



Comparative Genome Analysis and Spore Heat Resistance Assay Reveal a New Component to Population Structure and Genome Epidemiology Within *Clostridium perfringens* Enterotoxin-Carrying Isolates

Kaisa Jaakkola^{1*}, Kira Virtanen^{2,3}, Päivi Lahti⁴, Riikka Keto-Timonen¹, Miia Lindström¹ and Hannu Korkeala¹

OPEN ACCESS

Edited by:

Maria Elena Martino,
University of Padua, Italy

Reviewed by:

Jihong Li,
University of Pittsburgh, United States
Marjorie Pizarro-Guajardo,
Texas A&M University, United States

*Correspondence:

Kaisa Jaakkola
kaisa.w.jaakkola@helsinki.fi

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 30 May 2021

Accepted: 10 August 2021

Published: 08 September 2021

Citation:

Jaakkola K, Virtanen K, Lahti P,
Keto-Timonen R, Lindström M and
Korkeala H (2021) Comparative
Genome Analysis and Spore Heat
Resistance Assay Reveal a New
Component to Population Structure
and Genome Epidemiology Within
Clostridium perfringens
Enterotoxin-Carrying Isolates.
Front. Microbiol. 12:717176.
doi: 10.3389/fmicb.2021.717176

¹ Department of Food Hygiene and Environmental Health, University of Helsinki, Helsinki, Finland, ² Department of Bacteriology and Immunology, Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland, ³ Northern Finland Laboratory Centre NordLab, Oulu, Finland, ⁴ City of Helsinki, Unit of Environmental Services, Helsinki, Finland

Clostridium perfringens causes a variety of human and animal enteric diseases including food poisoning, antibiotic-associated diarrhea, and necrotic enteritis. Yet, the reservoirs of enteropathogenic enterotoxin-producing strains remain unknown. We conducted a genomic comparison of 290 strains and a heat resistance phenotyping of 30 *C. perfringens* strains to elucidate the population structure and ecology of this pathogen. *C. perfringens* genomes shared a conserved genetic backbone with more than half of the genes of an average genome conserved in >95% of strains. The *cpe*-carrying isolates were found to share genetic context: the *cpe*-carrying plasmids had different distribution patterns within the genetic lineages and the estimated pan genome of *cpe*-carrying isolates had a larger core genome and a smaller accessory genome compared to that of 290 strains. We characterize *cpe*-negative strains related to chromosomal *cpe*-carrying strains elucidating the origin of these strains and disclose two distinct groups of chromosomal *cpe*-carrying strains with different virulence characteristics, spore heat resistance properties, and, presumably, ecological niche. Finally, an antibiotic-associated diarrhea isolate carrying two copies of the enterotoxin *cpe* gene and the associated genetic lineage with the potential for the emergence of similar strains are outlined. With *C. perfringens* as an example, implications of input genome quality for pan genome analysis are discussed. Our study furthers the understanding of genome epidemiology and population structure of enteropathogenic *C. perfringens* and brings new insight into this important pathogen and its reservoirs.

Keywords: *Clostridium perfringens*, core genome, enterotoxin (CPE), spore heat resistance, gastrointestinal disease, comparative genome analysis

INTRODUCTION

Clostridium perfringens is an anaerobic, spore former causing gas gangrene, wound infections, and a variety of human and animal diseases involving the gastrointestinal (GI) system (Kiu and Hall, 2018). Based on the recent estimation, this pathogen originated 40–80 000 years ago, and strains similar to the current isolates have been isolated from a 12 000-year-old mummified puppy and a 5,000-year-old mummified human corpse (Lugli et al., 2017; Feng et al., 2020). Presently, *C. perfringens* is a common human enteric pathogen: approximately 15% of reported antibiotic-associated diarrhea (AAD) cases are caused by *C. perfringens* (Asha et al., 2006), and the amount of *C. perfringens* food-poisoning cases in the European Union has been projected to be between 850 000 and 5 million cases per year (Silva et al., 2015).

Genome sequencing and comparative analysis have shown that *C. perfringens* has a large accessory genome reflecting the wide variety of caused diseases and produced toxins (Petit et al., 1999; Rood et al., 2018). Recently, five phylogenetic lineages (I–V) of *C. perfringens* have been identified (Kiu et al., 2017; Feng et al., 2020; Abdel-Glil et al., 2021). There is currently no consensus on naming these lineages, and different nomenclature has been used in previous studies. Feng et al. (2020) estimated that the core genome (genes present in 95% of genomes) shared by these five lineages comprised only one-third of the full genome of each lineage, emphasizing the size and diversity of this species' accessory genomes.

Identified genetic lineages are not yet established in use for the classification of *C. perfringens* strains. Instead, five toxins are used to toxinotype strains to types A–G (Rood et al., 2018). Strains of different toxinotypes are associated with different host preferences and cause different diseases: for example, toxinotype B strains are associated with sheep dysentery and toxinotype G strains carrying NetB toxin with necrotic enteritis in chickens (Rood et al., 2018). Both food-poisoning and AAD cases of *C. perfringens* diarrhea are primarily caused by type F strains (previously *cpe*-carrying type A strains), which produce pore-forming *C. perfringens* enterotoxin (CPE) but not β -toxin, ϵ -toxin, or ι -toxin (Rood et al., 2018). The sole chromosomal gene in the toxinotyping scheme is alpha (*plc*), present in all *C. perfringens* strains, while the other toxinotyping toxins are carried on transposable elements or a family of conjugative plasmids (Li et al., 2013). Toxinotyping, therefore, does not reflect the phylogenetic lineage of strains. Recently, a virulence gene profile scheme (I–XV), including chromosomal genes for typing of foodborne isolates (Abdelrahim et al., 2019), has been proposed.

The current understanding is that all *C. perfringens* types can carry the *cpe* gene and produce CPE, but that only approximately 5% of the strains are *cpe*-positive (Miyamoto et al., 2006). The 319 aa *cpe* gene is located either in pCPF5603 or pCPF4969 plasmid (plasmid-mediated *cpe*, p-*cpe* strains) or on a transposable element Tn5565 integrated into the chromosome (chromosomal *cpe*, c-*cpe* strains) (Cornillot et al., 1995; Brynestad et al., 1997; Miyamoto et al., 2006; Li et al., 2010). Foodborne outbreaks are caused by both p-*cpe*

and c-*cpe* strains (Lahti et al., 2008). Additionally, a 325 aa variant of the *cpe* gene of unknown clinical relevance has been described in the pCPBB1 plasmid (Miyamoto et al., 2011). Known *cpe*-carrying plasmids are all conjugative and horizontally transferable (Miyamoto et al., 2006).

The current typing practices do not reveal the genetic lineage, and likewise, the source of enteric *C. perfringens* infection is often not identified. Transmission routes and ecology of *C. perfringens* are poorly understood, and the reservoirs of enteropathogenic strains remain unknown. *C. perfringens* is known as an environmental bacteria associated with soil, water, sewage, and dust, but also the GI tract of humans and animals (Hatheway, 1990). Suggested reservoirs for *cpe*-positive strains include healthy animals and humans, sludge, soil, and retail meat (Wen and McClane, 2004; Heikinheimo et al., 2006; Li et al., 2007; Lindström et al., 2011; Li et al., 2013; Lahti et al., 2012; Hu et al., 2018). C-*cpe* strains have been solely isolated from food items or food poisoning associated samples, while p-*cpe* strains are often associated with non-food-borne human disease and animal enteric disease (Collie and McClane, 1998; Sparks et al., 2001). Comparative genome analysis results have suggested that c-*cpe* strains are clonal (Deguchi et al., 2009; Xiao et al., 2012) and specialized to yet unknown environmental niche. The GI tract has been suggested to be the adapted niche for p-*cpe* strains (Lahti et al., 2012). Particularly, pCPF5603 plasmid-carrying strains are associated with care home outbreaks and AAD (Miyamoto et al., 2006; Kiu et al., 2019). A recent study on human enteric isolates suggested that *C. perfringens* strains from a single source might also persist and cause longitudinal outbreaks spanning several years (Kiu et al., 2019), supporting the hypothesis of healthy humans serving as reservoirs for *cpe*-carrying strains (Heikinheimo et al., 2006).

Some *C. perfringens* strains produce heat-resistant spores which enable them to proliferate in foods after cooking if the cold chain is compromised (Sarker et al., 2000; Li and McClane, 2006b; Lindström et al., 2011). The heat resistance of *C. perfringens* spores has been studied widely (Weiss and Strong, 1967; Ando et al., 1985; Sarker et al., 2000; Raju and Sarker, 2005; Li and McClane, 2006a,b, 2008; Xiao et al., 2015), and the main finding is that c-*cpe* strains produce heat-resistant spores with the $D_{99^{\circ}\text{C}}$ value average of 53.2 min (Weiss and Strong, 1967; Sarker et al., 2000). The known key mechanism for this spore heat resistance phenotype is the ability of strains to produce a variant of small, acid-soluble protein (Ssp4), which binds strongly to the spore DNA to protect it (Li et al., 2009). Contrarily, p-*cpe* strains and *cpe* negative strains produce heat-sensitive spores with low $D_{99^{\circ}\text{C}}$ values (on average 1.0 min) (Sarker et al., 2000). The link between genotype and phenotype has not been studied previously, and only three strains with known D values have been sequenced. The growth temperatures of different strains have not been widely studied.

To elucidate the ecology and reservoirs of *cpe*-carrying *C. perfringens* strains, we determined the heat resistance phenotypes and growth temperatures for 30 sequenced *C. perfringens* strains. To understand the population structure and genetic context of this pathogen, the pangenome, cgMLST typing, and phylogenetic tree of 290 *C. perfringens* strains

were created. Our results suggest that the presence of the *cpe* gene is associated with certain genetic backgrounds instead of just horizontally transferable enterotoxin. We also describe two distinct groups of chromosomal *cpe*-carrying strains with different virulence characteristics, spore heat resistance properties, and, presumably, ecological niche. The importance of input genome quality for pangenome analysis is discussed as a larger core genome, and a smaller pangenome of *C. perfringens* compared to recent estimations is presented.

MATERIALS AND METHODS

Bacterial Strains and Genomes

A total of 30 *C. perfringens* strains isolated from various sources during 1984–2007 were sequenced and used for heat resistance and growth temperature assays (**Supplementary Table 1**). The strains included food poisoning associated isolates from foods ($n = 8$) or feces ($n = 6$); human cases of antibiotic-associated diarrhea ($n = 3$); feces of healthy people ($n = 8$); and feces of healthy production animals ($n = 2$), soil ($n = 1$), and sludge ($n = 2$). The strains, their isolation source, and type of *cpe* locus have been previously reported (Heikinheimo et al., 2006; Lahti et al., 2008, 2012). Twenty-seven isolates carried the *cpe* gene, and three were *cpe*-negative. The *cpe* gene was present in the chromosome of 11 strains (c-*cpe*) and a plasmid in 16 strains (p-*cpe*). The *cpe* genetic location on p-*cpe* strains was either on pCPF4969/IS1470-like ($n = 9$) or pCPF5603/IS1151 ($n = 8$) plasmid (Heikinheimo et al., 2006; Lahti et al., 2008, 2012).

