Genomic investigations in animal enteric disease- associated Clostridium perfringens

Franzen, J.; Roder, T.; Larralde, M.; Espinosa, A.C.; Kittl, S.; Feyer, S.; ... ; Posthaus, H.

MICROBIAL GENOMICS

RESEARCH ARTICLE

Franzen *et al.*, *Microbial Genomics* 2025;11:001545

DOI 10.1099/mgen.0.001545

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Genomic investigations in animal enteric disease-associated *Clostridium perfringens*

Jan Franzen[1,2], Thomas Roder[3], Martin Larralde[4], Ana Carpio Espinosa[1,2], Sonja Kittl[5], Simon Feyer[5], Isabelle Brodard[5], Sam Nooij[4], Quinten Raymond Ducarmon[4], Faezeh Farhoosh[1,2], Marco Kreuzer[3,6], Rémy Bruggmann[3], Pamela Nicholson[6], Evy Goosens[7], Wiep Klaas Smits[4,8] and Horst Posthaus[1,*]

Abstract

The anaerobic bacterium *Clostridium perfringens* is commonly found in the intestinal tract of humans and animals. However, there are marked differences in virulence between isolates and toxinotypes, which largely depend on various virulence factors produced by these strains. Studying *C. perfringens* genomes has been limited by fragmented assemblies from short-read sequencing and incomplete clinical metadata. Here, we present a high-quality collection of 236 isolates from animal hosts that underwent detailed pathomorphological examination. From 220 of them, genomes were generated by PacBio long-read sequencing, enabling comprehensive, structural level analysis of their virulome, plasmids, conjugative elements and biosynthetic gene clusters (BGCs). One hundred forty isolates were collected from animals with signs of *C. perfringens*-associated enteric disease, 32 from animals with no signs of *C. perfringens*-associated disease and 64 from healthy animals. An additional 383 publicly available *C. perfringens* complete or draft genomes were included for comparative analyses. We discovered 2 previously undescribed pore-forming toxin (PFT) homologues and 11 novel haemolysin and aerolysin-type PFT variants. Both findings expand the known spectrum of the *C. perfringens* virulome. Moreover, we defined two novel putative plasmid conjugative loci in a collection of 888 here assembled and circularized plasmids. They may facilitate HGT, supporting the dissemination of virulence and metabolic traits. We predicted 414 BGCs that were frequently toxinotype specific and encoded centrally for a bacteriocin that could help their carriers to outcompete other bacteria in shared environments. Furthermore, comparative analysis of *C. perfringens* plasmids revealed three distinct clusters based on their conjugation system. Altogether, these findings significantly expand the landscape of animal-associated *C. perfringens* through high-quality genome data and highlight novel virulence-associated features that provide a foundation for future studies of this important pathogen.

Received 08 August 2025; Accepted 09 October 2025; Published 04 November 2025

Author affiliations: [1]Department of Infectious Diseases and Pathobiology, Institute of Animal Pathology, Vetsuisse Faculty, University of Bern, Bern, Switzerland; [2]Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland; [3]Interfaculty Bioinformatics Unit and Swiss Institute of Bioinformatics, University of Bern, Bern, Switzerland; [4]Leiden University Center of Infectious Diseases (LUCID), Leiden University Medical Center, Leiden, The Netherlands; [5]Department of Infectious Diseases and Pathobiology, Institute of Veterinary Bacteriology, Vetsuisse Faculty, University of Bern, Bern, Switzerland; [6]Next Generation Sequencing Platform, University of Bern, Bern, Switzerland; [7]Department of Pathobiology, Pharmacology and Zoological Medicine, Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium; [8]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, The Netherlands.

***Correspondence:** Horst Posthaus, horst.posthaus@unibe.ch

Keywords: biosynthetic gene clusters; *Clostridium perfringens*; plasmids; pore-forming toxins; whole genome sequences.

Abbreviations: AHDS, acute haemorrhagic diarrhoea syndrome; AMR, antimicrobial resistance; ANI, average NT identity; BGC, biosynthetic gene cluster; CPB, Clostridium perfringens beta toxin; CPE, Clostridium perfringens enterotoxin; EdpA/B/C, Epsilon domain protein A/B/C; ETX, epsilon toxin; FNE, foal necrotizing enteritis; GCFs, gene cluster families; HGT, horizontal gene transfer; ITX, iota toxin; NE, necrotic enteritis; NetB/E/F/G/H, necrotizing enteritis toxin B/E/F/G/H; ORF, open reading frame; PFT, pore-forming toxin; PGAP, procaryotic genome annotation pipeline; RiPP, ribosomally synthesized and post-translationally modified peptide; SMs, secondary metabolites; TMD, transmembrane domain; WT, wildtype.

Four supplementary tables and one supplementary file are available with the online version of this article. All supporting data, code and protocols have been provided within the article or through supplementary data files.

Impact Statement

Clostridium perfringens is a common gut bacterium in both humans and animals and is regularly isolated from healthy individuals. Nevertheless, some strains are more frequently associated with severe diseases and seem to be more virulent than others. Understanding why certain *C. perfringens* strains are highly pathogenic is challenging, in part because existing genome data is often fragmented and lacks detailed clinical context. In this study, we generated a high-quality genomic dataset of 236 *C. perfringens* isolates, many of which derived from animals that died from a confirmed *C. perfringens*-associated enteric disease. Long-read DNA sequencing provided mostly complete *C. perfringens* genomes, which enabled us to perform genomic investigations of *C. perfringens* on an unprecedented level. We discovered 2 previously unknown pore-forming toxin (PFT) genes and 11 variants of known PFTs. Both may directly contribute to disease. Additionally, we assembled 888 complete plasmids and identified 2 previously unrecognized DNA regions that might facilitate the spread of disease-related genes between bacteria. Lastly, we found over 400 gene clusters that may produce compounds allowing for bacteria to compete with other gut bacteria and eventually cause disease. These findings deepen our understanding of *C. perfringens* virulence and plasmid biology and provide a valuable resource for future in-depth studying of this important pathogen.

DATA SUMMARY

The authors confirm that all supporting data, code and protocols have been provided within the article or through supplementary data files.

The whole-genome shotgun project has been deposited at GenBank (BioProject PRJNA1294857) and OpenGenomeBrowser (https://pftclostridia.bioinformatics.unibe.ch) [1]. The BioSample IDs of all generated and used publicly available genomes are provided in Table S1 (available in the online Supplementary Material). Supporting data was summarized in additional supplementary tables. All generated and used commands and codes are shared via GitHub (https://github.com/HalfMoon168/perfringensmigenomics).

INTRODUCTION

Clostridium perfringens is a Gram-positive, spore-forming, obligate anaerobic bacterium with limited aerotolerance, belonging to the phylum *Bacillota* (formerly *Firmicutes*). It is well known to cause diverse diseases in animals and humans including wound infections, gas gangrene, septicaemia, enterotoxaemia, gastroenteritis or food poisoning. In particular, enteric diseases cause significant problems in animal husbandry worldwide as many species are affected by this pathogen [2].

Understanding and diagnosing *C. perfringens* infections is complicated due to its large strain diversity. *C. perfringens* is a natural gut commensal and widely present in the environment [3–6]. *C. perfringens* harbours a plethora of virulence factors, encompassing exotoxins, enzymes and adhesins, of which several have been identified as contributors to disease [7, 8].

C. perfringens is classified into toxinotypes A–G based on the presence of six typing toxins: alpha-toxin, beta-toxin (CPB), epsilon-toxin (ETX), iota-toxin (ITX), enterotoxin (CPE) and necrotizing enteritis toxin B (NetB), with alpha toxin (a phospholipase) being present in all toxinotypes [9]. Certain diseases are linked to specific toxinotypes, and their typing toxins, e.g. CPB (type C strains), are important in animal and human necrotic enteritis (NE) [7], ETX (type D strains) in ruminant enterotoxaemia [10], CPE (type F strains) in human food poisoning [11] and NetB (type G strains) in chicken NE [8]. Type A strains, producing alpha toxin as sole typing toxin, account for most enteric *C. perfringens* isolates. They, however, remain a heterogenous group of strains, impeding an unambiguous discrimination of strains that are associated with particular disease entities. Besides the differential expression of typing toxins mentioned above, *C. perfringens* differs in their capabilities of producing the 31 non-typing toxins identified to date, as well as other factors such as additional enzymes and adhesins that could potentially contribute to their pathogenicity. However, a distinct correlation of particular virulence factors to some animal diseases, such as haemorrhagic and necrotizing enteritis in calves and other species, is still lacking [12].

On a genomic level, *C. perfringens* possesses a small core genome, ensuring high plasticity and adaptation capabilities [5]. Previously, *C. perfringens* was categorized into four to eight clades, depending on phylogenetic comparison strategy and number of available strains [3, 4, 6, 13]. Genes promoting *C. perfringens* virulence were found to be enriched in some clades and depleted in others [3, 5, 13]. However, a distinct clade assignment cannot serve as a diagnostic criterium.

The recent discovery of novel virulence factor homologues suggests that our understanding of *C. perfringens* virulence is incomplete [6, 14, 15]. Therefore, additional, yet-undiscovered factors are potentially still to be discovered. In addition, variants of individual toxins might impact biological function, but a comprehensive sequence-based analysis for most of them is lacking so far in *C. perfringens* [16].

Toxin genes are mostly located on plasmids, which additionally harbour genes leading to antimicrobial resistance (AMR) [4]. *C. perfringens* strains carry up to ten plasmids, some of which can be transferred via conjugation enabling horizontal gene transfer (HGT) and boosting pathogenicity [17]. *C. perfringens* plasmids can be classified into four groups, based on size or conjugation capability [4, 18, 19]: (1) small plasmids (<10 kb); (2) non-conjugative pIP404 (encoding a bacteriocin gen) and phage-like plasmids (encoding phage derived genes); (3) large conjugative plasmids carrying one of the three known conjugation loci Tcp, Pcp or Bcp; and (4) unclassified plasmids not fitting into any of these categories.

Another important aspect of *C. perfringens* pathogenicity is its rapid growth under favourable conditions, which oftentimes occur during rapid environmental changes such as sudden dietary changes, particularly in grazing animals [20]. This can lead to increased virulence factor secretion and augmentation of pathogenicity [2]. Secondary metabolites (SMs), such as siderophores or bacteriocins, which facilitate iron uptake or adversely influence competing bacteria [21, 22], can also contribute to such rapid adaptations of bacteria in general. SMs are mainly encoded by biosynthetic gene clusters (BGCs). BGCs also regulate microbial interactions [23]; however, they are still incompletely studied in *C. perfringens* [21].

Here, we aim to improve our understanding of toxin, plasmid and BGC diversity in animal-derived *C. perfringens* isolates. We determined and analysed 236 high-quality whole-genome sequences including clinic-pathological metadata of *C. perfringens* isolates from so far underrepresented diseases. We provide an extensive comparison of small beta pore-forming toxin (PFT) variants and identify two putative novel toxin homologues and two novel plasmid-conjugation loci. In addition, we present a systematic characterization of *C. perfringens* BGCs, revealing six overrepresented gene cluster families (GCFs) which are associated with some toxinotypes.

