch R, editor. PLoS One [Internet]. 2015 Mar 26;10(3):e0120759. Available from: https://dx.plos.org/10.1371/journal.pone.0120759
41. Heuler J, Fortier L-C, Sun X. *Clostridioides difficile* phage biology and application. FEMS Microbiol Rev [Internet]. 2021 Sep 8;45(5):fuab012. Available from: https://academic.oup.com/femsre/article/doi/10.1093/femsre/fuab012/6134755
42. Scanlan PD, Bischofberger AM, Hall AR. Modification of *Escherichia coli* – bacteriophage interactions by surfactants and antibiotics *in vitro*. Smalla K, editor. FEMS Microbiol Ecol [Internet]. 2017 Jan;93(1):fiw211. Available from: https://academic.oup.com/femsec/article-lookup/doi/10.1093/femsec/fiw211
43. Garneau JR, Sekulovic O, Dupuy B, Soutourina O, Monot M, Fortier L-C. High Prevalence and Genetic Diversity of Large phiCD211 (phiCDIF1296T)-Like Prophages in *Clostridioides difficile*. Dozois CM, editor. Appl Environ Microbiol [Internet]. 2018 Feb;84(3):e02164-17. Available from: https://journals.asm.org/doi/10.1128/AEM.02164-17
44. Ramírez-Vargas G, Goh S, Rodríguez C. The Novel Phages phiCD5763 and phiCD2955 Represent Two Groups of Big Plasmidial Siphoviridae Phages of *Clostridium difficile*. Front Microbiol [Internet]. 2018 Jan 22;9:26. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2018.00026/full
45. Govind R, Fralick JA, Rolfe RD. Genomic organization and molecular characterization of *Clostridium difficile* bacteriophage ΦCD119. J Bacteriol. 2006;188(7):2568–77.
46. Filutowicz M. Plasmids, Bacterial. In: Encyclopedia of Microbiology [Internet]. Elsevier; 2009. p. 644–65. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780123739445000146
47. Pfeifer E, Moura de Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. Nucleic Acids Res [Internet]. 2021 Mar 18;49(5):2655–73. Available from:

https://academic.oup.com/nar/article/49/5/2655/6137301

48. Thomas P, Abdel-Glil MY, Eichhorn I, Semmler T, Werckenthin C, Baumbach C, et al. Genome Sequence Analysis of *Clostridium chauvoei* Strains of European Origin and Evaluation of Typing Options for Outbreak Investigations. Front Microbiol [Internet]. 2021 Sep 29;12:.732106. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.732106/full
49. Cowles KN, Goodrich-Blair H. Expression and activity of a *Xenorhabdus nematophila* haemolysin required for full virulence towards *Manduca sexta* insects. Cell Microbiol [Internet]. 2004 Sep 8;7(2):209–19. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1462-5822.2004.00448.x
50. Niu YD, Cook SR, Wang J, Klima CL, Hsu Y, Kropinski AM, et al. Comparative analysis of multiple inducible phages from *Mannheimia haemolytica*. BMC Microbiol [Internet]. 2015 Dec 30;15(1):175. Available from: http://dx.doi.org/10.1186/s12866-015-0494-5
51. Alkudmani ZSB. The identification and characterization of novel haemolysin genes from *Clostridium difficile* [Internet]. University College London; 2018. Available from: https://discovery.ucl.ac.uk/id/eprint/10078163/7/Alkudmani_00_Thesis_edited.pdf
52. Krogh S, Jørgensen ST, Devine KM. Lysis Genes of the *Bacillus subtilis* Defective Prophage PBSX. J Bacteriol [Internet]. 1998 Apr 15;180(8):2110–7. Available from: https://journals.asm.org/doi/10.1128/JB.180.8.2110-2117.1998
53. Orelle C, Mathieu K, Jault J-M. Multidrug ABC transporters in bacteria. Res Microbiol [Internet]. 2019 Nov;170(8):381–91. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0923250819300580
54. Hargreaves KR, Flores CO, Lawley TD, Clokie MRJ. Abundant and Diverse Clustered Regularly Interspaced Short Palindromic Repeat Spacers in *Clostridium difficile* Strains and Prophages Target Multiple Phage Types within This Pathogen. Paul J, Azam F, editors. MBio [Internet]. 2014 Oct 31;5(5):e01045-13. Available from: https://journals.asm.org/doi/10.1128/mBio.01045-13
55. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science (80-). 2007;315(March):1709–12.
56. Koonin E V, Makarova KS. CRISPR-Cas: an adaptive immunity system in prokaryotes. F1000 Biol Rep [Internet]. 2009 Dec 9;1(95). Available from: https://facultyopinions.com/prime/reports/b/1/95/

57. Gutiérrez B, Domingo-Calap P. Phage Therapy in Gastrointestinal Diseases. Microorganisms [Internet]. 2020 Sep 16;8(9):1420. Available from: https://www.mdpi.com/2076-2607/8/9/1420
58. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics [Internet]. 2014 Jul 15;30(14):2068–9. Available from: https://academic.oup.com/bioinformatics/article/30/14/2068/2390517
59. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev [Internet]. 2014 Jul;38(4):720–60. Available from: https://academic.oup.com/femsre/article-lookup/doi/10.1111/1574-6976.12058
60. Lao J, Lacroix T, Guédon G, Coluzzi C, Payot S, Leblond-Bourget N, et al. ICEscreen: a tool to detect Firmicute ICEs and IMEs, isolated or enclosed in composite structures. NAR Genomics Bioinforma [Internet]. 2022 Oct 6;4(4):lqac079. Available from: https://academic.oup.com/nargab/article/doi/10.1093/nargab/lqac079/6765278
61. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. Mol Microbiol [Internet]. 1998 Jul;29(2):527–43. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.1998.00947.x
62. Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature [Internet]. 2013 Feb 27;494(7438):489–91. Available from: https://www.nature.com/articles/nature11927
63. Duerkop BA, Clements C V., Rollins D, Rodrigues JLM, Hooper L V. A composite bacteriophage alters colonization by an intestinal commensal bacterium. Proc Natl Acad Sci [Internet]. 2012 Oct 23;109(43):17621–6. Available from: https://pnas.org/doi/full/10.1073/pnas.1206136109
64. Borodovich T, Shkoporov AN, Ross RP, Hill C. Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. Gastroenterol Rep [Internet]. 2022 Jan 25;10:goac012. Available from: https://academic.oup.com/gastro/article/doi/10.1093/gastro/goac012/6567682
65. Kleiner M, Bushnell B, Sanderson KE, Hooper L V., Duerkop BA. Transductomics: sequencing-based detection and analysis of transduced DNA in pure cultures and microbial communities. Microbiome [Internet]. 2020 Dec 15;8(1):158. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-020-00935-5

66. Chiang YN, Penadés JR, Chen J. Genetic transduction by phages and chromosomal islands: The new and noncanonical. Kline KA, editor. PLOS Pathog [Internet]. 2019 Aug 8;15(8):e1007878. Available from: https://dx.plos.org/10.1371/journal.ppat.1007878
67. Fillol-Salom A, Bacigalupe R, Humphrey S, Chiang YN, Chen J, Penadés JR. Lateral transduction is inherent to the life cycle of the archetypical *Salmonella* phage P22. Nat Commun [Internet]. 2021 Nov 8;12(1):6510. Available from: https://www.nature.com/articles/s41467-021-26520-4
68. Lang AS, Zhaxybayeva O, Beatty JT. Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol [Internet]. 2012 Jul 11;10(7):472–82. Available from: https://www.nature.com/articles/nrmicro2802
69. Tran F, Boedicker JQ. Genetic cargo and bacterial species set the rate of vesicle-mediated horizontal gene transfer. Sci Rep [Internet]. 2017 Aug 18;7(1):8813. Available from: https://www.nature.com/articles/s41598-017-07447-7
70. Bitto NJ, Chapman R, Pidot S, Costin A, Lo C, Choi J, et al. Bacterial membrane vesicles transport their DNA cargo into host cells. Sci Rep [Internet]. 2017 Aug 1;7(1):7072. Available from: https://www.nature.com/articles/s41598-017-07288-4
71. Fulsundar S, Harms K, Flaten GE, Johnsen PJ, Chopade BA, Nielsen KM. Gene Transfer Potential of Outer Membrane Vesicles of *Acinetobacter baylyi* and Effects of Stress on Vesiculation. Kivisaar M, editor. Appl Environ Microbiol [Internet]. 2014 Jun;80(11):3469–83. Available from: https://journals.asm.org/doi/10.1128/AEM.04248-13
72. Ebel-Tsipis J, Fox MS, Botstein D. Generalized transduction by bacteriophage P22 in *Salmonella typhimurium*. J Mol Biol [Internet]. 1972 Nov;71(2):449–69. Available from: https://linkinghub.elsevier.com/retrieve/pii/0022283672903622
73. Smits WK, Weese JS, Roberts AP, Harmanus C, Hornung B. A helicase-containing module defines a family of pCD630-like plasmids in *Clostridium difficile*. Anaerobe [Internet]. 2018 Feb;49:78–84. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1075996417302287
74. Roseboom AM, Ducarmon QR, Hornung BVH, Harmanus C, Crobach MJT, Kuijper EJ, et al. Carriage of three plasmids in a single human clinical isolate of *Clostridioides difficile*. Plasmid [Internet]. 2023 Jan;125:102669. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0147619X22000531
75. Whittle MJ, Bilverstone TW, van Esveld RJ, Lücke AC, Lister MM, Kuehne SA, et al.

A Novel Bacteriophage with Broad Host Range against *Clostridioides difficile* Ribotype 078 Supports SlpA as the Likely Phage Receptor. Auchtung JM, editor. Microbiol Spectr [Internet]. 2022 Feb 23;10(1):e0229521. Available from: https://journals.asm.org/doi/10.1128/spectrum.02295-21

76. Goh S, Ong PF, Song KP, Riley T V, Chang BJ. The complete genome sequence of *Clostridium difficile* phage ϕC2 and comparisons to ϕCD119 and inducible prophages of CD630. Microbiology [Internet]. 2007 Mar 1;153(3):676–85. Available from: https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.2006/002436-0
77. Hargreaves KR, Kropinski AM, Clokie MRJ. What Does the Talking?: Quorum Sensing Signalling Genes Discovered in a Bacteriophage Genome. Kaufmann GF, editor. PLoS One [Internet]. 2014 Jan 24;9(1):e85131. Available from: https://dx.plos.org/10.1371/journal.pone.0085131
78. Sekulovic O, Garneau JR, Néron A, Fortier L-C. Characterization of Temperate Phages Infecting *Clostridium difficile* Isolates of Human and Animal Origins. Griffiths MW, editor. Appl Environ Microbiol [Internet]. 2014 Apr 15;80(8):2555–63. Available from: https://journals.asm.org/doi/10.1128/AEM.00237-14
79. Fortier L-C, Moineau S. Morphological and Genetic Diversity of Temperate Phages in *Clostridium difficile*. Appl Environ Microbiol [Internet]. 2007 Nov 15;73(22):7358–66. Available from: https://journals.asm.org/doi/10.1128/AEM.00582-07
80. Hinc K, Kabała M, Iwanicki A, Martirosian G, Negri A, Obuchowski M. Complete genome sequence of the newly discovered temperate *Clostridioides difficile* bacteriophage phiCDKH01 of the family *Siphoviridae*. Arch Virol [Internet]. 2021 Aug 20;166(8):2305–10. Available from: https://link.springer.com/10.1007/s00705-021-05092-0
81. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST : faster than PHASTER, better than PHAST. Nucleic Acids Res. 2023;gkad382:1–8.

Supplementaries

Following supplementary material can be found on the enclosed CD/Supplements/Chapter_4.4:

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

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

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

Following supplementary material is included below:

Table S1, Fig S1, Table S2

Table S1. Detailed prophage prediction results. More detailed PHASTEST (81) results of the analyzed strains on predicted regions and completeness score.

Strain	Predicted region	completeness
***C. difficile* TS3_3**	1,437,930 – 1,465,233	incomplete
	1,680,766 – 1,736,741	intact
***C. difficile* B1_2**		
Chromosome	1,035,372 – 1,089,705	intact
	1,467,664 – 1,488,405	incomplete
	1,705,680 – 1,762,842	intact
	2,540,996 – 2,554,708	incomplete
ECE 1	19 – 41,921	intact
ECE 2	none	
***C. difficile* J2_1**		
Chromosome	1,439,999 – 1,467,302	incomplete
	1,682,838 – 1,738,813	intact
	2,503,885 – 2,518,076	incomplete
ECE	1 – 11,796	incomplete
	14,174 – 46,002	intact
***C. difficile* MA_1**		
Chromosome	1,375,651 – 1,402,933	incomplete
ECE	28 – 33,560	intact
***C. difficile* MA_2**		
Chromosome	355,247 – 406,180	intact
	1,466,266 – 1,487,446	incomplete
	1,532,128 – 1,599,211	intact
	2,506,900 – 2,563,763	intact
ECE	none	

Table S1 continued.