The *cpe*-negative *C. perfringens* strains ATCC 13124 and str. 13, and c-*cpe* strain SM101 were used as reference genomes (Shimizu et al., 2002; Myers et al., 2006). For phylogenetic and pangenome analyses also publicly available *C. perfringens* genomes ($n = 260$), representing all described genetic lineages, and including 56 *cpe*-carrying isolates, were included (**Supplementary Table 2**). Genomes were downloaded from the Patric database (Gillespie et al., 2011).

Genome Sequencing and Annotation

Genomic DNA of 30 *C. perfringens* strains was extracted (Keto-Timonen et al., 2006), and whole-genome sequencing was performed using PacBio RSII (Institute of Biotechnology, Helsinki, Finland). Sequenced genomes were assembled using HGAP3 and checked for circularity using Gap4 (Staden et al., 2003; Chin et al., 2013). To improve the draft assembly, Illumina MiSeq reads and Pilon tool were used for genome polishing (Walker et al., 2014). Sequenced *C. perfringens* genomes were deposited in the GenBank (accession numbers listed in **Supplementary Table 3**). Both sequenced and downloaded genomes were annotated using Prokka (Seemann, 2014). Protein functional annotation and conserved domain predictions were refined using CD-search and CDD (Li and Godzik, 2006; Marchler-Bauer et al., 2009).

cgMLST

To determine the cgMLST target gene set and to create a genome-wide gene-by-gene comparison, ChewBBACA was used

(Silva et al., 2018). Both custom-made schema and schema created by Abdel-Glil (2019) available at <https://www.cgmlst.org> (accessed 10.01.2021) were used. The custom-made schema was created using closed *C. perfringens* genome ATCC 13124 as a training genome and selected genomes as initial query (**Supplementary Table 1**). Orthologous coding sequences (CDSs) were called using blast score ratio equal to or greater than 0.6 as the threshold. Paralogs, genes shorter than 200 bp, and genes without start or stop codons were excluded, and only the longest allele was kept in the list. Listed genes were used to call alleles from query genomes (**Supplementary Table 2**). In both called schemes, a threshold of 95 was selected for further analysis (9 genomes removed). Called alleles were once again pruned for paralogs, and the cgMLST gene list was created by filtering out alleles present in less than 95% of genomes.

Pangenome and Comparative Genome Analyses

For the core genome and pangenome analysis, Roary (Page et al., 2015) and Panaroo (Tonkin-Hill et al., 2020) pipelines were used. The genomes were submitted to the Roary pangenome pipeline v3.6.1 using parameters to split paralogs and 95% minimum percentage identity for blastp. The Panaroo pipeline was run in strict mode and using the default parameter to split paralogs. Similar parameters were used for the creation of pangenomes for 86 *cpe*-carrying strains and 63 lineage IV strains. The core genome clusters were functionally annotated with EggNog 5.0 (Huerta-Cepas et al., 2019). Scoary was used to associate gene clusters with observed phenotypes (Brynildsrud et al., 2016).

Initially, the Roary pipeline produced a noticeably small core genome, and to produce comparable results with the Panaroo pipeline, a set of quality criteria for input genomes was applied. Genomes were required to meet two or more quality indicators to be included in pangenome analysis (Parrello et al., 2019). Used quality criteria were (1) Patric Genome Quality coarse consistency over 97.0%; (2) Patric Genome Quality fine consistency over 92.0%; (3) less than 100 alleles over 20% shorter than reference sequence in Abdel-Glil cgMLST scheme (**Supplementary Tables 2, 10**). In total, seven genomes (BER-NE33, W1319, PBS5, PBD1, PC5, K473, and T3381) were excluded from the Roary analysis of 283 strains based on these criteria.

Additional genome comparisons were performed using bi-directional BLASTP with the standard scoring matrix BLOSUM62 and an initial E-value cut-off of $1e^{-05}$. Two genes were acknowledged orthologs if a reciprocal best blast hit existed among them, and both hits had over 90% similarity and 80% coverage. On average identified bi-directional best hits between any two strains shared over 95% identity (data not shown). When manual curation revealed putative orthologs with conserved synteny but sequence identity below 80%, a term divergent ortholog is used. Genomes were aligned and visually compared with progressive Mauve (Darling et al., 2010) and SEED Viewer 2.0 (Overbeek et al., 2014). Comparison tools available at Patric database (Gillespie et al., 2011; Wattam et al., 2016) were also used for manual curation.

Phylogenetic Analysis

Maximum likelihood phylogenetic trees were constructed using fasttree 2 (Price et al., 2010). Minimal spanning tree was calculated with Grapetree (Zhou et al., 2018) using MSTreeV3. Phylogenetic trees based on Roary calculated core genomes and cgMLST were visualized and annotated with Grapetree and iTOL v3¹.

Heat-Resistance Assay for Spores and Vegetative Cells

The heat-resistance phenotype and estimated D values of 30 *C. perfringens* spores and vegetative cells were determined according to established procedures (Weiss and Strong, 1967; Ando et al., 1985; Sarker et al., 2000). All strains were prepared as described by Duncan and Strong (1968) and Ando et al. (1985). Briefly, Duncan-Strong (DS) medium cultures grown for 24 h at 37°C were heat-shocked at 75°C for 15 min to kill the vegetative cells and facilitate spore germination. A 100 µl sample was withdrawn, diluted, and plated to determine the initial spore count. The remainder of each heat-shocked DS medium culture was then heated at either 89 or 99°C for 1 min to 4 h or until the spore count reached 0 (depending on the individual isolate and the temperature being used). Sampling occurred every 30 s for 0–5 min, then every 60 s until 10 min, and 11–240 min every 30 min. At each time point, culture was mixed and a 100 µl time point sample was withdrawn and diluted (dilution range, 10⁻² to 10⁻⁷). For vegetative cells, FTG (fluid thioglycolate) medium cultures grown for 24 h at 37°C were heated to 60°C for a time ranging from 1 min to 2 h. The time point samples were withdrawn as explained above. Finally, to determine the number of viable spores or cells present per milliliter (CFU/ml), the dilutions were plated on BHI. The logarithmic counts of viable spores and cells of every isolate were graphed against the heating time to determine the slope of the survival curve. The estimation of the D value was determined from this curve. P values were calculated using ANOVA analysis of variance (IBM SPSS Statistics for Macintosh, Version 27.0). Strains (955/85, CPI 57K-1, C269) did not produce spores with the methods used here.

Maximum and Minimum Growth Temperatures

The growth temperatures of 30 *C. perfringens* strains were examined using the Gradiplate W10 incubator (Biodata Oy, Helsinki, Finland) placed in an anaerobic workstation (MK III, Don Whitley Scientific, Ltd., Shipley, United Kingdom) to determine the minimum and maximum growth temperatures (Korkeala et al., 1990; Derman et al., 2011). Briefly, strains were cultured in FTG at 37°C for 24 h. Strains were refreshed and grown at 37°C until turbidity OD₆₀₀ reached 0.6–0.9. At least two biological replicates were cultured and analyzed. Cultures were diluted 1:10 in peptone water, and a sample of 25 microliters was transferred to a Gradiplate cuvette. Tested temperature gradients for minimum and maximum growth temperatures were

8–18°C and 47–57°C, respectively. Strains were incubated for 2 days for maximum temperatures and 21 days for minimum temperatures. Growth boundaries were determined by using a stereomicroscope, and the growth temperature threshold was determined as the boundary where dense bacterial growth was discontinued. Minimum and maximum growth temperatures were calculated as described by Korkeala et al. (1990) and Derman et al. (2011) using the formula $T = T_{low} + d * g$, where d is the distance (in millimeters) of the growth boundary to the measurement point of T_{low} and g is the temperature gradient.

RESULTS

Phylogenetic Analysis Revealed Dispersal Patterns of *cpe*-Plasmids

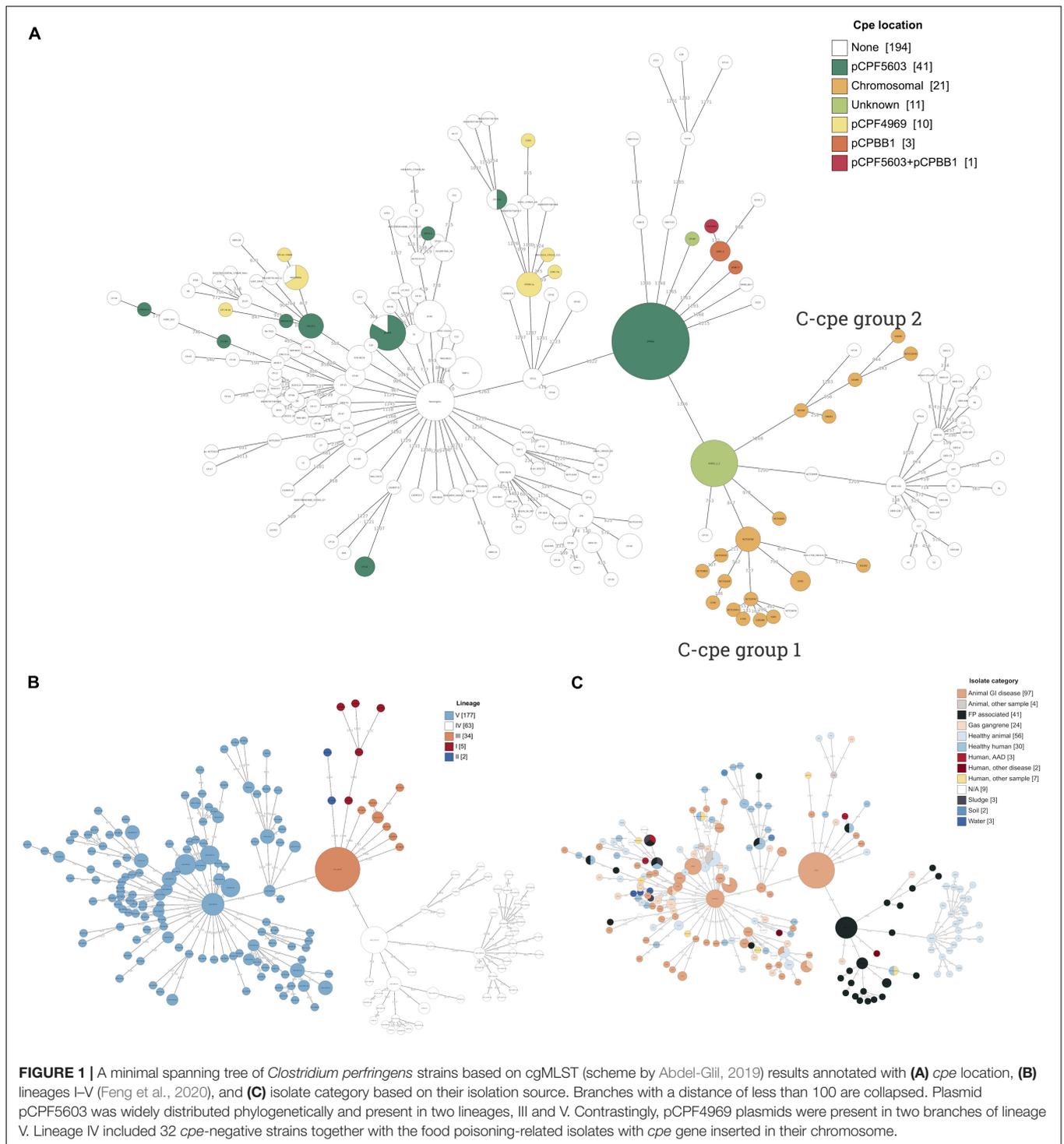
Analyzed strains (*n* = 290) represented all five genetic lineages (I–V) and were isolated from a variety of animals as well as from humans and the environment (Supplementary Tables 1, 2). Here, lineages are numbered based on tree topology and are consistent with that used by Feng et al. (2020).

