











# Particular genomic and virulence traits associated with preterm infant-derived toxigenic *Clostridium perfringens* strains

Received: 3 August 2021

Accepted: 17 April 2023

Published online: 25 May 2023

 Check for updates

Raymond Kiu <sup>1</sup>, Alexander G. Shaw<sup>2</sup>, Kathleen Sim<sup>2</sup>, Antia Acuna-Gonzalez <sup>1</sup>, Christopher A. Price <sup>1</sup>, Harley Bedwell<sup>1</sup>, Sally A. Dreger <sup>1</sup>, Wesley J. Fowler<sup>1</sup>, Emma Cornwell<sup>2</sup>, Derek Pickard<sup>3</sup>, Gusztav Belteki<sup>4</sup>, Jennifer Malsom<sup>1</sup>, Sarah Phillips<sup>1</sup>, Gregory R. Young <sup>5</sup>, Zoe Schofield<sup>1</sup>, Cristina Alcon-Giner<sup>1</sup>, Janet E. Berrington<sup>6,7</sup>, Christopher J. Stewart <sup>6,7</sup>, Gordon Dougan<sup>3</sup>, Paul Clarke <sup>8,9</sup>, Gillian Douce <sup>10</sup>, Stephen D. Robinson <sup>1,11</sup>, J. Simon Kroll<sup>2</sup> & Lindsay J. Hall <sup>1,9,12</sup> ✉

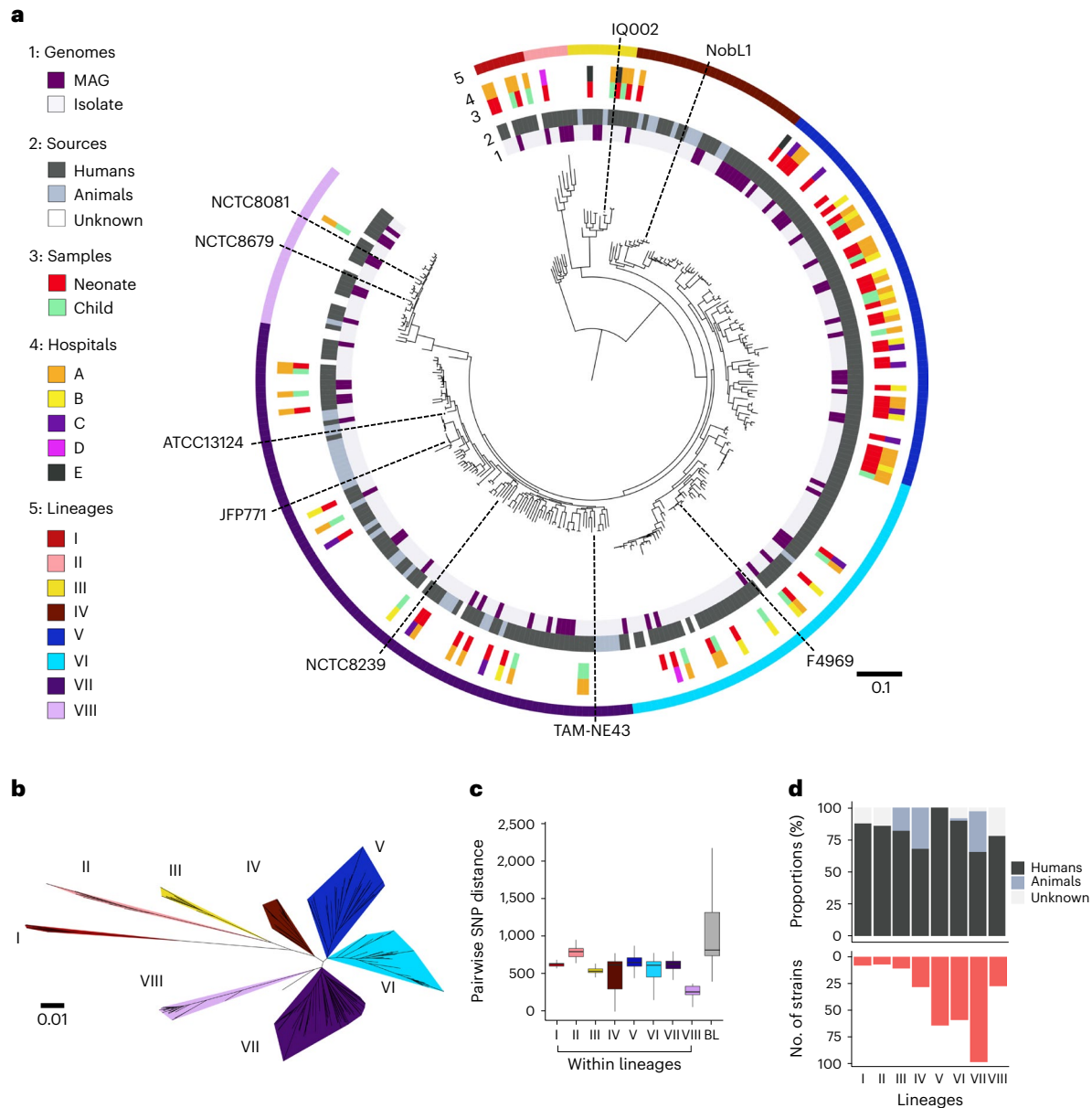
*Clostridium perfringens* is an anaerobic toxin-producing bacterium associated with intestinal diseases, particularly in neonatal humans and animals. Infant gut microbiome studies have recently indicated a link between *C. perfringens* and the preterm infant disease necrotizing enterocolitis (NEC), with specific NEC cases associated with overabundant *C. perfringens* termed *C. perfringens*-associated NEC (CPA-NEC). In the present study, we carried out whole-genome sequencing of 272 *C. perfringens* isolates from 70 infants across 5 hospitals in the United Kingdom. In this retrospective analysis, we performed in-depth genomic analyses (virulence profiling, strain tracking and plasmid analysis) and experimentally characterized pathogenic traits of 31 strains, including 4 from CPA-NEC patients. We found that the gene encoding toxin perfringolysin O, *pfoA*, was largely deficient in a human-derived hypovirulent lineage, as well as certain colonization factors, in contrast to typical *pfoA*-encoding virulent lineages. We determined that infant-associated *pfoA*<sup>+</sup> strains caused significantly more cellular damage than *pfoA*<sup>-</sup> strains in vitro, and further confirmed this virulence trait in vivo using an oral-challenge C57BL/6 murine model. These findings suggest both the importance of *pfoA*<sup>+</sup> *C. perfringens* as a gut pathogen in preterm infants and areas for further investigation, including potential intervention and therapeutic strategies.

*C. perfringens* has been frequently linked to various intestinal diseases in both humans and animals, particularly in neonates<sup>1</sup>. It is known to secrete >20 toxins, including several pore-forming toxins, for example,  $\beta$ -toxin, perfringolysin O (PFO), NetB and *C. perfringens* enterotoxin (CPE), alongside encoding other virulence factors associated with different facets of pathophysiology<sup>2–6</sup>. Although whole-genome

sequencing (WGS) has advanced our understanding of adult-associated *C. perfringens*, gastroenteritis-associated strains, association studies with other intestinal diseases in younger age groups, including carriage in healthy individuals, are very limited<sup>7–9</sup>.

NEC, the most severe and lethal neonatal gastrointestinal (GI) emergency worldwide, is an acquired inflammatory gut disease that

A full list of affiliations appears at the end of the paper. ✉ e-mail: [lindsay.hall@quadram.ac.uk](mailto:lindsay.hall@quadram.ac.uk)



**Fig. 1 | Strain-level population genomics of *C. perfringens* reveals eight distinct phylogenetic lineages.** **a**, Mid-point rooted, maximum-likelihood phylogenetic tree of 302 representative *C. perfringens* strains (derived from a collection of 673 *C. perfringens* genomes) inferred from 8,506 SNPs (from 858 single-copy core genes), aligned with metadata including genome types, sources, sample hosts, hospitals and lineages assigned using hierBAPS. Black dashed lines, type strain ATCC13124 and several representative strains were labelled in this phylogenetic tree. Branch lengths are indicative of the estimated nucleotide substitution per site (SNPs). MAG, metagenome-assembled genomes (from

MGnify study, human gut microbiome-derived *C. perfringens* genomes only). **b**, Unrooted phylogenetic tree coloured by assigned lineages. **c**, Pairwise SNP distances among strains located within lineages I ( $n = 8$ ), II ( $n = 7$ ), III ( $n = 11$ ), IV ( $n = 28$ ), V ( $n = 57$ ), VI ( $n = 66$ ), VII ( $n = 98$ ) and VIII ( $n = 27$ ) and between lineages. The box in the boxplot represents 50% of the central data, in between the lower and upper quartiles, with a central line representing the median, whereas the whiskers show the most extreme data points. BL, between lineages. **d**, Proportions of human and animal-derived *C. perfringens* genomes in each lineage.

manifests as tissue necrosis in the GI tract<sup>10,11</sup>. NEC typically affects 5–15% of very-low-birthweight (<1,500 g) preterm infants, leading to high mortality rates (~40%) and severe longer-term complications<sup>12–15</sup>. Clinical observations support a microbial role in NEC, including gas cyst formation in the gut wall (pneumatosis intestinalis), and antimicrobial treatment often resolves symptoms<sup>16–18</sup>. Several pathogens including *Klebsiella* spp. and *C. perfringens* have been linked to clinical NEC<sup>19,20</sup>, with microbiota-based, preterm infant cohort studies implicating overabundance before NEC<sup>21,22</sup>. *C. perfringens* has been consistently reported as a potential pathogenic agent associated with NEC<sup>1,23,24</sup>, with CPA-NEC symptoms often more fulminant and severe compared with ‘classic’ NEC<sup>19,25,26</sup>.