Strain	Predicted region	completeness
***C. difficile* DSM 28196**	1,439,999 – 1,467,302	incomplete
	1,682,838 – 1,738,813	intact
	2,503,885 – 2,518,076	incomplete
***C. difficile* SC084-01-01**		
Chromosome	1,151,245 – 1,220,747	intact
	1,357,429 – 1,392,299	intact
	1,521,667 – 1,542,408	incomplete
	2,537,975 – 2,551,687	incomplete
ECE 1	136 – 16,524	incomplete
	19,825 – 46,846	intact
ECE 2	736 – 130,763	intact
***C. difficile* SC083-01-01**		
Chromosome	1,434,660 – 1,462,107	incomplete
	1,678,611 – 1,735,029	intact
	2,173,498 – 2,243,863	intact
	2,578,831 – 2,592,871	incomplete
ECE	58 – 45,180	intact
***C. difficile* DSM 29747**	1,375,725 – 1,403,007	incomplete

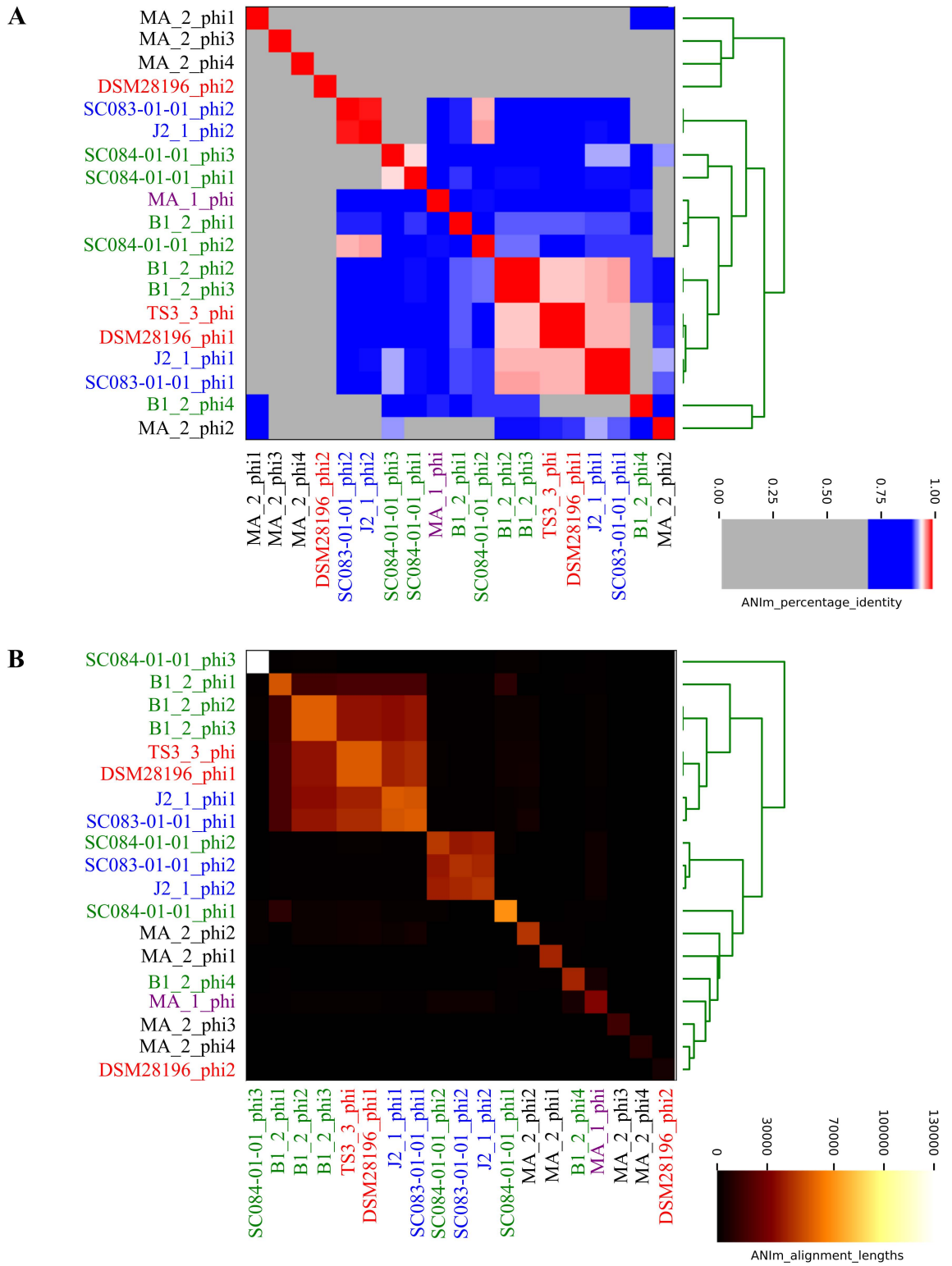


Fig S1. ANIm analysis of the active prophage regions. Heatmaps depict the (A) ANI values and (B) alignment lengths among the various active regions. The active regions are color-coded according to their ST for comparison better of analogous phages: red = ST1, green = ST3, blue = ST8, purple = ST11, black = ST340.

Table S2. Genome-based prediction of morphological family. Sequence length of the tail tape measure protein and presence of baseplate proteins were used to predict a phage of belonging to the *Myo-* (M) or *Siphoviridae* (S).

Phage	Tail length tape measure protein length (aa)	Presence Baseplate	Family prediction
TS3_3_phi	770	+	M
DSM28196_phi1	770	+	M
DSM28196_phi2	–	–	–
B1_2_phi1	764	+	M
B1_2_phi2	797	+	M
B1_2_phi3	1,767	–	S
B1_2_phi4	–	–	–
SC084-01-01_phi1	1,416	+	M
SC084-01-01_phi2	1,130	+	M
SC084-01-01_phi3	2,000*	–	S
J2_1_phi1	797	+	M
J2_1_phi2	1,129	+	M
SC083-01-01_phi1	797	+	M
SC083-01-01_phi2	1,129	+	M
MA_1_phi	584	+	M
MA_2_phi1	2,226	–	S
MA_2_phi2	1,839	–	S
MA_2_phi3	–	–	–
MA_2_phi4	–	–	–

*SC084-01-01_phi1 possessed two putative tail length tape measure proteins in close proximity. This is similar to the recent entry of phiCD211 (NC_029048.2), whereas the smaller protein is annotated as minor tail protein in the old entry LN681537.2.

5 GENERAL DISCUSSION

As a major contributor to nosocomial, antibiotic-associated diarrheic infections worldwide, the pathogen *Clostridioides difficile* is extensively studied by researchers all across the world (Smits et al. 2016; Balsells et al. 2019). Divergent attempts work on unravelling *C. difficile* virulence, and various, virulence-related aspects of its biology were thereby identified (Hunt and Ballard 2013; Lewis et al. 2017; Taggart et al. 2021). One characteristic of *C. difficile* is its mobile genome, which e.g. allows fast adaptation via acquisition of advantageous genes (Sebaihia et al. 2006). *C. difficile* harbors a diverse set of MGEs that affect strain fitness and virulence in different ways. Despite being a global pathogen, *C. difficile* also occurs as a natural inhabitant of human or mammalian intestines without disease manifestation (Eyre et al. 2013; Weese 2020), and it was also found in several environmental surroundings (Janezic et al. 2016; Dharmasena and Jiang 2018; Y. Zhou et al. 2021). Although researchers investigated diverse features to elucidate *C. difficile* virulence, the majority of them analyzed strains with clinical origin and antibiotic-based isolation. We therefore assumed a bias in *C. difficile* isolation and following analyses. In this context, this thesis aimed to investigate differences between clinical and non-clinical *C. difficile* strains with a focus on MGEs that are potentially linked to strain virulence. This research goal was approached by initial isolation of non-clinical *C. difficile* strains from environmental samples (Chapter 4.2 & 4.3), whereby antibiotic-free isolation attempts were pursued to overcome the presumable isolation bias. To support these and future isolation attempts, a *C. difficile*-specific detection PCR was established (Chapter 4.1). This PCR allows assessing the feasibility of environmental samples to isolate *C. difficile*. These non-clinical strains were compared to clinical reference strains in thorough genomic examinations (Chapter 4.3). They were further investigated on holistic prophage activity, an important aspect in *C. difficile* biology (Chapter 4.4). These experiments followed phage isolation under more natural conditions than commonly applied, which approximated actual phage activity of *C. difficile in vivo*.

5.1 Establishment of a *C. difficile*-specific detection PCR promotes isolation attempts and supports the hypothesis of an antibiotic-based isolation bias

Own *C. difficile* strains were isolated from environmental samples to obtain non-clinical isolates that would be subjected to different analyses within this thesis. The environmental samples (Chapter 4.1 Table 1 & S1) were collected at different locations, and mainly comprised soil or fecal samples from diverse animal species that are already described for *C. difficile*

carriage (Weese 2020). Although no direct contact of the sampling sites to antibiotics existed, the possibility of antibiotic resistances in these settings cannot be ruled out, as resistances can spread in diverse environment through various transmission routes (Kunhikannan et al. 2021). We assumed a bias in *C. difficile* isolation due to the general employment of antibiotic-based isolation procedures. Therefore, the non-clinical strains should preferably be isolated without antibiotic treatment to verify and overcome this bias. In support of this project, the establishment of a *C. difficile*-specific PCR to detect this bacterial species in metagenomic DNA from environmental samples with sequencing-based deduction of phylogenetic information was intended (Chapter 4.1). Different PCRs targeting *C. difficile* already exist, but these are either restricted to detect toxigenic strains (Bélanger et al. 2003) or were only validated against various Clostridia and applied on samples with already enriched *C. difficile* (Lemee et al. 2004; Stone et al. 2016; van Rossen et al. 2021). Consequently, these PCRs were not designed and evaluated in an environmental context. As novel PCR target, a region within the *hpdBCA*-operon was selected, which encodes enzymes for the transformation of tyrosine to *p*-cresol (Elsden, Hilton, and Waller 1976). This operon is present in all *C. difficile* strains but not widespread in other bacterial species (Passmore et al. 2018), and which was demonstrated to allow phylogenetic assignment based on sequence variants. This potential for phylogenetic classification was estimated by ANI analyses with *in silico*-retrieved detection PCR sequences from all complete *C. difficile* genomes available on NCBI. Comparison of this detection ANI analysis (Chapter 4.1 Fig. 3) to an ANI of all complete genomes on chromosome level (Chapter 4.1 Fig. 1) revealed similar clustering into the known phylogenetic clades of *C. difficile* (Knight et al. 2021), and thereby verified our detection PCR sequence as indicator for strain phylogeny. This indication however has its limits. As apparent in the detection ANI (Chapter 4.1 Fig. 2 & 3), phylogenetic assignment worked well on clade-level but could not always distinguish between different STs of the same clade. Thus, an initial classification of isolated strains and a rough determination of strain diversity in environmental samples is possible with our detection PCR. For final assignment of strain phylogeny, further analyses such as detailed MLST assessment is necessary. A proof of principle of the detection PCR was approached in two ways. First, true-positive and true-negative PCR results were verified on genomic DNA of different *C. difficile* strains and of bacterial species with potential cross reactivity as determined by BLASTn analysis (Chapter 4.1 Fig. S1), respectively. All these PCR results were correct. Second, detection PCR amplicons from various environmental samples were sequenced with NGS technique and inspected for *C. difficile* origin and phylogenetic information to validate PCR specificity. This assured *C. difficile* identity only and revealed five amplicon sequence

variants (ASV) belonging two three phylogenetic clades (Chapter 4.1 Fig. 4 & Table 2). Clade 1 and 5 dominated among the ASVs. Both clades comprise *C. difficile* strains that are prevalent in animal species (Janezic et al. 2014; Rodriguez Diaz, Seyboldt, and Rupnik 2018; Weese 2020), which accorded with the origin of the environmental samples (Chapter 4.1 Table 1). Clade 1 is further the globally most widespread and heterogeneous *C. difficile* clade according to the abundance of isolated and described strains (Knight et al. 2021). Therefore, the ASV-based phylogenetic analyses of the environmental samples reflected current phylogenetic dissemination. Additional phylogenetic assessments of various environmental samples and habitats by detection PCR amplicon sequencing would be interesting to further validate this global phylogenetic distribution. In the course of examining NGS amplicon data, detection PCR amplicons from *C. difficile* strains isolated from some of the environmental samples were included as controls for proper data processing. Interestingly, two isolates did not coincide with their individual ASV to the corresponding environmental ASVs (Chapter 4.1 Fig. 4). This observation indicated an underrepresentation of these two isolates in their respective environmental sample, so that the detection PCR captured other, more abundant *C. difficile* strains instead. Isolation of these underrepresented strains pointed to their selective enrichment over the other strains. This was indeed true since both strains were isolated under antibiotic treatment (Chapter 4.2 & 4.3 file S1). These results supported our initial assumption of an isolation bias due to antibiotic treatment.

The sensitivity of the detection PCR was evaluated in comparison to the common microbiome analysis via 16S rRNA gene sequencing (Chapter 4.1) (Klindworth et al. 2013). This demonstrated that *C. difficile* abundance in the environmental samples was below the detection limit of the 16S rRNA gene sequencing analysis (Chapter 4.1 Fig. S5 & data file S2), but sufficient for detection by the detection PCR. By that, the superior sensitivity of the detection PCR compared to the common 16S-based microbiome analysis was proven. Conclusively, the detection PCR is a simple but reliable method to identify a *C. difficile* isolate, and to assess *C. difficile* presence and diversity in environmental samples. It can therefore support future isolation attempts of *C. difficile* from non-clinical environments, and promote investigations on *C. difficile* diversity and dissemination worldwide, all without prior enrichments (Stone et al. 2016).