METHODS

Biobank

Sequences of isolates were deposited at the NCBI under BioProject number PRJNA1294857. Before sequencing, bacteria were stored in 10% glycerol stock solution at −80 °C, thawed and plated on 5% sheep blood agar plates and incubated overnight at 37 °C under anaerobic conditions. Single colonies demonstrating characteristic biphasic zones of haemolysis were subsequently sub-cultivated in 10 ml brain heart infusion broth (Sigma-Aldrich). High-molecular-weight DNA was extracted using the phenol-chloroform extraction method described by Pitcher *et al.* [24], with adaptations.

Sequencing, genome assemblies and annotations

A detailed summary of the conducted sequencing can be found in File S1. Briefly, gDNA was assessed for quantity, quality and purity using a Qubit 4.0 fluorometer (Qubit dsDNA Assay kit; Q32851, Thermo Fisher Scientific), an Advanced Analytical FEMTO Pulse instrument (Genomic DNA 165 kb Kit; FP-1002–0275, Agilent) and a Denovix DS-11 UV-Vis spectrophotometer, respectively. SMRTbell libraries for 220 genomes were prepared using the PacBio SMRTbell prep kit 3.0, following the manufacturer's protocol with minor adaptations to the DNA shearing using a Covaris g-TUBE. A PacBio microbial multiplexing workflow was used to pool, size select and control the quality of the libraries. Sequencing was performed on both the Sequel II and Revio platforms applying 15 and 24 h as adaptive and standard movie times, respectively. Circular consensus sequences were generated on the instrument with subsequent demultiplexing using SMRT Link (v13) for sequencing on the Sequel IIe/Revio exactly according to the PacBio guidelines. The remaining 16 genomes were prepared using a plexWell 96 Library Preparation Kit for Illumina Sequencing Platform. The library pool was sequenced 150 bp paired-end using a NovaSeq 6000 SP reagents (300 cycles) kit (Illumina, 20028400) on an Illumina NovaSeq 6000 instrument. All base call files were demultiplexed and converted using Illumina bcl2fastq conversion software (v2.20). The long reads were *de novo* assembled using SMRT Tools (v13.0.0.207600), Flye (v2.9.3-b1797) [25] and LJA (v0.2) [26], and a consensus assembly was created manually to obtain the highest number of circularized contigs and to remove spurious contigs and contamination. The short reads were *de novo* assembled using SPAdes (v3.14.0) [27]. The assemblies were annotated using the NCBI PGAP (2023-10-03.build7061) [28]. Quality control was performed using BUSCO (v5.3.2; including database=clostridiales_odb10) [29].

Phylogenetic analysis

The pan- and core genome was determined using panaroo (v1.4.2) [30]. Genes shared by 95% of the isolates were considered core genes. A maximum likelihood phylogenetic tree was inferred from the aligned core genome using IQ-TREE (v2.0.3) including Modelfinder Plus and 1,000 ultrafast bootstrap iterations for branch support [31–33]. GTR+F+R2 was determined as the best-fitting evolutionary model. The resulting phylogenetic tree was annotated using iTOL [34]. To determine the distribution of virulence factors, virulence factor releasing genes, sporulation initiators and spore protection factors, we have created a customized protein data set (makeblastdb) (Table S3) for the use in the TBLASTN option implemented in BLAST+ (v2.14.1) [35–37]. Minimum 80% identity and a difference in length of $\leq$ 5% compared to the query sequence were considered sufficient for presence. To discriminate similar sequences of NetB/E,

perfringolysin O (PfoA) and alveolysin, the minimum identity threshold was set to 90%. Genes for AMR were predicted using ABRicate (v1.0.1) with implemented CARD database (downloaded on 16 November 2022) (https://github.com/tseemann/abricate) [38].

Toxin and toxin variant identification

To search for putative toxin and conjugation protein homologues, we built profile hidden Markov models implemented in HMMER (v3.3.2, hmmer.org) [39]. Therefore, we used the AA sequences of seven proteins (CPB, ETX, CPE, TcpA, TcpF, PcpD4 and PcpB4; Table S3) and created protein profiles using jackhammer running in five iterations over the TrEMBL database (2024_01). Our dataset was then searched with the resulting profiles using the hmmsearch function. Hits were further analysed using SignalP (v6.0) and the conserved domain database [40, 41]. *In silico* structure predictions were performed by AlphaFold3 server [42]. Toxin variants were identified and characterized by multiple sequence alignments using CLUSTAL OMEGA (v1.2.4) [43, 44].

Plasmids

Plasmids were defined as closed extrachromosomal contigs encoding a *rep* gene, as described previously [4], and as putative plasmids in the case of an absent *rep* gene. To assess plasmid relatedness, average NT identity (ANI) was computed amongst all plasmids and putative plasmids using ANIclustermap (v2.0.0), a wrapper that uses FastANI for pairwise comparisons [45]. The resulting Newick-formatted tree was used to infer a dendrogram that was annotated using iTOL [34]. Transmembrane domains (TMDs) were predicted using deepTMHMM (v1.0.42) [46]. Remote homology and protein structure similarity were predicted using HHpred [47, 48]. Phage-derived genes were predicted using VIBRANT (v1.2.1) [49]. The gene synteny of both described conjugation loci was visualized using clinker (v0.0.29) [50].

Biosynthetic gene clusters

Genome assemblies were parsed to GECCO (v0.9.10) to predict present BGCs [51]. Similar BGCs were summarized to GCFs using IGUA (v0.1.0) [52]. GCFs received arbitrary identifiers. Selected BGCs were confirmed with antiSMASH (v8.0) [53]. Ribosomally synthesized and post-translationally modified peptides (RiPPs) were analysed using RiPPMiner (http://www.nii.ac.in/rippminer.html) [54]. BGC hits were compared to the MiBiG database using diamond (v0.9.13) [55, 56]. Cluster gene synteny was then visualized using clinker (v0.0.29) [50].

Data visualization was performed using GraphPad Prism 9, BioRender (accessed in June 2025) and Adobe Illustrator v29.5.1, Adobe Inc.

RESULTS

Biobank and toxinotypes

A total of 236 *C. perfringens* isolates were collected from nine different species (pigs, horses, dogs, cats, alpacas, parrots, cattle, sheep and goats). Two hundred twenty were sequenced using PacBio long-read technology (Assembly levels: 150 complete, 69 chromosome and 1 scaffold). The mean DNA fragment length of 16 samples was significantly below our set threshold of 2.5 kb for PacBio; thus, they were sequenced using Illumina short-read sequencing technology (assembly levels: 3 scaffold and 13 contig). One hundred sixty-seven isolates were generated from animals submitted to our diagnostic necropsy service for evaluation of the cause of disease or death, respectively. Of these, 13 isolates are derived from five different outbreaks of NE in 12 coconut lorikeets (*Trichoglossus haematodus*) in a Swiss Zoo and a private aviary [57]. In addition, we have collected a total of 69 isolates (50 ovine, 7 porcine, 10 bovine and 2 canine) from different sources. In detail, the 50 ovine isolates were collected from stool samples from healthy sheep collected at a Swiss slaughterhouse. The seven porcine isolates were isolated from rectal swabs obtained from healthy piglets during a previous study [58]. The ten bovine (necro haemorrhagic enteritis) and 2 canine (haemorrhagic gastroenteritis) isolates were collected from carcasses submitted to the necropsy service of the Faculty of Veterinary Medicine, Ghent University, Belgium.

The clinical metadata was used to categorize isolates into three different host health statuses: *C. perfringens*-associated diseases, non-*C. perfringens*-associated disease and healthy host (Table S1). For contextualization purposes and for some of the downstream analysis, a total of 383 publicly available *C. perfringens* genomes was included, of which 343 were on draft level and 37 complete. They are derived from at least 25 distinct sources across 23 countries and five continents. All 619 isolates represent all 7 toxinotypes (type A, 358; type B, 10; type C, 34; type D, 62; type E, 2; type F, 113; and type G, 40). Overall, genes encoding 39 different virulence factors, 26 AMR genes, 10 genes involved in sporulation and 9 genes involved in virulence factor secretion were detected (Table S1).

Phylogenetic analysis

To determine the phylogenetic placement of the newly sequenced genomes, the core genome similarity in comparison to previously published *C. perfringens* genomes was compared. The pangenome of all 619 genomes was calculated, identifying a total of 21,021 genes, of which 1,950 are present in more than 95% of the isolates and thus classified as core genes (Fig. 1a, left). A Heap's law model fitted to the mean pan-genome curve yielded an α of 0.279, indicating an open genome, which is consistent with previous findings [5]. This was further supported by the empirical new gene discovery curve, which gradually declines but shows persistently one approximate new gene per three new genomes on average (Fig. 1a, right). Important elements of the defined core genome are six genes encoding for virulence factors *cpa*, *ccp*, *ccpA*, *fbpB*, *nanH* and *mprF* and all analysed genes that are involved in *C. perfringens* sporulation and virulence factor secretion pathways. The core genome was used as a basis to infer a phylogenetic tree of newly sequenced and publicly available genomes (Fig. 1b). The tree can be divided into five distinct phylogenetic clades, consistent with previous findings [3]. The newly introduced genomes of this study are grouped to all five clades, most often to clades III–V, yet without a clear disease-clade association. Newly sequenced type C strains (26 pigs and 1 cattle) display a similar core genome and are allocated to clade IV. In contrast, publicly available type C genomes (n=7) belong to clades I, III, IV and V. Only the two porcine isolates (JGS1495 and NCTC10719) cluster together with our genomes. Porcine type A genomes provided by us (n=7) and publicly available ones (n=39) can be grouped into all clades except clade V, with most of them in clade I. We provide 19 additional equine and canine type F strain sequences from animals with foal necrotizing enteritis (FNE) and canine acute haemorrhagic diarrhoea syndrome (AHDS), respectively. NetF- and alveolysin-positive strains clustered with previously reported clade V isolates [3]. In contrast, four isolates from a single dog lacking both genes grouped within clade II. Overall, canine or equine strains positive for netF, alveolysin or enterotoxin were significantly more often assigned to clade V (Fisher's exact test; all P<0.05). Ten Belgian isolates from eight veal calves with necro-haemorrhagic enteritis belonged to toxinotypes A, C, D or F. Despite similar disease metadata, they clustered into three distinct clades (III–V). Type D genomes of small ruminants provided by us (n=45) plus publicly available (n=8) cluster entirely into clade III. Remaining small ruminant-derived *C. perfringens* genomes (89 newly sequenced and 30 public) encompass toxinotypes A, B, C and F and cluster in clades II to V. All 13 coconut lorikeet-derived type A genomes are dispersed over clades II–V; roughly half of them (n=7) harboured a *beta2* toxin variant. This frequency is like that observed in further type A isolates (172/345) and did not show a significant host-associated enrichment (Fisher's exact test, P=1).