Three of the genetic lineages were well presented in sequenced strains: III (34 strains), IV (63 strains), and V (183 strains), and all *cpe*-carrying isolates belonged to these three lineages (Table 1). The majority of sequenced genomes (56.9%) were clinical isolates associated with veterinary (126/290) or human disease (39/290), and only 2.8% of genomes (eight, including the three environmental samples sequenced in this study) had been isolated from an environmental source such as soil, water, and sludge (Figure 1). Veterinary clinical isolates were scattered across all lineages, while all environmental isolates belonged to lineage V. Considering how common the AAD caused by *C. perfringens* is, it is surprising assemblies or closed genomes of AAD strains have not been published previously. In this study, three AAD isolates were included. The *c-cpe* genomes and one *cpe*-negative strain sequenced in this study belonged to lineage IV (*n* = 11), one AAD isolate to lineage III, and the rest to lineage V (*n* = 18). The minimal spanning tree of *C. perfringens* strains and their *cpe* type and isolation source based on cgMLST results are represented in Figure 1.

TABLE 1 | Distribution of gastrointestinal (GI) disease isolates of human or animal origin and *cpe*-carrying isolates within lineages I–V.

Lineage	No. of genomes/ isolates	% of isolates associated with GI disease	% of <i>cpe</i> -carrying genomes	<i>cpe</i> types (no. of genomes)
I	8	0	0	None (0)
II	2	50	0	None (0)
III	34	88	91	III/pCPPB1 (4), III/pCPF5603 (27)
IV	63	41	49	Chromosomally inserted <i>cpe</i> (21), unconfirmed <i>cpe</i> location on a draft genome (10)
V	183	44	14	V/pCPF4969 (10), V/pCPF5603 (15)

¹<http://itol.embl.de/>



C-cpe strains ($n = 31$) all belonged to lineage IV. Phylogenetic analysis further revealed two subgroups of c-cpe strains (from now on referred to as the c-cpe group 1 and 2) and a closely related subgroup of *cpe*-negative isolates. In total, lineage IV included 32 (51%) *cpe*-negative isolates isolated from a variety of sources and geographic locations. These included 26 isolates from healthy swine and chicken (including the here sequenced

European swine isolate 2a), stool of a healthy human (North America), two cases of gas gangrene in animals (Asia), and two food poisoning isolates (Europe). All c-cpe genomes had been isolated from food or stool samples linked to food poisonings in Europe.

Contrastingly, p-cpe strains ($n = 52$) belonged to lineages III and V and had been isolated from Europe, Asia, and North

America. **Figure 1A** shows the dispersal patterns of *cpe*-carrying plasmids (pCPF5603, pCP4969, and pCPBB1) within genetic lineages. Plasmid pCPF5603 ($n = 42$) was the most widely distributed and present in both lineages III and V. Plasmid pCP4969 ($n = 10$) was present in two branches of lineage V, which were phylogenetically related with each other's and with lineage IV (**Supplementary Figure 1**). All strains carrying pCPBB1 plasmid ($n = 4$) with a variant of *cpe* toxin clustered together within lineage III (**Table 1** and **Figure 1A**). The presence of toxin plasmids in other sequenced bacteria was looked up, but all three toxin plasmids were only found from *C. perfringens* genomes.

Core Genome and Pangenome of *C. perfringens*

The core genome of the 290 strains consisted of, depending on the clustering method used, 1170–1660 coding sequences (CDSs), and the pangenome amounted to 14 306–23 148 CDSs (**Table 2**). The pangenome was open, and after 270 genomes had been

added, the pangenome still expanded at a pace of approximately 10 new genes per added genome. The number of unique CDSs per genome varied from 133 (CP-21) to 0 (several genomes). The core genome presented 7–12% of the pangenome and 38–53% of the CDSs of an average genome (3120 CDSs).

The pangenomes of *cpe*-carrying strains ($n = 86$) and 290 strains were compared (**Figure 2** and **Table 2**). For the *cpe*-carrying strains, the core genome covered 51–62% (1577–1941/3120) of the average genome CDSs and the pangenome amounted to 7721 CDSs. The core genome of *cpe*-carrying genomes was 17% (core genome present in 99% of strains) or 88% (core genome present in 100% of strains) larger compared to the core genome of all strains. The accessory genome of analyzed *cpe*-carrying genomes was also remarkably constrained (5780 CDSs) compared to that of all strains (12 646 CDSs). The most frequent gene cluster of the *cpe*-carrying genomes was the core genome (25%) (**Figure 2**).

Most of the genes (94%, 1044/1114) associated with *cpe*-carrying ($p < 0.001$, Benjamini-Hochberg adjusted) represented the accessory genome rather than shared characteristics between

TABLE 2 | Core genome and pangenome estimations of *Clostridium perfringens*.

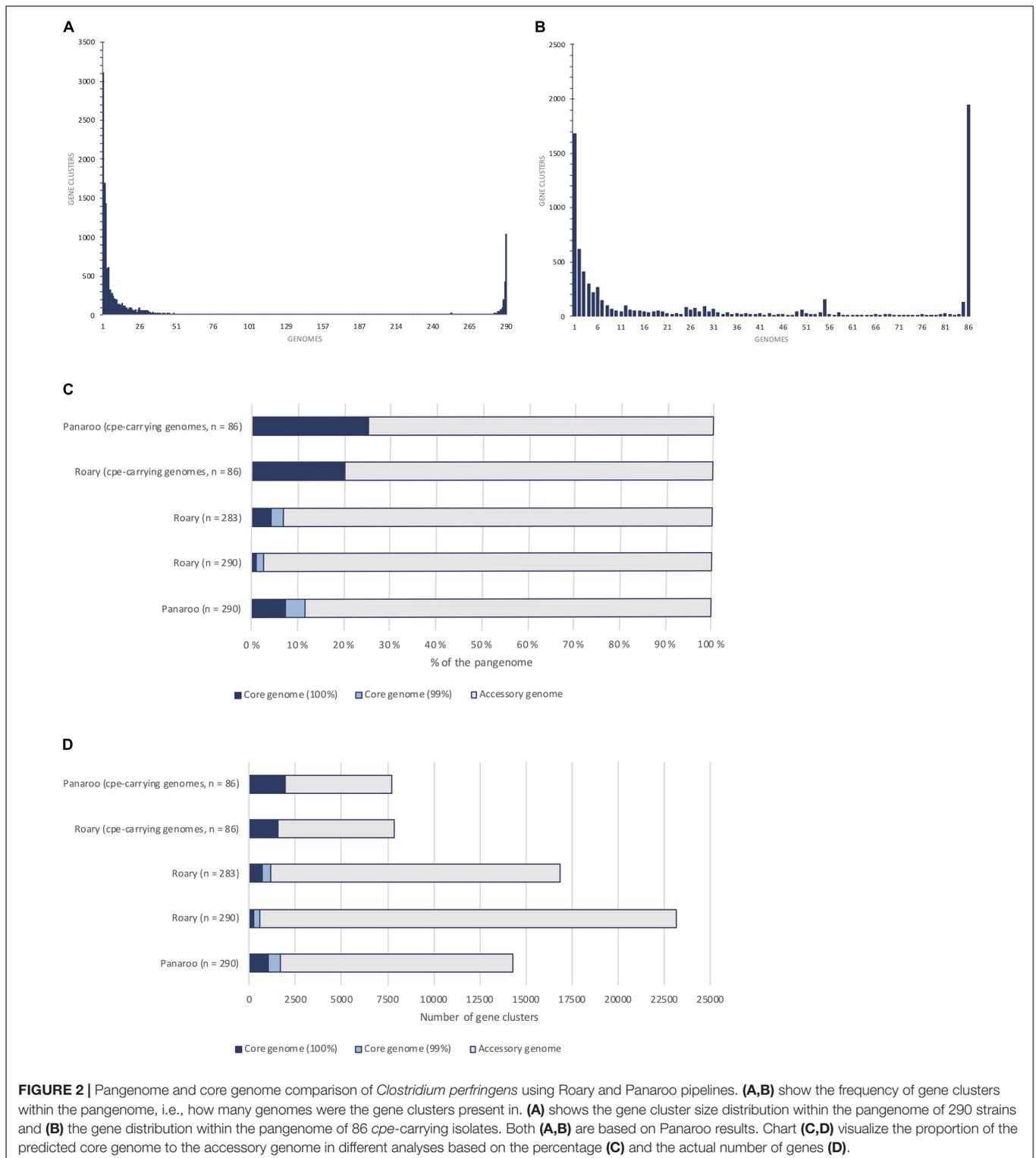
Study	No. of input strains (subgroup)	Pangenome Pipeline and chosen parameters	No. of core CDSs per core threshold Present in % of genomes			No. of CDSs in the pangenome	The proportion (%) of core genome (99%) of the pangenome and the average genome ^b
			100	99	95 ^a		
This study	290	Panaroo, strict	1034	1660	2059	14306	11.6, 53.2
This study	290	Roary, 95% identity, split paralogs	231	590	1078	23148	2.5, 18.9
This study	283 ^c	Roary, 95% identity, split paralogs	696	1170	1581	16875	6.9, 37.5
This study	86 (<i>cpe</i> -carrying strains)	Roary, 95% identity, split paralogs	1577	1577	1797	7835	20.1, 50.7
This study	86 (<i>cpe</i> -carrying strains)	Panaroo, strict	1941	1941	2101	7721	25.1, 63.2
This study	63 (lineage IV strains)	Panaroo, strict	1840	1840	2090	5917	31.0, 65.4
Kiu et al., 2019	110 (human enteric isolates)	Roary, 90%, don't split paralogs	N/A	1965	N/A	~6300	31.2, 63.2
Kiu et al., 2017	56	Roary, 95% identity, split paralogs	1470	1470	N/A	11667	12.6, 47.3
Abdel-Guil, 2019	76	Roary, 90%, split paralogs	N/A	N/A	2057	10098	20.3, 66.1 ^d
Feng et al., 2020	173	Roary, parameters not specified	N/A	N/A	1020	26954	3.8, 32.8 ^d

^aTo enable comparison with previous results the core genome present in 95% of strains is given. Due to the concerns explained in the text, these should not be considered biologically relevant

^bThe average CDSs of analyzed genomes: 3120 CDSs for 290 strains, 3,071 for *cpe*-carrying strains, and 2814 for lineage IV strains.