5.2 Antibiotic-free isolation attempts of non-clinical *C. difficile* strains

Various attempts for antibiotic-free isolation of *C. difficile* from environmental samples that were promising according to detection PCR analysis (Chapter 4.1) were performed, but only two strains could be isolated under these conditions (Chapter 4.3 file S1). Three further strains were isolated from additional isolation procedures under antibiotic treatment that is commonly used in *C. difficile* isolation (Chapter 4.2 & 4.3 file S1) (Aspinall and Hutchinson 1992; Dharmasena and Jiang 2018). Throughout all isolation attempts, both with and without antibiotics, various other bacteria were isolated as well, belonging to the genera of *Clostridium* sensu stricto 1, *Romboutsia*, *Terrisporobacter*, and *Bacillus* (data not shown). This was no surprise since the microbial community in the environmental samples comprised a multitude of different species, as determined during the sensitivity evaluation of the detection PCR in comparison to microbiome analysis via 16S rRNA gene sequencing (Chapter 4.1). The 16S amplicon-based community assessment verified bacterial diversity in all environmental samples (Chapter 4.1 Fig. S4-S6 & data file S2). Coinciding with the bacterial species that were isolated alongside *C. difficile* strains, ASVs representing bacteria of the aforementioned genera were found in the 16S amplicon data in significant abundance (Chapter 4.1 Fig. S5 & data file S2). In contrast, none of the environmental samples contained detectable reads corresponding to *C. difficile*. These results showed that *C. difficile* was prevailed by various other bacterial species, which produce bacterial spores and thrive under the same basic cultivation conditions as employed during isolations, such as moderate temperature and exclusion of oxygen (Schleifer 2009; Nakano and Zuber 1998; Gerritsen et al. 2014). It is therefore reasonable that isolation of *C. difficile* under these premises without antibiotic-based cultivation is challenging. In the absence of a particular selective pressure as growth advantage for *C. difficile*, all prevailing bacteria with similar cultivation requirements will further dominate during enrichment and isolation cultivations (Wan et al. 2023). In fact, different steps to promote *C. difficile* growth during the enrichment and isolation procedures were employed. First, *C. difficile* spore germination was supported by addition of taurocholic acid (Chapter 4.3 file S1) (Sorg and Sonenshein 2008). This should facilitate *C. difficile* to out-compete other spore-forming bacteria. Second, enrichment cultivation in amino acid-based medium was done to select for bacteria that can grow by performing the Stickland reaction, a fundamental ability of *C. difficile* (Mead 1971; Jackson et al. 2006). Therefore, a defined medium with amino acids as sole carbon source (Chapter 4.3 file S1) that was specifically developed for *C. difficile* cultivation and only contained essential components was used (Yamakawa et al. 1996;

Karasawa et al. 1995). However, these physiological and metabolic benefits were not selective enough to establish an effective, antibiotic-free *C. difficile* isolation protocol. The issue here is the various bacteria accompanying *C. difficile* as mentioned previously, which can also ferment amino acids (Mead 1971; Amaretti et al. 2019). Together, all these findings explain the challenge of antibiotic-free *C. difficile* isolation and thereby highlight the benefit of antibiotic treatment as selective factor. However, additional antibiotic-based isolation attempts (Chapter 4.2 & 4.3 file S1) were not significantly more successful. In reference to the antibiotic-driven isolation bias addressed in the detection PCR project (Chapter 4.1), where un-detected strains were isolated in contrast to the detected ones (Chapter 4.1 Fig. 4), further attempts to establish an antibiotic-free isolation procedure for complementary application with antibiotic-based isolation protocols would be advisable. This would help to overcome this bias and ensure isolation of all *C. difficile* strains present in environmental samples.

5.3 Comparative genome analyses of non-clinical and clinical *C. difficile* strains support genomic adaptations of clinical strains with relevance for strain virulence

Comprehensive whole-genome comparisons of the self-isolated non-clinical strains TS3_3, B1_2, J2_1, and MA_2 (Chapter 4.3 file S1) to clinical reference strains of the same ST were conducted to reveal possible connections to the clinical background (Chapter 4.3). Analyses focused on MGEs and their putative influence on strain virulence. An initial assessment of the genomic virulence potentials of corresponding clinical and non-clinical strains was performed by investigating virulence factors that are critical in *C. difficile* pathogenesis (Chapter 4.3 Fig. 1). The *C. difficile* toxins as major virulence factors alongside their regulatory genes were likewise identified in the corresponding strains at the same genomic position, and nucleotide sequence comparison ascertained identical gene sequences (Chapter 4.3 Fig. 1 & 2). Variations of toxin gene sequences are known to alter strain virulence (Lanis et al. 2013; Q. Dong et al. 2023). The other analyzed virulence factors that are involved in adherence, exoenzymatic reaction, motility, and sporulation were compared on protein sequence level. Almost no differences between corresponding strains were detected besides few cases of deviating amino acids (Chapter 4.3 Fig. 1). No connection to the clinical background regarding these sequence deviations was observed. Moreover, these slight differences could not be linked to a virulence-related phenotype since studies about some of the concerned proteins only addressed their modularity (Tulli et al. 2013) or investigated deletion or disruption of the entire gene (Q. Zhou et al. 2022). Evaluating the potential effects of these amino acid deviations on strain phenotype would require experimental evidence. Since no significant differences in the analyzed virulence

factors between clinical and non-clinical strains were detected, the corresponding strains possessed similar genomic virulence potential.

Pairwise core/pan genome analyses of corresponding strains revealed a higher number of accessory genes in the clinical strains, which was further supported in regard of the relative proportion of unique genes to the total content of coding sequences (CDS) (Chapter 4.3 Fig. 3B). The accessory genes were assigned to functional categories of clusters of orthologous groups (COG) to examine putative connections of the clinical strains to more unique genes of particular COGs. In this context, the proportions of COGs among the accessory genes were calculated in relation to total CDS amount, and the differences of COG proportions between corresponding clinical and non-clinical strains were determined (Chapter 4.3 Fig. 4). This revealed unique genes without successful COG assignment or with assignment to the COG categories S ("Function unknown"), K ("Transcription"), and L ("Replication, recombination and repair") in higher abundance in the clinical strains. Among the genes of COG category S, the potential virulence factors haemolysin XhlA and the virulence-associated protein E were encoded, which are both known virulence factors in other bacterial species (Thomas et al. 2021; Ji et al. 2016). Their influence on virulence in *C. difficile* is not examined, yet. Interestingly, *C. difficile* is typically not described as hemolytic, but its hemolytic activity was demonstrated under specific conditions (Alkudmani 2018). In the course of the prophage activity project (Chapter 4.4), the haemolysin XhlA genes were identified in genomes of active phages. This would imply the possibility of phage-mediated gene transfer of this virulence factor between *C. difficile* strains with potential impact on strain virulence, as addressed in the pathogen *Mannheimia haemolytica* (Niu et al. 2015). But haemolysin XhlA might also exhibit a different function by involvement in the phage lysis module, which was described for instance in a *Bacillus subtilis* prophage (Krogh, Jørgensen, and Devine 1998). Thus, the actual role of haemolysin XhlA and the virulence-associated protein E remain elusive without further experiments. However, no connection of these potential virulence factors to the clinical background was apparent.

Accessory genes belonging to the previously mentioned functions (COG S, K, L) were linked to higher strain virulence in *in vivo* experiments (Lewis et al. 2017) and in comparative investigations of historical and hypervirulent RT027 strains (Stabler et al. 2009). Referring to the higher abundance of these genes in the clinical strains (Chapter 4.3 Fig. 4B), this might suggest likewise higher virulence with association to the clinical background. Verification of this assumption needs experimental evidence *in vivo*. The higher abundance of unique genes

observed in the clinical strains, in particular with transcriptional-related functions, might also promote rapid physiological adaptation to changing environments (Parter, Kashtan, and Alon 2007). This aspect was already addressed during genomic investigations of *C. difficile* 630 (Sebaihia et al. 2006). Faster adaptation to environmental changes supports strain survival, which ultimately effects strain virulence as well. Another clinical-related trend was observed for accessory genes involved in conjugal DNA transfer (COG category U) (Chapter 4.3 Fig. 4B). This is not only relevant regarding the dissemination of ARGs (de la Cruz and Davies 2000) but also the potential transfer of the *C. difficile* toxin locus, which allows previously non-toxigenic strains to produce the disease-causing toxins (Brouwer et al. 2013). Conjugation-related genes are also interesting in the context of biofilm formation, a fundamental aspect in the *C. difficile* life style (Dapa and Unnikrishnan 2013). The close proximity of cells within a biofilm facilitates cell-to-cell contact for conjugative activity, so that a significant impact of biofilm formation on conjugation efficiency was already shown in other bacterial species (Lécuyer et al. 2018; Ghigo 2001). A connection between biofilm formation and conjugation was so far not addressed in *C. difficile* but would be worth investigating. This assumption might also shed a new light on the previously identified amino acid deviations in adherence-related virulence factors (Chapter 4.3 Fig. 1), which could be involved in the interaction of biofilm formation and conjugative activity. Overall, the results from the pan genome analyses indicated clinical-related adaptations that result in higher reactivity, which can promote strain virulence.

The genome analyses further included detection of ARGs (Chapter 4.3 Fig. 5) and MGEs (Chapter 4.3 Fig. 6). No particular gene or element was characteristic for the clinical strains, but some were restricted to clinical or non-clinical strains, respectively. For example, the ARGs *erm*(B) and *tet*(M) were only detected in clinical strains (ST1/ST3 and ST11). These ARGs are often found in transposons, which allows horizontal transfer of these resistances (D. Dong et al. 2014; Kartalidis et al. 2021). Corresponding transposons in the form of Integrative and Conjugative or Mobilizable Elements (ICE/IME) were indeed identified in the respective strains (Chapter 4.3 Fig. 6), which could imply a connection between these ARGs and MGEs. An overall trend of more ARGs and MGEs in total was observed in relation to clinical background.

Pairwise genome comparisons combined all previous analyses within genomic context (Chapter 4.3 Fig. 7). This verified conjunctions between specific ARGs and MGEs, such as *erm*(B) (Chapter 4.3 Fig. 7A & B) or *tet*(M) (Chapter 4.3 Fig. 7D) with ICE Tn*916* as previously assumed and also observed elsewhere (D. Dong et al. 2014; Kartalidis et al. 2021). Some ARGs

did not reside within but close to a MGE, like *cfr*(E) to IME MOB_V (Chapter 4.3 Fig. 7A). This raises the question of a connection between this ARG and MGE with potential to gene transfer, a speculation worth investigating when addressing the spread of antibiotic resistances. Noteworthy, the ARG *cfr*(E) was already identified in *C. difficile* within an undescribed MGE (Stojković et al. 2019). Besides association with ARGs, MGEs were also found to bear the majority of the unique genes assigned to COG categories S, K, and L. Since accessory genes encoding these functions correlated with higher strain virulence (Stabler et al. 2009; Lewis et al. 2017) and consequently suggested higher virulence in the clinical strains (Chapter 4.3), the MGEs associated with these genes represent important agents of strain virulence. In this context, the higher abundance of transcriptional-related accessory genes (COG K) in clinical strains with their suggested association to higher physiological reactivity on environmental fluctuations is also connected to MGEs. This might imply the potential transfer of this reactivity via HGT, which ultimately allows fast dissemination of this virulence-relevant aspect among *C. difficile* strains. All these findings highlight the relevance of MGEs in *C. difficile* virulence.

5.4 Phage analyses confirmed inducing effect of physiological DCA, highlighted the relevance of spontaneous phage activity and revealed novel mechanisms of HGT in *C. difficile*

Phenotypic investigations of the self-isolated non-clinical strains and the clinical reference strains addressed spontaneous prophage activity and further examined the secondary bile acid DCA as putative phage-inducing agent (Chapter 4.4). Since DCA is a naturally occurring stress factor for *C. difficile* (Sorg and Sonenshein 2008; Theriot, Bowman, and Young 2016) and can trigger the bacterial SOS-response (Kandell and Bernstein 1991), it is highly relevant regarding *C. difficile* phage activity *in vivo*. Initial tolerance assessments of the strains to varying DCA concentrations demonstrated no significant differences between corresponding clinical and non-clinical strains of the same ST, which implied no relation to clinical background (Chapter 4.4 Fig. 1). This entailed similar stress levels of the corresponding strains upon exposure to DCA, which could also indicate a similar phage-inducing effect.