In the present study, we describe an extensive genomic study of *C. perfringens* clinical preterm isolates from non-CPA-NEC (95%) and CPA-NEC patients (5%). We probed virulence factors and plasmids associated with infant/CPA-NEC *C. perfringens* strains and traced putative hospital transmission. Additional experiments on a selection of *C. perfringens* strains and an oral-challenge infection model allowed us to assay key virulence and pathology readouts.

## Results

### Eight distinct phylogenetic lineages of *C. perfringens*

We isolated and performed WGS on 272 *C. perfringens* isolates from infant stool samples collected longitudinally from 70 individuals

(including 4 CPA-NEC infants), obtained from 5 UK neonatal intensive care units (NICUs) (and at-home collections) between 2011 and 2016 (Supplementary Table 1), which indicated an ~24% incidence. A strain-level, core-genome, global phylogenetic tree (isolates and metagenome-assembled genomes; Fig. 1a) clustered into eight major lineages (Fig. 1b,c). The relatively distant lineage VIII consisted of mostly typical human gastroenteritis isolates, whereas lineage VII included mixed human and animal sources, indicating a potential zoonotic signature. Lineage V comprised exclusively human-derived strains (Fig. 1d), with more than half obtained from infants and hospital linked.

### Genomic investigation reveals a hypovirulent lineage

Genome-based, strain-level virulence profiling was performed on all 80 infant-associated *C. perfringens* representative strains (derived from original 272, including reference strains; Fig. 2a). Toxinome analysis indicated that lineage III encoded more toxin genes ( $n = 9$ ) than other lineages, whereas lineage V had a reduced presence of the pore-forming toxin gene *pfoA* (Fig. 2b) and mostly consisted of neonatal isolates (Fig. 2c). Overall, *C. perfringens* strains from lineages III, VI and VII carried significantly more virulence genes versus strains from lineages I and V (Fig. 2d). Significantly fewer toxin genes including *pfoA* and *cpb2* were encoded within lineage V versus other lineages (Fig. 2e), with colonization-associated genes sialidases (*nan*), hyaluronidase (*nag*) and adhesin (*cna*) overrepresented in lineages III, VI and VII when compared with lineages I and V (Extended Data Fig. 1a).

CPA-NEC strains ( $n = 5$ ) nested within lineages I, IV, VI and VII, respectively (Fig. 2f), and encoded similar toxinome profiles. No CPA-NEC strains nested within lineage V, which predominantly contained strains with fewer virulence potentials (including apparent lack of *pfoA*), indicative of a potentially hypovirulent or commensal-like lineage (Fig. 2g).

### SNP analysis indicates potential strain dissemination

We investigated circulation of *C. perfringens* strains based on 210 longitudinal isolates ( $\geq 2$  sampling time points) obtained from 31 infants from 3 hospitals (Fig. 3a,b). Strain-level SNP distance was determined to be -0.2 SNPs, whereas between-strain SNP distance was estimated at -544 SNPs (Fig. 3c).

Sublineage Vif represented a clonal cluster consisting of 31 identical isolates at core-SNP level (0 SNPs) obtained from 6 infants residing in sister hospitals A and B (Fig. 3d,e). Clinical metadata linkage indicated that the index case of strain Vif was in infant N18 in hospital A. Later, this strain was detected in several individuals in hospital A (N17, N18 and N20), one of whom (N20) was diagnosed with CPA-NEC. Infant N20, who was also diagnosed with a condition predisposing to GI mucosal hypoxaemia, subsequently succumbed to NEC 10 d after strain Vif was first detected. N17 and N18 (index case for strain Vif) had an episode of possible NEC which led to administration of antibiotics

(broad-range  $\beta$ -lactam + vancomycin) for 2 d, with strain Vif first detected -8 d after antibiotic administration finished in N18. Strain Vif was also detected in individuals residing in hospital B (across the same 100-d window). This suggests potential spread of the *C. perfringens* strain between hospitals, whether transferred on/in patients, their parents or equipment or healthcare staff. We also observed that strain Ib in the index case N3 (who succumbed on day 27; Fig. 3f) was again detected in N10 (after a 60-d gap), suggesting possible horizontal transmission and persistence of *C. perfringens* strains, potentially in the NICU environment. Additional transmissions events were also implicated in four twin pairs carrying identical strains (Fig. 3g).

### Virulence plasmid carriage among infant-associated isolates

We next investigated plasmid acquisition of two key *C. perfringens* (virulence) conjugative plasmid families, pCW3 (ref. 27) and pCPI3 (ref. 28) (Fig. 4a). Plasmid pCW3 (with *tcp* system) encoded toxin genes such as *cpb2* and *cpe*, iota binary toxin genes *iap* and *ibp* and tetracycline resistance determinants *tetA(P)* and *tetB(P)* (Fig. 4b). The adhesin gene *cnaC* was also detected in most pCW3 plasmids, potentially enhancing colonization capacity. NEC-associated isolates from sublineage Vif primarily carried pCW3 plasmids in either -50-kb or -70-kb sizes, encoding *tetA(P)* and *tetB(P)*, *cnaC* and toxin gene *cpb2*, with the 50-kb plasmids missing -20 genes due to transposases encoded by *CPR\_0630* on both flanks. Plasmid Vif, encoded by strain Vif, was not detected in strains from other lineages apart from a lineage V strain IQ074 (obtained from infant N10; Fig. 4c), indicating potential plasmid transmissible events between *C. perfringens* strains, because the NICU stay of N10 overlapped with infants carrying Vif strains (Fig. 4d).

The pCPI3 plasmids (-45–60 kb) in our dataset harboured *cpb2*, *becAB* and *edpA* toxin genes<sup>29</sup>. Adhesin *cnaB* was detected in all extracted pCPI3 plasmids. Importantly, sublineage IIIb strains carried pCPI3 plasmids (54,591 bp) which encoded *becAB* binary toxin genes, previously associated with human gastroenteritis, with further analysis confirming identical plasmids from three individual infants (Fig. 4e and Extended Data Fig. 1b).

Another pCPI3 plasmid (-62 kb) was detected in nine *C. perfringens* identical isolates (Fig. 4f), obtained from four individuals, that nested in three distinct sublineages in lineage V, highlighting the common transferable nature of conjugative plasmids among *C. perfringens* strains (Fig. 4g). Both plasmids pCW3 and pCPI3 were less prevalent in *C. perfringens* strains in lineage V (19%); in contrast plasmid carriage in other lineages ranged from 50% to 86% (Fig. 4h). Plasmid carriage in CPA-NEC strains was 80% versus non-CPA-NEC strains at 44% (Fig. 4i).

### Toxicity assays links toxin gene *pfoA* to epithelial damage

From the genomic data, it was apparent that *C. perfringens* strains from lineage V encoded significantly fewer virulence factors, most clearly evidenced by the absence of the *pfoA* toxin gene. We selected

### Fig. 2 | Genomic analysis of infant-associated *C. perfringens* virulence potentials unravels a hypovirulent lineage that does not commonly encode toxin genes *pfoA* and *cpb2*.

**a**, A mid-point-rooted, maximum-likelihood phylogenetic tree of 80 infant-associated *C. perfringens* representative strains, derived from 272 *C. perfringens* isolate genomes using an ANI cutoff of 99.9% (strain-level) for dereplication of genomes. This tree was inferred from 175,479 core SNPs (in 1,995 single-copy core genes) aligned with clinical metadata and virulence profiles of each strain including toxin genes, colonization factors and antimicrobial resistance genes, in addition to virulence gene counts and genome sizes, as indicated. Branches are colour coded according to the lineages assigned in Fig. 1. AMR, antimicrobial resistance. **b**, Toxinome profiles of each lineage. **c**, Sample source proportions in each lineage. **d**, Virulence count (per strain) comparison across lineages. Significance was pairwise compared. The box in the boxplot represents 50% of the central data, in between the lower and upper quartiles, with a central line representing the median, whereas the whiskers show the most extreme data points. There was no difference between lineages I and