^cConfounding strains BER-NE33, K473, T3381, W1319, PC5, PBD1, and PBS5 were excluded from the analysis of 283 strains.

^d95% core threshold was used instead of 99%.



all *cpe*-carrying strains. Functional comparison of core genomes of 290 strains and *cpe*-carrying strains still revealed that the latter group shared more genes in 18 out of 24 clusters of orthologous groups (COGs). Especially genes belonging to COGs linked with amino acid metabolism and transport

($n = 12$), transcription ($n = 18$), signal transduction ($n = 10$), and unknown ($n = 38$) functions were more conserved in the core genome of *cpe*-carrying strains than the core genome of all *C. perfringens* strains. Operons related to riboflavin biosynthesis and sucrose metabolism, and three genes

associated with drug resistance, bacitracin resistance gene *uppP* (CPR_1021), *mepA* (CPR_1218), and vancomycin resistance gene *yoaR* (CPR_1209), were fixed within the core genome of *cpe*-carrying strains, but not within the core genome (99%) of all *C. perfringens* strains.

The Panaroo pipeline produced a larger core genome with a smaller pangenome compared to the Roary pipeline in all settings (Figure 2 and Table 2). Based on manual curation of results, the presence of split CDSs or significantly truncated alleles generated most of this difference. Based on the conserved synteny the Panaroo clustered the broken sequences together with the conserved alleles, resulting in higher conservation counts for genes. In the clustering step for the Roary pipeline, the highly identical split CDSs fractionated the entire protein cluster to two or more separate clusters and consequently dropped these outside both the core genome of 99% strains as well as the soft-core present in 95% strains. Due to this, Roary did not return a reliable count of soft-core genes (present in 95% of genomes) for *C. perfringens* and stringent 99% or 100% core gene estimations were considered instead of 95% results (Table 2). Seven of the genomes had a notably high number of split reading frames/truncated alleles, and once these were removed from the analysis, the Roary pipeline produced a 98% larger core genome (83 CDSs per removed genome) (Figure 2 and Table 2).

Conserved *cpe* Gene and a *C. perfringens* Strain With Two Plasmidial Copies of *cpe*

All but one of the strains (82/83) with a 319 aa *cpe* carried an identical copy of the enterotoxin gene either chromosomally or on a pCPF5603 or pCPF4969 plasmid. A single nucleotide mutation or sequencing error resulting in one amino acid change was observed in a previously sequenced strain (JFP981). Additionally, in all assembled *cpe* genomes ($n = 21$), the *cpe* gene was inserted between uracil-xanthine permease and *nadC* genes together with mobile elements, a transposase, and a hypothetical protein (75 aa) (Brynstad et al., 1997). Our study revealed no deletion scars of the *cpe*-carrying mobile element in the corresponding intergenic region of *cpe*-negative lineage IV strains.

In addition to 83 genomes carrying the 319 aa *cpe* gene, 5 genomes carried a 325 aa variant of the *cpe* gene. Noteworthy, one sequenced strain (AAD1903) contained both the 319 aa *cpe* gene and a 325 aa variant of the *cpe* gene. The 319 aa *cpe* was harbored on a pCPF5603 plasmid and the variant *cpe* on a pCPPB1 plasmid. Genetic regions surrounding the enterotoxin genes on plasmids pCPF5603 and pCPPB1 have been previously described (Miyamoto et al., 2011; Li et al., 2013) and are not repeated here. In addition to AAD1903, strains Q061.2 (pCPPB1), CP-09, a508.17 (pCPPB1), and a515.17 (pCPPB1) also harbored a 325 aa variant of *cpe* but did not carry the 319 aa *cpe* gene.

All five strains with the variant *cpe* clustered together in phylogenetic analysis (Figure 1A) and belonged to lineage III. Lineage III included 27 veterinary clinical isolates, mainly from

necrotizing enteritis cases from foals and dogs from North America ($n = 24$), in addition to six other isolates (China, Great Britain, Finland, and France). Lineage III strains had larger genomes (p -value <0.001 , ANOVA), lower GC% (p -value <0.001 , ANOVA), and more CDSs (p -value <0.01 , ANOVA) compared to lineages IV and V. All lineage III strains carried two copies of perfringolysin (PfoA). This variant of *pfoA* has an 85.7% nucleotide identity to the typical *pfoA* (Abdel-Glil et al., 2021).

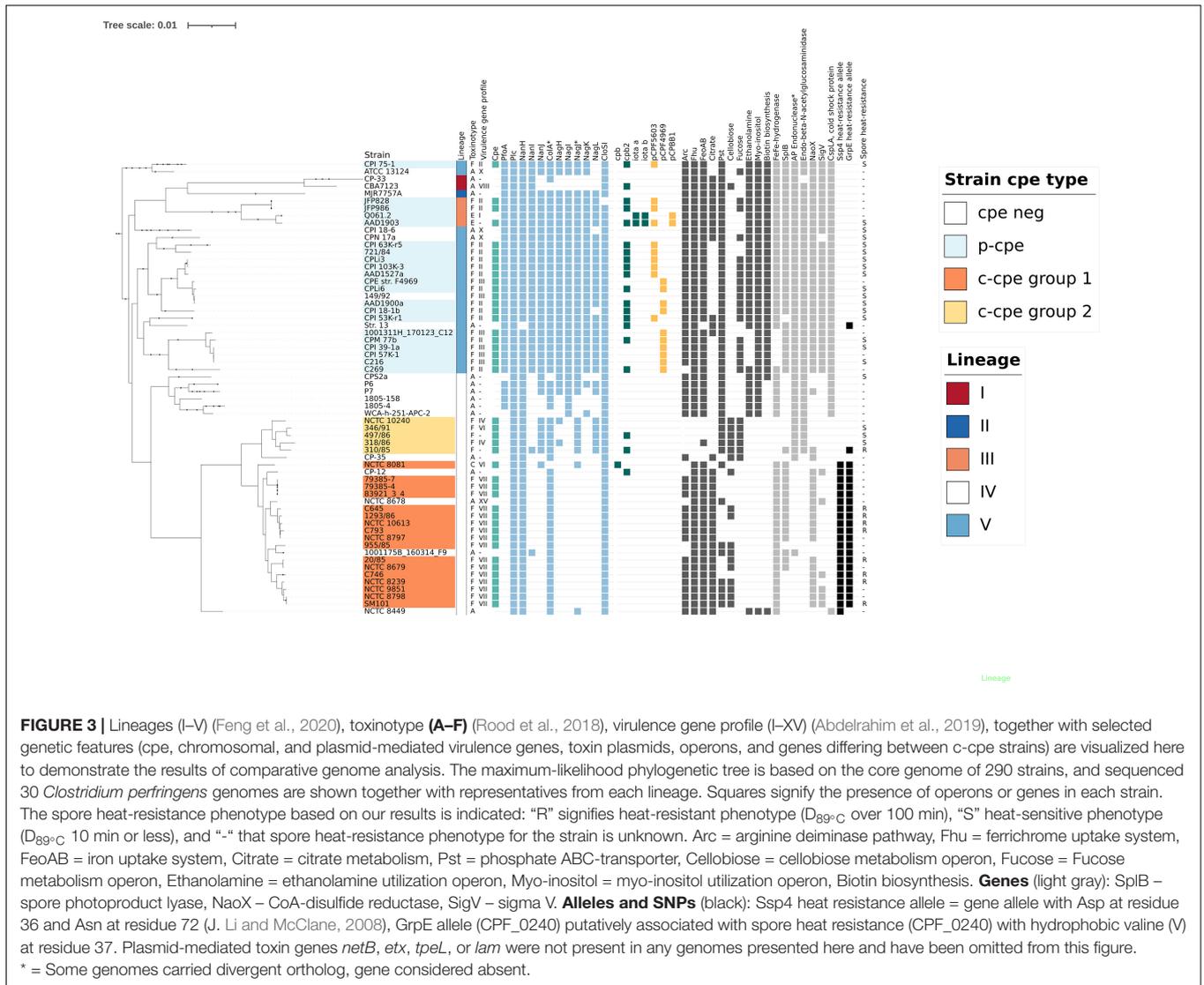
Lineage IV Showed Signs of Reductive Evolution, and the Strains Represented Two Distinct Gene Profiles

Lineage IV strains ($n = 63$) had a smaller genome (3.04 Mb, p -value <0.001 , ANOVA) with fewer coding sequences (2814 ± 160 genes, p -value <0.001 , ANOVA) than other lineages ($n = 227$, 3.43 Mb, 3204 ± 305 CDSs). Operons and genes absent from lineage IV were common in other lineages, suggesting that the ancestors of *cpe* strains carried these genetic characteristics and the genes have been lost on the course of evolution. Mobile genetic elements were more prevalent in lineage IV as annotation of 30 here sequenced strains revealed that lineage IV strains carried on average 99 mobile elements against the 27 found in *p-cpe* and related strains ($p < 0.001$, ANOVA), 172 repeat elements against 87 ($p < 0.001$, ANOVA), and 17 transposon sequences against 6 ($p < 0.001$, ANOVA) (Supplementary Table 3).

Comparative genome analysis showed that *cpe* strains carried fewer virulence genes than other strains. Within lineage V and III, the chromosomal virulence gene profiles were similar (Figure 3), but two groups of *cpe* strains (*cpe* group 1 and 2) had distinguishable gene profiles different from other lineages and the other *cpe* group. *Cpe* group 2 encoded a more diverse virulence gene profile than *cpe* group 1 (Figure 3). Sialidase gene *nanJ*, μ -toxin homolog *nagJ*, glycoside hydrolase endo- β -N-acetylglucosaminidase, and hyaluronidase *nagL* are present in *cpe* group 2 but absent from *cpe* group 1. These genes were located chromosomally, and the genetic context around these genes was otherwise conserved in *cpe* and *p-cpe* strains, suggesting these genes had been deleted from *cpe* group 1.