Prophage activity was analyzed by sequencing of particle-protected, double-stranded DNA and mapping of sequencing reads onto corresponding host genomes, thereby implementing technical normalization of differing read amounts as well as OD-normalization to compare phage abundance of the same cell density. Sequencing library preparation of clinical ST11 strain DSM 29747 failed, so that that this strain was missing for analysis and comparisons of prophage activity. The mapped sequencing data revealed spontaneous prophage activity in

all other strains, often involving multiple phages and extrachromosomal elements (ECE) as episomal phages (Chapter 4.4 Fig. 2-6). This underlined the prevalence of active, extrachromosomal phages (Fortier 2018), and the importance of spontaneous prophage activity in *C. difficile*. The mechanism of spontaneous prophage activity (also termed spontaneous prophage induction, in short SPI) is still under elucidation, but the release of phage particles and the accompanying cell lysis benefits the host population by promoting biofilm formation, release of virulence-associated proteins, HGT, or elimination of non-lysogenic competitors (Nanda, Thormann, and Frunzke 2015; Cortes, Krog, and Balázsi 2019). Noteworthy, a relevant connection between biofilm formation and HGT was already pointed out in the comprehensive genome comparisons (Chapter 4.3) and is now extended by the beneficial involvement of SPI. Almost all regions exhibited a distinctly higher activity after DCA-treatment with divergent magnitudes of increase. This implies differential inductions of the various regions, but overall proves the inducing effect of DCA. Since spontaneous phage activity was already pointed out to support biofilm formation (Nanda, Thormann, and Frunzke 2015), and DCA was shown in *C. difficile* to induce the production of a protective biofilm (Dubois et al. 2019), a connection between DCA-related phage induction and biofilm formation seems plausible. The relation between DCA-induced phage activity and corresponding biofilm formation would be worth investigating to elucidate this virulence aspect of *C. difficile*. No significant difference regarding carriage of active prophages was observed between clinical and non-clinical strains. This prompted to ANI-based similarity assessments of analogous phages (Chapter 4.4 Fig. S1). These comparisons revealed the phage of ST1-strains to be identical, which made it especially interesting for activity comparisons between the corresponding strains. Although the analogous phages were identical and exhibited no distinctly different spontaneous activity, the DCA-induced activity increase differed remarkably (Chapter 4.4 Fig. 2). This indicated no phage- but host-dependent reaction, resulting in a stronger phage-induction in the clinical strain upon DCA exposure. The clinical-related difference appears in a new light when referring to the genome comparisons (Chapter 4.3), which indicated higher reactivity of the clinical strains to environmental changes, such as addition of DCA during exponential growth within the phage induction experiments. Faster reaction by the clinical ST1-strain on stressful DCA exposure might explain the divergent, strain-dependent phage inductions (Chapter 4.4 Fig. 1). Moreover, this reaction could possibly result in enhanced biofilm formation upon increased phage activity as protective mechanism against the adverse effects of DCA.

No further identical phages between corresponding clinical and non-clinical strains existed, so that additional comparisons could not provide clear support for the assumed

connection between faster reactivity of clinical strains (Chapter 4.3) and stronger DCA-related phage induction, as described for ST1 strains (Chapter 4.4 Fig. 2). Analogous phages of ST8 strains were similar to each other, whereas the remaining phages were rather unique among all analyzed phages (Chapter 4.4 Fig. S1). Interestingly, the analogous, chromosomal phage of the ST8 strains exhibited a similar trend of stronger DCA-induction in the clinical strain (Chapter 4.4 Fig. 4). A host-dependent mechanism cannot be ruled out, but the genomic differences between the analogous phages might be responsible as well. This phage-dependency is further supported by the striking activity pattern of the chromosomal phage from non-clinical ST8-strain J2_1 (phage J2_1_phi1), where coverage up- and downstream of the actual phage region and drastic differences in signal intensity within the phage genome were observed (Chapter 4.4 Fig. 4A). The upstream region could be identified as an ICE, which points to the transducing mechanism of hitchhiking ICEs as already observed in other bacteria (Lindsay et al. 1998; Duerkop et al. 2012; Seed et al. 2013) and also described once in *C. difficile* (Goh et al. 2013). Therefore, this observation here is the second description of this specific gene transfer mechanism in *C. difficile* so far. The downstream region comprised only bacterial genes without indication for any MGE, but several proteins with potential involvement in antibiotic resistance (ABC-transporters, multidrug efflux system ATP-binding protein) or virulence were encoded. This also indicates transduction of bacterial DNA. Based on the location of the transduced DNA adjacent to phage J2_1_phi1, and the transduction frequency with signal strength comparable to the phage activity, lateral transduction is the most plausible mechanism to explain these observations (Borodovich et al. 2022; Kleiner et al. 2020). Accordingly, the DNA-packaging strategy of phage J2_1_phi1 determined via protein sequence alignments of the large terminase subunit was identified as P22-like headful (Chapter 4.4 Fig. 7). The headful mechanism is required for transduction activity (Borodovich et al. 2022), and the original phage of the P22-like strategy, P22 in *Salmonella enterica*, was already shown to employ lateral transduction (Fillol-Salom et al. 2021). This is the first evidence of lateral transduction in *C. difficile*, and further investigations on this phage and its non-transducing analogue in the clinical ST8-strain could provide insight into this transduction strategy. The genomic differences between both phages might be responsible for the transduction activity in only one of the two phages. These comprised divergent excisionase and integrase genes as well as slightly differing protein sequences of the large terminase subunits, which perform key steps in lateral transduction (Borodovich et al. 2022). The action of lateral transduction might further explain the drastic activity difference within the genome of phage J2_1_phi1 (Chapter 4.4 Fig 4A), since this mechanism involves *in situ* replication of the phage genome and late excision from the bacterial

chromosome (Borodovich et al. 2022), which could possibly result in erroneously excised and therefore truncated phage genomes.

The remaining pair of clinical and non-clinical strains of ST3 exhibited spontaneous phage activity with DCA-induced increase, but no clinical-related trend could be observed (Chapter 4.4 Fig. 3). However, these strains shared no analogue phages (Chapter 4.4 Fig. S1), so that activity comparisons as performed in ST8- and ST1-strains in particular are not appropriate. The non-clinical ST3-strain was interesting though, as it possessed substantial activity on its second ECE (Chapter 4.4 Fig. 3A) that was however no phage according to its gene content. Similar non-phage elements with distinct activity (spontaneous and DCA-induced) were also observed in the clinical ST1-strain DSM 28196 (Chapter 4.4 Fig. 2A; DSM28196_phi2) and the ST340 strain MA_2 (Chapter 4.4 Fig. 6; MA_2_phi_3/4). Genomic examination of these regions showed several genes with relation to MGEs, which did not comprise encoded structural proteins for particle assembly to envelop the respective DNA. However, envelopment of these elements is certain based on the experimental procedure for isolation of particle-protected DNA, which prompts to the involvement of un-related particles. Further phage-mediated transductions as previously described in strain J2_1 would coincide with the P22-like headful DNA-packaging mechanism determined for the co-existing phages (Chapter 4.4 Fig. 7). However, the headful packaging strategy implies transduced DNA of similar phage genome size (Borodovich et al. 2022), which did not apply to the significantly smaller non-phage elements compared to the co-existing phages (Chapter 4.4 Table 3). Sizes of the non-phage elements did not match to phage-mediated transduction, but would accord with packaging by gene transfer agents (GTA) (Borodovich et al. 2022). GTAs are phage-like particles encoded in the host chromosome that encapsidate random bacterial DNA un-related to their encoding region (Borodovich et al. 2022). This random DNA packaging does not accord with the distinct active non-phage elements, though. Particles protecting the non-phage elements might also be vesicles, which are already described in other bacteria as important DNA vehicle (Fulsundar et al. 2014; Bitto et al. 2017; Tran and Boedicker 2017), allowing even gene transfer between different species (Fulsundar et al. 2014). Although of relevance for HGT, this specific type of genetic exchange is so far not described in *C. difficile*. All above-described mechanism might be involved in the envelopment of the diverse non-phage elements, so that further experiments are required for final conclusions.

Nucleotide BLAST analyses of the active phages showed partially similarity to known *C. difficile* phages, but overall demonstrated their uniqueness (Chapter 4.4 Table 4 & data file S1).

Therefore, this work substantially broadens the knowledge on *C. difficile* phages. Significant matches against *C. difficile* genomes further indicate the presence of putative prophages, which might be active as well and, thus, could be investigated for further insights into *C. difficile* phages. The non-phage elements showed similarity to *C. difficile* chromosomes or plasmids (Chapter 4.4 Table 4 & data file S1), also here indicating the prevalence of similar elements in other *C. difficile* strains. The plasmids represent different classes that show similar organization and were found in various *C. difficile* strains (Smits et al. 2018; Roseboom et al. 2023). This raises the question, if the similar plasmids are also inducible and particle-protected, thereby shedding another light on their transfer mechanism.

6 CONCLUSION

This thesis provided novel findings that support investigations on the global pathogen *C. difficile* and its virulence and encourages further research of this pathogen.

The established detection PCR can support future isolation attempts of *C. difficile* from various sources, and also allows an initial phylogenetic assessment via amplicon NGS. This gives an overview on the *C. difficile* diversity in environmental samples, which promotes holistic strain isolation and analysis of strain dissemination. Amplicon sequencing results in conjunction with the actually isolated strains further supported the initially stated assumption of an isolation bias due to the common antibiotic-based isolation techniques. This encourages further work on antibiotic-free isolation in combination with the already established antibiotic-based protocols to isolate each strain present in the sample of interest. Different approaches to overcome the assumed isolation bias were performed with antibiotic-free attempts, but efficient *C. difficile* isolation was demonstrated to be challenging, so that further work is required to establish an effective protocol. In total, five divergent strains could be obtained from environmental samples with and without antibiotic-based isolation.

These non-clinical isolates were compared to clinical reference strains in comprehensive genome analyses to elucidate clinical-related features with virulence association, thereby addressing MGEs. The clinical and non-clinical strains did not significantly differ in their fundamental virulence equipment, as determined genomic analysis of virulence factors relevant for *C. difficile* pathogenesis. Clinical strains possessed overall more ARGs, MGEs, and accessory genes than the corresponding non-clinical strains. These accessory genes comprised in particular genes encoding hypothetical proteins and proteins involved in unknown functions, transcription, or replication, recombination, and repair. All these functions were encoded in the accessory genome of strains with higher virulence. Based on these results, it is recommended to extend comparative studies on *C. difficile* virulence beyond strains of clinical origin. Further, the higher abundance of accessory genes, especially with transcriptional association, might allow faster adaptation to changing environments, thereby supporting virulence in the clinical strains. These genes were mainly associated with MGEs, which underlines the importance of MGEs in *C. difficile* fitness and virulence.

Phenotypic investigations on prophage activity in all non-clinical strains isolated within this thesis and the clinical reference strains revealed the relevance of spontaneous phage induction, comprising almost all analyzed strains as well as multiple phages within the same

host. Further, the phage-inducing effect of the secondary bile salt DCA, a natural stressor for *C. difficile*, was also confirmed. Divergent induction intensities showed evidence for involvement of both phage- and host-dependent mechanisms, but overall underlined the importance of phage research under conditions that rather resemble the natural habitat of *C. difficile* than commonly used in phage isolation. No clear feature or activity was linked to the clinical strains. However, comparison of DCA-induced phage activity in one pair of clinical and non-clinical strain indicated a clinical-related trend. This trend might connect the observed stronger phage induction in the clinical strain to its assumed higher reactivity by faster adaptation to the sudden DCA exposure. The experiments further revealed active genomic regions without phage affiliation but nevertheless some kind of DNA-envelopment, which sheds light on so far undescribed gene transfer mechanisms in *C. difficile*. One of this mechanism could be identified as lateral transduction, which was the first description in *C. difficile*. This study significantly broadened the knowledge about *C. difficile* phages and set the basis for novel research attempts to elucidate this pivotal aspect in *C. difficile* biology and virulence.

7 GENERAL REFERENCES

Abad-Fau, Ana, Eloísa Sevilla, Inmaculada Martín-Burriel, Bernardino Moreno, and Rosa Bolea. 2023. "Update on Commonly Used Molecular Typing Methods for *Clostridioides Difficile*." *Microorganisms* 11 (7): 1752. https://doi.org/10.3390/microorganisms11071752.

Alkudmani, Zeina Subhi B. 2018. "The Identification and Characterization of Novel Haemolysin Genes from *Clostridium Difficile*." University College London. https://discovery.ucl.ac.uk/id/eprint/10078163/7/Alkudmani_00_Thesis_edited.pdf.

Allegretti, J. R., S. Kearney, N. Li, E. Bogart, K. Bullock, G. K. Gerber, L. Bry, C. B. Clish, E. Alm, and J. R. Korzenik. 2016. "Recurrent *Clostridium Difficile* Infection Associates with Distinct Bile Acid and Microbiome Profiles." *Alimentary Pharmacology & Therapeutics* 43 (11): 1142–53. https://doi.org/10.1111/apt.13616.

Amaretti, Alberto, Caterina Gozzoli, Marta Simone, Stefano Raimondi, Lucia Righini, Vicente Pérez-Brocal, Rodrigo García-López, Andrés Moya, and Maddalena Rossi. 2019. "Profiling of Protein Degraders in Cultures of Human Gut Microbiota." *Frontiers in Microbiology* 10 (November): 2614. https://doi.org/10.3389/fmicb.2019.02614.

Aspinall, S T, and D N Hutchinson. 1992. "New Selective Medium for Isolating *Clostridium Difficile* from Faeces." *Journal of Clinical Pathology* 45 (9): 812–14. https://doi.org/10.1136/jcp.45.9.812.

Baban, Soza T., Sarah A. Kuehne, Amira Barketi-Klai, Stephen T. Cartman, Michelle L. Kelly, Kim R. Hardie, Imad Kansau, Anne Collignon, and Nigel P. Minton. 2013. "The Role of Flagella in *Clostridium Difficile* Pathogenesis: Comparison between a Non-Epidemic and an Epidemic Strain." Edited by Michel R. Popoff. *PLoS ONE* 8 (9): e73026. https://doi.org/10.1371/journal.pone.0073026.

Balsells, Evelyn, Ting Shi, Callum Leese, Iona Lyell, John Burrows, Camilla Wiuff, Harry Campbell, Moe H Kyaw, and Harish Nair. 2019. "Global Burden of *Clostridium Difficile* Infections: A Systematic Review and Meta-Analysis." *Journal of Global Health* 9 (1): 010407. https://doi.org/10.7189/jogh.09.010407.