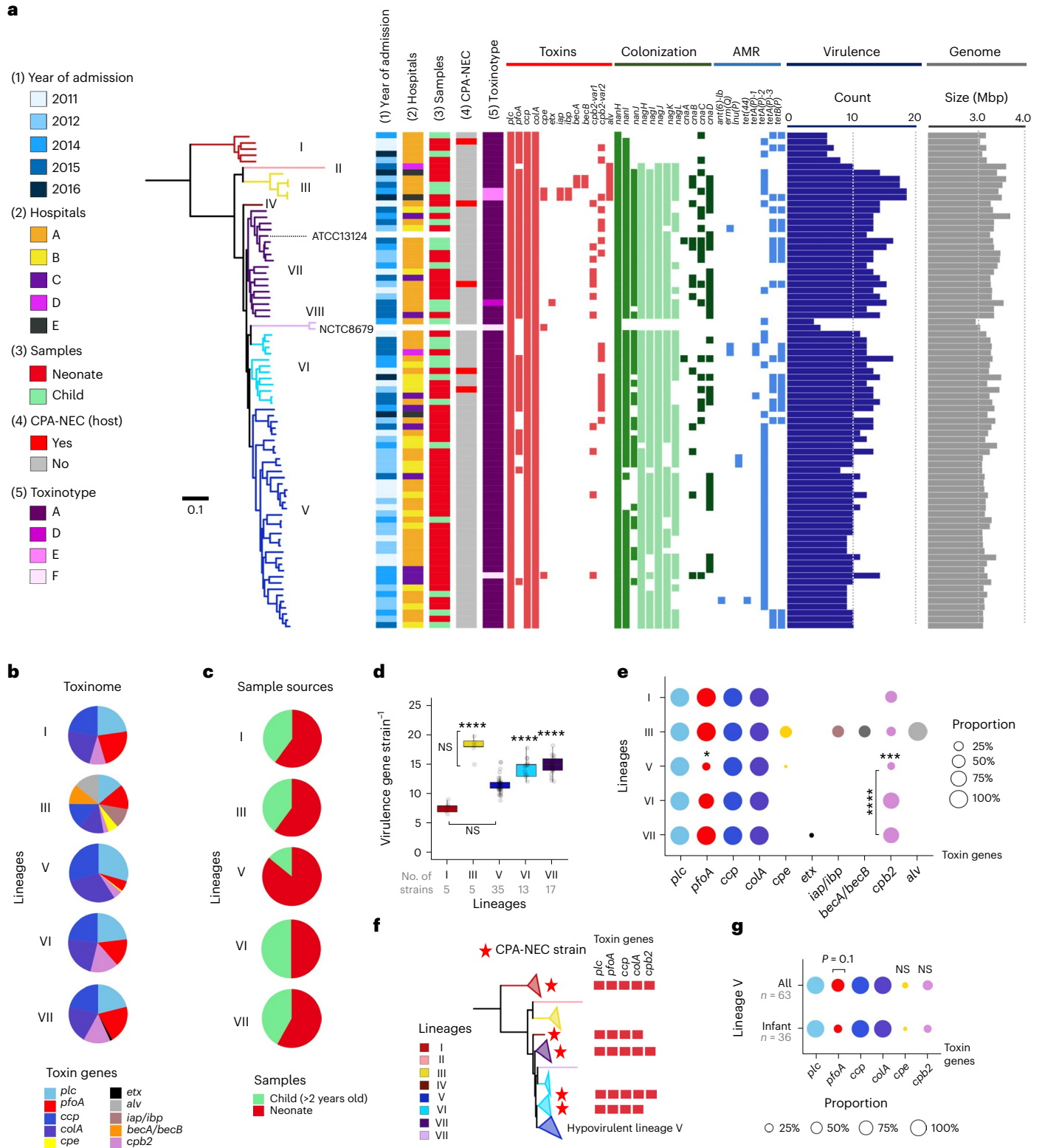
V, whereas lineages III, VI and VII (no difference among these three lineages) are significantly different from lineages I and V. \*\*\*\* $P < 0.0001$ . Statistical analysis was performed using Kruskal–Wallis test and post-hoc Dunn's test. **e**, Proportion of each toxin gene across all lineages. The significance was pairwise compared within individual toxin genes. The significant difference was indicated in *cpb2* genes across lineages V, VI and VII, and pairwise comparison indicated that lineage V has reduced proportionally in *cpb2* genes compared with lineages VI and VII. \*\*\* $P = 0.003$ , \*\*\*\* $P < 0.0001$ , whereas proportion of *pfoA* gene is significantly reduced in lineage V versus all other lineages: \* $P < 0.05$ . **f**, Full toxin gene profiles of five CPA-NEC *C. perfringens* strains. **g**, Toxin gene proportion in isolates within global lineage V (which includes MAGs from human microbiome project) and infant-associated lineage V isolates. NS, non-significant (Note: lineages II, IV and VIII were not included in statistical investigations due to these lineages only encompassing singletons.) In **e** and **g**, statistical analysis was performed using Fisher's test (two sided).

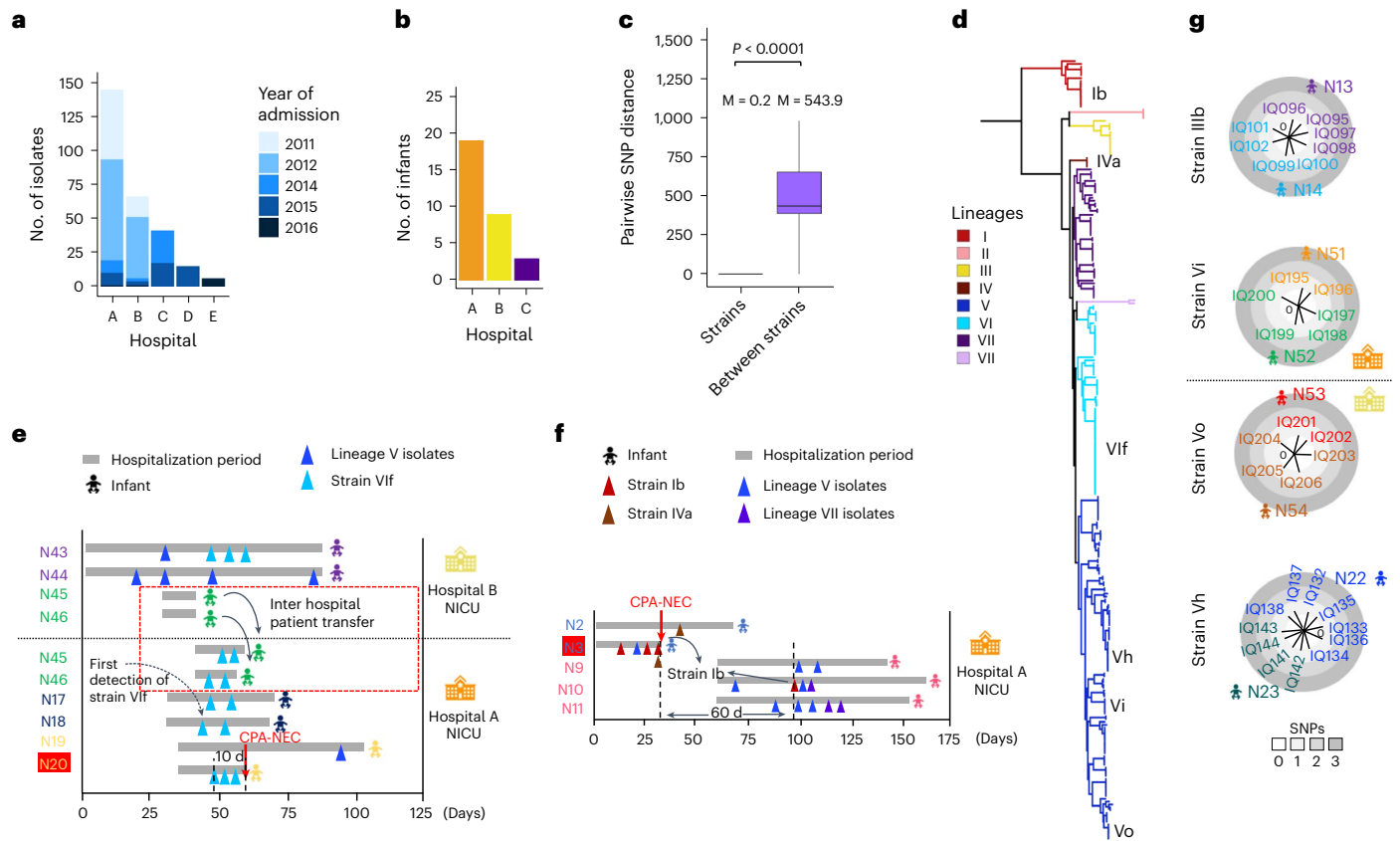
30 infant-associated *C. perfringens* strains from all lineages, including *pfoA*<sup>+</sup> and *pfoA*<sup>-</sup> strains for in vitro tests (Fig. 5a).

Survivability/transmissibility assays indicated that lineage VII strains appeared to be most oxygen tolerant (Fig. 5b), surviving significantly better than strains from lineage V from 48 h to 192 h (Extended Data Fig. 2a), although *C. perfringens* strains from lineage VI produced significantly more spores than strains nested in lineage V (Fig. 5c).