In addition to differences in known virulence factors, *cpe* group 1 and 2 genomes were distinguishable by several genes and operons with predicted functions. The comparison revealed that *cpe* group 2 had lost, amongst others, arginine deiminase pathway operon, arsenic resistance operon, citrate metabolism operon, a paralog of iron uptake system FeoAB, iron transporting operon Fhu, and one homolog of [FeFe]-hydrogenase (Figure 3). *Cpe* group 1 strains, on the other hand, had lost the above-mentioned virulence genes and also enzymes, fucose utilization operon, and carried a broken version of DNA repair mechanism (AP endonuclease family 2 protein) (Figure 3).

Comparative genome analysis by Lahti et al. (2012) revealed that *cpe* strains do not carry myo-inositol and ethanolamine utilization operon and also lack an operon for biotin synthesis. Our results add to this list a cold shock protein (CspLA, CPF_1452) and sigma V (SigV, CPF_0539) factor. SigV was absent from 25/31 *cpe* strains and 100% conserved in *p-cpe*



strains (52/52). CspLA was conserved in 100% of p-cpe strains but absent from all c-cpe strains. Both c-cpe group 1 and group 2 carried very few phylogroup-specific accessory genes, 30 and 0, respectively. Genes unique to, and conserved within, c-cpe strains included 106 coding sequences, of which 26 were mobile elements or transposases, 49 hypothetical proteins, and 12 broken or truncated copies of genes (Supplementary Table 4). The most notable set of specific genes for lineage IV was the cellobiose utilization operon carried by 30% (19/63) of lineage IV strains and but absent from all but one lineage V and III strains. Cellobiose operon was flanked by phage and transposon-related genetic elements.

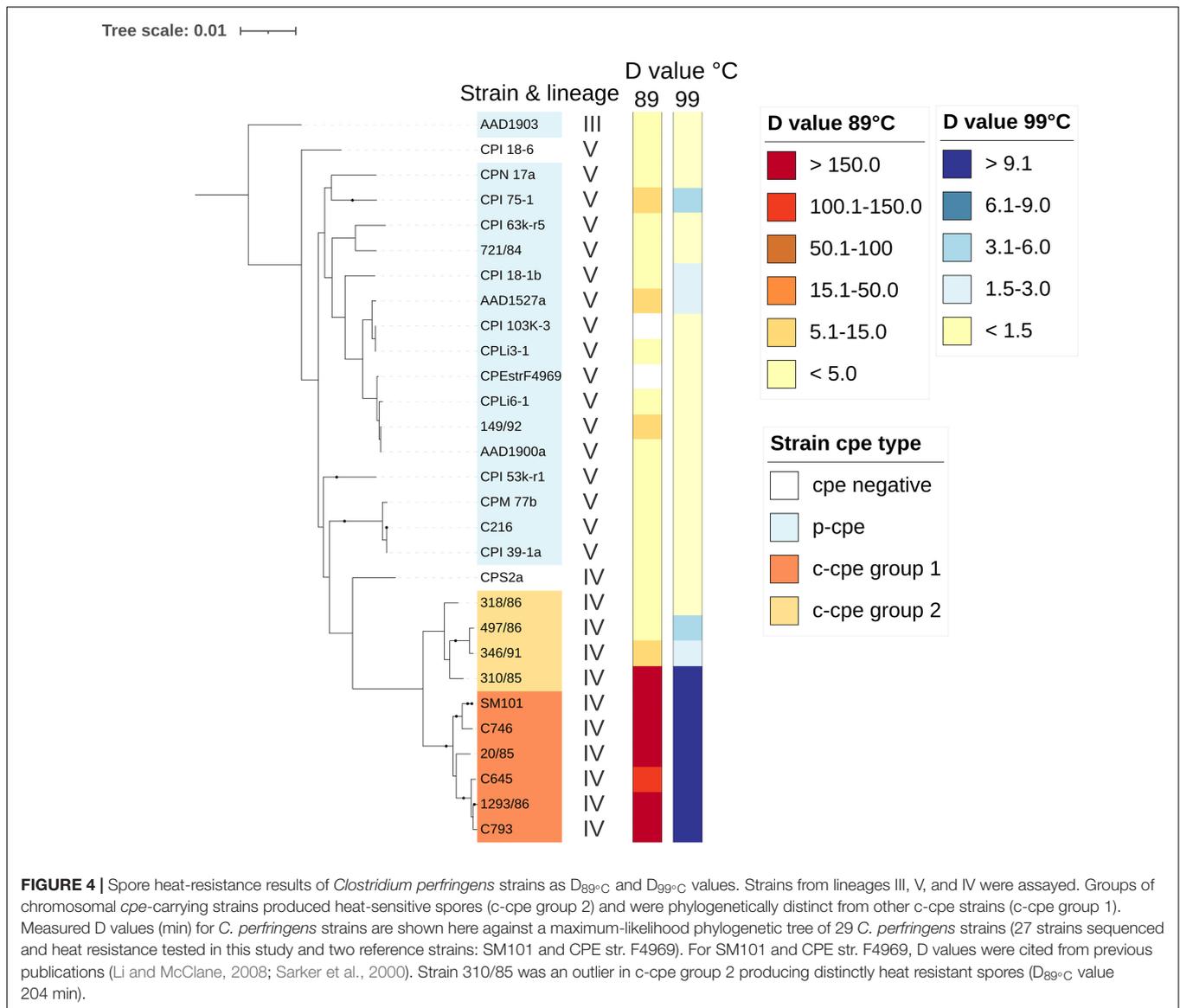
Group of Chromosomal cpe-Carrying Strains in Lineage IV Produced Heat-Sensitive Spores

To characterize the phenotypes of two different c-cpe gene profiles (Figure 3), the sequenced strains were assayed for heat

resistance properties of their spores. C-cpe group 1 strains produced highly heat resistant spores with $D_{89^{\circ}\text{C}}$ values of over 100 min (average 226 min) and $D_{99^{\circ}\text{C}}$ values with an average of 36 min (Figure 4). C-cpe group 2 strains produced relatively heat-sensitive spores with $D_{89^{\circ}\text{C}}$ values of 3–9 min (average without strain 310/85 5.5 min) and $D_{99^{\circ}\text{C}}$ values with an average of 2.7 min. Strain 310/85 was an outlier in c-cpe group 2 producing distinctly heat resistant spores ($D_{89^{\circ}\text{C}}$ value 204 min). One representative of cpe negative group 3 strains was assayed and produced heat-sensitive spores. P-cpe strains and cpe-negative strains from lineage V and III produced spores with $D_{89^{\circ}\text{C}}$ values below 8 min (average 2.8 min) (Figure 4). Calculated D values of strains are available in Supplementary Table 6.

Genes Associated With the Spore Heat-Resistance Phenotype

All c-cpe group 1 genomes carried the known heat resistance allele of Ssp4 (Asp at residue 36, Asn at residue 72), and all



tested c-cpe group 1 strains produced heat-resistant spores. None of the tested strains that produced heat-sensitive spores carried the heat resistance allele of *Ssp4* in their genome (Figure 3). However, one c-cpe group 2 strain (310/85) produced highly heat-resistant spores ($D_{89^{\circ}\text{C}}$ value 204 min) despite carrying the heat sensitivity associated allele (Gly instead of Asp, Lys instead of Asn) of *Ssp4*. It is, therefore, possible for *C. perfringens* to produce extremely heat-resistant spores without the heat resistance allele of *Ssp4*.

Differences between heat-sensitive and heat-resistant c-cpe strains were further analyzed. Orthologs that differed the most between heat sensitive and heat resistant c-cpe strains are summarized in Supplementary Table 7. Most of the sporulation-associated genes were highly conserved between strains, and the most divergent sporulation genes between heat sensitive and heat resistant c-cpe strains are listed in Supplementary Table 8. HGT elements were not associated with heat resistance.

These comparisons revealed a cluster of 5 genes present in heat resistant strains (CPR_0713-0717) but absent from heat-sensitive c-cpe strains. Genes in this cluster encode for CoA-disulfide reductase *NaoX* and related permease and putative transcriptional regulator (Figure 3). Additionally, heat-sensitive c-cpe strains had lost a copy of *GraRS* two-component system, and type IV pilus operon genes *pilW*, *pilA*, and *pulG* were divergent (identity below 80%) between heat sensitive and heat resistant c-cpe strains.

Single amino acid substitutions (SNP) between c-cpe group 1 and group 2 strains were present in many sporulation and heat resistance-related genes (*ssp1*, *ssp2*, *spp3*, *clpX*, *clpC*, *clpP*, *grpE*, *groESL*, ribosome heat shock protein, *dnaJ*, *clpB*, spore coat protein S, *cotS*, *cotJC*, sporulation protein B) in addition to *ssp4*. Observed SNPs between c-cpe group 1 and 2 strains were consistent with phylogeny and did not differentiate 310/85 from the other c-cpe group 2 strains (data not shown). One

exception to this was an SNP within a heat shock protein *grpE* (CPF_0240) that had an alanine substituted with hydrophobic valine on position 37 in all strains producing heat resistant spores including outlier strain 310/85 (Figure 3).

Heat Resistance, Minimum, and Maximum Growth Temperatures of Vegetative Cells

Phenotypes of vegetative cells were characterized for heat resistance and growth temperatures. $D_{60^{\circ}\text{C}}$ values for vegetative cells ranged from 1 to 3.8 min (average 1.3 min), with no significant differences in heat resistance between c-cpe strains and p-cpe strains, or between the two c-cpe groups (Supplementary Table 6). The observed minimum and maximum growth temperatures for the strains ranged from 13.1 to 17.7°C and 49.6 to 52.0°C (Supplementary Table 6), respectively. None of the observed differences between genetic lineages or strain Cpe types were statistically significant.

DISCUSSION

In this study, we performed phenotypic, phylogenetic, and genomic analysis of enteropathogenic *C. perfringens* and showed that plasmidial *cpe*-gene is present in strains with certain genetic backgrounds. We also identified a new group of c-cpe strains and provided insight into the reservoirs and origin of c-cpe strains.

All *C. perfringens* strains have been presumed to be able to acquire horizontally transferable enterotoxin gene and produce enterotoxin. Based on our results, the *cpe*-carrying isolates can be divided into three main groups with two subgroups each: the lineage V isolates carrying either pCPF4969 or pCPF5603 plasmids, the lineage III isolates carrying either pCPF5063 or pCPPB1, and the lineage IV isolates with c-cpe group 1 and group 2 isolates. Isolates carrying plasmids pCPF4969 and pCPPB1 were phylogenetically close to their kin. Additionally, the accessory genome within the *cpe*-carrying isolates was smaller and the core genome larger than that of all strains. These results suggest that not all *C. perfringens* strains are as likely or perhaps even able to carry enterotoxin. One possible explanation for the here observed genetic context of *cpe*-carrying could be a biased transfer rate of conjugative plasmids favoring the plasmid transfer to the kin. The bias of the transfer rate of plasmids toward related strains has been described within natural populations of *E. coli* (Dimitriu et al., 2019). Alternatively, the drug resistance genes and other genes fixed within the core genome of *cpe*-carrying isolates might reflect the capacities required for strains to flourish at the reservoirs where toxin plasmids are shared. Further studies on enterotoxin plasmids and their incompatibilities in *C. perfringens* are warranted to understand the dispersal of the *cpe* gene between different strains.