Collectively, this supports the genomic organization of *C. perfringens* into five phylogenetic clades. Whereas some genomes with similar metadata cluster together and suggest close phylogenetic relationships, others unexpectedly demonstrate significant differences and lack conserved virulence factor genes in their accessory genome.

Identification of putative PFT homologues

One obvious differentiation of *C. perfringens* toxinotypes B, C, D, F and G, which are clearly associated with particular enteric disease, is their capability of secreting one or more exotoxins belonging to the group of small beta PFTs [9]. Yet, novel βPFT homologues are routinely discovered [6, 14, 15]. We thus speculated that potentially not yet discovered members of the haemolysin and aerolysin βPFT families might be found in our isolates. We searched all available draft and complete genome sequences (619) for potentially novel aerolysin or haemolysin βPFT homologues. We could identify two putative PFT homologues in the strains sequenced in this study, one of which was also identified in publicly available sequences. The first homologue is encoded in the newly sequenced isolate 138_2_JF2-p1.1 (small intestine, healthy sheep collected at slaughter, type A, Switzerland) where it is located on a circular PCP plasmid (138_2_JF2-p1_scf5). This haemolysin βPFT homologue consists of 304 AA, including a 24-residue signal peptide, followed by a leukocidin domain of the leukocidin superfamily (PFAM07968) from residues 31–303. The sequence shares 61.2% identity with NetF and 62% with the NCTC8081_14145 haemolysin homologue (Fig. 2a). Based on their sequence similarity to known haemolysin βPFTs and its placement in a maximum likelihood phylogeny, the protein was designated as necrotizing enteritis toxin H (NetH) (Fig. 2a). The gene is flanked by a putative transposon and a recombinase family member. Several genes on this plasmid were predicted to belong to a lysogenic prophage (Table S2). The second homologue was identified in the newly sequenced isolate 92_JF2-p1.1 (Rectal swab, healthy piglet, type A, Switzerland) on a circularized TCP-type plasmid (92_JF2-p1_scf7). The aerolysin βPFT homologue consists of 281 AA that includes a PFM-epsilon-toxin-like domain (cd20223) spanning residues 50–182. There was no signal peptide predicted. The sequence shares 46% identity with epsilon domain protein A (EdpA) and 40% with EdpB (Fig. 2d). Based on its sequence similarity to these known aerolysin βPFTs and its placement in a maximum likelihood phylogeny, we refer to this protein as EdpC (Fig. 2d). Several neighbouring genes are annotated as transposons and recombinases. Furthermore, this plasmid harbours a bacteriocin gene from the circularin A/uberolysin family (Table S2). Interestingly, we identified homologues but not annotated EdpC sequences in genomes of faeces-derived type A strains from four healthy pigs from China (1805-113, 1805-50, 1P32 and 1P6) [59]. They are localized on similar, yet linear contigs of ~6–10 kb size that could not be circularized using circlator [60].

We predicted the *in silico* structures of both NetH and EdpC using AlphaFold 3 (Fig. 2b, e). Both monomers are dominated by beta barrels and indicate a domain organization as described for the archetypal *C. perfringens* beta-toxin (haemolysin) and epsilon toxin (aerolysin), respectively [61, 62].

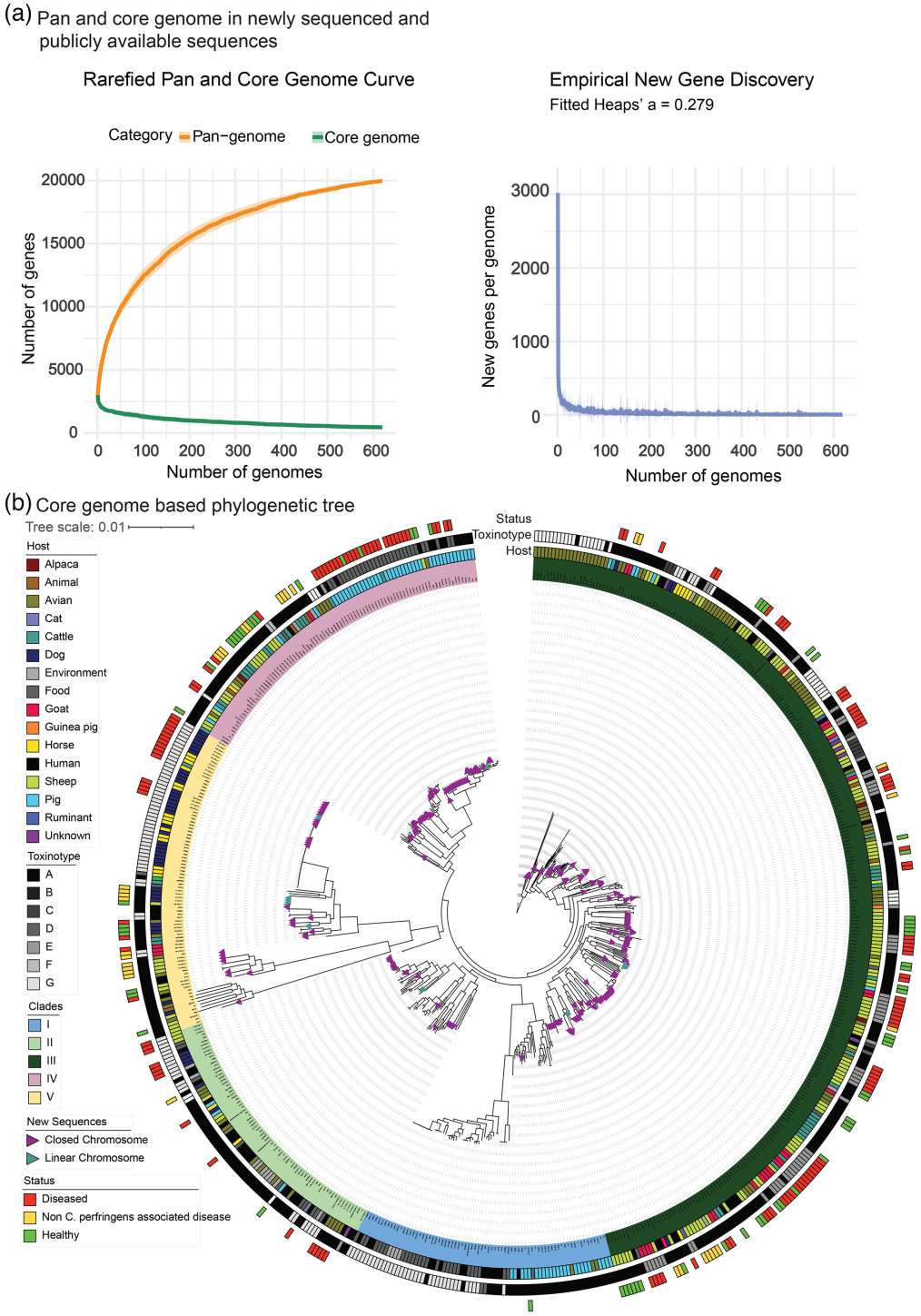

Fig. 1. Core genome-based phylogenetic tree. (a) Pan and core genome comparisons in all 619 evaluated sequences. (Left) The number of pan genome (dark orange) and core genome (sea green) genes was calculated across 30 randomly sampled genome orderings. The resulting curves are shown with shaded ribbons representing ±1 SD. (Right) Average number of new genes contributed per genome across the same replicates. A Heap's law model fitted to the pan-genome curve yielded α=0.279, consistent with an open genome model. (b) Phylogenetic tree of 619 *C. perfringens* genomes including the 236 newly sequenced isolates of this study. Circularized and linear genomes sequenced in this study are indicated by purple or teal triangles, respectively. Clade affiliation is indicated by isolate identifier colour code (I, light blue; II, light green; III, dark green; IV, rose; and V, yellow). The tree was annotated with the host of each isolate (inner ring), toxinotype (second ring) and disease status for all newly sequenced isolates (outermost ring). Tree scale shows bp substitution per sequence site.

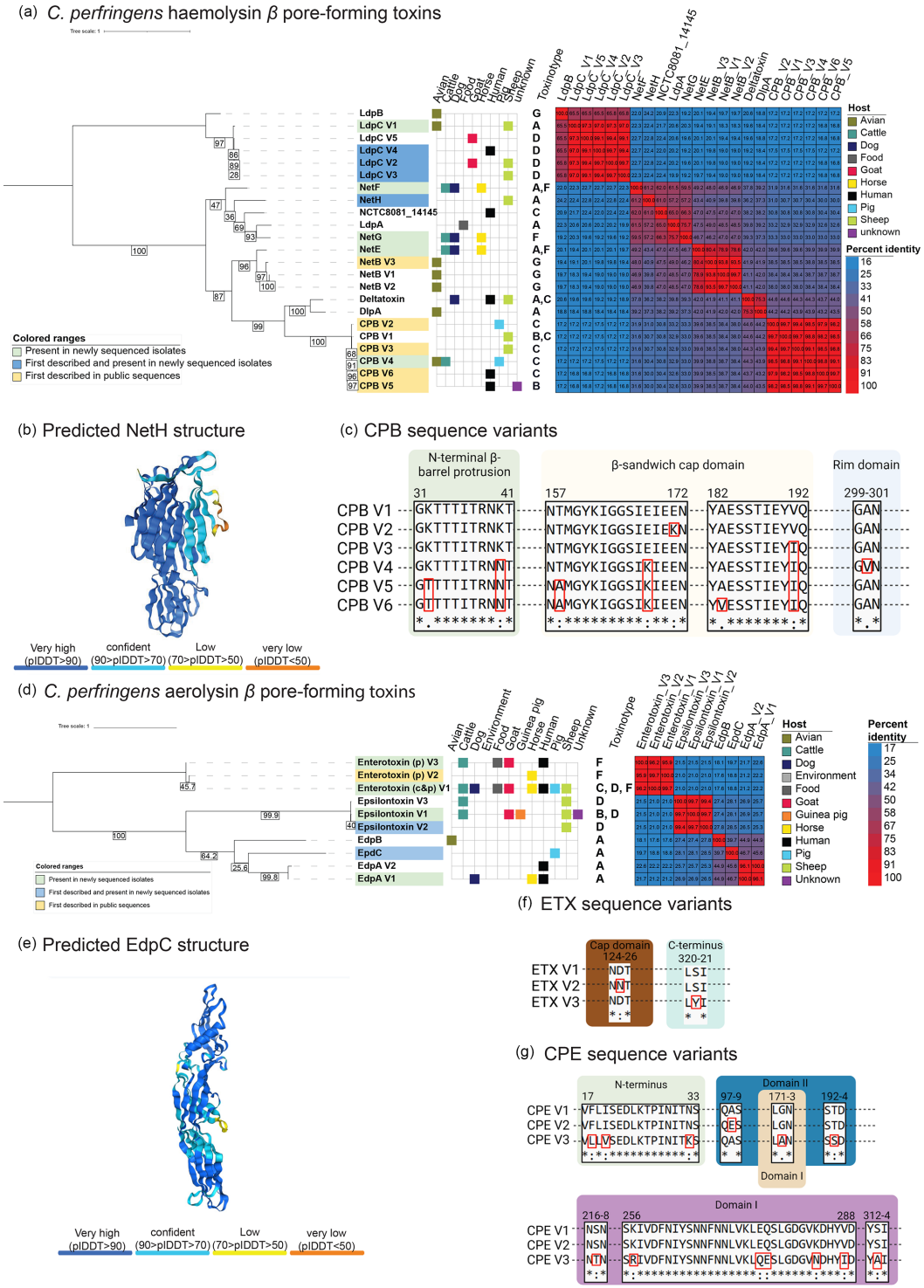

Fig. 2. *C. perfringens* haemolysin and aerolysin *β*PFTs. (a) Pairwise comparisons of known and newly identified *C. perfringens* haemolysin *β*PFT sequences were used to generate a maximum likelihood tree and percent identity matrix. Trees are annotated with host species and toxinotypes. Newly identified sequences are highlighted in light blue (found in our isolates) or yellow (from public genomes only); previously described sequences present in our genomes are shown in light green. Species colour code as in Fig. 1. Branch lengths are supported by 1,000 bootstrap iterations. (b) Structure of NetH predicted using AlphaFold 3. (c) AA sequence variants of beta toxin, highlighting affected domains. (d) Pairwise comparisons of known and newly identified *C. perfringens* aerolysin *β*PFT sequences analysed and depicted as in (a). (e) Structural appearance of EdpC predicted using AlphaFold. (f) AA sequence variants of epsilon toxin. (g) AA sequence variants of enterotoxin.