*C. perfringens*-associated cellular toxicity assays showed supernatants of strains in lineages I, III, VI and VII inducing significant cell death

in epithelial monolayers, whereas lineage V strains did not (Fig. 5d). Grouping by presence of the *pfoA* pore-forming toxin gene indicated a clear cytotoxicity signal, whereas cell death induced by *pfoA*<sup>-</sup> strains was no different statistically from untreated controls (Fig. 5e). Cellular toxicity phenotypes correlated with the presence of *pfoA*, and to a lesser extent toxin gene *cpb2* (Fig. 5f), with a potential synergistic effect observed in strains encoding both toxins (Fig. 5g). It is interesting that *pfoA*<sup>+</sup> and CPA-NEC strains also exhibited significantly higher cell toxicity versus non-CPA-NEC strains (Fig. 5h). All *pfoA*<sup>+</sup> strains also





**Fig. 3 | In-depth strain-level SNP analysis in tracking putative transmission of infant-associated *C. perfringens*.** **a**, Number of *C. perfringens* isolates sequenced from five hospitals. **b**, Number of infants with two or more longitudinal sampling time points. **c**, Pairwise SNP distance in longitudinally sampled isolates ( $n = 209$ ): strains (ANI > 99.9%) versus between strains (ANI < 99.9%). M, mean (SNPs). The statistical analysis was performed using Wilcoxon’s test (two sided). The box in the boxplot represents 50% of the central data, in between the lower and upper quartiles, with a central line representing the median, whereas the whiskers show the most extreme data points. **d**, Phylogenetics of 272 infant-associated *C. perfringens* isolates with branches colour coded according to the lineage

assigned. The key strains, defined as isolates having highly similar ANI > 99.9%, as indicated on the tree, are further described in **e–g**. **e**, Tracking multi-host *C. perfringens* strain VIf (31 isolates) which involves six individuals in two sister hospitals, A and B, based on clinical metadata. Twin pairs are colour coded. A CPA-NEC infant (N20) was highlighted in red. **f**, Putative intrahospital circulation of strain Ib. Strain Ib was not detected in a 60-d window in the infant cohort. A CPA-NEC infant (N3) was highlighted in red. **g**, Comparison of SNP distance trees between isolates from twin pairs, suggesting dissemination of *C. perfringens*.

exhibited complete haemolysis on blood agar, whereas *pfoA*<sup>-</sup> strains displayed partial haemolysis (Fig. 5i). One *pfoA*<sup>-</sup> strain (LH019) induced higher cell death (Fig. 5f), probably due to the presence of alveolysin (*alv*) which shares ~86% homology (nucleotide identity) with *pfoA*<sup>30</sup>.

***C. perfringens* strains induce murine colonic histopathology**

We developed an oral-challenge murine model to explore pathology related to *pfoA* using six *C. perfringens* (*pfoA*<sup>+</sup>/*pfoA*<sup>-</sup>) strains (Fig. 6a). Colonization rates indicated that most strains were completely cleared by day 4 (Fig. 6b and Extended Data Fig. 2b), although microbiota genus-level diversity profiles between *pfoA*<sup>+</sup> and *pfoA*<sup>-</sup> groups did not differ statistically (Fig. 6c and Extended Data Fig. 2c,d). Differences in daily weight changes between groups were not observed (Fig. 6d); however, overall weight change indicated that the *pfoA*<sup>-</sup> groups had significantly less weight gain versus controls (Fig. 6e).

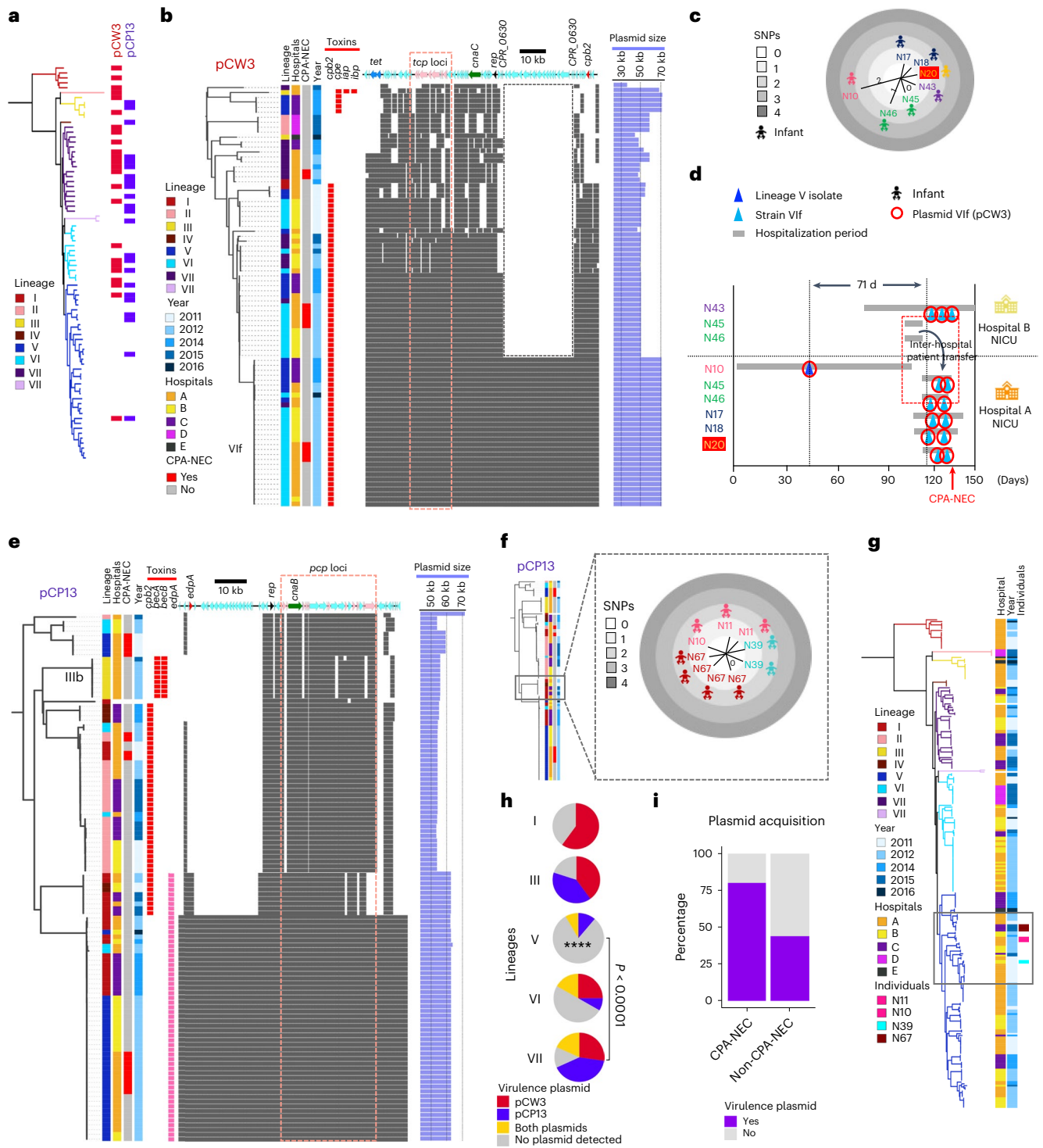
Both *C. perfringens*-infected groups (*pfoA*<sup>+</sup> and *pfoA*<sup>-</sup>) exhibited significantly higher histopathological scores for intestinal inflammation compared with control groups, with colonic sections from *pfoA*<sup>+</sup> groups displaying more severe damage/pathology versus *pfoA*<sup>-</sup> groups (Fig. 6f). Distal colonic sections of both infected groups demonstrated similar pathological features, including marked goblet cell loss and inflammatory cell infiltrate (Fig. 6g and Extended Data Fig. 3), with *pfoA*<sup>+</sup> groups having significantly longer crypt lengths versus

*pfoA*<sup>-</sup> and control groups (Fig. 6h), although not when comparing colon weights (Fig. 6i). Immune profiling of colonic tissue showed significantly increased levels of tumour necrosis factor (TNF)- $\alpha$ , KC/GRO and interleukin (IL)-1 $\beta$  in the *pfoA*<sup>+</sup> group versus controls (Extended Data Fig. 4) and, although not significant, elevated levels of IL-22 and IL-6 were observed.