Bartlett, J G, N S Taylor, T Chang, and J Dzink. 1980. "Clinical and Laboratory Observations in Clostridium Difficile Colitis." *The American Journal of Clinical Nutrition* 33 (11): 2521–26. https://doi.org/10.1093/ajcn/33.11.2521.

Bartlett, John G., Te Wen Chang, Marc Gurwith, Sherwood L. Gorbach, and Andrew B. Onderdonk. 1978. "Antibiotic-Associated Pseudomembranous Colitis Due to Toxin-

Producing Clostridia." *New England Journal of Medicine* 298 (10): 531–34. https://doi.org/10.1056/NEJM197803092981003.

Bélanger, Simon D., Maurice Boissinot, Natalie Clairoux, François. J. Picard, and Michel G. Bergeron. 2003. "Rapid Detection of *Clostridium Difficile* in Feces by Real-Time PCR." *Journal of Clinical Microbiology* 41 (2): 730–34. https://doi.org/10.1128/JCM.41.2.730-734.2003.

Bitto, Natalie J., Ross Chapman, Sacha Pidot, Adam Costin, Camden Lo, Jasmine Choi, Tanya D'Cruze, et al. 2017. "Bacterial Membrane Vesicles Transport Their DNA Cargo into Host Cells." *Scientific Reports* 7 (1): 7072. https://doi.org/10.1038/s41598-017-07288-4.

Bonneaud, Camille, and Ben Longdon. 2020. "Emerging Pathogen Evolution." *EMBO Reports* 21 (9): e51374. https://doi.org/10.15252/embr.202051374.

Borodovich, Tatiana, Andrey N Shkoporov, R Paul Ross, and Colin Hill. 2022. "Phage-Mediated Horizontal Gene Transfer and Its Implications for the Human Gut Microbiome." *Gastroenterology Report* 10 (January): goac012. https://doi.org/10.1093/gastro/goac012.

Brouwer, Michael S.M., Adam P. Roberts, Haitham Hussain, Rachel J. Williams, Elaine Allan, and Peter Mullany. 2013. "Horizontal Gene Transfer Converts Non-Toxigenic *Clostridium Difficile* Strains into Toxin Producers." *Nature Communications* 4 (1): 2601. https://doi.org/10.1038/ncomms3601.

Buddle, Jessica E., and Robert P. Fagan. 2023. "Pathogenicity and Virulence of *Clostridioides Difficile*." *Virulence* 14 (1): 2150452. https://doi.org/10.1080/21505594.2022.2150452.

Chandrasekaran, Ramyavardhanee, and D. Borden Lacy. 2017. "The Role of Toxins in *Clostridium Difficile* Infection." *FEMS Microbiology Reviews* 41 (6): 723–50. https://doi.org/10.1093/femsre/fux048.

Chaves-Olarte, Esteban, Peter Löw, Enrique Freer, Thomas Norlin, Manfred Weidmann, Christoph von Eichel-Streiber, and Monica Thelestam. 1999. "A Novel Cytotoxin from *Clostridium Difficile* Serogroup F Is a Functional Hybrid between Two Other Large Clostridial Cytotoxins." *Journal of Biological Chemistry* 274 (16): 11046–52. https://doi.org/10.1074/jbc.274.16.11046.

Chen, John, Nuria Quiles-Puchalt, Yin Ning Chiang, Rodrigo Bacigalupe, Alfred Fillol-Salom, Melissa Su Juan Chee, J. Ross Fitzgerald, and José R. Penadés. 2018. "Genome Hypermobility by Lateral Transduction." *Science* 362 (6411): 207–12. https://doi.org/10.1126/science.aat5867.

Chiang, Yin Ning, José R. Penadés, and John Chen. 2019. “Genetic Transduction by Phages and Chromosomal Islands: The New and Noncanonical.” Edited by Kimberly A. Kline. *PLOS Pathogens* 15 (8): e1007878. https://doi.org/10.1371/journal.ppat.1007878.

Cortes, Michael G., Jonathan Krog, and Gábor Balázsi. 2019. “Optimality of the Spontaneous Prophage Induction Rate.” *Journal of Theoretical Biology* 483 (December): 110005. https://doi.org/10.1016/j.jtbi.2019.110005.

Curry, Scott R., Jane W. Marsh, Carlene A. Muto, Mary M. O’Leary, A. William Pasculle, and Lee H. Harrison. 2007. “*TcdC* Genotypes Associated with Severe TcdC Truncation in an Epidemic Clone and Other Strains of *Clostridium Difficile*.” *Journal of Clinical Microbiology* 45 (1): 215–21. https://doi.org/10.1128/JCM.01599-06.

Czepiel, Jacek, Mirosław Dróżdż, Hanna Pituch, Ed J. Kuijper, William Perucki, Aleksandra Mielimonka, Sarah Goldman, Dorota Wultańska, Aleksander Garlicki, and Grażyna Biesiada. 2019. “*Clostridium Difficile* Infection: Review.” *European Journal of Clinical Microbiology & Infectious Diseases* 38 (7): 1211–21. https://doi.org/10.1007/s10096-019-03539-6.

Dapa, Tanja, and Meera Unnikrishnan. 2013. “Biofilm Formation by *Clostridium Difficile*.” *Gut Microbes* 4 (5): 397–402. https://doi.org/10.4161/gutm.25862.

de la Cruz, Fernando de, and Julian Davies. 2000. “Horizontal Gene Transfer and the Origin of Species: Lessons from Bacteria.” *Trends in Microbiology* 8 (3): 128–33. https://doi.org/10.1016/S0966-842X(00)01703-0.

Dharmasena, Muthu, and Xiuping Jiang. 2018. “Improving Culture Media for the Isolation of *Clostridium Difficile* from Compost.” *Anaerobe* 51 (June): 1–7. https://doi.org/10.1016/j.anaerobe.2018.03.002.

Dion, Moïra B., Frank Oechslin, and Sylvain Moineau. 2020. “Phage Diversity, Genomics and Phylogeny.” *Nature Reviews Microbiology* 18 (3): 125–38. https://doi.org/10.1038/s41579-019-0311-5.

Dong, Danfeng, Xu Chen, Cen Jiang, Lihua Zhang, Gang Cai, Lizhong Han, Xuefeng Wang, Enqiang Mao, and Yibing Peng. 2014. “Genetic Analysis of Tn916-like Elements Conferring Tetracycline Resistance in Clinical Isolates of *Clostridium Difficile*.” *International Journal of Antimicrobial Agents* 43 (1): 73–77. https://doi.org/10.1016/j.ijantimicag.2013.09.004.

Dong, Qiwen, Huaiying Lin, Marie-Maude Allen, Julian R. Garneau, Jonathan K. Sia, Rita C. Smith, Fidel Haro, et al. 2023. “Virulence and Genomic Diversity among Clinical Isolates of ST1 (BI/NAP1/027) *Clostridioides Difficile*.” *Cell Reports* 42 (8): 112861.

https://doi.org/10.1016/j.celrep.2023.112861.

Drudy, Denise, Séamus Fanning, and Lorraine Kyne. 2007. "Toxin A-Negative, Toxin B-Positive *Clostridium Difficile*." *International Journal of Infectious Diseases* 11 (1): 5–10. https://doi.org/10.1016/j.ijid.2006.04.003.

Dubois, Thomas, Yannick D. N. Tremblay, Audrey Hamiot, Isabelle Martin-Verstraete, Julien Deschamps, Marc Monot, Romain Briandet, and Bruno Dupuy. 2019. "A Microbiota-Generated Bile Salt Induces Biofilm Formation in *Clostridium Difficile*." *Npj Biofilms and Microbiomes* 5 (1): 14. https://doi.org/10.1038/s41522-019-0087-4.

Ducarmon, Quinten R., Tjomme van der Bruggen, Céline Harmanus, Ingrid M.J.G. Sanders, Laura G.M. Daenen, Ad C. Fluit, Rolf H.A.M. Vossen, Susan L. Kloet, Ed J. Kuijper, and Wiep Klaas Smits. 2023. "*Clostridioides Difficile* Infection with Isolates of Cryptic Clade C-II: A Genomic Analysis of Polymerase Chain Reaction Ribotype 151." *Clinical Microbiology and Infection* 29 (4): 538.e1-538.e6. https://doi.org/10.1016/j.cmi.2022.12.003.

Duerkop, Breck A., Charmaine V. Clements, Darcy Rollins, Jorge L. M. Rodrigues, and Lora V. Hooper. 2012. "A Composite Bacteriophage Alters Colonization by an Intestinal Commensal Bacterium." *Proceedings of the National Academy of Sciences* 109 (43): 17621–26. https://doi.org/10.1073/pnas.1206136109.

EClinicalMedicine. 2021. "Antimicrobial Resistance: A Top Ten Global Public Health Threat." *EClinicalMedicine* 41 (November): 101221. https://doi.org/10.1016/j.eclinm.2021.101221.

Elsden, Sidney R., Martin G. Hilton, and Janet M. Waller. 1976. "The End Products of the Metabolism of Aromatic Amino Acids by Clostridia." *Archives of Microbiology* 107 (3): 283–88. https://doi.org/10.1007/BF00425340.

Eyre, David W., David Griffiths, Alison Vaughan, Tanya Golubchik, Milind Acharya, Lily O'Connor, Derrick W. Crook, A. Sarah Walker, and Tim E. A. Peto. 2013. "Asymptomatic *Clostridium Difficile* Colonisation and Onward Transmission." Edited by Yung-Fu Chang. *PLoS ONE* 8 (11): e78445. https://doi.org/10.1371/journal.pone.0078445.

Fatima, Rawish, and Muhammad Aziz. 2019. "The Hypervirulent Strain of Clostridium Difficile: NAP1/B1/027 - A Brief Overview." *Cureus* 11 (1): e3977. https://doi.org/10.7759/cureus.3977.

Fekety, R, J Silva, RA Browne, GD Rifkin, and JR Ebright. 1979. "Clindamycin-Induced Colitis." *The American Journal of Clinical Nutrition* 32 (1): 244–50.

https://doi.org/10.1093/ajcn/32.1.244.

Fillol-Salom, Alfred, Rodrigo Bacigalupe, Suzanne Humphrey, Yin Ning Chiang, John Chen, and José R. Penadés. 2021. "Lateral Transduction Is Inherent to the Life Cycle of the Archetypical *Salmonella* Phage P22." *Nature Communications* 12 (1): 6510. https://doi.org/10.1038/s41467-021-26520-4.

Fortier, Louis-Charles. 2018. "Bacteriophages Contribute to Shaping *Clostridioides (Clostridium) Difficile* Species." *Frontiers in Microbiology* 9 (August): 2033. https://doi.org/10.3389/fmicb.2018.02033.

Fortier, Louis-Charles, and Sylvain Moineau. 2007. "Morphological and Genetic Diversity of Temperate Phages in *Clostridium Difficile*." *Applied and Environmental Microbiology* 73 (22): 7358–66. https://doi.org/10.1128/AEM.00582-07.

Fortier, Louis-Charles, and Ognjen Sekulovic. 2013. "Importance of Prophages to Evolution and Virulence of Bacterial Pathogens." *Virulence* 4 (5): 354–65. https://doi.org/10.4161/viru.24498.

Frost, Laura S., Raphael Leplae, Anne O. Summers, and Ariane Toussaint. 2005. "Mobile Genetic Elements: The Agents of Open Source Evolution." *Nature Reviews Microbiology* 3 (9): 722–32. https://doi.org/10.1038/nrmicro1235.

Frost, Lucy R., Jeffrey K. J. Cheng, and Meera Unnikrishnan. 2021. "*Clostridioides Difficile* Biofilms: A Mechanism of Persistence in the Gut?" Edited by Deborah A. Hogan. *PLOS Pathogens* 17 (3): e1009348. https://doi.org/10.1371/journal.ppat.1009348.

Fulsundar, Shweta, Klaus Harms, Gøril E. Flaten, Pål J. Johnsen, Balu Ananda Chopade, and Kaare M. Nielsen. 2014. "Gene Transfer Potential of Outer Membrane Vesicles of *Acinetobacter Baylyi* and Effects of Stress on Vesiculation." Edited by M. Kivisaar. *Applied and Environmental Microbiology* 80 (11): 3469–83. https://doi.org/10.1128/AEM.04248-13.

Gerding, Dale N., Susan P. Sambol, and Stuart Johnson. 2018. "Non-Toxigenic *Clostridioides* (Formerly *Clostridium*) *Difficile* for Prevention of *C. Difficile* Infection: From Bench to Bedside Back to Bench and Back to Bedside." *Frontiers in Microbiology* 9 (July): 1700. https://doi.org/10.3389/fmicb.2018.01700.

Gerding, Dale N, Stuart Johnson, Maja Rupnik, and Klaus Aktories. 2014. "*Clostridium Difficile* Binary Toxin CDT." *Gut Microbes* 5 (1): 15–27. https://doi.org/10.4161/gmic.26854.

Gerritsen, Jacoline, Susana Fuentes, Wieke Grievink, Laura van Niftrik, Brian J. Tindall, Harro M. Timmerman, Ger T. Rijkers, and Hauke Smidt. 2014. "Characterization of

Romboutsia Ilealis Gen. Nov., Sp. Nov., Isolated from the Gastro-Intestinal Tract of a Rat, and Proposal for the Reclassification of Five Closely Related Members of the Genus *Clostridium* into the Genera *Romboutsia* ." *International Journal of Systematic and Evolutionary Microbiology* 64 (Pt 5): 1600–1616. https://doi.org/10.1099/ijs.0.059543-0.