Together with recent publications discovering *C. perfringens* toxin candidates *in silico* and *in vivo* [6, 14, 15], this underscores the versatility of toxins secreted by different *C. perfringens* isolates.

Toxin variants of small *β*PFTs

We next sought to comprehensively analyse the sequence variability of known haemolysin (*n*=11) and aerolysin (*n*=4) *β*PFTs in *C. perfringens*. Therefore, all available sequences (*n*=619) were evaluated. Sixteen sequences were found that retained the overall structure of a known toxin but differed at least by one AA from their WT sequence (referred to as V1) and thus were defined as variants of a toxin (referred to as VN). The presence of the single known variants of CPB, NetB, CPE, ETX and EdpA was confirmed. Interestingly, a total of 11 novel variants were found in the examined sequences: 4 for CPB and LdpC and 1 for each NetB, CPE and ETX. In contrast, LdpB, NCTC8081_14145, LdpA, NetEFG, deltatoxin, DlpA and EdpB were highly conserved.

Haemolysin *β*PFTs

Five sequence variants (V2-6) of CPB were detected. Porcine type C strains encoded V2 and V4, while human type C strains harboured V5 and V6. V4, which has been described previously [16], is the most widely distributed variant found in birds, cattle and pigs. Remaining variants were only found in one species (Fig. 2a). Several mutations were present in key structural domains (Fig. 2c). In the N-terminal protrusion, variants V5 and V6 exhibited Lys32Thr, while V4–V6 displayed a Lys40Asn mutation. Five mutation sites were observed in the cap domain, which is important for oligomerization: Thr158Ala (V5-6), Glu168Lys (V4-6), Glu171Lys (V2), Ala182Val (V6) and Val191Ile (V3-6) [61]. Lastly, one mutation (Ala300Val) was identified in the rim domain of V4, the region likely to be involved in receptor binding and membrane interaction.

Two variants were detected for NetB, all in chicken-derived isolates. V2 differs from V1 by a single substitution (Ala168Thr), identical to a known variant retaining cytotoxicity [63]. V3 displays a total of 20 substitutions affecting several domains: signal peptide (Ile9Thr and Val16Ile), beta-sandwich domain (Ile57Phe, Lys68Arg, Trp70Leu, His83Asn, Val122Ile, Asn266Asp, Glu269Asn, Ser302Asn and Asn305Thr), rim domain (Asp100Gly, Phe104Tyr, Asp241Ala, Leu298Pro and Thr301Ser), pre-stem domain (Asp146Asn and Ser150Thr) and in a disordered region aside the pre-stem domain (Gly167Ser and Asn175Ser). Additionally, there was one deletion of Ser300. V3 was present on four plasmid-encoding contigs (CHD9634P_seq17, CHD9638P_seq17, CHD9628P_seq18 and CHD9644P_seq17), with all four isolates also carrying WT NetB (V1). None of the affected residues is in close vicinity of residues described to be key for cytopathic effects [63].

LdpC had four variants, and the WT V1 was detected in type A strains, whereas its variants V2–5 were detected in type D strains. Interestingly, all variants differed in the same seven residues from the WT, namely, Phe19Leu, Ser77Asn, MetAla85-86ThrAsp, Glu92Ala, Ile195Val and Lys226Arg. In addition, there were variant-specific substitutions for V2 (Leu157Ser and Lys312Glu), V3 (LeuPro157-158SerSer and Lys312Glu), V4 (Asn110Asp, Leu157Ser and Lys312Glu) and in V5 (Lys104Glu and Leu157Ser). V2 was found on two TCP plasmids (118_JF2-p1_scf10 and 17_JF2-p1_scf11; goat/sheep, enterotoxaemia). V3 was present on linear contigs (57_3_JF2-p1.1_scf2 and 57_1_JF2-p1_scf2; sheep, enterotoxaemia), containing TCP conjugation genes. V4 was located on a circular TCP plasmid (pCN3842etx), from a human-derived type D strain. Interestingly, V2, V3 and V4 reside within a RiPP cluster with an enterocin as precursor peptide (GCF1). V5 was located on a linear contig derived from a caprine type D with several TCP conjugation genes.

Aerolysin *β*PFTs

We detected two CPE sequence variants. The archetypal V1 was located on the chromosome and on plasmids. V2 (Ala98Glu), affecting an alpha helix of domain II, was located on a linear 6 kb contig derived from a horse with FNE (JFP981). Ala98Glu resides near two glutamic acid residues (Glu94 and Glu110), which determine the inter-monomer conduit [64, 65]. V3 (Phe18Leu, Ile20Val, Asn32Lys, Gly172Ala, Thr193Ser, Ser216Thr, Lys257Arg, Lys283Asn, Val287Ile, Ser313Ala and one insertion GluGln276GlnGlu), encompassed all domains and the N-terminus. This variant has been previously described in type E isolates [66]. It is located on two clonal Pcp-type plasmids derived from a goat which was diagnosed with coccidiosis (160_1_JF2-p1.1; 160_4_JF2-p1.1). The same plasmid contig additionally harbours the virulence factor genes *iap* and *ilpB* as well as the GCF17 (see 9.5). Further isolates with V3 variants were derived from cattle (a508.17), human (a515.17; Q061.2) and food (CP_PB-1).

Two ETX sequence variants were detected. V2 (Asp125Asn, beta hairpin domain) was located on two type D sheep-derived linear contigs (57_3_JF2-p1.1_scf3 and 57_1_JF2-p1_scf3) containing TCP genes [62]. Mutations in this domain of the archetype aerolysin toxin have been described to lead to a compromised domain unfolding and membrane insertion [67, 68]. Interestingly, the same isolate (sheep, enterotoxaemia) harboured LdpC V3. ETX V3 (Ser321Tyr, C-terminus, previously reported) was found in ruminant (NCTC8346) and ovine (CN1675) derived strains [69].

In addition to the continued expansion of *C. perfringens* toxin homologues, we show a great variability of toxin primary structures. Different variants present in isolates from different host species might influence their overall biology through altered biochemical and physical properties, e.g. in the interplay with their cellular receptor.

Plasmids

Accurately predicting circularized plasmids from short-read data remains a challenging task [70]. In the case of the common virulence-promoting *C. perfringens* plasmids, there has been a constant encroachment for in-depth research. Less than 100 *C. perfringens* plasmids were unequivocally established, with 55 that could be circularized recently using long-read sequencing [4]. In our extensive long-read sequencing efforts of 220 isolates, 801 closed, extra-chromosomal contigs carrying one *rep* gene were identified, classifying them as plasmids (Table S2). Additionally, 87 closed extra-chromosomal contigs lacked identifiable *rep* genes but were considered putative plasmids based on their structure. Up to ten plasmids are found simultaneously in one isolate. Isolates that were typed as C, D and F harbour significantly more plasmids, compared to type A isolates. Especially type C strains harbour ≥5 plasmids. Since we had only one type E isolate, no statistically reliable statement could be made for this toxinotype (Fig. 3a).

Plasmid size ranged from 1.5 to 163.9 kb, with two outliers (putative plasmids with phage-derived genes) measuring up to 916 kb (Fig. 3b).

Sixteen virulence factors, including both newly identified toxin homologues and three AMR loci (*lnuP*, *tetAB* and *ermQ*), were carried on plasmids, except for two clonal *lnup* and one *ermq* encoding putative plasmids. 43.5% of all plasmids' ORFs are annotated as hypothetical proteins based on our PGAP annotation. Phage-derived genes and bacteriocins were identified in 145 and 187 contigs, respectively. Fourteen distinct BGCs were present on 93 plasmids. All plasmids and putative plasmids were categorized into six previously published categories, namely, 'TCP', 'PCP', 'BCP', 'phage-like', 'pIP404' and 'small' (Fig. 3c) [4].

To assess plasmid relatedness, we compared all plasmids and putative plasmids with previously published circularized plasmids [4], using ANI to construct a genomic distance dendrogram (Fig. 3c). Three clusters of plasmids strongly linked to category as defined before [4] were found. Clusters I and II were dominated by Tcp and Pcp/pIP404 plasmids, respectively, and showed high intra-cluster conservation (mean ANI: 93.2% and 80.7%). In contrast, cluster III is heterogeneous, containing PCP, BCP, phage-like, small and unclassified plasmids, including 24 unique plasmids.

Recently, BCP was described as a promising new system for *C. perfringens* plasmid conjugation. It is based upon a pair of VirD4 and VirB4 homologues as core conjugative genes, both only distantly related (<30% sequence identity) to their TCP and PCP counterparts. Instead, they showed a higher similarity to VirD4/B4 homologues of a *Clostridium botulinum* plasmid, which also served as a category template [4]. This discovery indicated that there might still be undiscovered conjugation systems in *C. perfringens* plasmids. We thus scrutinized our plasmid collection for pairs of VirD4/B4, distantly related to Tcp, Pcp or Bcp homologues. We discovered seven and one plasmids in cluster III with a VirD4/VirB4 pair each, matching these criteria. By analogy to the BCP conjugation system, we propose designating these putative conjugation loci as BaCP (similar to *Clostridium baratii* locus) and TeCP (similar to *Clostridium tertium* locus), representing potentially novel conjugation systems in *C. perfringens*.