**Discussion**

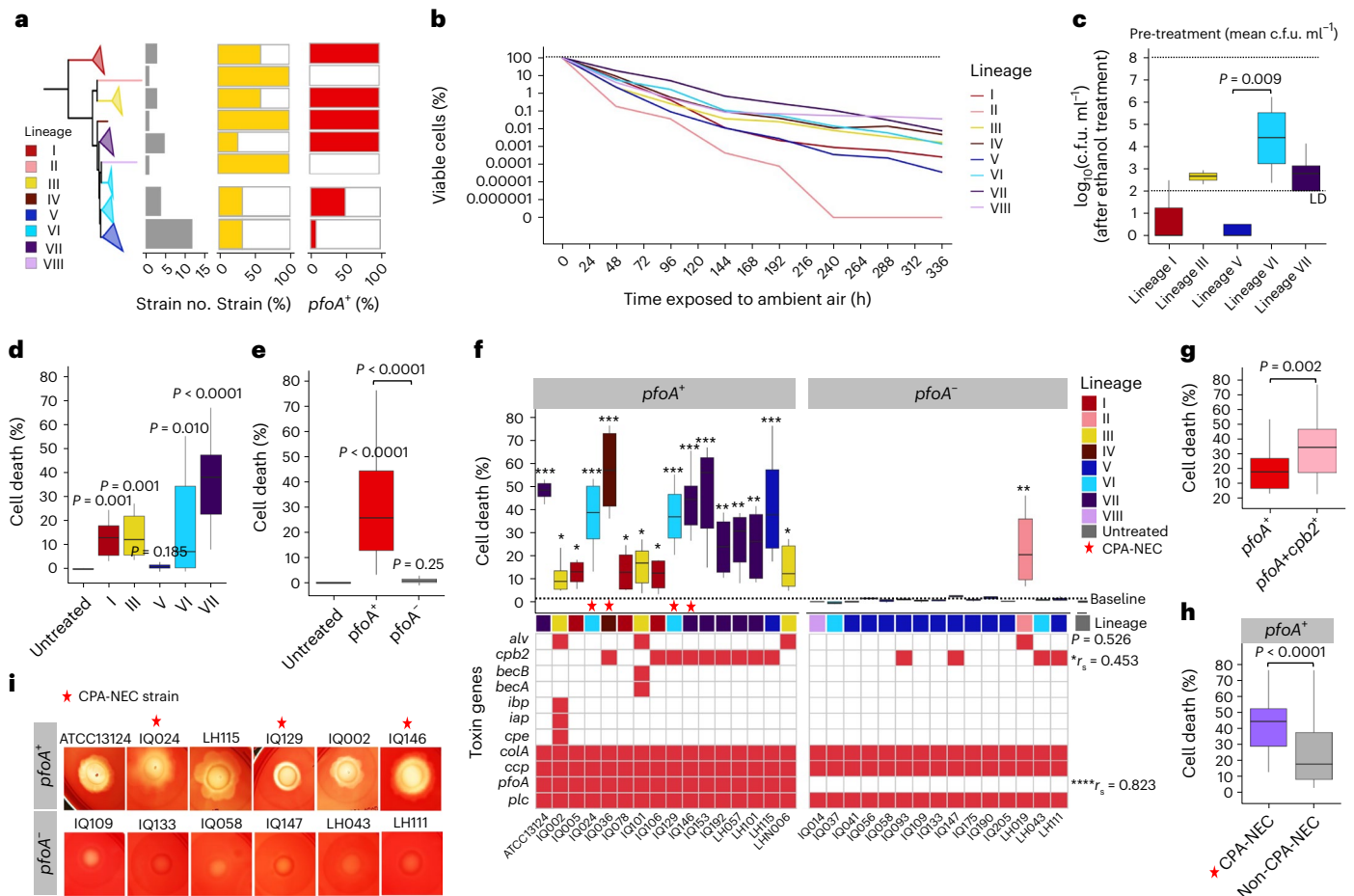
Genomic analysis of preterm infant-associated *C. perfringens* strains suggests distinct lineages that comprise potentially hypovirulent strains versus typical virulent strains linked with possible NICU circulation and carriage of virulence-carrying mobile elements, including CPA-NEC strains. Certain genotype-linked phenotypical traits, for example, *pfoA*<sup>-</sup>-associated cytotoxicity, appear to correlate with increased intestinal cell pathology, which may be associated with infection/disease outcomes.

Currently, public genome databases are biased towards diseased-associated isolates, with only a limited number of *C. perfringens* genomes obtained from healthy individuals (~5%)<sup>31,32</sup>. *C. perfringens* is also considered as a normal gut microbiota member<sup>33,34</sup> and we observed a potentially hypovirulent or ‘commensal-like’, human-derived monophyletic lineage (V). Strains encoded significantly fewer virulence factors and plasmids and were thus distinct from more



**Fig. 4 | Computational analysis of key conjugative virulence plasmids pCW3 and pCP13 encoded by infant-associated *C. perfringens* isolates.** **a**, Carriage/distribution of pCW3 and pCP13 conjugative virulence plasmids among infant-associated *C. perfringens* strains indicated in the infant *C. perfringens* phylogeny. **b**, Plasmid-coding sequences mapped to reference plasmid sequence (pLH112) to present a comparative overview of pCW3 plasmids computationally extracted from infant-associated *C. perfringens* isolates ( $n = 85$ ). Data/sequences were aligned with a newly constructed plasmid phylogeny. **c**, SNP distance tree of plasmids from strain Vif (from seven individuals). **d**, Probable plasmid circulation map of plasmid Vif (identical plasmids encoded in strains Vif). **e**, Plasmid-coding sequences mapped to reference plasmid sequence (pLH112) to present a comparative overview of pCP13 plasmids computationally extracted from infant-associated *C. perfringens* isolates ( $n = 111$ ). Data/sequences were aligned with a newly constructed plasmid phylogeny. **f**, SNP distance tree of identical plasmids

encoded in nine isolates obtained from four individuals. **g**, Isolates' positions from the four individuals indicated in **f** within the infant-associated *C. perfringens* phylogenetic tree. These four individuals harboured *C. perfringens* isolates ( $n = 9$ ; from different hospitals) that encoded identical pCP13 conjugative virulence plasmids nested within distinct sublineages of lineage V. **h**, Proportions of conjugative virulence plasmids encoded by *C. perfringens* strains across lineages. Proportion of strain-level virulence plasmid carriage was significantly lower in lineage V compared with lineage VII ( $****P < 0.0001$ ). Statistical analysis was performed using Fisher's test (two sided). **i**, Proportions of infant-associated *C. perfringens* strains carrying one or more conjugative virulence plasmid(s), pCW3 and/or pCP13: a comparison across lineage V ( $n = 36$ ), CPA-NEC strains ( $n = 5$ ) and non-CPA-NEC strains ( $n = 73$ ). Statistical analysis was performed using Fisher's test (two sided). The proportional difference was not found to be significant.



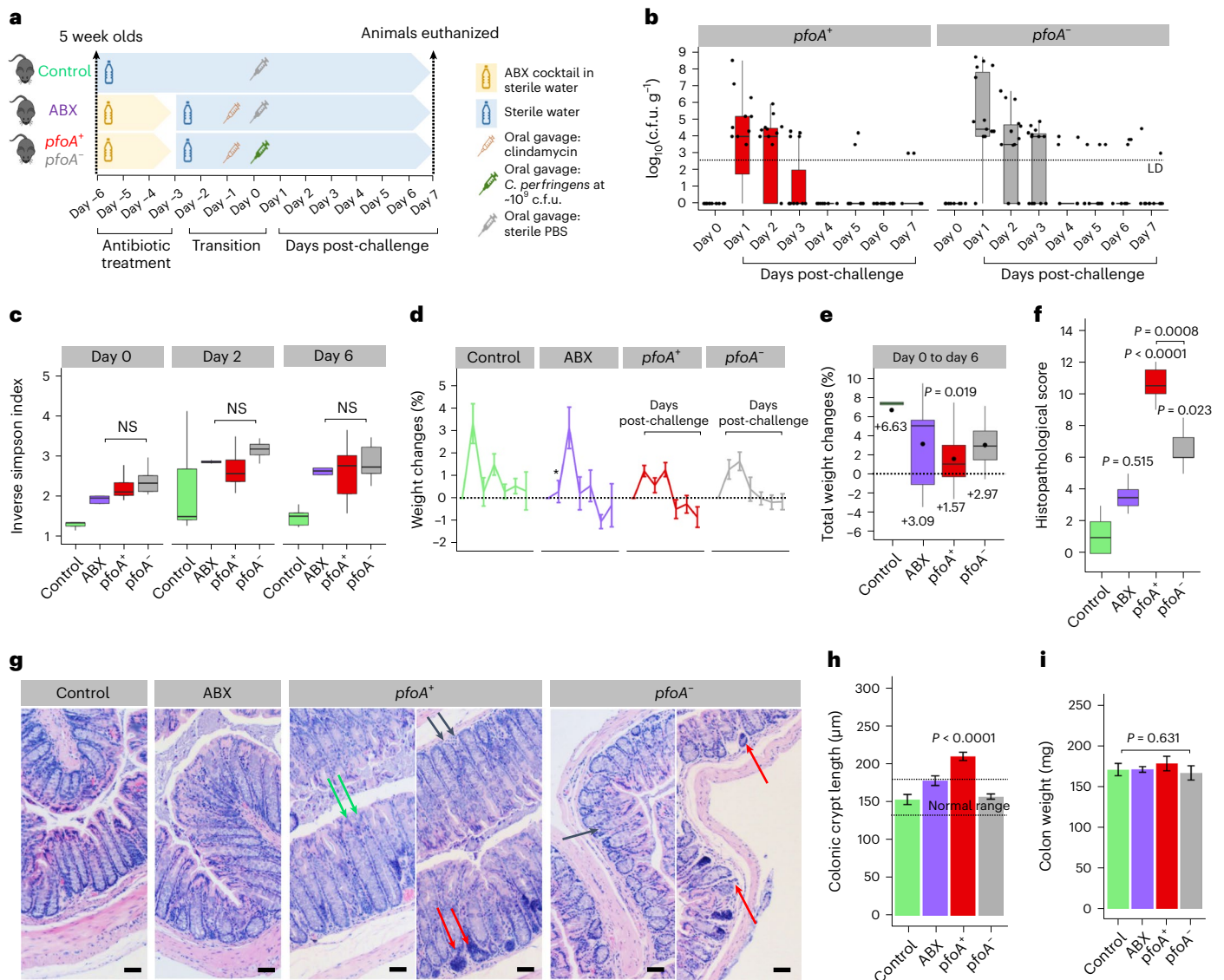
**Fig. 5 | Phenotypic virulence assessment of 30 infant-associated *C. perfringens* strains suggests *pfoA* as a key virulence gene in gut epithelial cell death.** **a**, Schematic of strain selection ( $n = 31$ ). **b**, Viability of *C. perfringens* strains ( $n = 31$ ) exposed to ambient air over 14 d (data stratified by lineage). A lineage II strain was undetectable after 10 d. Data are the mean. **c**, Sporulation capacity of *C. perfringens* strains ( $n = 28$ ) assessed using ethanol pre-treatment (4 h) to eliminate vegetative cells before culturing, with  $n = 3$  biologically independent samples for each strain. The statistical significance was pairwise compared across the groups. LD, limit of detection. **d**, Toxicity of sterile-filtered supernatants ( $n = 31$ ) determined by 2-h co-culturing with colonic cell line Caco-2 ( $n = 6$  in two independent experiments). Cell death comparison was between lineages. The significance was compared with the untreated group. **e**, Cell toxicity of *C. perfringens* strains compared between *pfoA*<sup>+</sup> ( $n = 15$ ) and *pfoA*<sup>-</sup> groups ( $n = 15$ ). The *pfoA*<sup>+</sup> strains encode *pfoA* toxin gene; the *pfoA*<sup>-</sup> strains do not encode *pfoA*. The significance was compared with the untreated group, and pairwise between *pfoA*<sup>+</sup> and *pfoA*<sup>-</sup> groups as indicated. **f**, Cell death percentage of individual strains aligned with toxin gene profiles. The presence of both toxin genes *pfoA* ( $r_s = 0.823$ ) and *cpb2* ( $r_s = 0.453$ ) was positively