Interestingly, c-cpe group 1 and c-cpe group 2 included also *cpe*-negative isolates, and the sequence comparison revealed that all c-cpe genomes had the *cpe* gene inserted in the same genetic location. This supports the notion of the clonal nature of c-cpe strains and suggests that *cpe* insertion to chromosome precedes

diversification of c-cpe group 1 and group 2 strains. The *cpe*-negative strains have either persisted together with *cpe*-positive strains or, at some point, lost the chromosomal *cpe* gene. Since chromosomal *cpe* insert is carried in *Tn5565* (Brynestad et al., 1997), the latter explanation seems more probable. Proof for deletion of a *cpe*-carrying mobile element was not observed in the region between uracil-xanthine permease and *nadC* genes in *cpe*-negative lineage IV strains, but this should be further researched. For now, we hypothesize *Tn5565* has been cleanly removed through circulation from chromosomes of *cpe*-negative c-cpe group 1 and group 2 strains.

Noteworthy, here sequenced antibiotic-associated diarrhea strain AAD1903 carried two copies of plasmid-mediated *cpe* gene, one the well-known *cpe* gene of 319 aa sequence length (in pCPF5603 plasmid) and the other a 325aa variant (in pCPPB1 plasmid). The *cpe* variant in pCPPB1 plasmid is expressed (Miyamoto et al., 2011), but the clinical relevance of this variant *cpe* gene remains unknown. The presence of both versions of *cpe* in strain AAD1903 suggests these variants of enterotoxin might have different biological roles. To our understanding, this is the first time a single *C. perfringens* strain has been reported to carry two copies of the enterotoxin gene and two different enterotoxin plasmids. As all pCPPB1 plasmids were sequenced from isolates belonging to double *pfoA*-carrying lineage III, there seems to be potential within lineage III for the emergence of other double *cpe*, double *pfoA* strains similar to the strain AAD1903 described here.

C. perfringens genomes share a conserved genetic backbone, and more than half of the genes of an average genome are present in >95% of other strains. Two different methods for core genome and pangenome estimation were used in this study to probe the pangenome of this microbe. Pseudogenes and truncated alleles seem to be a typical mode of reductive evolution for *C. perfringens*, and the presence of broken CDSs creates issues when clustering-based algorithms are used to identify orthologs. Our results show that the exclusion of confounding genomes dramatically improved the pangenome prediction results. The Panaroo pipeline analyzing synteny handled broken CDSs better in the presence of confounding strains but risked the overestimation of functional core genome as pseudogenes and significantly truncated alleles were predicted as present. Despite the emergence of pangenome pipelines such as Roary, no standardized exclusion methods for poor-quality genomes are available. Here, we utilized genome sequence quality (Parrello et al., 2019) and frequency of truncated alleles in the cgMLST schema as inclusion criteria. In particular, truncated allele counts were useful in detecting genomes with a high frequency of broken reading frames (Supplementary Table 10).

The phylogenetic analysis here revealed that swine and chicken farm isolates sequenced by Li et al. (2021) are closely related with c-cpe strains and with a swine isolate sequenced here. Previously, c-cpe strains have been isolated solely from food and food poisoning related human samples, and their reservoirs and origin have been unknown. These findings are not yet directly indicative of the c-cpe reservoirs, but it is clear that the *cpe*-negative swine and chicken commensal strains are the

closest known relatives of *c-cpe* strains. *Cpe*-negative members of lineage IV have not gathered much attention previously, and they have been rarely reported. However, the sequencing results of Li et al. (2021) suggest that lineage IV *cpe*-negative strains colonize the GI tract of domestic animals and are not rare, but rather underrepresented as they have been classified as type A strains. Sequencing of type A strains will likely help to uncover the reservoirs of *c-cpe* strains. Our results also show that *p-cpe* strains carrying pCPF4969 plasmid are phylogenetically closest to lineage IV (**Supplementary Figure 1**). Particularly, a group of lineage V strains mainly isolated from healthy humans (11/15), and carrying pCPF4969 (6/15), and having a smaller genome compared to other lineage V strains ($p < 0.05$, ANOVA) were phylogenetically close, suggesting that lineage IV and these putative members of human commensal GI population share ancestry.

Since the diversification from other lineages, lineage IV has undergone reductive evolution and niche adaptation. Supporting this observation, the *c-cpe* genomes have smaller genomes, fewer CDSs, few lineage-specific CDSs, and more HGT material compared to the other strains. Our results show that evolution within *c-cpe* strains has taken at least two directions and the two groups of *c-cpe* strains described here are likely adapted to different niches. *C-cpe* group 1 strains, but not group 2 strains, carry operons or genes linked to survival in harsh conditions and autolytic lifestyle (Arc operon, [FeFe]-hydrogenase, arsenic resistance/detoxification operon), competitive iron uptake (Fhu operon, FeoAB), and growth on citrate (Cit operon). An abundance of citrate is known to inhibit iron uptake through normal transport and instead, high-affinity siderophores such as ferrichrome (Fhu operon) are used (Pollack et al., 1970). In general, operons such as Fhu, Cit, and Arc are associated with acidophilic lifestyle, and paralogs of [FeFe]-hydrogenases are known to provide an advantage in changing pH conditions. *C-cpe* group 1 strains seem to be, compared to *c-cpe* group 2, better equipped for changing pH and sometimes acidic, high-temperature environments where iron uptake is competitive, and citrate utilization is beneficial. Previously, the reservoir for *c-cpe* group 1 strains has been proposed to be compost or sludge by Lahti et al. (2012). Both sludge and compost are acidic, prone to pH and temperature changes, abundant in citrate and iron sparse fitting to the profile described here. On the other hand, the loss of the aforementioned genes and operons makes it easy to speculate that pH changes and acidic stress are not characteristic of the adapted niche of *c-cpe* group 2.

Both *c-cpe* groups have lost virulence genes and genes common in intestinal bacteria and important for host colonization (Stahl et al., 2011; Becerra et al., 2015). *C-cpe* group 2 has, compared to *c-cpe* group 1, retained some genetic traits that likely provide a competitive advantage for host colonization and commensal lifestyle (fucose utilization operon, hyaluronidases *nagJ*, *nagL*, and sialidase *nanJ*). Adaptation to intestinal colonization is one possible explanation why the arginine deiminase pathway (Arc operon), despite its many uses in stressful conditions, has been deleted from *c-cpe* group 2 strains. Similar deletion of Arc operon has been suggested to be

a trade-off between adaptation to urinary tract compared to the intestine in ESBL *E. coli*; Arc operon was beneficial for strains in the urinary tract, but a burden for mouse intestine colonization (Billard-Pomares et al., 2019). *C-cpe* group 2 genomes suggest adaptation to a less acidic environment, where the improved acquisition of anaerobic iron does not provide a competitive advantage. Their more diverse virulence gene profile is likely beneficial for commensal life and host colonization and group 2 strains may have retained a niche within the gastrointestinal tract of yet unknown host reservoir(s).

One possible big theme for reductive evolution within *c-cpe* strains could be an adaptation to planktonic growth. Soncini et al. (2020) have studied the gene expression of *C. perfringens* str. 13 growing on a plate, and many of the identified differentially expressed genes have been deleted from *c-cpe* strains. This means that the transcription of *c-cpe* strains growing on a plate differs greatly from those researched by Soncini et al. (2020). *C-cpe* strains also lack *pfoA*, important for biofilm formation (Vidal et al., 2015), and type IV pilus genes, important for quorum sensing and adhesion, are highly divergent between *p-cpe*, *c-cpe* group 1, and *c-cpe* group 2 (**Supplementary Table 7**). We hypothesize that *c-cpe* group 1 strains are adapted to environmental liquid niches and prefer planktonic growth. The preferences for liquid versus plated growth between *C. perfringens* strains should be further studied.

We studied heat-resistance and growth temperatures of *cpe*-carrying isolates and concluded that only *c-cpe* group 1 produced heat-resistant spores, while *c-cpe* group 2 and *p-cpe* spores are heat-sensitive. *C-cpe* strains have been characterized as heat resistant and clonal in previous studies, and to our understanding, this is the first time two separate groups of *c-cpe* strains with different virulence and functional characteristics have been described.

We also report one *c-cpe* strain to produce extremely heat-resistant spores without the presence of heat resistance associated allele of Ssp4. Li and McClane have previously shown that a single amino acid substitution in small acid-soluble spore protein beta or Ssp4 significantly affects spore heat resistance (Li and McClane, 2008; Li et al., 2009). Furthermore, genome analysis of 290 strains revealed that this heat resistance allele of Ssp4 was only found from *c-cpe* group 1 strains indicating this allele has most likely originated within this phylogroup. The capacity to produce highly heat-resistant spores seems to be characteristic and a putative adaptation for *c-cpe* group 1 strains. However, our results also indicate that other strains have the potential to acquire the capacity to produce highly heat-resistant spores and that heat resistance is not reliant on one allele but more likely to be affected by several genes.

The number of isolates in this study is not sufficient to deduce clear associations between heat-resistant phenotype and genotype within *c-cpe* strains. We identified two genes potentially associated with spore heat resistance in *c-cpe* strains: a certain allele of GrpE and the presence of CoA-disulfide reductase NaoX. GrpE in strains producing heat-resistant spores had a point mutation in the N-terminal part before helix structure. This N-terminal part has been proposed to act as a thermometer in *E. coli* (Gelinas et al., 2002), and alterations in this part may

affect the thermal response of strain. CoA-disulfide reductase gene *naoX* has been deleted from *c-cpe* strains producing heat-sensitive spores. CoA is a common substance in dormant spores, and CoA-disulfide reductases have also been implied to have a role in the heat resistance of dormant spores (Setlow and Setlow, 1977; Hamilton et al., 2014).

To understand the ecology and epidemiology of this important pathogen, lineage-distinguishing typing method to identify closely related toxin-negative strains is needed. Lack of sequenced or lineage-typed environmental isolates compared to clinical isolates is likely contorting our view of *C. perfringens*, which after all is an environmental bacterium common in soil and water.

CONCLUSION

Our study adds valuable genomic and phenotypic data to the publicly available collection of assembled and complete *C. perfringens* strains. The genomes include antibiotic-associated diarrhea isolates, *cpe*-positive environmental isolates, antibiotic-associated diarrhea isolate carrying double enterotoxin gene, and food poisoning isolates including strains representing the here described new subgroup of chromosomal *cpe*-carrying (*c-cpe*) isolates.