BaCP was identified on seven plasmids (112_1_JF2-p2_scf2, 48_2_JF2-p1_scf6, 48_3_JF2-p1_scf1, 48_4_JF2-p1_scf2, 48_5_JF2-p1_scf2, 41_JF2-p1_scf7 and 122_JF2-p1_scf2) (Fig. 3d). Comparative analysis revealed moderate sequence conservation to the *C. baratii* BoNT/F7 neurotoxin-encoding plasmid pNPD11_1 (CP014203) [71]. Specifically, the *C. perfringens* VirD4 and VirB4 homologues, essential for plasmid conjugation, shared 43–44% and 47.6–49.1% sequence identity, respectively, with their *C. baratii* counterparts. Further essential genes possibly forming a type IV secretion system were present in this locus, namely, TcpE and TcpH (VirB6) [19, 72–75]. Notably, the identified TcpH homologue differed from the canonical TcpH, as it possessed seven TMDs but lacked the conserved $_{242}$VQQPW$_{246}$ motif [74]. The TcpE homologue has two TMDs and a TcpE domain (PF12648) [74, 75]. A C40 family peptidase was located downstream of VirB4. Structural analysis suggested that it encoded a putative peptidoglycan hydrolase (*E*-value: 2.3e-10), which may function analogously to TcpG or TcpI, both of which act as peptidoglycan hydrolases [76]. Upstream of TcpE, three pilins (PF18895) were identified, each containing two TMDs. Two of these resembled TrbC/VirB2 pilins (PF04956, *E*-values: 8.3e-7 and 1.7e-5), suggesting a potential role in host cell adhesion, similar to the VirB2 adhesin [77]. Downstream of VirD4, a hypothetical protein belonging to the replication-relaxation family (PF13814) was detected. This family includes proteins essential for plasmid replication and DNA relaxation [78], suggesting that this protein might function analogously to the TCP relaxase TcpM [79]. Further downstream, a class D sortase was identified, a feature also found in the *C. perfringens cnac* region of the TCP locus [80]. Nearby, a SpaA isopeptide-forming pilin-related protein was detected. Based on the presence of clumping factor A domains (COG4932), this protein likely belongs to the MSCRAMM family and may function similarly to CnaC [80].

Taken together, we found an elaborate conjugative module, with distinct VirD4/VirB4 homologues accompanied by several accessory proteins and structural genes.

In contrast to BaCP, TeCP was detected only on a single ovine-derived plasmid (107_JF2-p1_scf3). We did not find publicly available (NCBI *nr* database; 20.11.2024) plasmid contigs, simultaneously encoding for comparable VirD4/B4 homologues. The VirD4 homologue showed the highest similarity (45.2%) to a type IV secretion system conjugative DNA transfer protein on a linear *Clostridium cylindrosporum* contig (WP_048570114.1). The VirB4 homologue showed the highest similarity to a FtsK/SpoIIE domain-containing protein from a linear *C. tertium* contig (50.8%; WP_250456040.1) (Fig. 3e). In addition, the locus

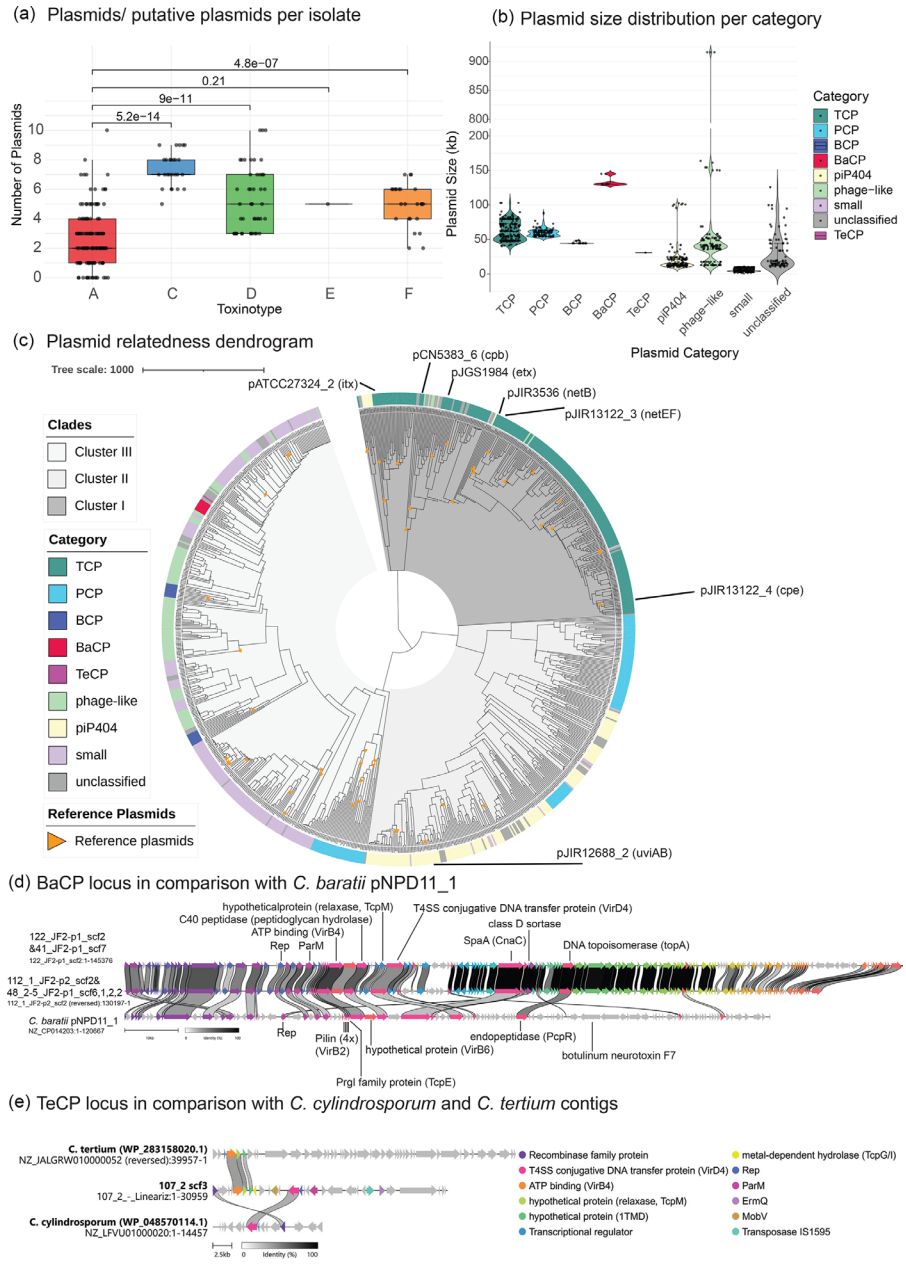

Fig. 3. Analysis of plasmids of *C. perfringens*. (a) Distribution of the total 888 plasmid/putative plasmid counts per isolate found in this study. Each box plot states the interquartile range with the median as a black line. The plasmid counts for each individual isolate are indicated as jittered dots. Statistical comparison between toxinotype A and the remaining groups was performed using the Wilcoxon rank-sum test. Resulting *P*-values were added on top and in between the compared boxes. The y-axis displays the number of plasmids in a single isolate. All analyses were performed using R (v4.4.0) with implemented ggplot2 and ggubr packages. (b) Plasmid size distribution by category, according to the classification of Gulliver et al. [4]. Categories are shown on the x-axis and contig sizes (in kb) on the y-axis. Each distribution is represented as a violin plot, with the median (horizontal bar) and interquartile range (2nd to 3rd quartile; vertical bar) indicated. Individual data points are shown as jittered dots. The y-axis includes a break between 200 and 800 kb to improve visualization. The plot was generated in RStudio (v4.4.0) using the ggplot2, ggbreak and dplyr packages. (c) A dendrogram displaying genomic distance of all 888 detected plasmids and putative plasmids including 47 publicly available reference plasmids (orange triangles) based on ANI as described in methods. Plasmid categories are plotted as coloured rings, with the novel conjugation loci indicated in red (BaCP) and magenta (TeCP), respectively. Tree scale shows bb substitution per sequence site. (d) Clinker (50) representation of the novel conjugative locus BaCP in comparison with plasmid pNPD11_1 (CP014203) of *Clostridium baratii*. Homologous genes are connected by bars, showing sequence similarity. (e) Clinker (50) representation of the novel conjugative locus TeCP in comparison with two contigs of *Clostridium tertium* (WP_250456040.1) and *Clostridium cylindrosporum* (WP_048570114.1). Homologous genes are connected by bars, showing sequence similarity.

encoded a putative metal-dependent hydrolase (PF04307.19, *E*-value: 1.3e-16) with four TMDs, which may function similarly to TcpG/I peptidoglycan hydrolases during DNA transfer. A MobV family protein (PF01076.24, *E*-value: 3.3e-28) was additionally found and could facilitate mobilization, similar to the MobC homologue proposed in BCP [4]. Lastly, a hypothetical protein belonging to the replication-relaxation family (PF13814) was present, which could function analogously to the relaxase TcpM that processes the origin of transfer in TCP. Upstream of these putative conjugative genes, *ermQ*, an antibiotic resistance gene, flanked by a IS1595 family transposase, was present. A ParM variant was detected in both BaCP and TeCP, indicating the presence of a type II or ParMRC plasmid partitioning system, as commonly observed in conjugative *C. perfringens* plasmids (Fig. 3d) [81].

In contrast to BaCP, TeCP is more minimalistic but retains the core and some accessory genes to potentially enable conjugation.

Altogether, our comprehensive analysis expands the known diversity of *C. perfringens* plasmids and reveals the presence of previously uncharacterized conjugation systems. The identification of BaCP and TeCP loci, with distinct VirD4/VirB4 homologues and associated mobilization and structural genes, suggests the existence of novel conjugation machinery that may drive HGT. These findings provide new insights into plasmid evolution and conjugation dynamics in *C. perfringens*, with important implications for understanding virulence and AMR spread in this pathogen.

Biosynthetic gene clusters

In our plasmid collection, we have noted numerous BGCs and bacteriocins (Table S2). Given the critical role of plasmid-mediated pathogenicity in *C. perfringens,* we further investigated the distribution of BGCs and GCFs. All available draft and complete genomes (*n*=619) were screened for BGCs using GECCO [51]. In total, we identified 414 distinct BGCs across 314 genomes, including some that were present twice per strain (Tables S1 and S4). Of these, 189 BGCs were present in newly sequenced strains. They were categorized into four classes: RiPPs (*n*=240), saccharides (*n*=146), non-ribosomal peptides (*n*=17) and 11 unclassified clusters. Based on gene synteny, we defined 32 GCFs, none of which exhibited >50% identity to characterized clusters [56].

Next, we wanted to evaluate potential associations between GCFs and toxinotypes. To avoid technical biases through differing sequencing technologies, the analysis was restricted to the high-quality PacBio genomes generated in this study (*n*=220). Within these, 194 BGCs were assigned to 18 GCFs. While the majority of these GCFs consist of few (≤5) BGCs, 162 BGCs (84%) occurred in only 6 GCFs. We were interested in whether these clusters showed a toxinotype preference and analysed their distribution throughout four toxinotypes (A, C, D and F). GCF 12 was present in all toxinotypes. Type A and C strains harboured to some extent GCF12, 14 and 23, type D strains encoded GCF6, 12 and 22 and in type F strains, we encountered GCF12, 14 and 17. Regarding GCF14 and GCF23, most occurrences were restricted to type C strains. Most of these clusters encoded bacteriocin-related genes. The proportion of strains within each toxinotype that harbour a given GCF was visualized as a heatmap (Fig. 4a).