correlated with the increase in Caco-2 cell death ( $P < 0.05$ ). A correlation analysis was performed using the point-biserial-correlation test (Spearman's rank correlation). The significance was compared with the untreated group: \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ . **g**, Cell death comparison between *pfoA*<sup>+</sup> strains ( $n = 7$ ) and *pfoA*<sup>+</sup> *cpb2*<sup>-</sup> strains ( $n = 9$ ) showing potential synergistic effects of *cpb2* toxin gene alongside *pfoA*. **h**, Cell death comparison between CPA-NEC *pfoA*<sup>+</sup> strains ( $n = 4$ ) and non-CPA-NEC *pfoA*<sup>+</sup> strains ( $n = 26$ ). **i**, Images of *C. perfringens* colonies (10  $\mu$ l of confluent cultures per spot) on CBA after 24 h of anaerobic incubation. All representative *pfoA*<sup>+</sup> strains ( $n = 6$ ) exhibit complete haemolysis on CBA whereas *pfoA*<sup>-</sup> strains ( $n = 6$ ) display partial haemolysis. In **c-h**, the box in the boxplot represents 50% of the central data, in between the lower and upper quartiles, with a central line representing the median, whereas the whiskers show the most extreme data points. All statistical analyses were performed using the Kruskal–Wallis test and post-hoc Dunn's test of multiple comparisons, with  $P$  values adjusted using the Benjamini–Hochberg method, unless otherwise stated. Significance was tested against the untreated group, otherwise significance brackets were used in pairwise comparison.

typically virulent *C. perfringens*, including CPA-NEC strains. Indeed, potentially hypovirulent lineage V strains may not encode the necessary virulence traits required to cause overt disease such as severe intestinal colitis or CPA-NEC.

A genomic signature apparently enriched in typical, virulent, infant-associated *C. perfringens* isolates (including CPA-NEC strains) was the presence of the gene *pfoA* that transcribes the pore-forming haemolytic toxin PFO. We consistently observed a direct correlation between the presence of *pfoA* and significantly enhanced cell toxicity in intestinal cell lines and complete haemolysis, when compared with *pfoA*<sup>-</sup> strains regardless of lineage (Fig. 5e,i). PFO ( $\theta$ -toxin), a typical cholesterol-dependent cytolysin, is known as a pore-forming toxin<sup>35</sup>,

which has been heavily associated with the pathogenesis of myonecrosis (gas gangrene), haemorrhagic enteritis in calves and more recently septicaemia (intravascular haemolysis) in humans, including neonates<sup>36–40</sup>. PFO can act synergistically with  $\alpha$ -toxin in the pathology of gas gangrene, which is a disease that shares a high degree of symptom similarity with CPA-NEC in preterm infants, and has also been called 'intestinal gas gangrene'<sup>19,41,42</sup>. The haemolysis rates of *C. perfringens* septicaemia clinical isolates were shown to strongly correlate with PFO expression alone, regardless of  $\alpha$ -toxin production (whereas a *pfoA*-deficient strain did not induce haemolysis), with proinflammatory cytokine production (TNF- $\alpha$ , interleukin IL-5 and IL-6) significantly stronger than  $\alpha$ -toxin<sup>40</sup>. Nevertheless, the mechanistic role of PFO in



**Fig. 6 | *C. perfringens* strains induce histopathological changes in murine colons.** This experiment was based on eight groups of mice ( $n = 5$  for each group) including control (untreated) and antibiotic groups, and six experimental groups (three groups of *pfoA*<sup>+</sup> strains and three groups of *pfoA*<sup>-</sup> strains). **a**, Schematic of in vivo infection model experimental design. PBS, phosphate-buffered saline. **b**, Intestinal colonization of *C. perfringens* monitored over a 7-d post-challenge *pfoA*<sup>+</sup> ( $n = 15$ ) versus *pfoA*<sup>-</sup> ( $n = 15$ ). **c**, Inverse Simpson's index: gut microbiome diversity of mice on days 0, 2 and 6, respectively, showing no significant genus-level differences when comparing groups *pfoA*<sup>+</sup> ( $n = 15$ ) and *pfoA*<sup>-</sup> ( $n = 15$ ) to antibiotics ( $n = 5$ ). The statistical analysis was performed with ANOVA and Tukey's test (post-hoc pairwise). NS, non-significant ( $P > 0.05$ ). **d**, Daily weight changes (%) of mice over experimental period across groups. Data are mean  $\pm$  s.e.m.  $^*P = 0.034$ . **e**, Total average weight changes from day 0 to day 6. The *pfoA*<sup>+</sup> group ( $n = 15$ ) has a significantly lower mean weight change compared with the control group ( $P < 0.05$ ). The statistical analysis was performed with ANOVA and Tukey's test (post-hoc pairwise). **f**, Histopathological scores for intestinal inflammation of murine distal colons (day 7 post-challenge), comparison between groups.

Scoring was based on two H&E-stained images per distal colon. Increments in score value denote increasing severity. **g**, Representative H&E-stained murine distal colonic sections of control, antibiotics and experimental groups (*pfoA*<sup>+</sup> and *pfoA*<sup>-</sup>), respectively, showing epithelial changes. The green arrow shows erosion and crypt hyperplasia, the black arrow loss of goblet cells and the red arrow immune cell infiltration. Scale bars, 50  $\mu$ m. **h**, Distal colonic crypt length comparison. Measurement was performed based on 20 crypts per representative image. Data are mean  $\pm$  s.e.m. Normal range for distal colonic crypt length is 133–175  $\mu$ m. **i**, Colon weight (whole-colon) comparison across all groups. Data are mean  $\pm$  s.e.m. In **b**, **c**, **e** and **f**, the box in the boxplot (where applicable) represents 50% of the central data, in between the lower and upper quartiles, with a central line representing the median, whereas the whiskers show the most extreme data points. All statistical analyses were performed using the Kruskal–Wallis test and a post-hoc Dunn's test of multiple comparisons and  $P$  values adjusted using the Benjamini–Hochberg method, unless otherwise stated. The significance was tested against the control group; otherwise significance brackets were used in pairwise comparison.

disease development is not completely defined, although previous studies have indicated additional modulation of host responses<sup>36,43–45</sup>. Indeed, our in vivo studies suggest a high degree of colonic immune cell infiltration and a higher proinflammatory cytokine milieu (and associated pathology) during *pfoA*<sup>+</sup> *C. perfringens* infection. Although previous studies have indicated that the *pfoA* toxin gene is universally

encoded in *C. perfringens*, aside from gastroenteritis-associated toxinotype F strains (nested in lineage VIII in Fig. 1a), we did not observe this in lineage V strains<sup>8,9,46,47</sup>. This prior assumption may be due to skewing in public databases for disease-associated isolates, so our data highlight the important role of this toxin in disease involvement, including CPA-NEC.



A concern with typical virulent *pfoA*<sup>+</sup> *C. perfringens* strains is nosocomial transmission in at-risk populations, which is a huge issue in related *Clostridioides difficile*<sup>48,49</sup>. We observed a circulating strain Vif, which was detected in six individuals during their NICU stay, and seemingly circulated between two sister hospitals that regularly exchange patients. Transmission of *C. perfringens* has also been demonstrated in similar confined settings such as within elderly care homes (more vulnerable communities), where non-foodborne outbreaks of toxigenic *C. perfringens* take place frequently<sup>8,50</sup>. Notably, the index case of strain Vif was detected in an individual who had been treated with broad-spectrum antibiotics, suggesting that creation of a more 'favourable' niche may have allowed *C. perfringens* to flourish in the microbiota-depleted gut. This strain (IQ129) also nested within lineage VI linking to a stronger spore-forming capacity (Fig. 5c), a trait that may provide protection against antibiotics, as demonstrated by its detection 8 d after termination of antibiotic treatment. Although six infants harboured strain Vif, only N20 developed CPA-NEC and ultimately succumbed to the infection. This preterm infant had been previously diagnosed with congenital heart disease which may have caused reduced blood flow and oxygen supply to the gut, potentially altering the gut microbiota and barrier permeability that may favour the growth of certain virulent *C. perfringens* strains<sup>51–53</sup>. Further studies exploring wider microbiome dynamics longitudinally in preterm infant cohorts, complemented by in vitro studies (for example, model colon systems), are required to tease apart potential mechanisms.