Ghigo, Jean-Marc. 2001. "Natural Conjugative Plasmids Induce Bacterial Biofilm Development." *Nature* 412 (6845): 442–45. https://doi.org/10.1038/35086581.

Goh, Shan, Barbara J. Chang, and Thomas V. Riley. 2005. "Effect of Phage Infection on Toxin Production by Clostridium Difficile." *Journal of Medical Microbiology* 54 (2): 129–35. https://doi.org/10.1099/jmm.0.45821-0.

Goh, Shan, Haitham Hussain, Barbara J. Chang, Warren Emmett, Thomas V. Riley, and Peter Mullany. 2013. "Phage ΦC2 Mediates Transduction of Tn *6215* , Encoding Erythromycin Resistance, between *Clostridium Difficile* Strains." Edited by Andrew B. Onderdonk. *MBio* 4 (6): 1–7. https://doi.org/10.1128/mBio.00840-13.

Govind, Revathi, and Bruno Dupuy. 2012. "Secretion of *Clostridium Difficile* Toxins A and B Requires the Holin-like Protein TcdE." Edited by Ambrose Cheung. *PLoS Pathogens* 8 (6): e1002727. https://doi.org/10.1371/journal.ppat.1002727.

Govind, Revathi, Govindsamy Vediyappan, Rial D. Rolfe, Bruno Dupuy, and Joe A. Fralick. 2009. "Bacteriophage-Mediated Toxin Gene Regulation in *Clostridium Difficile*." *Journal of Virology* 83 (23): 12037–45. https://doi.org/10.1128/JVI.01256-09.

Griffiths, David, Warren Fawley, Melina Kachrimanidou, Rory Bowden, Derrick W. Crook, Rowena Fung, Tanya Golubchik, et al. 2010. "Multilocus Sequence Typing of *Clostridium Difficile*." *Journal of Clinical Microbiology* 48 (3): 770–78. https://doi.org/10.1128/JCM.01796-09.

Hall, Ivan C., and Elizabeth O'Toole. 1935. "INTESTINAL FLORA IN NEW-BORN INFANTS." *American Journal of Diseases of Children* 49 (2): 390. https://doi.org/10.1001/archpedi.1935.01970020105010.

Hargreaves, K. R., H. V. Colvin, K. V. Patel, J. J. P. Clokie, and M. R. J. Clokie. 2013. "Genetically Diverse *Clostridium Difficile* Strains Harboring Abundant Prophages in an Estuarine Environment." *Applied and Environmental Microbiology* 79 (20): 6236–43. https://doi.org/10.1128/AEM.01849-13.

Hertel, Robert, David Pintor Rodríguez, Jacqueline Hollensteiner, Sascha Dietrich, Andreas Leimbach, Michael Hoppert, Heiko Liesegang, and Sonja Volland. 2015. "Genome-Based Identification of Active Prophage Regions by Next Generation Sequencing in *Bacillus Licheniformis* DSM13." Edited by Raymond Schuch. *PLOS ONE* 10 (3):

e0120759. https://doi.org/10.1371/journal.pone.0120759.

Heuler, Joshua, Louis-Charles Fortier, and Xingmin Sun. 2021. "*Clostridioides Difficile* Phage Biology and Application." *FEMS Microbiology Reviews* 45 (5): fuab012. https://doi.org/10.1093/femsre/fuab012.

Humphrey, C D, C W Condon, J R Cantey, and F E Pittman. 1979. "Partial Purification of a Toxin Found in Hamsters with Antibiotic-Associated Colitis. Reversible Binding of the Toxin by Cholestyramine." *Gastroenterology* 76 (3): 468–76. http://www.ncbi.nlm.nih.gov/pubmed/34553.

Hunt, Jonathan J., and Jimmy D. Ballard. 2013. "Variations in Virulence and Molecular Biology among Emerging Strains of *Clostridium Difficile*." *Microbiology and Molecular Biology Reviews* 77 (4): 567–81. https://doi.org/10.1128/MMBR.00017-13.

Jackson, Sarah, Mary Calos, Andrew Myers, and William T. Self. 2006. "Analysis of Proline Reduction in the Nosocomial Pathogen *Clostridium Difficile*." *Journal of Bacteriology* 188 (24): 8487–95. https://doi.org/10.1128/JB.01370-06.

Janezic, Sandra, Mojca Potocnik, Valerija Zidaric, and Maja Rupnik. 2016. "Highly Divergent *Clostridium Difficile* Strains Isolated from the Environment." Edited by Daniel Paredes-Sabja. *PLOS ONE* 11 (11): e0167101. https://doi.org/10.1371/journal.pone.0167101.

Janezic, Sandra, Valerija Zidaric, Bart Pardon, Alexander Indra, Branko Kokotovic, Jose Luis Blanco, Christian Seyboldt, et al. 2014. "International *Clostridium Difficile* Animal Strain Collection and Large Diversity of Animal Associated Strains." *BMC Microbiology* 14 (December): 173. https://doi.org/10.1186/1471-2180-14-173.

Ji, Xue, Yang Sun, Jun Liu, Lingwei Zhu, Xuejun Guo, Xulong Lang, and Shuzhang Feng. 2016. "A Novel Virulence-Associated Protein, VapE, in Streptococcus Suis Serotype 2." *Molecular Medicine Reports* 13 (3): 2871–77. https://doi.org/10.3892/mmr.2016.4818.

Kandell, Risa L., and Carol Bernstein. 1991. "Bile Salt/Acid Induction of DNA Damage in Bacterial and Mammalian Cells: Implications for Colon Cancer." *Nutrition and Cancer* 16 (3–4): 227–38. https://doi.org/10.1080/01635589109514161.

Karasawa, T., S. Ikoma, K. Yamakawa, and S. Nakamura. 1995. "A Defined Growth Medium for *Clostridium Difficile*." *Microbiology* 141 (Pt 2): 371–75. https://doi.org/10.1099/13500872-141-2-371.

Kartalidis, Philip, Anargyros Skoulakis, Katerina Tsilipounidaki, Zoi Florou, Efthymia Petinaki, and George C. Fthenakis. 2021. "*Clostridioides Difficile* as a Dynamic Vehicle for the Dissemination of Antimicrobial-Resistance Determinants: Review and In Silico

Analysis." *Microorganisms* 9 (7): 1383. https://doi.org/10.3390/microorganisms9071383.

Kleiner, Manuel, Brian Bushnell, Kenneth E. Sanderson, Lora V. Hooper, and Breck A. Duerkop. 2020. "Transductomics: Sequencing-Based Detection and Analysis of Transduced DNA in Pure Cultures and Microbial Communities." *Microbiome* 8 (1): 158. https://doi.org/10.1186/s40168-020-00935-5.

Klindworth, Anna, Elmar Pruesse, Timmy Schweer, Jörg Peplies, Christian Quast, Matthias Horn, and Frank Oliver Glöckner. 2013. "Evaluation of General 16S Ribosomal RNA Gene PCR Primers for Classical and Next-Generation Sequencing-Based Diversity Studies." *Nucleic Acids Research* 41 (1): e1. https://doi.org/10.1093/nar/gks808.

Knight, Daniel R, Korakrit Imwattana, Brian Kullin, Enzo Guerrero-Araya, Daniel Paredes-Sabja, Xavier Didelot, Kate E Dingle, David W Eyre, César Rodríguez, and Thomas V Riley. 2021. "Major Genetic Discontinuity and Novel Toxigenic Species in *Clostridioides Difficile* Taxonomy." *ELife* 10 (June): e64325. https://doi.org/10.7554/eLife.64325.

Koonin, Eugene V, and Kira S Makarova. 2009. "CRISPR-Cas: An Adaptive Immunity System in Prokaryotes." *F1000 Biology Reports* 1 (95). https://doi.org/10.3410/B1-95.

Krogh, Susanne, Steen T. Jørgensen, and Kevin M. Devine. 1998. "Lysis Genes of the *Bacillus Subtilis* Defective Prophage PBSX." *Journal of Bacteriology* 180 (8): 2110–17. https://doi.org/10.1128/JB.180.8.2110-2117.1998.

Kropinski, Andrew M., Amanda Mazzocco, Thomas E. Waddell, Erika Lingohr, and Roger P. Johnson. 2009. "Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay." In *Methods in Molecular Biology*, 69–76. https://doi.org/10.1007/978-1-60327-164-6_7.

Kunhikannan, Shalini, Colleen J. Thomas, Ashley E. Franks, Sumana Mahadevaiah, Sumana Kumar, and Steve Petrovski. 2021. "Environmental Hotspots for Antibiotic Resistance Genes." *MicrobiologyOpen* 10 (3): e1197. https://doi.org/10.1002/mbo3.1197.

Labrie, Simon J., Julie E. Samson, and Sylvain Moineau. 2010. "Bacteriophage Resistance Mechanisms." *Nature Reviews Microbiology* 8 (5): 317–27. https://doi.org/10.1038/nrmicro2315.

Lanis, Jordi M., Latisha D. Heinlen, Judith A. James, and Jimmy D. Ballard. 2013. "*Clostridium Difficile* 027/BI/NAP1 Encodes a Hypertoxic and Antigenically Variable Form of TcdB." Edited by Kenneth A. Bradley. *PLoS Pathogens* 9 (8): e1003523. https://doi.org/10.1371/journal.ppat.1003523.

Lawson, Paul A., Diane M. Citron, Kerin L. Tyrrell, and Sydney M. Finegold. 2016. "Reclassification of *Clostridium Difficile* as *Clostridioides Difficile* (Hall and O'Toole 1935) Prévot 1938." *Anaerobe* 40 (August): 95–99. https://doi.org/10.1016/j.anaerobe.2016.06.008.

Lécuyer, Frédéric, Jean-Sébastien Bourassa, Martin Gélinas, Vincent Charron-Lamoureux, Vincent Burrus, and Pascale B. Beauregard. 2018. "Biofilm Formation Drives Transfer of the Conjugative Element ICE *Bs1* in *Bacillus Subtilis*." Edited by Craig D. Ellermeier. *MSphere* 3 (5): e00473. https://doi.org/10.1128/mSphere.00473-18.

Lemee, Ludovic, Anne Dhalluin, Sabrina Testelin, Marie-Andre Mattrat, Karine Maillard, Jean-François Lemeland, and Jean-Louis Pons. 2004. "Multiplex PCR Targeting *Tpi* (Triose Phosphate Isomerase), *TcdA* (Toxin A), and *TcdB* (Toxin B) Genes for Toxigenic Culture of *Clostridium Difficile*." *Journal of Clinical Microbiology* 42 (12): 5710–14. https://doi.org/10.1128/JCM.42.12.5710-5714.2004.

Levett, P. N. 1985. "Effect of Antibiotic Concentration in a Selective Medium on the Isolation of *Clostridium Difficile* from Faecal Specimens." *Journal of Clinical Pathology* 38 (2): 233–34. https://doi.org/10.1136/jcp.38.2.233.

Lewis, Brittany B., Rebecca A. Carter, Lilan Ling, Ingrid Leiner, Ying Taur, Mini Kamboj, Erik R. Dubberke, Joao Xavier, and Eric G. Pamer. 2017. "Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to *Clostridium Difficile* Virulence." Edited by Jimmy D. Ballard. *MBio* 8 (4): e00885-17. https://doi.org/10.1128/mBio.00885-17.

Lindsay, Jodi A., Alexey Ruzin, Hope F. Ross, Natasha Kurepina, and Richard P. Novick. 1998. "The Gene for Toxic Shock Toxin Is Carried by a Family of Mobile Pathogenicity Islands in *Staphylococcus Aureus*." *Molecular Microbiology* 29 (2): 527–43. https://doi.org/10.1046/j.1365-2958.1998.00947.x.

Loo, Vivian G., Louise Poirier, Mark A. Miller, Matthew Oughton, Michael D Libman, Sophie Michaud, Anne-marie Bourgault, et al. 2005. "A Predominantly Clonal Multi-Institutional Outbreak of *Clostridium Difficile* –Associated Diarrhea with High Morbidity and Mortality." *New England Journal of Medicine* 353 (23): 2442–49. https://doi.org/10.1056/NEJMoa051639.

Lyerly, D M, D E Lockwood, S H Richardson, and T D Wilkins. 1982. "Biological Activities of Toxins A and B of *Clostridium Difficile*." *Infection and Immunity* 35 (3): 1147–50. https://doi.org/10.1128/iai.35.3.1147-1150.1982.

Mani, Nagraj, and Bruno Dupuy. 2001. "Regulation of Toxin Synthesis in *Clostridium Difficile* by an Alternative RNA Polymerase Sigma Factor." *Proceedings of the National*

Academy of Sciences 98 (10): 5844–49. https://doi.org/10.1073/pnas.101126598.

Martínez-Meléndez, Adrián, Flora Cruz-López, Rayo Morfin-Otero, Héctor J. Maldonado-Garza, and Elvira Garza-González. 2022. "An Update on *Clostridioides Difficile* Binary Toxin." *Toxins* 14 (5): 305. https://doi.org/10.3390/toxins14050305.

Matamouros, Susana, Patrick England, and Bruno Dupuy. 2007. "*Clostridium Difficile* Toxin Expression Is Inhibited by the Novel Regulator TcdC." *Molecular Microbiology* 64 (5): 1274–88. https://doi.org/10.1111/j.1365-2958.2007.05739.x.