GCF6 was predicted to encode a hybrid saccharide-RiPP BGC. It harboured multiple genes encoding sugar-modifying enzymes, alongside genes indicative of lasso peptide synthesis (RiPP). These include a putative core peptide and two adjacent lasso-peptide effector genes, one of which encodes a lasso peptide biosynthesis chaperone of the PqqD family with a stand-alone lasso RiPP recognition element. We hypothesized that GCF6 may produce a sugar-modified lasso peptide. It is predominantly found in type A and D strains and absent in type C and F strains (Fig. 4b). The cluster is only present on chromosome equivalents. Apart from GCF6, there are 23 BGCs in 5 GCFs predicted as saccharide clusters, without obvious toxinotype preference.

GCF12 consisted of RiPP BGCs with two central precursor peptides containing small double-glycine leader motifs (Fig. 4c). One peptide belongs to the Pfam family PF10439 and shares 60.34% sequence identity with a not further characterized bacteriocin of *Clostridioides difficile* (EJO5349538.1). The second, a small hydrophobic peptide with a GG motif, is encoded adjacent to the first, consistent with a class IIb bacteriocin that requires two-component activity [82]. GCF12 is present in all four toxinotypes but is relatively enriched in type C and F strains, compared to type A and D strains. Amongst nine porcine type C strains with NE, GCF12 was detected on pIP404-type plasmids in 89% of cases. In type F strains, GCF12 was found in isolates from dogs with AHDS (*n*=11), horses with FNE (*n*=4) and cattle with necrohaemorrhagic enteritis (*n*=2), with 65% located on pIP404-type plasmids. In type F strains, two additional double-glycine leader peptides, annotated as hypothetical proteins, were also encoded, suggesting possible auxiliary or alternative bacteriocin functions.

GCF14 represented a more complex variant of GCF12, incorporating a ComC/BlpC-family peptide pheromone/bacteriocin and a response regulator (Fig. 4c). It was enriched in porcine type C strains, except for single occurrences in type A and F strains. Most GCF14 clusters were located on pIP404-type plasmids (15/20, 75%) that appeared to lack additional known virulence factors.

GCF22 was another variant of GCF12 and GCF14, containing a conserved ABC transporter and recombinase family gene. The central precursor belonged to the PF10439 family, as seen in GCF12 and GCF14. In contrast, its upstream hypothetical protein had three instead of two TMDs and lacked the double-glycine motif (Fig. 4c). It was exclusively found in a few type D strains, 83% of which carried it on pIP404-type plasmids.

GCF17 encoded a class I lanthipeptide cluster, comprising a predicted LanA precursor peptide (49 AA), a dehydratase (LanB) and cyclase (LanC). The precursor was from the gallidermin/nisin family and exhibited low sequence similarity to *streptin* from

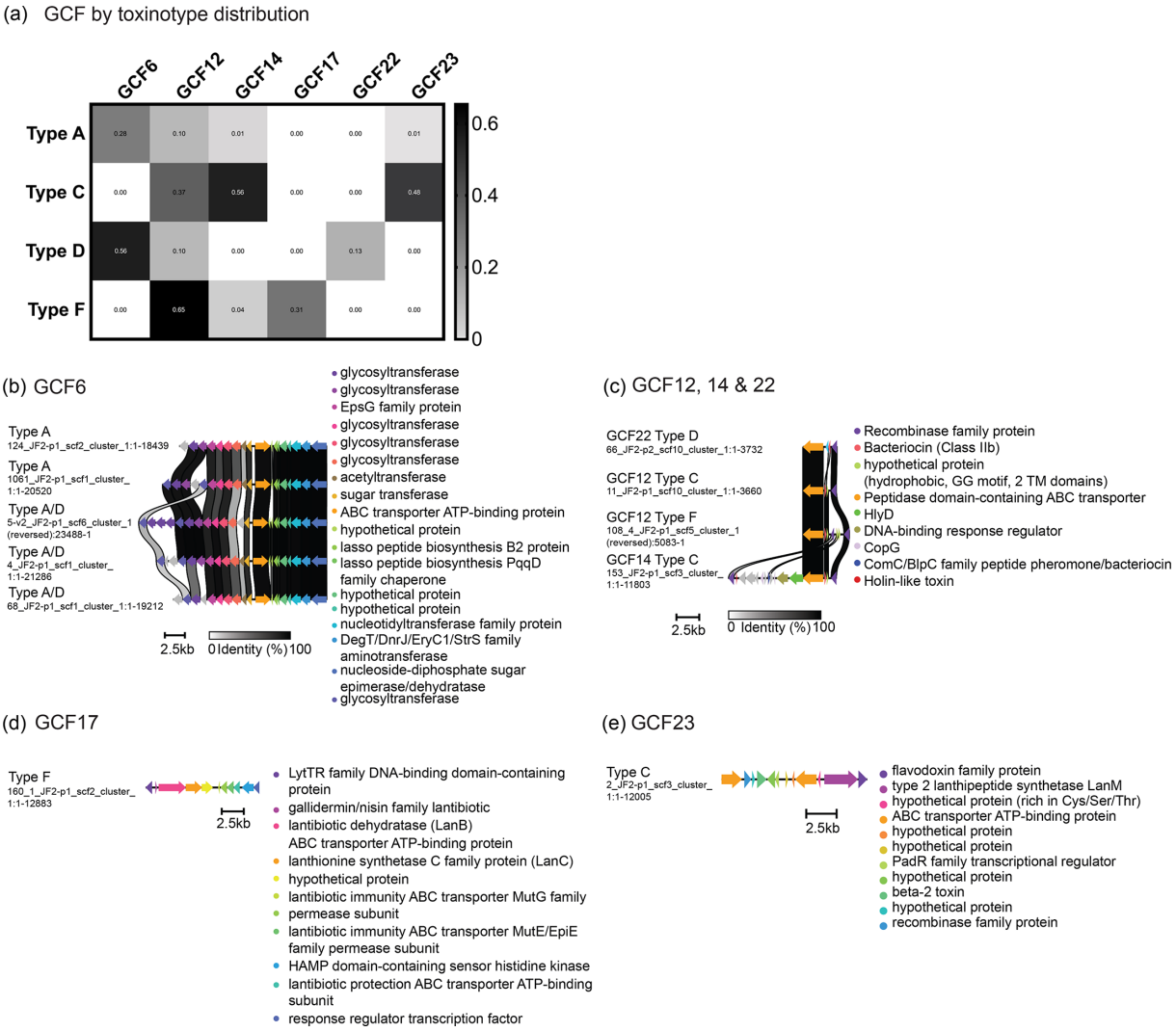

Fig. 4. BGCs of *C. perfringens*. BGCs were grouped into GCFs and GCFs occurring in >5 complete genomes are shown. (a) Heatmap of GCF distribution per toxinotype. Scale indicates proportion of GCF-positive isolate per toxinotype. (b–e) Clinker (50) representation of gene synteny of the respective NT representatives of each GCF for (b) GCF6; (c) GCF12,14 and 22; (d) GCF17; and (e) GCF23. Where applicable, best match per cent identity of conserved genes is indicated by bars.

Streptococcus pyogenes (39.02% identity; P0C0H8) and *nisin A* of *Lactococcus lactis* (34.3% identity, BGC0000535) (Fig. 4d). It was mainly found in type F strains and resides on TCP-type plasmids (7/8, 88%), often alongside the plasmid-encoded *cpe* gene and phage-derived genes. Two carrier plasmids (160_1_JF2-p1_scf2 and 160_4_JF2-p1_scf2) also encoded the virulence genes *iap* and *ilpB*.

GCF23 encoded a class II lanthipeptide cluster with three putative lanthipeptide precursors and a predicted LanM synthetase. Notably, the *cpb2* gene was present, encoding the consensus *β*2-toxin variant (Fig. 4e). The cluster was mostly restricted to porcine type C strains (n=12), where it was located on PCP-type plasmids in 92% of cases. These strains originated from pigs across seven farms and were isolated over a 16-year period (1999–2015). Remaining occurrences of GCF23 were in two closely related bovine type C strains and a porcine type A strain (associated with diarrhoea).

Together, these findings highlight a diverse and previously uncharacterized landscape of BGCs in *C. perfringens*, some of which are enriched in specific toxinotypes and frequently plasmid borne. Several GCFs, particularly GCF12, GCF14 and GCF23, encode putative class IIb bacteriocins and are highly associated with the pIP404 plasmid backbone, suggesting a potential role in strain competitiveness or host adaptation. Moreover, the discovery of sugar-modified lasso peptides (GCF6) and both class I and II lanthipeptides (GCF17 and GCF23) underscores the biosynthetic versatility of *C. perfringens* and its capacity to evolve complex antimicrobial systems.

DISCUSSION

In this study, we provide the genomes of 236 new *C. perfringens* isolates from the intestinal tract of different animal hosts, most of which are circularized, signifying a valuable increase of complete publicly available animal-associated *C. perfringens* genomes. Recent publications noted a bias of publicly available *C. perfringens* isolates towards clinically relevant strains [13]. Therefore, we chose to add a substantial amount (*n*=96) of isolates derived from healthy or animals with diseases not caused by *C. perfringens*. Our phylogenetic analysis highlights relatedness between some isolates with similar metadata: NE in pigs, AHDS and FNE in dogs and horses and enterotoxaemia in small ruminants. Despite a relatively small geographical origin of these strains, this demonstrates their role as disease-causing pathogens in economically important diseases. Isolates from haemorrhagic enteritis in calves and NE in coconut lorikeets (*T. haematodus*) turned out to be distantly allocated, supporting the hypothesis that different *C. perfringens* strains contribute to the development of these diseases and that other consistent factors are likely to be involved in these entities [83]. Thus, our results confirm previous publications [3–5] that phylogenetic placement of the genome does not serve as a robust distinction between commensal and pathogenic *C. perfringens* isolates.

The prediction of two potential toxin candidates (NetH, a haemolysin, and EdpC, an aerolysin) should encourage to scrutinize upcoming *C. perfringens* sequences for novel toxin candidates, as more toxin homologues could still be discovered. The presence of transposon, recombinase and prophage encoding genes flanking NetH and EdpC suggests that the toxin homologues were introduced into these regions via HGT. We could identify EdpC only in porcine-derived *C. perfringens* isolates; however, more sequences should be investigated to substantiate the hypothesis that EdpC might be a porcine-associated *C. perfringens* gene/virulence factor.