The widespread dissemination of certain strains may be partly attributed to several traits. Typical *pfoA*<sup>+</sup> *C. perfringens* strains (including those associated with CPA-NEC) appeared to be more oxygen tolerant and have enhanced spore-forming/germination ability, which may facilitate spread between infants and NICUs. This may include resilience against standard disinfectant practices, thus enabling persistence in even adverse 'sterile' hospital environments, and subsequent germination once spores enter the infant gut<sup>54,55</sup>. These traits pose a significant challenge for infection control measures, especially those wards where extremely low-birth-weight, preterm infants reside who have a heightened risk of developing CPA-NEC<sup>56</sup>. The presence of circulating strains also indicates that routine genomic surveillance of *C. perfringens* may be helpful in NICUs to monitor and prevent circulation of virulent strains.

Certain plasmids (for example, pCW3 and pCP13) have long been recognized as key mobile genetic elements for transfer of virulence genes in *C. perfringens*<sup>27,28</sup>. We observed nine isolates from three distinctive sublineages in lineage V carrying identical pCP13 conjugative plasmids, suggesting frequent conjugative plasmid transfer<sup>28</sup>. Alongside virulence factors, certain colonization factors are encoded on these plasmids such as adhesin genes *cnaB* and *cnaC*<sup>27</sup>, with collagen adhesin critical in pathogenesis by conferring additional cell-binding ability in host strains<sup>57</sup>. Notably, adhesin gene *cnaC*, also required for conjugative transfer, has previously been correlated with virulent poultry-NEC *C. perfringens* strains, whereas a pCP13 plasmid that encodes *cnaB* and *becAB* was recently linked to acute gastroenteritis outbreaks<sup>6,27,30,58</sup>. Carriage of these plasmids was found to be more common in all lineages except hypovirulent lineage V, suggesting that plasmid-encoded virulence traits represent key genomic signatures linked to pathophysiology. This potential plasmid circulation links with recent findings that phylogenetically distinct type-F *C. perfringens* isolates were found to harbour identical CPE-encoding plasmids in a single gastroenteritis outbreak<sup>8</sup>. However, as our results are based on short-read sequence data (albeit at very high sequencing coverage), further studies using long-read-based sequencing with plasmid reconstruction/characterization are required to probe these findings in detail.

We identified *pfoA* as the major toxin gene positively associated with intestinal cell injury, whereas *pfoA*<sup>+</sup> CPA-NEC strains also correlated with significantly higher cell death compared with *pfoA*<sup>-</sup> non-CPA-NEC strains. Notably, overt infection and subsequent production of

virulence factors in many cases will be prevented by gut microbiota colonization resistance mechanisms. However, the preterm infant gut microbiota is known to be significantly disrupted compared with full-term infants (in which NEC is extremely rare), which may allow overgrowth of *C. perfringens*<sup>1</sup>. The use of probiotics has been proposed as a prophylactic strategy to beneficially modulate the preterm gut microbiota and reduce incidence of NEC, and previous studies have indicated that supplementation with the early life microbiota genus *Bifidobacterium* spp. correlated with a reduction in *Clostridium* spp. and associated NEC<sup>59,60</sup>.

We added an antibiotic regimen (five-antibiotic cocktail + clindamycin) to deplete the murine microbiota, and more closely mimic the clinical scenario in preterm neonates, who are routinely exposed to empirical antibiotic therapy<sup>59</sup>. Distinctive disease symptoms were detected: *pfoA*<sup>+</sup> strains induced moderate colonic inflammation versus milder colonic inflammation observed in the *pfoA*<sup>-</sup> groups, suggesting that PFO may act as a major virulence factor linked to intestinal pathology. The pathological changes observed in distal colonic sections also correlated with diarrhoea, which is similar to *Citrobacter rodentium* infection, a model for human enteropathogenic *Escherichia coli*<sup>61,62</sup>. It also appeared that immune-mediated pathology plays a role during *pfoA*<sup>+</sup> *C. perfringens* infection, as evidenced by mucosal erosions, goblet cell reduction, immune cell infiltration and increased levels of the pro-inflammatory cytokine TNF- $\alpha$ , when compared with *pfoA*<sup>-</sup> infected animals. This implicates the importance of the potential immune-driven aspects of *C. perfringens* intestinal infection, which warrants further investigation. Although *pfoA*<sup>+</sup> CPA-NEC strains were used in this infection model, we did not observe severe CPA-NEC symptoms including complete intestinal necrosis and abdominal distension. This is probably due to differences between mice and humans and factors such as diet. The standard chow diet may not favour this protein-hungry pathogen, the use of adolescent mice (rather than neonatal mice) may lead to more efficient *C. perfringens* clearance through immune-mediated pathways, and murine rather than preterm gut microbiota may have impacted infection kinetics. Further optimization studies, including the use of 'humanized' mouse models, may allow development of a more clinically relevant CPA-NEC model<sup>63</sup>.

There are certain limitations associated with the present study. First, only five CPA-NEC strains from four CPA-NEC patients from two sister hospitals were included. Given the very rare occurrence of CPA-NEC (subset of NEC cases), we had access to a limited number of patients (including issues/restrictions related to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic), and we were unable to reconstruct any additional *C. perfringens* metagenome-assembled genomes (MAGs) from previously published preterm infant microbiome studies, which was largely due to lack of sequencing depth<sup>64</sup>. Within our dataset/cohort we had ~6% NEC incidence. A rate of between 5% and 15% NEC is often reported, but its incidence varies widely between NICUs, with NEC cases often reported in 'outbreaks' that may link to emergence of a specific virulent strain in the hospital environment, a hypothesis supported by our SNP analysis<sup>60,65</sup>. This is particularly important, because previous terminology for the disease 'intestinal gas gangrene' signals the more severe and fulminant condition of CPA-NEC, compared with classic NEC, often leading to surgery within 24 h and associated higher mortality rates (78%)<sup>19,66</sup>. Larger (and longer) surveillance studies that incorporate numerous NICUs across a wider geographical region would be required to capture a larger number of samples from CPA-NEC-diagnosed infants. We also fully recognize NEC as a multifactorial intestinal disease that has been associated with several bacteria including *C. perfringens* and *Klebsiella* and *Enterococcus* spp.<sup>21,22,67</sup>. Indeed, two preterm neonates had previously been diagnosed with *Klebsiella*-associated NEC in the previous cohort study<sup>22</sup> and, although they each harboured *C. perfringens* strains, these isolates exhibited a 'commensal' genomic signature (within hypovirulent lineage V and did not encode *pfoA*). We used

Caco-2 cells to examine whether the secreted toxins of *C. perfringens* caused epithelial damage. We appreciate the usual caveats of using an immortal cell line and over-extrapolation of results; however, our further strain testing in vivo does strengthen our findings in terms of cell death and links to pathology. Future work is required to pinpoint the mechanistic role of the virulence factors identified, such as *pfoA*, potentially in a more clinically relevant, preterm infant enteroid model<sup>68</sup> and/or using neonatal mice, although ethical and regulatory considerations limit these types of studies.

Overall, we have identified a potentially hypovirulent lineage of *C. perfringens*, characterized by lack of the *pfoA* toxin gene, colonization factors and survival-related capacities, in contrast to the typical *pfoA*<sup>+</sup> virulent lineages of *C. perfringens* (including CPA-NEC strains), which linked to in vitro and in vivo pathological traits. Potential dissemination of virulent *C. perfringens* strains was observed between and within preterm infants and hospitals (including a case of CPA-NEC), highlighting the potential value of routine surveillance and enhanced infection control measures.

## Methods

### Clinical samples and bacterial isolation work

We conducted a retrospective genomic analysis on *C. perfringens* isolates obtained from faecal samples of 70 neonatal patients admitted to NICUs at hospitals A, B, C, D and E, respectively, in the UK between February 2011 and March 2016 (Supplementary Table 2). All cohort data and sequenced microbiota profiles can be accessed in the original clinical study publications; most samples were sequenced via 16S ribosomal RNA or metagenomic sequencing as per the original studies/publications in which the colonization dynamics had already been reported and discussed<sup>22,59,69</sup>. Dates of hospital admission and transfers were extracted electronically. Faecal samples collected were stored at  $-80^{\circ}\text{C}$  in a freezer before the experimental process to isolate *C. perfringens* using the ethanol-shock method (50% ethanol in Robertson's cooked meat medium), followed by plating on fastidious anaerobic agar supplemented with defibrinated sheep blood and 0.1% sodium taurocholate; alternatively, faecal samples were plated directly on tryptose–sulfite–cycloserine egg yolk agar (TSC-EYA) before  $37^{\circ}\text{C}$  anaerobic incubation for 18–24 h (ref. 70). Multiple or single distinct colonies on the plates were purified and maintained as pure isolates in autoclaved brain–heart infusion (BHI) broth with 30% glycerol for cryopreservation at  $-80^{\circ}\text{C}$ . Subjects without *C. perfringens* were defined as *C. perfringens* negative by culturing. We predefined CPA-NEC in the present study as definite NEC diagnosis (Bell stage II/III) with over-abundant *C. perfringens* before NEC onset (Bell's staging classification system: Bell stage II/III is considered to be definite NEC whereas Bell stage I covers non-specific signs) as described in Sim et al.<sup>22</sup>. This was a 2-year microbiota study that involved a cohort of >350 infants, with 4 CPA-NEC cases who showed statistically enriched *C. perfringens* before NEC onset in longitudinal samples (by comparing with samples from aged-matched non-NEC preterm infants), although CPA-NEC strains were the isolates collected at the final time points before disease onset.