Mead, G. C. 1971. "The Amino Acid-Fermenting Clostridia." *Journal of General Microbiology* 67 (1): 47–56. https://doi.org/10.1099/00221287-67-1-47.

Meessen-Pinard, Mathieu, Ognjen Sekulovic, and Louis-Charles Fortier. 2012. "Evidence of *In Vivo* Prophage Induction during *Clostridium Difficile* Infection." *Applied and Environmental Microbiology* 78 (21): 7662–70. https://doi.org/10.1128/AEM.02275-12.

Monot, Marc, Catherine Eckert, Astrid Lemire, Audrey Hamiot, Thomas Dubois, Carine Tessier, Bruno Dumoulard, et al. 2015. "*Clostridium Difficile*: New Insights into the Evolution of the Pathogenicity Locus." *Scientific Reports* 5 (1): 15023. https://doi.org/10.1038/srep15023.

Mullany, Peter, Elaine Allan, and Adam P. Roberts. 2015. "Mobile Genetic Elements in *Clostridium Difficile* and Their Role in Genome Function." *Research in Microbiology* 166 (4): 361–67. https://doi.org/10.1016/j.resmic.2014.12.005.

Muto, Carlene A., Marian Pokrywka, Kathleen Shutt, Aaron B. Mendelsohn, Kathy Nouri, Kathy Posey, Terri Roberts, et al. 2005. "A Large Outbreak of *Clostridium Difficile* –Associated Disease with an Unexpected Proportion of Deaths and Colectomies at a Teaching Hospital Following Increased Fluoroquinolone Use." *Infection Control & Hospital Epidemiology* 26 (3): 273–80. https://doi.org/10.1086/502539.

Nakano, Michiko M., and Peter Zuber. 1998. "ANAEROBIC GROWTH OF A 'STRICT AEROBE' (*BACILLUS SUBTILIS*)." *Annual Review of Microbiology* 52 (1): 165–90. https://doi.org/10.1146/annurev.micro.52.1.165.

Nale, Janet Y., Jinyu Shan, Peter T. Hickenbotham, Warren N. Fawley, Mark H. Wilcox, and Martha R. J. Clokie. 2012. "Diverse Temperate Bacteriophage Carriage in *Clostridium Difficile* 027 Strains." Edited by Michel R. Popoff. *PLoS ONE* 7 (5): e37263. https://doi.org/10.1371/journal.pone.0037263.

Nale, Janet Y., Janice Spencer, Katherine R. Hargreaves, Anthony M. Buckley, Przemysław Trzepiński, Gillian R. Douce, and Martha R. J. Clokie. 2016. "Bacteriophage Combinations Significantly Reduce *Clostridium Difficile* Growth In Vitro and

Proliferation In Vivo." *Antimicrobial Agents and Chemotherapy* 60 (2): 968–81. https://doi.org/10.1128/AAC.01774-15.

Nanda, Arun M., Kai Thormann, and Julia Frunzke. 2015. "Impact of Spontaneous Prophage Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions." Edited by W. Margolin. *Journal of Bacteriology* 197 (3): 410–19. https://doi.org/10.1128/JB.02230-14.

Natarajan, Mukil, Seth T. Walk, Vincent B. Young, and David M. Aronoff. 2013. "A Clinical and Epidemiological Review of Non-Toxigenic *Clostridium Difficile*." *Anaerobe* 22 (734): 1–5. https://doi.org/10.1016/j.anaerobe.2013.05.005.

Niu, Yan D., Shaun R. Cook, Jiaying Wang, Cassidy L. Klima, Yu-hung Hsu, Andrew M. Kropinski, Dann Turner, and Tim A. McAllister. 2015. "Comparative Analysis of Multiple Inducible Phages from *Mannheimia Haemolytica*." *BMC Microbiology* 15 (1): 175. https://doi.org/10.1186/s12866-015-0494-5.

Olszak, Tomasz, Agnieszka Latka, Bartosz Roszniowski, Miguel A. Valvano, and Zuzanna Drulis-Kawa. 2017. "Phage Life Cycles Behind Bacterial Biodiversity." *Current Medicinal Chemistry* 24 (36): 3987–4001. https://doi.org/10.2174/0929867324666170413100136.

Ozaki, Eijiro, Haru Kato, Hiroyuki Kita, Tadahiro Karasawa, Tsuneo Maegawa, Youko Koino, Kazumasa Matsumoto, et al. 2004. "*Clostridium Difficile* Colonization in Healthy Adults: Transient Colonization and Correlation with Enterococcal Colonization." *Journal of Medical Microbiology* 53 (2): 167–72. https://doi.org/10.1099/jmm.0.05376-0.

Parter, Merav, Nadav Kashtan, and Uri Alon. 2007. "Environmental Variability and Modularity of Bacterial Metabolic Networks." *BMC Evolutionary Biology* 7 (December): 169. https://doi.org/10.1186/1471-2148-7-169.

Passmore, Ian J., Marine P. M. Letertre, Mark D. Preston, Irene Bianconi, Mark A. Harrison, Fauzy Nasher, Harparkash Kaur, et al. 2018. "*Para*-Cresol Production by *Clostridium Difficile* Affects Microbial Diversity and Membrane Integrity of Gram-Negative Bacteria." Edited by Theresa M. Koehler. *PLOS Pathogens* 14 (9): e1007191. https://doi.org/10.1371/journal.ppat.1007191.

Pépin, J., Nathalie Saheb, M.-A. Coulombe, M.-E. Alary, M.-P. Corriveau, Simon Authier, Michel Leblanc, et al. 2005. "Emergence of Fluoroquinolones as the Predominant Risk Factor for *Clostridium Difficile*-Associated Diarrhea: A Cohort Study during an Epidemic in Quebec." *Clinical Infectious Diseases* 41 (9): 1254–60.

https://doi.org/10.1086/496986.

Pépin, J., Louis Valiquette, Marie-Eve Alary, Philippe Villemure, Anick Pelletier, Karine Forget, Karine Pépin, and Daniel Chouinard. 2004. "*Clostridium Difficile*-Associated Diarrhea in a Region of Quebec from 1991 to 2003: A Changing Pattern of Disease Severity." *Canadian Medical Association Journal* 171 (5): 466–72. https://doi.org/10.1503/cmaj.1041104.

Perelle, Sylvie, Maryse Gibert, Pierre Bourlioux, Gerard Corthier, and Michel R. Popoff. 1997. "Production of a Complete Binary Toxin (Actin-Specific ADP-Ribosyltransferase) by *Clostridium Difficile* CD196." *Infection and Immunity* 65 (4): 1402–7. https://doi.org/10.1128/iai.65.4.1402-1407.1997.

Ramírez-Vargas, Gabriel, Shan Goh, and César Rodríguez. 2018. "The Novel Phages PhiCD5763 and PhiCD2955 Represent Two Groups of Big Plasmidial *Siphoviridae* Phages of *Clostridium Difficile*." *Frontiers in Microbiology* 9 (January): 26. https://doi.org/10.3389/fmicb.2018.00026.

Ramírez-Vargas, Gabriel, Diana López-Ureña, Adriana Badilla, Josué Orozco-Aguilar, Tatiana Murillo, Priscilla Rojas, Thomas Riedel, et al. 2018. "Novel Clade C-I *Clostridium Difficile* Strains Escape Diagnostic Tests, Differ in Pathogenicity Potential and Carry Toxins on Extrachromosomal Elements." *Scientific Reports* 8 (1): 13951. https://doi.org/10.1038/s41598-018-32390-6.

Ridlon, Jason M., Dae-Joong Kang, and Phillip B. Hylemon. 2006. "Bile Salt Biotransformations by Human Intestinal Bacteria." *Journal of Lipid Research* 47 (2): 241–59. https://doi.org/10.1194/jlr.R500013-JLR200.

Rodriguez Diaz, Cristina, Christian Seyboldt, and Maja Rupnik. 2018. "Non-Human *C. Difficile* Reservoirs and Sources: Animals, Food, Environment." In *Advances in Experimental Medicine and Biology*, 1050:227–43. https://doi.org/10.1007/978-3-319-72799-8_13.

Roseboom, Anna M., Quinten R. Ducarmon, Bastian V.H. Hornung, Céline Harmanus, Monique J.T. Crobach, Ed J. Kuijper, Rolf H.A.M. Vossen, Susan L. Kloet, and Wiep Klaas Smits. 2023. "Carriage of Three Plasmids in a Single Human Clinical Isolate of *Clostridioides Difficile*." *Plasmid* 125 (January): 102669. https://doi.org/10.1016/j.plasmid.2022.102669.

Rossen, Tessel M. van, Joffrey van Prehn, Alex Koek, Marcel Jonges, Robin van Houdt, Rosa van Mansfeld, Ed J. Kuijper, Christina M. J. E. Vandenbroucke-Grauls, and Andries E. Budding. 2021. "Simultaneous Detection and Ribotyping of *Clostridioides Difficile*, and

Toxin Gene Detection Directly on Fecal Samples." *Antimicrobial Resistance & Infection Control* 10 (1): 23. https://doi.org/10.1186/s13756-020-00881-9.

Sayers, Eric W, Evan E Bolton, J Rodney Brister, Kathi Canese, Jessica Chan, Donald C Comeau, Ryan Connor, et al. 2022. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 50 (D1): D20–26. https://doi.org/10.1093/nar/gkab1112.

Schleifer, Karl-Heinz. 2009. "Phylum XIII. Firmicutes Gibbons and Murray 1978, 5 (Firmacutes [Sic] Gibbons and Murray 1978, 5)." In *Systematic Bacteriology*, 19–1317. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-68489-5_3.

Sebaihia, Mohammed, Brendan W Wren, Peter Mullany, Neil F Fairweather, Nigel Minton, Richard Stabler, Nicholas R Thomson, et al. 2006. "The Multidrug-Resistant Human Pathogen *Clostridium Difficile* Has a Highly Mobile, Mosaic Genome." *Nature Genetics* 38 (7): 779–86. https://doi.org/10.1038/ng1830.

Seed, Kimberley D., David W. Lazinski, Stephen B. Calderwood, and Andrew Camilli. 2013. "A Bacteriophage Encodes Its Own CRISPR/Cas Adaptive Response to Evade Host Innate Immunity." *Nature* 494 (7438): 489–91. https://doi.org/10.1038/nature11927.

Shan, Jinyu, Krusha V. Patel, Peter T. Hickenbotham, Janet Y. Nale, Katherine R. Hargreaves, and Martha R.J. Clokie. 2012. "Prophage Carriage and Diversity within Clinically Relevant Strains of Clostridium Difficile." *Applied and Environmental Microbiology* 78 (17): 6027–34. https://doi.org/10.1128/AEM.01311-12.

Smits, Wiep Klaas, Dena Lyras, D. Borden Lacy, Mark H. Wilcox, and Ed J. Kuijper. 2016. "*Clostridium Difficile* Infection." *Nature Reviews Disease Primers* 2 (1): 16020. https://doi.org/10.1038/nrdp.2016.20.

Smits, Wiep Klaas, J. Scott Weese, Adam P. Roberts, Céline Harmanus, and Bastian Hornung. 2018. "A Helicase-Containing Module Defines a Family of PCD630-like Plasmids in *Clostridium Difficile*." *Anaerobe* 49 (February): 78–84. https://doi.org/10.1016/j.anaerobe.2017.12.005.

Sorg, Joseph A., and Abraham L. Sonenshein. 2008. "Bile Salts and Glycine as Cogerminants for *Clostridium Difficile* Spores." *Journal of Bacteriology* 190 (7): 2505–12. https://doi.org/10.1128/JB.01765-07.

Spigaglia, Patrizia. 2016. "Recent Advances in the Understanding of Antibiotic Resistance in *Clostridium Difficile* Infection." *Therapeutic Advances in Infectious Disease* 3 (1): 23–42. https://doi.org/10.1177/2049936115622891.

Stabler, Richard A., Miao He, Lisa Dawson, Melissa Martin, Esmeralda Valiente, Craig

Corton, Trevor D. Lawley, et al. 2009. “Comparative Genome and Phenotypic Analysis of *Clostridium Difficile* 027 Strains Provides Insight into the Evolution of a Hypervirulent Bacterium.” *Genome Biology* 10 (9): R102. https://doi.org/10.1186/gb-2009-10-9-r102.

Stojković, Vanja, María Fernanda Ulate, Fanny Hidalgo-Villeda, Emmanuel Aguilar, Camilo Monge-Cascante, Marjorie Pizarro-Guajardo, Kaitlyn Tsai, et al. 2019. “*Cfr*(B), *Cfr*(C), and a New *Cfr*-Like Gene, *Cfr*(E), in *Clostridium Difficile* Strains Recovered across Latin America.” *Antimicrobial Agents and Chemotherapy* 64 (1): e01074-19. https://doi.org/10.1128/AAC.01074-19.

Stone, Nathan E., Lindsay C. Sidak-Loftis, Jason W. Sahl, Adam J. Vazquez, Kristin B. Wiggins, John D. Gillece, Nathan D. Hicks, et al. 2016. “More than 50% of *Clostridium Difficile* Isolates from Pet Dogs in Flagstaff, USA, Carry Toxigenic Genotypes.” Edited by Hsin-Chih Lai. *PLOS ONE* 11 (10): e0164504. https://doi.org/10.1371/journal.pone.0164504.

Sullivan, N M, S Pellett, and T D Wilkins. 1982. “Purification and Characterization of Toxins A and B of *Clostridium Difficile*.” *Infection and Immunity* 35 (3): 1032–40. https://doi.org/10.1128/iai.35.3.1032-1040.1982.