We also comprehensively compared 22 toxin sequence variants. Based on our analysis, the six CPB variants suggest variant-to-species adaptation. A previously characterized CPB variant (corresponding to V4 in this study) containing an A300V mutation showed significantly increased cytotoxicity in HUVEC cells in comparison to WT CPB (CPB_V1) [16]. Additional mutations in this variant were found to influence trypsin sensitivity, highlighting the functional impact of CPB sequence variation [16]. In *Staphylococcus aureus* leukocidins, it is known that toxin and receptor variability significantly influence toxin–receptor interactions [84, 85]. Notably, the CPB receptor, CD31, exhibits interspecies variation, and in humans, different haplotypes with synonymous mutations have been described [86]. It thus remains to be investigated whether sequence variants of CPB relate to receptor variability in different hosts.

In contrast, CPE (affecting a wider range of species by targeting a range of eight different claudins) is relatively conserved [87]. CPE V3 showed many mutated AA close to key receptor binding residues [64, 66, 88–91]. Whether this variant is still functional remains to be evaluated. Other toxin sequences, such as NetE, F and G, found in closely related canine and equine type F strains and a bovine strain, are highly conserved, lacking non-synonymous mutations confirming previous findings [92]. The 11 sequence variants of aerolysin and haemolysin *β*PFTs described here underscore the importance of genomic analyses as a basis for further *in vitro* validation to assess the functional relevance of these variants.

We additionally contribute a plasmid dataset of 888 circularized extra-chromosomal contigs, which signifies a substantial increase in well characterized *C. perfringens* plasmids. We show evidence to propose BaCP and TeCP, possibly signifying the fourth and fifth identified *C. perfringens* plasmid conjugation system. In analogy to the recently reported BCP [4], they indicate shared common plasmid ancestors of *C. tertium*, *C. cylindrosporum* and *C. baratii* with *C. perfringens*. These three clostridia are relatively closely related cluster I *Clostridia*, supporting this hypothesis [93–95].

Ultimately, we predicted 414 BGCs clustered into 32 GCFs, of which 6 displayed differential representation in newly sequenced isolates. Interestingly, none of the predicted BGCs showed a significant resemblance to characterized clusters, underpinning the need to further investigate *C. perfringens* encoded BGCs and their SMs [56]. Most precursors were found to be yet uncharacterized bacteriocins. Other *C. perfringens* bacteriocins have been found as potent SMs providing advantages for *C. perfringens* growth and pathogenicity [21]. Our work therefore also provides a starting point for the characterization of potentially novel antimicrobial compounds. Aside from our analysis, we noted the divergence of BGC enrichment amongst type F strains, notably the enriched GCF12 in animal type F strains, whereas the here included publicly available human and food-derived type F strains often harbour GCF17, both with a distinct bacteriocin gene. Due to the discussed technical bias in differing sequencing technologies, this however should be verified in subsequent sequencing experiments. Saccharide BGCs are the most abundant family in intestinal bacteria [23]. Therefore, it is remarkable that type C and F strains are completely devoid of any clusters containing sugar modifying enzymes. Uptake of exogenous saccharides might be essential for these toxinotypes. Efficient glycan degradation has been described in *netF*-positive type F strains [96] and in type G strains related to NE in poultry [97], which similarly are devoid of saccharide BGCs (Table S1).

In summary, this study provides a comprehensive collection of circularized animal-associated *C. perfringens* genomes and plasmids, most of them from economically important diseases. By integrating phylogenetics, toxin gene diversity, plasmid architecture and BGCs, we uncover novel virulence factors, plasmid conjugation systems and toxinotype-associated GCFs. The inclusion of healthy host isolates addresses a long-standing bias in publicly available animal-derived data. This potentially enables future

comparative analyses. Together, these contributions enhance our understanding of *C. perfringens* biology and open new avenues for future *in silico* and *in vitro* research.

Funding information
J.F. was funded by an MD/PhD stipend from the "Forschungsstiftung Zinkernagel" and the Swiss National Science Foundation (grant Nr. 323630_221864/1) and A.C.E. by a Swiss Government Excellence Scholarships for Foreign Students (ESKAS).The study was also supported by the Swiss National Science Foundation Fund 310030_212837 (H.P., salary for F.F.). The project was also supported by a grant of the Vetsuisse Faculty Specialization Committee.

Acknowledgements
Most calculations were performed on UBELIX (https://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern. The Galaxy server used for some calculations is partly funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg. We used OpenAI's ChatGPT-4o (chat.openai.com) to assist with scripting bioinformatics analyses and refining the manuscript text. The authors reviewed and verified all AI-generated content. All steps post gDNA extraction were performed at the Next Generation Sequencing Platform, University of Bern, Switzerland.

Author contributions
J.F.: conceptualization, formal analysis, funding acquisition, investigation, methodology, software, validation, visualization and writing – original draft. T.R.: data curation, formal analysis, methodology, resources and software. M.L.: data curation, formal analysis, investigation, methodology, resources, software, validation and visualization. A.C.E.: investigation. S.K.: investigation. S.F.: investigation. I.B.: investigation. S.N.: conceptualization. Q.D.: conceptualization. F.F.: formal analysis. M.K.: data curation and formal analysis. R.B.: supervision and writing – review and editing. P.N.: data curation, resources and software. E.G.: investigation. W.K.S.: conceptualization, resources, supervision and writing – review and editing. H.P.: conceptualization, funding acquisition, project administration, resources, supervision and writing – review and editing.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
All samples derived from dead animals and were taken for diagnostic purposes, from stool samples at slaughterhouses or in previous research studies. The Institute of Animal Pathology, Vetsuisse Faculty at the University of Bern does not require ethical approval for research using diagnostic samples.

References

1. Roder T, Oberhänsli S, Shani N, Bruggmann R. OpenGenome-Browser: a versatile, dataset-independent and scalable web platform for genome data management and comparative genomics. *BMC Genomics* 2022;23:855.
2. Kiu R, Hall LJ. An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerg Microbes Infect* 2018;7:141.
3. Geier RR, Rehberger TG, Smith AH. Comparative genomics of *Clostridium perfringens* reveals patterns of host-associated phylogenetic clades and virulence factors. *Front Microbiol* 2021;12:649953.
4. Gulliver EL, Adams V, Marcelino VR, Gould J, Rutten EL, *et al*. Extensive genome analysis identifies novel plasmid families in *Clostridium perfringens*. *Microb Genom* 2023;9.
5. Feng Y, Fan X, Zhu L, Yang X, Liu Y, *et al*. Phylogenetic and genomic analysis reveals high genomic openness and genetic diversity of *Clostridium perfringens*. *Microb Genom* 2020;6:mgen000441.
6. Abdel-Glil MY, Thomas P, Linde J, Busch A, Wieler LH, *et al*. Comparative in silico genome analysis of *Clostridium perfringens* unravels stable phylogroups with different genome characteristics and pathogenic potential. *Sci Rep* 2021;11:6756.
7. Bruggisser J, Tarek B, Wyder M, Müller P, von Ballmoos C, *et al*. CD31 (PECAM-1) serves as the endothelial cell-specific receptor of *Clostridium perfringens* β-toxin. *Cell Host Microbe* 2020;28:69–78.
8. Keyburn AL, Boyce JD, Vaz P, Bannam TL, Ford ME, *et al*. NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. *PLoS Pathog* 2008;4:e26.
9. Rood JI, Adams V, Lacey J, Lyras D, McClane BA, *et al*. Expansion of the *Clostridium perfringens* toxin-based typing scheme. *Anaerobe* 2018;53:5–10.
10. Alves GG, Machado de Ávila RA, Chávez-Olórtegui CD, Lobato FCF. *Clostridium perfringens* epsilon toxin: the third most potent bacterial toxin known. *Anaerobe* 2014;30:102–107.
11. Freedman JC, Shrestha A, McClane BA. *Clostridium perfringens* enterotoxin: action, genetics, and translational applications. *Toxins* 2016;8:73.
12. Svobodová I, Hulánková R. Nontyping virulence factors of *Clostridium perfringens*. *Acta Vet Brno* 2024;93:93–103.
13. Kiu R, Shaw AG, Sim K, Acuna-Gonzalez A, Price CA, *et al*. Particular genomic and virulence traits associated with preterm infant-derived toxigenic *Clostridium perfringens* strains. *Nat Microbiol* 2023;8:1160–1175.
14. Mehdizadeh Gohari I, Parreira VR, Nowell VJ, Nicholson VM, Oliphant K, *et al*. A novel pore-forming toxin in type A *Clostridium perfringens* is associated with both fatal canine hemorrhagic gastroenteritis and fatal foal necrotizing enterocolitis. *PLoS One* 2015;10:e0122684.
15. Lacey JA, Johanesen PA, Lyras D, Moore RJ. *In silico* identification of novel toxin homologs and associated mobile genetic elements in *Clostridium perfringens Pathogens* 2019;8:16.
16. Theoret JR, Uzal FA, McClane BA. Identification and characterization of *Clostridium perfringens* beta toxin variants with differing trypsin sensitivity and in vitro cytotoxicity activity. *Infect Immun* 2015;83:1477–1486.
17. Freedman JC, Theoret JR, Wisniewski JA, Uzal FA, Rood JI, *et al*. *Clostridium perfringens* type A-E toxin plasmids. *Res Microbiol* 2015;166:264–279.
18. Watts TD, Vidor CJ, Awad MM, Lyras D, Rood JI, *et al*. pCP13, a representative of a new family of conjugative toxin plasmids in *Clostridium perfringens*. *Plasmid* 2019;102:37–45.
19. Wisniewski JA, Rood JI. The Tcp conjugation system of *Clostridium perfringens*. *Plasmid* 2017;91:28–36.
20. Mehdizadeh Gohari I, A Navarro M, Li J, Shrestha A, Uzal F, *et al*. Pathogenicity and virulence of *Clostridium perfringens Virulence* 2021;12:723–753.
21. Pahalagedara ASNW, Flint S, Palmer J, Brightwell G, Gupta TB. Antimicrobial production by strictly anaerobic *Clostridium* spp. *Int J Antimicrob Agents* 2020;55:105910.
22. Khan A, Singh P, Srivastava A. Synthesis, nature and utility of universal iron chelator - Siderophore: a review. *Microbiol Res* 2018;212–213:103–111.

23. Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, *et al*. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* 2014;158:1402–1414.
24. Pitcher DG, Saunders NA, Owen RJ. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol* 1989;8:151–156.
25. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
26. Bankevich A, Bzikadze AV, Kolmogorov M, Antipov D, Pevzner PA. Multiplex de Bruijn graphs enable genome assembly from long, high-fidelity reads. *Nat Biotechnol* 2022;40:1075–1081.
27. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
28. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
29. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;31:3210–3212.
30. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, *et al*. Producing polished prokaryotic pangenomes with the panaroo pipeline. *Genome Biol* 2020;21:180.
31. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, *et al*. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–1534.
32. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35:518–522.
33. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
34. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.
35. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
36. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, *et al*. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
37. Cock PJA, Chilton JM, Grüning B, Johnson JE, Soranzo N. NCBI BLAST+ integrated into galaxy. *Gigascience* 2015;4:39.
38. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, *et al*. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2017;45:D566–D573.
39. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;7:e1002195.
40. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, *et al*. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 2022;40:1023–1025.
41. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, *et al*. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 2015;43:D222–6.
42. Abramson J, Adler J, Dunger J, Evans R, Green T, *et al*. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;630:493–500.
43. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, *et al*. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* 2022;50:W276–W279.
44. Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, *et al*. The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res* 2024;52:W521–W525.
45. Shimoyama Y. ANIclustermap: a tool for drawing ANI clustermap between all-vs-all microbial genomes. 2022.
46. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, *et al*. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* 2022:2022. DOI: 10.1101/2022.04.08.487609.
47. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, *et al*. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 2018;430:2237–2243.
48. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, *et al*. RCSB protein data bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res* 2021;49:D437–D451.
49. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 2020;8:90.
50. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 2021;37:2473–2475.
51. Carroll LM, Larralde M, Fleck JS, Ponnudurai R, Milanese A, *et al*. Accurate *de novo* identification of biosynthetic gene clusters with GECCO. *BioRxiv* 2021:03. DOI: 10.1101/2021.05.03.442509.
52. Larralde M, Blom J, Gourlé H, Carroll LM, Zeller G. Fast, flexible gene cluster family delineation with IGUA. *bioRxiv* 2025:15. DOI: 10.1101/2025.05.15.654203.
53. Blin K, Shaw S, Vader L, Szenei J, Reitz ZL, *et al*. antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. *Nucleic Acids Res* 2025;53:W32–W38.
54. Agrawal P, Khater S, Gupta M, Sain N, Mohanty D. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res* 2017;45:W80–W88.
55. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 2021;18:366–368.
56. Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, *et al*. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res* 2023;51:D603–D610.
57. Grau-Roma L, Navarro M, Blatter S, Wenker C, Kittl S, *et al*. *Clostridium perfringens*-associated necrotic enteritis-like disease in coconut lorikeets (*Trichoglossus haematodus*). *Vet Pathol* 2021;58:423–427.
58. Wollschläger N, Zimmermann W, Brodard I, Albini S, Doherr M, *et al*. Occurrence of *Clostridium perfringens* type A and type C in piglets of the Swiss swine population. *Schweiz Arch Tierheilkd* 2009;151:377–382.
59. Li J, Zhou Y, Yang D, Zhang S, Sun Z, *et al*. Prevalence and antimicrobial susceptibility of *Clostridium perfringens* in chickens and pigs from Beijing and Shanxi, China. *Vet Microbiol* 2021;252:108932.
60. Hunt M, Otto TD, Parkhill J, Keane JA, *et al*. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015;16:294.
61. Bruggisser J, Iacovache I, Musson SC, Degiacomi MT, Posthaus H, *et al*. Cryo-EM structure of the octameric pore of *Clostridium perfringens β*-toxin. *EMBO Rep* 2022;23:e54856.
62. Savva CG, Clark AR, Naylor CE, Popoff MR, Moss DS, *et al*. The pore structure of *Clostridium perfringens* epsilon toxin. *Nat Commun* 2019;10:2641.
63. Keyburn AL, Yan X-X, Bannam TL, Van Immerseel F, Rood JI, *et al*. Association between avian necrotic enteritis and *Clostridium perfringens* strains expressing netb toxin. *Vet Res* 2010;41:21.
64. Kitadokoro K, Nishimura K, Kamitani S, Fukui-Miyazaki A, Toshima H, *et al*. Crystal structure of *Clostridium perfringens*

enterotoxin displays features of beta-pore-forming toxins. *J Biol Chem* 2011;286:19549–19555.

65. Briggs DC, Naylor CE, Smedley JG III, Lukoyanova N, Robertson S, *et al*. Structure of the food-poisoning *Clostridium perfringens* enterotoxin reveals similarity to the aerolysin-like pore-forming toxins. *J Mol Biol* 2011;413:138–149.
66. Miyamoto K, Yumine N, Mimura K, Nagahama M, Li J, *et al*. Identification of Novel *Clostridium perfringens* type E strains that carry an iota toxin plasmid with a functional enterotoxin gene. *PLoS One* 2011;6:e20376.
67. Degiacomi MT, Iacovache I, Pernot L, Chami M, Kudryashev M, *et al*. Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nat Chem Biol* 2013;9:623–629.
68. Cirauqui N, Abriata LA, van der Goot FG, Dal Peraro M. Structural, physicochemical and dynamic features conserved within the aerolysin pore-forming toxin family. *Sci Rep* 2017;7:13932.
69. Ma Y, Sannino D, Linden JR, Haigh S, Zhao B, *et al*. Epsilon toxin–producing *Clostridium perfringens* colonize the multiple sclerosis gut microbiome overcoming CNS immune privilege. *J Clin Invest* 2023;133.
70. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb Genom* 2017;3:e000128.
71. Halpin JL, Hill K, Johnson SL, Bruce DC, Shirey TB, *et al*. Finished whole-genome sequences of clostridium butyricum toxin subtype E4 and *Clostridium baratii* toxin subtype F7 strains. *Genome Announc* 2017;5:00375–17.
72. Teng WL, Bannam TL, Parsons JA, Rood JI. Functional characterization and localization of the TcpH conjugation protein from *Clostridium perfringens*. *J Bacteriol* 2008;190:5075–5086.
73. Mary C, Fouillen A, Bessette B, Nanci A, Baron C. Interaction via the N terminus of the type IV secretion system (T4SS) protein VirB6 with VirB10 is required for VirB2 and VirB5 incorporation into T-pili and for T4SS function. *J Biol Chem* 2018;293:13415–13426.
74. Wisniewski JA, Teng WL, Bannam TL, Rood JI. Two novel membrane proteins, TcpD and TcpE, are essential for conjugative transfer of pCW3 in *Clostridium perfringens*. *J Bacteriol* 2015;197:774–781.
75. Bannam TL, Teng WL, Bulach D, Lyras D, Rood JI. Functional identification of conjugation and replication regions of the tetracycline resistance plasmid pCW3 from *Clostridium perfringens*. *J Bacteriol* 2006;188:4942–4951.
76. Bantwal R, Bannam TL, Porter CJ, Quinsey NS, Lyras D, *et al*. The peptidoglycan hydrolase TcpG is required for efficient conjugative transfer of pCW3 in *Clostridium perfringens*. *Plasmid* 2012;67:139–147.
77. Backert S, Fronzes R, Waksman G. VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? *Trends Microbiol* 2008;16:409–413.
78. Núñez B, De La Cruz F. Two atypical mobilization proteins are involved in plasmid CloDF13 relaxation. *Mol Microbiol* 2001;39:1088–1099.
79. Wisniewski JA, Traore DA, Bannam TL, Lyras D, Whisstock JC, *et al*. TcpM: a novel relaxase that mediates transfer of large conjugative plasmids from *Clostridium perfringens*. *Mol Microbiol* 2016;99:884–896.
80. Revitt-Mills SA, Watts TD, Lyras D, Adams V, Rood JI. The ever-expanding tcp conjugation locus of pCW3 from *Clostridium perfringens*. *Plasmid* 2021;113:102516.
81. Adams V, Watts TD, Bulach DM, Lyras D, Rood JI. Plasmid partitioning systems of conjugative plasmids from *Clostridium perfringens*. *Plasmid* 2015;80:90–96.
82. Nissen-Meyer J, Oppegård C, Rogne P, Haugen HS, Kristiansen PE. Structure and mode-of-action of the two-peptide (Class-IIb) bacteriocins. *Probiotics Antimicro Prot* 2010;2:52–60.
83. Goossens E, Valgaeren BR, Pardon B, Haesebrouck F, Ducatelle R, *et al*. Rethinking the role of alpha toxin in *Clostridium perfringens*-associated enteric diseases: a review on bovine necrohaemorrhagic enteritis. *Vet Res* 2017;48:9.
84. Perelman SS, James DBA, Boguslawski KM, Nelson CW, Ilmain JK, *et al*. Genetic variation of staphylococcal LukAB toxin determines receptor tropism. *Nat Microbiol* 2021;6:731–745.
85. Spaan AN, van Strijp JAG, Torres VJ. Leukocidins: staphylococcal bi-component pore-forming toxins find their receptors. *Nat Rev Microbiol* 2017;15:435–447.
86. Robbins F-M, Hartzman RJ. CD31/PECAM-1 genotyping and haplotype analyses show population diversity. *Tissue Antigens* 2007;69:28–37.
87. Mitchell LA, Koval M. Specificity of Interaction between *Clostridium perfringens* enterotoxin and claudin-family tight junction proteins. *Toxins* 2010;2:1595–1611.
88. Horiguchi Y, Akai T, Sakaguchi G. Isolation and function of a *Clostridium perfringens* enterotoxin fragment. *Infect Immun* 1987;55:2912–2915.
89. Katahira J, Inoue N, Horiguchi Y, Matsuda M, Sugimoto N. Molecular cloning and functional characterization of the receptor for *Clostridium perfringens* enterotoxin. *J Cell Biol* 1997;136:1239–1247.
90. Veshnyakova A, Piontek J, Protze J, Waziri N, Heise I, *et al*. Mechanism of *Clostridium perfringens* enterotoxin interaction with claudin-3/-4 protein suggests structural modifications of the toxin to target specific claudins. *J Biol Chem* 2012;287:1698–1708.
91. Takahashi A, Komiya E, Kakutani H, Yoshida T, Fujii M, *et al*. Domain mapping of a claudin-4 modulator, the C-terminal region of C-terminal fragment of *Clostridium perfringens* enterotoxin, by site-directed mutagenesis. *Biochem Pharmacol* 2008;75:1639–1648.
92. Mehdizadeh Gohari I, Kropinski AM, Weese SJ, Whitehead AE, Parreira VR, *et al*. NetF-producing *Clostridium perfringens*: clonality and plasmid pathogenicity loci analysis. *Infection, Genetics and Evolution* 2017;49:32–38.
93. Moore RJ, Lacey JA. Genomics of the pathogenic clostridia. *Microbiol Spectr* 2019;7.
94. Stackebrandt E, Kramer I, Swiderski J, Hippe H. Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*. *FEMS Immunol Med Microbiol* 1999;24:253–258.
95. Kiu R, Caim S, Alcon-Giner C, Belteki G, Clarke P, *et al*. Preterm infant-associated *Clostridium tertium*, *Clostridium cadaveris*, and *Clostridium paraputrificum* strains: genomic and evolutionary insights. *Genome Biol Evol* 2017;9:2707–2714.
96. Mehdizadeh Gohari I, Brefo-Mensah EK, Palmer M, Boerlin P, Prescott JF. Sialic acid facilitates binding and cytotoxic activity of the pore-forming *Clostridium perfringens* netf toxin to host cells. *PLoS One* 2018;13:e0206815.
97. MacMillan JL, Vicaretti SD, Noyovitz B, Xing X, Low KE, *et al*. Structural analysis of broiler chicken small intestinal mucin o-glycan modification by *Clostridium perfringens*. *Poult Sci* 2019;98:5074–5088.