We established the signs of reductive evolution within lineage IV containing the *c-cpe* strains and elucidated their potential evolutive origin. *C-cpe* strains share ancestry with contemporary commensal strains isolated from swine and chicken. Comparative genome analysis revealed niche adaptation within the lineage of *c-cpe* strains. One subgroup produced heat-resistant spores and retained a genetic toolset suitable for survival in harsh environments. Meanwhile, the other subgroup produced heat-sensitive spores, lacked relevant operons for stress survival, but had conserved genes associated with host colonization and commensal lifestyle.

Focus on the rarely reported *cpe*-negative strains of lineage IV is likely to provide more insight into the ecology of *c-cpe* strains. Currently, the lack of genomic information about environmental isolates and the toxinotyping based on mobile elements is hindering research on ecology and reservoirs of *C. perfringens*.

Based on our results, it is apparent that spore heat resistance of *C. perfringens* strains is likely affected by multiple genes and the capacity to produce heat-resistant spores has developed primarily within one subgroup of chromosomal *cpe*-carrying strains.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

HK and KJ designed the study. KV performed the experiments. KJ performed the genome and data analysis and drafted the manuscript. KJ, PL, ML, and HK contributed to the data analysis and interpretation. KV, PL, RK-T, ML, and HK contributed to the manuscript revision. All authors have read and approved the final manuscript.

FUNDING

This work was funded by the Finnish Foundation of Veterinary Research.

ACKNOWLEDGMENTS

We thank Yagmur Derman and Kirsi Ristkari from the University of Helsinki and Pia Laine from the Institute of Biotechnology for technical assistance and advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.717176/full#supplementary-material>

Supplementary Figure 1 | Phylogenetic tree of 290 *C. perfringens* strains.

Supplementary Table 1 | *Clostridium perfringens* strains sequenced ad reference genomes used in this study: the *cpe* genotype, isolation source, and isolation year.

Supplementary Table 2 | Genomes included in the pangenome study.

Supplementary Table 3 | Selected key factors of sequenced *Clostridium perfringens* genomes including plasmids and deposition keys.

Supplementary Table 4 | Genes unique and conserved for *c-cpe* strains (bi-directional best hit orthologs).

Supplementary Table 5 | Genes unique and conserved for *c-cpe* strain subgroups (bi-directional best hit orthologs).

Supplementary Table 6 | Phenotyping results, heat-resistance assay results including D values for spores and vegetative cells, and minimum and maximum growth temperatures.

Supplementary Table 7 | Genes divergent between in heat-resistant and heat-sensitive *c-cpe* strains.

Supplementary Table 8 | Sporulation related genes (Xiao et al., 2015) divergent between *c-cpe Clostridium perfringens* strains and their identity against reference genes from *C. perfringens* SM101.

Supplementary Table 9 | Presentation of genetic lineages within analyzed genomes, GC%, and genome size per lineage.

Supplementary Table 10 | Number of truncated alleles per genome compared to cgMLST scheme.

REFERENCES

- Abdel-Glil, M. (2019). *In Silico Genome Analysis and Molecular Typing of Clostridium Perfringens*. Berlin: Freie Univ., Diss., Mensch & Buch Verlag.
- Abdel-Glil, M. Y., Thomas, P., Linde, J., Busch, A., Wieler, L. H., Neubauer, H., et al. (2021). Comparative in silico genome analysis of *Clostridium perfringens* unravels stable phylogroups with different genome characteristics and pathogenic potential. *Sci. Rep.* 11:6756. doi: 10.1038/s41598-021-86148-8
- Abdelrahim, M. A., Radomski, N., Delannoy, S., Djellal, S., Le Négrate, M., Hadjab, K., et al. (2019). Large-scale genomic analyses and toxinotyping of *Clostridium perfringens* implicated in foodborne outbreaks in France. *Front. Microbiol.* 10:777. doi: 10.3389/fmicb.2019.00777
- Ando, Y., Toshiyumi, T., Sunagawa, H., and Oka, S. (1985). Heat resistance, spore germination, and enterotoxigenicity of *Clostridium perfringens*. *Microbiol. Immunol.* 29, 317–326. doi: 10.1111/j.1348-0421.1985.tb00830.x
- Asha, N. J., Tompkins, D., and Wilcox, M. H. (2006). Comparative analysis of prevalence, risk factors, and molecular epidemiology of antibiotic-associated diarrhea due to *Clostridium difficile*, *Clostridium perfringens*, and *Staphylococcus aureus*. *J. Clin. Microbiol.* 44, 2785–2791. doi: 10.1128/JCM.00165-06
- Becerra, J. E., Yebra, M. J., and Monedero, V. (2015). An L-fucose operon in the probiotic *Lactobacillus rhamnosus* GG is involved in adaptation to gastrointestinal conditions. *Appl. Environ. Microbiol.* 81, 3880–3888. doi: 10.1128/AEM.00260-15
- Billard-Pomares, T., Clermont, O., Castellanos, M., Magdoud, F., Royer, G., Condamine, B., et al. (2019). The arginine deiminase operon is responsible for a fitness trade-off in extended-spectrum- β -lactamase-producing strains of *Escherichia coli*. *Antimicrob. Agents Chemother.* 63:e00635-19. doi: 10.1128/AAC.00635-19
- Brynstad, S., Synstad, B., and Granum, P. E. (1997). The *Clostridium perfringens* enterotoxin gene is on a transposable element in type a human food poisoning strains. *Microbiology* 143, 2109–2115. doi: 10.1099/00221287-143-7-2109
- Brynilsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17:238. doi: 10.1186/s13059-016-1108-8
- Chin, C.-S., David, H. A., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Collie, R. E., and McClane, B. A. (1998). Evidence that the enterotoxin gene can be episomal in *Clostridium perfringens* isolates associated with non-food-borne human gastrointestinal diseases. *J. Clin. Microbiol.* 36, 30–36. doi: 10.1128/jcm.36.1.30-36.1998
- Cornillot, E., Saint-Joanis, B., Daube, G., Katayama, S., Granum, P. E., Canard, B., et al. (1995). The enterotoxin gene (Cpe) of *Clostridium perfringens* can be chromosomal or plasmid-borne. *Mol. Microbiol.* 15, 639–647. doi: 10.1111/j.1365-2958.1995.tb02373.x
- Darling, A. E., Mau, B., and Perna, N. T. (2010). Progressivemaue: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147
- Deguchi, A., Miyamoto, K., Kuwahara, T., Miki, Y., Kaneko, I., Li, J., et al. (2009). Genetic characterization of type a enterotoxigenic *Clostridium perfringens* strains. *PLoS One* 4:e0005598. doi: 10.1371/journal.pone.0005598
- Derman, Y., Lindström, M., Selby, K., and Korkeala, H. (2011). Growth of group II *Clostridium botulinum* strains at extreme temperatures. *J. Food Protect.* 74, 1797–1804. doi: 10.4315/0362-028X.JFP-11-187
- Dimitriu, T. A., Lauren, M., Angus, B., and Ben, R. (2019). Bacteria from natural populations transfer plasmids mostly towards their kin. *Proc. R. Soc. B* 286:20191110. doi: 10.1098/rspb.2019.1110
- Duncan, C. L., and Strong, D. H. (1968). Improved medium for sporulation of *Clostridium perfringens*. *Appl. Microbiol.* 16, 82–89. doi: 10.1128/aem.16.1.82-89.1968
- Feng, Y., Fan, X., Zhu, L., Yang, X., Liu, Y., Gao, S., et al. (2020). Phylogenetic and genomic analysis reveals high genomic openness and genetic diversity of *Clostridium perfringens*. *Microb. Genom.* 6:e000441. doi: 10.1099/mgen.0.000441
- Gelinas, A. D., Langsetmo, K., Toth, J., Bethoney, K. A., Stafford, W. F., and Harrison, C. J. (2002). A Structure-based interpretation of *E. coli* grpe thermodynamic properties. *J. Mol. Biol.* 323, 131–142. doi: 10.1016/S0022-2836(02)00915-4
- Gillespie, J. J., Wattam, A. R., Cammer, S. A., Gabbard, J. L., Shukla, M. P., Dalay, O., et al. (2011). PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect. Immun.* 79, 4286–4298. doi: 10.1128/IAI.00207-11
- Hamilton, C. J., Arbach, M., Groom, M. (2014). “Beyond glutathione: Different low molecular weight thiols as mediators of redox regulation and other metabolic functions in lower organisms”, *Recent Advances in Redox Active Plant and Microbial Products: From Basic Chemistry to Widespread Applications in Medicine and Agriculture* (Netherlands: Springer), 291–320.
- Hatheway, C. L. (1990). Toxigenic clostridia. *Clin. Microbiol. Rev.* 3, 66–98. doi: 10.1128/cmr.3.1.66-98.1990
- Heikinheimo, A., Lindström, M., Granum, P. E., and Korkeala, H. (2006). Humans as reservoir for enterotoxin gene-carrying *Clostridium perfringens* type A. *Emerg. Infect. Dis.* 12, 1724–1729. doi: 10.3201/eid1211.060478
- Hu, W. S., Kim, H., and Koo, O. K. (2018). Molecular genotyping, biofilm formation and antibiotic resistance of enterotoxigenic *Clostridium perfringens* isolated from meat supplied to school cafeterias in South Korea. *Anaerobe* 52, 115–121. doi: 10.1016/j.anaerobe.2018.06.011
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085
- Keto-Timonen, R., Heikinheimo, A., Eerola, E., and Korkeala, H. (2006). Identification of *Clostridium* species and DNA fingerprinting of *Clostridium perfringens* by amplified fragment length polymorphism analysis. *J. Clin. Microbiol.* 44, 4057–4065. doi: 10.1128/JCM.01275-06
- Kiu, R., Caim, S., Alexander, S., Pachori, P., and Hall, L. J. (2017). Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant pangenome diversity, and a diverse array of virulence factors. *Front. Microbiol.* 8:2485. doi: 10.3389/fmicb.2017.02485
- Kiu, R., Caim, S., Painset, A., Pickard, D., Swift, C., Dougan, G., et al. (2019). Phylogenomic analysis of gastroenteritis-associated *Clostridium perfringens* in England and Wales over a 7-year period indicates distribution of clonal toxigenic strains in multiple outbreaks and extensive involvement of enterotoxin-encoding (Cpe) plasmids. *Microb. Genom.* 5:e000297. doi: 10.1099/mgen.0.000297
- Kiu, R., and Hall, L. J. (2018). An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerg. Microbes Infect.* 7:141. doi: 10.1038/s41426-018-0144-8
- Korkeala, H., Mäkelä, P., and Suominen, H. (1990). Growth temperatures of ropy slime-producing lactic acid bacteria. *J. Food Protect.* 53, 793–794. doi: 10.4315/0362-028X-53.9.793
- Lahti, P., Heikinheimo, A., Johansson, T., and Korkeala, H. (2008). *Clostridium perfringens* type a strains carrying a plasmid-borne enterotoxin gene (Genotype IS1151-Cpe or IS1470-like-Cpe) as a common cause of food poisoning. *J. Clin. Microbiol.* 46, 371–373. doi: 10.1128/JCM.01650-07
- Lahti, P., Lindström, M., Somervuo, P., Heikinheimo, A., and Korkeala, H. (2012). Comparative genomic hybridization analysis shows different epidemiology of chromosomal and plasmid-borne Cpe-carrying *Clostridium perfringens* type A. *PLoS One* 7:e46162. doi: 10.1371/journal.pone.0046162
- Li, J., Adams, V., Bannam, T. L., Miyamoto, K., Garcia, J. P., Uzal, F. A., et al. (2013). Toxin plasmids of *Clostridium perfringens*. *Microbiol. Mol. Biol. Rev.* 77, 208–233. doi: 10.1128/mmb.00062-12
- Li, J., and McClane, B. A. (2006a). Comparative effects of osmotic, sodium nitrite-induced, and pH-induced stress on growth and survival of *Clostridium perfringens* type a isolates carrying chromosomal or plasmid-borne enterotoxin genes. *Appl. Environ. Microbiol.* 72, 7620–7625. doi: 10.1128/AEM.01911-06