Taggart, Megan G., William J. Snelling, Patrick J. Naughton, Roberto M. La Ragione, James S. G. Dooley, and Nigel G. Ternan. 2021. “Biofilm Regulation in *Clostridioides Difficile*: Novel Systems Linked to Hypervirulence.” Edited by Antje Blumenthal. *PLOS Pathogens* 17 (9): e1009817. https://doi.org/10.1371/journal.ppat.1009817.

Thanissery, Rajani, Jenessa A. Winston, and Casey M. Theriot. 2017. “Inhibition of Spore Germination, Growth, and Toxin Activity of Clinically Relevant *C. Difficile* Strains by Gut Microbiota Derived Secondary Bile Acids.” *Anaerobe* 45 (June): 86–100. https://doi.org/10.1016/j.anaerobe.2017.03.004.

Theriot, Casey M., Alison A. Bowman, and Vincent B. Young. 2016. “Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium Difficile* Spore Germination and Outgrowth in the Large Intestine.” Edited by Craig D. Ellermeier. *MSphere* 1 (1): e00045-15. https://doi.org/10.1128/mSphere.00045-15.

Thomas, Prasad, Mostafa Y. Abdel-Glil, Inga Eichhorn, Torsten Semmler, Christiane Werckenthin, Christina Baumbach, Wybke Murmann, et al. 2021. “Genome Sequence Analysis of *Clostridium Chauvoei* Strains of European Origin and Evaluation of Typing Options for Outbreak Investigations.” *Frontiers in Microbiology* 12 (September):

.732106. https://doi.org/10.3389/fmicb.2021.732106.

Tran, Frances, and James Q. Boedicker. 2017. “Genetic Cargo and Bacterial Species Set the Rate of Vesicle-Mediated Horizontal Gene Transfer.” *Scientific Reports* 7 (1): 8813. https://doi.org/10.1038/s41598-017-07447-7.

Tulli, Lorenza, Sara Marchi, Roberto Petracca, Helen Alexandra Shaw, Neil F. Fairweather, Maria Scarselli, Marco Soriani, and Rosanna Leuzzi. 2013. “CbpA: A Novel Surface Exposed Adhesin of *Clostridium Difficile* Targeting Human Collagen.” *Cellular Microbiology* 15 (10): 1674–87. https://doi.org/10.1111/cmi.12139.

Valiente, E., M.D. Cairns, and B.W. Wren. 2014. “The *Clostridium Difficile* PCR Ribotype 027 Lineage: A Pathogen on the Move.” *Clinical Microbiology and Infection* 20 (5): 396–404. https://doi.org/10.1111/1469-0691.12619.

Wan, Xuchun, Qianqian Yang, Xiangfeng Wang, Yun Bai, and Zhi Liu. 2023. “Isolation and Cultivation of Human Gut Microorganisms: A Review.” *Microorganisms* 11 (4): 1080. https://doi.org/10.3390/microorganisms11041080.

Weese, J. Scott. 2020. “*Clostridium* (*Clostridioides*) *Difficile* in Animals.” *Journal of Veterinary Diagnostic Investigation* 32 (2): 213–21. https://doi.org/10.1177/1040638719899081.

Weinbauer, Markus G. 2004. “Ecology of Prokaryotic Viruses.” *FEMS Microbiology Reviews* 28 (2): 127–81. https://doi.org/10.1016/j.femsre.2003.08.001.

Whittle, M. J., T. W. Bilverstone, R. J. van Esveld, A. C. Lücke, M. M. Lister, S. A. Kuehne, and N. P. Minton. 2022. “A Novel Bacteriophage with Broad Host Range against *Clostridioides Difficile* Ribotype 078 Supports SlpA as the Likely Phage Receptor.” Edited by Jennifer M. Auchtung. *Microbiology Spectrum* 10 (1): e0229521. https://doi.org/10.1128/spectrum.02295-21.

Williamson, Charles Hall Davis, Nathan E. Stone, Amalee E. Nunnally, Chandler C. Roe, Adam J. Vazquez, Samantha A. Lucero, Heidie Hornstra, et al. 2022. “Identification of Novel, Cryptic *Clostridioides* Species Isolates from Environmental Samples Collected from Diverse Geographical Locations.” *Microbial Genomics* 8 (2): 000742. https://doi.org/10.1099/mgen.0.000742.

Yamakawa, K., T. Karasawa, S. Ikoma, and S. Nakamura. 1996. “Enhancement of Clostridium Difficile Toxin Production in Biotin-Limited Conditions.” *Journal of Medical Microbiology* 44 (2): 111–14. https://doi.org/10.1099/00222615-44-2-111.

Zhou, Qingshuai, Fengqin Rao, Zhenghong Chen, Yumei Cheng, Qifang Zhang, Jie Zhang, Zhizhong Guan, et al. 2022. “The *Cwp66* Gene Affects Cell Adhesion, Stress Tolerance,

and Antibiotic Resistance in *Clostridioides Difficile*." Edited by Cezar M. Khursigara. *Microbiology Spectrum* 10 (2): e0270421. https://doi.org/10.1128/spectrum.02704-21.

Zhou, Yanzi, Wangxiao Zhou, Tingting Xiao, Yunbo Chen, Tao Lv, Yuan Wang, Shuntian Zhang, et al. 2021. "Comparative Genomic and Transmission Analysis of *Clostridioides Difficile* between Environmental, Animal, and Clinical Sources in China." *Emerging Microbes & Infections* 10 (1): 2244–55. https://doi.org/10.1080/22221751.2021.2005453.

8 APPENDIX

8.1 Supplements

Table S1. All 43 *C. difficile* phage genomes available on NCBI, on 29.08.2023.

GenBank accession	Phage name
CP067347.1	*Clostridioides* phage ES-S-0107-01
CP067352.1	*Clostridioides* phage ES-S-0173-01
CP069347.1	*Clostridioides* phage pCD5401_3
CP069348.1	*Clostridioides* phage pCD1602_4
MT193276.1	*Clostridioides* phage JD033
MW512570.1	*Clostridioides* phage CD1801
MW512571.1	*Clostridioides* phage CD2301
MW512572.1	*Clostridioides* phage phiCD08011
MW512573.1	*Clostridioides* phage phiCD418
OQ703261.1	*Clostridioides* phage AR1075-1
AY855346.1	*Clostridium* phage phi CD119
CP011970.1	*Peptoclostridium* phage phiCDIF1296T
CP103805.1	*Clostridioides difficile* strain plasmid unnamed1
CP103806.1	*Clostridioides* phage Hain-Saunders-2022a
CP103976.1	*Clostridioides* phage HGP05
DQ466086.1	*Clostridioides* phage phiC2
HG796225.1	*Clostridium* phage phiCDHM13
HG798901.1	*Clostridium* phage phiCDHM11
HM568888.1	*Clostridium* phage phiCD38-2
JX145341.1	*Clostridium* phage phiMMP02
JX145342.1	*Clostridium* phage phiMMP04
KU057941.1	*Clostridium* phage CDSH1
KX228399.1	*Clostridium* phage CDKM9
KX228400.1	*Clostridium* phage CDKM15
KX905163.1	*Clostridioides* phage phiSemix9P1
LK985321.1	*Clostridium* phage phiCDHM14

Table S1 continued.

GenBank accession	Phage name
LK985322.1	*Clostridium* phage phiCDHM19
LN681534.1	*Clostridium* phage phiCD24-1
LN681535.1	*Clostridium* phage phiCD111
LN681536.1	*Clostridium* phage phiCD146
LN681538.1	*Clostridium* phage phiCD481-1
LN681540.1	*Clostridium* phage phiCD506
LN681541.1	*Clostridium* phage phiMMP01
LN681542.1	*Clostridium* phage phiMMP03
MF547662.1	*Clostridioides* phage LIBA6276
MF547663.1	*Clostridioides* phage LIBA2945
MK473382.1	*Clostridium* phage JD032
MN718463.1	*Clostridium* phage phiCDKH01
NC_011398.1	*Clostridioides* phage phiCD27
NC_015262.1	*Clostridium* phage phiCD6356
NC_024144.1	*Clostridium* phage CDMH1
NC_028764.1	*Clostridium* phage phiCD505
NC_029048.2	*Clostridium* phage phiCD211

8.2 Publications & conference poster

Poehlein, A., Funkner, K., **Schüler, M. A.**, & Daniel, R. (2017). First Insights into the Genome Sequence of the Cellulolytic Bacterium *Clostridium hungatei* DSM 14427. *Genome Announcements*, 5(20), e00363-17. https://doi.org/10.1128/genomeA.00363-17

Schüler, Miriam A, Stegmann, B. A., Poehlein, A., Daniel, R., & Dürre, P. (2020). Genome sequence analysis of the temperate bacteriophage TBP2 of the solvent producer *Clostridium saccharoperbutylacetonicum* N1-4 (HMT, ATCC 27021). *FEMS Microbiology Letters*, 367(14), fnaa103. https://doi.org/10.1093/femsle/fnaa103

Schüler, Miriam Antonia, Schneider, D., Poehlein, A., & Daniel, R. (2023). Direct and culture-independent detection of low-abundant *Clostridioides difficile* in environmental DNA via PCR. *BioRxiv*, 1–36. https://doi.org/10.1016/j.tim.2018.09.004

Schüler, Miriam Antonia, Daniel, R., & Poehlein, A. (2023). Novel insights into phage biology of the pathogen *Clostridioides difficile* based on the active virome. *BioRxiv*, 1–45. https://doi.org/10.1101/2023.09.27.559748.

Schüler, Miriam Antonia, Daniel, R., & Poehlein, A. (2022). Differential prophage activity within *Clostridioides difficile* isolates. *VAAM*, online. Poster.

9 ACKNOWLEDGEMENTS

I am very grateful for all the support and love that accompanied me through my entire Ph.D. I have learned so many things, not only in the scientific way but also about myself. Thank you to everybody who was a part of this journey.

I thank Prof. Dr. Rolf Daniel, who made my Ph.D. possible in the first place by giving me the project position. You were also my first examiner throughout this adventure, and already allowed me to do my master thesis in your department.

In this sense, I also want to thank PD Dr. Michael Hoppert, who was my second examiner not only for my Ph.D. but already for my master thesis. I always appreciated your enthusiasm in all topics of microbiology.

Of course, I further thank the other members of my examination board, Prof. Dr. Gerhard Braus, Prof. Dr. Stefanie Pöggeler, Prof. Dr. Kai Heimel, and Prof. Dr. Jan de Vries, who contribute significantly to my Ph.D. by reading and evaluating my thesis and the upcoming defense.

Another important companion of my Ph.D. was my supervisor Dr. Anja Poehlein, who introduced me to the fascinating world of anaerobes. Thank you for your commitment, especially on the final days. You always advocate for your students and try to make things possible.

I am also very grateful for the technical assistance and expertise that I obtained from our marvelous TAs. Melanie Heinemann, thank you for always helping when something was urgent and even working overtime. Mechthild Bömeke, thank you for everything I have learned from you, the lab would not be the same without your organization. Thanks go also to Christiane Wohlfeil, who always helped me and us with our plethora of culture flasks, despite the awful smell of my cultures. At this point, I also want to thank Sarah Schüssler, who I met as a TA in our department and is now a very precious friend.

I big thank you goes to my heroes in our institute workshop. Dennis Stromburg und Patrick Regin, you always me helped with diverse technical problems, and cheered me up and made me laugh whenever I needed it. Olaf Waase and Jarek Sobkowiak also belong to my workshop heroes, who bailed me out several times.

Not all heroes work in the workshop, and my next thanks go to all the IT guys who solved my computer problems and fulfilled my software wishes if possible. Thank you to Tarek Morsi

and Marc Dornieden, who left long time ago but still belong to our department for me. Another thanks to Mirco Stolte and Paul Westermann, who left not so long ago and fortunately come over from time to time for a little chat and IT support. Thank you to Johannes Wittenberg and Julian Rack, who did their best to support me in the final Ph.D. phase of extensive bioinformatics work and problems. Though they just started in our department, I also thank Sebastian Grosmann and Jim for their IT support. Of course, I also thank the IT chef PD Dr. Heiko Liesegang.

I am also very grateful for the expertise and support from all the former and current Postdocs, Dr. Heiko Nacke, Dr. Birgit Pfeiffer, and especially Dr. Robert Hertel, Dr. Dominik Schneider, and Dr. Jacqueline Hollensteiner. You always had an open ear whenever I needed it. I also want to highlight Jacky's commitment for the well-being of everybody in the department.

Everything is easier to bear with a good companionship, so I want to thank all my Ph.D. colleagues over the last years. Thanks to those who have already left, Richard Egelkamp, Genis Castilllo, Dirk Berkelmann, Tatiana Murillo, Amélie Dukunde, Rahadian Pratama, Inka Willms, Cynthia Chibani, and Avril von Hoyningen-Huene. And thanks to those, who are still here, Stefani Diaz, Alina Lüschen, Ines Friedrich, and Tim Böer. Tim and Ines deserve special thanks for their support, both in the lab and mentally.

Special thanks go also to my friends, who had to be very patient this year and were always supportive and understanding. I have missed a lot, and I am looking forward to making up for this now.

Finally, I cannot put into words how grateful I am for the unconditional love and support from my family, especially from my parents and my significant other Björn. You helped me to relax from time to time in the last months of constant and hard work. You accompanied in joy and sorrow, and I could and can always count on you. Thank you so much!