- Li, J., and McClane, B. A. (2006b). Further comparison of temperature effects on growth and survival of *Clostridium perfringens* type a isolates carrying a chromosomal or plasmid-borne enterotoxin gene. *Appl. Environ. Microbiol.* 72, 4561–4568. doi: 10.1128/AEM.00177-06
- Li, J., and McClane, B. A. (2008). A novel small acid soluble protein variant is important for spore resistance of most *Clostridium perfringens* food poisoning isolates. *PLoS Pathog.* 4:e1000056. doi: 10.1371/journal.ppat.1000056
- Li, J., Miyamoto, K., Sayeed, S., and McClane, B. A. (2010). Organization of the Cpe locus in CPE-positive *Clostridium perfringens* type C and D isolates. *PLoS One* 5:e0010932. doi: 10.1371/journal.pone.0010932
- Li, J., Paredes-Sabja, D., Sarker, M. R., and McClane, B. A. (2009). Further characterization of *Clostridium perfringens* small acid soluble protein-4 (Ssp4) properties and expression. *PLoS One* 4:e6249. doi: 10.1371/journal.pone.0006249
- Li, J., Sayeed, S., and McClane, B. A. (2007). Prevalence of enterotoxigenic *Clostridium perfringens* isolates in pittsburgh (pennsylvania) area soils and home kitchens. *Appl. Environ. Microbiol.* 73, 7218–7224. doi: 10.1128/AEM.01075-07
- Li, J., Zhou, Y., Yang, D., Zhang, S., Sun, Z., Wang, Y., et al. (2021). Prevalence and antimicrobial susceptibility of *Clostridium perfringens* in chickens and pigs from Beijing and Shanxi, China. *Vet. Microbiol.* 24:252. doi: 10.1016/j.vetmic.2020.108932
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Lindström, M., Heikinheimo, A., Lahti, P., and Korkeala, H. (2011). Novel insights into the epidemiology of *Clostridium perfringens* type a food poisoning. *Food Microbiol.* 28, 192–198. doi: 10.1016/j.fm.2010.03.020
- Lugli, G. A., Milani, C., Mancabelli, L., Turroni, F., Ferrario, C., Duranti, S., et al. (2017). Ancient bacteria of the ötzi's microbiome: a genomic tale from the copper age. *Microbiome* 5:5. doi: 10.1186/s40168-016-0221-y
- Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., et al. (2009). CDD: specific functional annotation with the conserved domain database. *Nucleic Acids Res.* 37, D205–D210.
- Miyamoto, K., Fisher, D. J., Li, J., Sayeed, S., Akimoto, S., and McClane, B. A. (2006). Complete sequencing and diversity analysis of the enterotoxin-encoding plasmids in *Clostridium perfringens* type a non-food-borne human gastrointestinal disease isolates. *J. Bacteriol.* 188, 1585–1598. doi: 10.1128/JB.188.4.1585-1598.2006
- Miyamoto, K., Yumine, N., Mimura, K., Nagahama, M., Li, J., McClane, B. A., et al. (2011). Identification of novel *Clostridium perfringens* type e strains that carry an iota toxin plasmid with a functional enterotoxin gene. *PLoS One* 6:e0020376. doi: 10.1371/journal.pone.0020376
- Myers, G. S., Rasko, D. A., Cheung, J. K., Ravel, J., Seshadri, R., DeBoy, R. T., et al. (2006). Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. *Genome Res.* 16, 1031–1040. doi: 10.1101/gr.5238106
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421
- Parrello, B., Butler, R., Chlenski, P., Olson, R., Overbeek, J. C., Pusch, G. D., et al. (2019). A machine learning-based service for estimating quality of genomes using PATRIC. *BMC Bioinformatics* 20:486. doi: 10.1186/s12859-019-3068-y
- Petit, L., Gibert, M., and Popoff, M. R. (1999). *Clostridium perfringens*: toxinotype and genotype. *Trends Microbiol.* 7, 104–110. doi: 10.1016/S0966-842X(98)01430-9
- Pollack, J. R., Ames, B. N., and Neilands, J. B. (1970). Iron transport in *Salmonella typhimurium*: mutants blocked in the biosynthesis of enterobactin. *J. Bacteriol.* 104, 635–639. doi: 10.1128/jb.104.2.635-639.1970
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Raju, D., and Sarker, M. R. (2005). Comparison of the levels of heat resistance of wild-type, Cpe knockout, and Cpe plasmid-cured *Clostridium perfringens* type a strains. *Appl. Environ. Microbiol.* 71, 7618–7620. doi: 10.1128/AEM.71.11.7618-7620.2005
- Rood, J. I., Adams, V., Lacey, J., Lyras, D., McClane, B. A., Melville, S. B., et al. (2018). Expansion of the *Clostridium perfringens* toxin-based typing scheme. *Anaerobe* 53, 5–10. doi: 10.1016/j.anaerobe.2018.04.011
- Sarker, M. R., Shivers, R. P., Sparks, S. G., Juneja, V. K., and McClane, B. A. (2000). Comparative experiments to examine the effects of heating on vegetative cells and spores of *Clostridium perfringens* isolates carrying plasmid genes versus chromosomal enterotoxin genes. *Appl. Environ. Microbiol.* 66, 3234–3240. doi: 10.1128/AEM.66.8.3234-3240.2000
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Setlow, B., and Setlow, P. (1977). Levels of acetyl coenzyme A, reduced and oxidized coenzyme A, and coenzyme A in disulfide linkage to protein in dormant and germinated spores and growing and sporulating cells of *Bacillus megaterium*. *J. Bacteriol.* 132, 444–452. doi: 10.1128/jb.132.2.444-452.1977
- Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., et al. (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc. Natl. Acad. Sci. U.S.A.* 99, 996–1001. doi: 10.1073/pnas.022493799
- Silva, F., Da, M. T., Hald, T., Liebana, E., Allende, A., Hugas, M., et al. (2015). Risk ranking of pathogens in ready-to-eat unprocessed foods of non-animal origin (FoNAO) in the EU: initial evaluation using outbreak data (2007–2011). *Int. J. Food Microbiol.* 195, 9–19. doi: 10.1016/j.ijfoodmicro.2014.11.005
- Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., et al. (2018). ChewBBACA: a complete suite for gene-by-gene schema creation and strain identification. *Microb. Genom.* 4:e000166. doi: 10.1099/mgen.0.000166
- Soncini, S. R., Hartman, A. H., Gallagher, T. M., Camper, G. J., Jensen, R. V., and Melville, S. B. (2020). Changes in the expression of genes encoding type IV pili-associated proteins are seen when *Clostridium perfringens* is grown in liquid or on surfaces. *BMC Genomics* 21:45. doi: 10.1186/s12864-020-6453-z
- Sparks, S. G., Carman, R. J., Sarker, M. R., and McClane, B. A. (2001). Genotyping of enterotoxigenic *Clostridium perfringens* fecal isolates associated with antibiotic-associated diarrhea and food poisoning in North America. *J. Clin. Microbiol.* 39, 883–888. doi: 10.1128/JCM.39.3.883-888.2001
- Staden, R., Judge, D. P., and Bonfield, J. K. (2003). Managing sequencing projects in the GAP4 environment. *Introductions. Bioinform.* 4, 327–344. doi: 10.1007/978-1-59259-335-4_20
- Stahl, M., Friis, L. M., Nothaft, H., Liu, X., Li, J., Szymanski, C. M., et al. (2011). L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7194–7199. doi: 10.1073/pnas.1014125108
- Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J., et al. (2020). Producing polished prokaryotic pangenomes with the *Panaroo pipeline*. *Genome Biol.* 21:180. doi: 10.1186/s13059-020-02090-4
- Vidal, J. E., Shak, J. R., and Canizalez-Roman, A. (2015). The CpAL quorum sensing system regulates production of hemolysins CPA and PFO to build *Clostridium perfringens* biofilms. *Infect. Immun.* 83, 2430–2442. doi: 10.1128/IAI.00240-15
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2016). Improvement to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* 4, D535–D542. doi: 10.1093/nar/gkw1017
- Weiss, K. F., and Strong, D. H. (1967). Some properties of heat-resistant and heat-sensitive strains of *Clostridium perfringens*. I. Heat resistance and toxigenicity. *J. Bacteriol.* 93, 21–26. doi: 10.1128/jb.93.1.21-26.1967

- Wen, Q., and McClane, B. A. (2004). Detection of enterotoxigenic *Clostridium perfringens* type A isolates in American retail foods. *Appl. Environ. Microbiol.* 70, 2685–2691. doi: 10.1128/AEM.70.5.2685-2691.2004
- Xiao, Y., Van Hijum, S. A. F. T., Abee, T., and Wells-Bennik, M. H. J. (2015). Genome-wide transcriptional profiling of *Clostridium perfringens* SM101 during sporulation extends the core of putative sporulation genes and genes determining spore properties and germination characteristics. *PLoS One* 10:e0127036. doi: 10.1371/journal.pone.0127036
- Xiao, Y., Wagendorp, A., Moezelaar, R., Abee, T., and Wells-Bennik, M. H. J. (2012). A wide variety of *Clostridium perfringens* type a food-borne isolates that carry a chromosomal Cpe gene belong to one multilocus sequence typing cluster. *Appl. Environ. Microbiol.* 78, 7060–7068. doi: 10.1128/AEM.01486-12
- Zhou, Z., Alikhan, N. F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., et al. (2018). Grapetree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* 28, 1395–1404. doi: 10.1101/gr.232397.117

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Jaakkola, Virtanen, Lahti, Keto-Timonen, Lindström and Korkeala. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.