### Genomic DNA extraction and WGS

Pure isolates were cultured anaerobically overnight at  $37^{\circ}\text{C}$  in BHI broth (–10–15 h) for genomic DNA extraction using either the phenol–chloroform extraction method or FastDNA SPIN Kit for soil according to the manufacturer's instructions (hospital E samples only; MP Bio-medicals). For the phenol–chloroform extraction method<sup>71</sup>, briefly, overnight 10 ml of pure cultures in BHI was harvested, followed by resuspension of bacterial pellets in 2 ml of 25% sucrose in 10 mM Tris and 1 mM EDTA, pH 8.0. Cells were then enzymatically lysed using 50  $\mu\text{l}$  of lysozyme, 100  $\text{mg ml}^{-1}$  (Roche). Next, 100  $\mu\text{l}$  of Proteinase K, 20  $\text{mg ml}^{-1}$  (Roche), 30  $\mu\text{l}$  of RNase A, 10  $\text{mg ml}^{-1}$  (Roche), 400  $\mu\text{l}$  of 0.5 M EDTA, pH 8.0 and 250  $\mu\text{l}$  of 10% Sarkosyl NL30 (Thermo Fisher Scientific) were added accordingly into the lysed suspension. This

was then followed by 1 h of ice incubation and a  $50^{\circ}\text{C}$  water bath overnight. On the second day, three rounds of phenol–chloroform–isoamyl alcohol (Merck) extraction were performed using 15 ml of gel-lock tubes (QIAGEN). Chloroform–isoamyl alcohol (Merck) extraction was then carried out followed by ethanol precipitation and 70% ethanol wash (once). DNA pellets (usually visible) were finally resuspended in 200–300  $\mu\text{l}$  of 10 mM Tris, pH 8.0 and stored at  $-20^{\circ}\text{C}$  until further analysis. WGS of each isolate sample was performed on Illumina HiSeq 2500 to generate 101 125-bp paired-end reads<sup>72</sup>. Isolate samples from hospital E were sequenced on Illumina NextSeq 500 to generate 150-bp paired-end reads.

### De novo genome assemblies and metagenome-assembled genomes

Raw sequence reads (FASTQ) generated from sequencers were quality filtered ( $-q\ 20$ ) with fastp v.0.20.0 before de novo genome assembly using SPAdes v.3.14.1 (ref. 73) at default parameters. Contigs of <500 bp were discarded in each isolate genome assembly before subsequent analyses. Genome assembly statistics were generated via sequence-stats v.1.0 (ref. 74) (Supplementary Table 3). All infant-associated genome assemblies were subjected to contamination check via CheckM v.1.1.3 (ref. 75) and contaminated (contamination >5%) and/or incomplete (completeness <90%) genome assemblies were excluded from further analysis ( $n = 2$ ). A total of 117 human-gut-associated, metagenome-assembled genomes and 17 isolate genomes of *C. perfringens* was retrieved from the Unified Human Gastrointestinal Genome v.2.0 (UHGG)<sup>76</sup> collection (genome completeness >80% and contamination <5%).

### Phylogenetic analyses

A total of 673 *C. perfringens* genome assemblies was investigated, including 171 public isolate genomes retrieved from GenBank, 272 infant-associated draft genomes newly generated in the present study, 96 food-poisoning-linked isolate genomes (from the UK) published previously and 117 high-quality metagenome-assembled genomes, alongside 17 isolate genomes from UHGG (human-gut-associated *C. perfringens* genomes)<sup>8,76</sup>. All genomes investigated in the present study passed contamination checks via CheckM v.1.1.3 (ref. 75) (completeness >80% and contamination <5%) and taxonomically assigned as *C. perfringens* by gtdb-tk v.1.5.1 (ref. 77) (average nucleotide identity (ANI) > 95% against type strain).

These high-quality *C. perfringens* genome assemblies ( $n = 673$ ) were subsequently dereplicated at strain level (ANI > 99.9%) using dRep v.3.2.2 (ref. 78) before reconstruction of a core-gene alignment. A core-gene alignment (comprising 858 single-copy core genes) of 310 *C. perfringens* genomes based on Prokka v.1.14.6 (ref. 79) annotation of coding sequences (general feature format) was next generated using Panaroo v.1.2.8 (ref. 80) with options `--merge_paralogs`, `--clean-mode strict` and `--aligner mafft`<sup>81</sup>. In a similar manner, infant-associated, core-gene alignment (1,995 single-copy core genes) was generated using dereplicated strain-level genomes ( $n = 80$ ) via Panaroo v.1.2.8. SNPs were next extracted from core-gene alignments using snp-sites v.2.3.3 (ref. 82) and further dereplicated at core-gene level via in-house script, resulting in 302 representative genomes in the global phylogenetic tree (Fig. 1) and 80 representative genomes in the infant-associated phylogenetic tree (Fig. 2; with two reference genomes, type strains ATCC13124 and NCTC8679). Maximum-likelihood trees were constructed based on the extracted and dereplicated core-gene SNP alignment using IQ-TREE v.2.0.5 (ref. 83) with ultra-fast bootstrap replicates -B 1000 and automatic evolutionary model selection option `-m TEST` (best fit model for both trees was determined as GTR + F + ASC + G4).

Phylogenetic lineages were assigned using R library RhoBAPS v.1.1.3 (ref. 84). Tree topology was checked and validated using mash-distance trees generated by Mashree v.1.2.0 (ref. 85). Tree annotation was performed via iTOL v.6.5.8 (ref. 86).

### Toxinotype assignment and virulence profiling

Toxinotypes A–G were assigned to each *C. perfringens* sample via database TOXlper v.1.1 (ref. 87). Virulence-related genes and colonization factor sequence search were performed via ABRicate v.1.0.1 with options --minid=90 and --mincov=80 based on in-house sequence databases (Supplementary Table 5). ResFinder v.4.0 database was used via ABRicate for profiling antimicrobial resistance genes<sup>88</sup>.

### SNP analysis

SNPs were extracted from core-gene alignment with snp-sites v.2.3.3 (ref. 82) and snp-dists v.0.7 (ref. 89) was utilized for computing pairwise SNP distances. Transmission analysis was limited to hospital-associated strains (from 70 individuals) and within closest genetic distance, identified by phylogenetic topologies and pairwise SNP distances within 8,506 core-gene SNPs in 858 single-copy core-gene alignment. Probable transmission dynamics were predicted based on clinical metadata of the patients and the time of the individuals' stay within the hospitals. SNP distance was further explored to determine the exact genetic distance to infer transmission.

### Plasmid investigation

Plasmid analysis was limited to infant-associated *C. perfringens* genomes. Conjugative plasmids, both pCW3 and pCP13 families (known to encode multiple virulence factors), were extracted based on the identification of the *tcp* (pCW3 plasmids<sup>27</sup>;  $n = 12$ ) and *pcp* (pCP13 (ref. 28) plasmids;  $n = 20$ ) genes in the conjugative systems via ABRicate v.1.0.1 (ref. 90). Briefly, *tcp* and *pcp* loci and plasmid replication protein gene *rep* were comprehensively searched on 274 genomes; only single contigs within genome assemblies comprising more than five conjugative genes (*rep* is compulsory) were extracted and assumed as functional conjugative plasmids for further analyses. Predicted genes were mapped to reference plasmids LH112 (the largest plasmid size in both families) to allow genome architecture comparison of plasmid contents across both pCW3 and pCP13 plasmid families. Virulence factors were identified via ABRicate v.1.0.1 with in-house databases as described in the previous section. Plasmid sequences were aligned with MAFFT v.7.305b, SNP distance was compared via snp-dists v.0.7 (ref. 81). Easyfig v.2.2.2 (ref. 91) was utilized for visualization of plasmid sequence comparison.

### Maintenance of cell line Caco-2

Frozen stocks were resuscitated/resuspended in pre-warmed (37 °C) Dulbecco's Modified Eagle Medium (DMEM; Thermo Fisher Scientific), supplemented with 20% fetal bovine serum (FBS; Thermo Fisher Scientific). Cells were immediately transferred to T75 sterile culture flasks (Corning) and incubated at 37 °C in a 5% CO<sub>2</sub> incubator for 24 h. Spent medium was removed after 48–72 h, replaced with 20–25 ml of fresh warm medium as described and incubated until confluency was reached at ~5–8 d. Caco-2 cell line passages 53–54 were used for the present study, cells were split at ratios 1:5–10 for each passage. Cells were counted on a haemocytometer (Neubauer Chamber).

### In vitro phenotypic characterization

**Cell toxicity assay.** The Caco-2 cell line was maintained in liquid nitrogen at Quadram Institute Bioscience (American Type Culture Collection (ATCC), catalogue no. HTB-37). Cells (passages 53–54) were seeded at 20,000 cells per well in tissue-treated 96-well plates (Corning Costar). After 6–7 d of 5% CO<sub>2</sub> incubation at 37 °C, a confluent monolayer was formed (manually confirmed using microscopy) and subjected to further experiments. Briefly, phenol-red DMEM was removed, cells were washed twice gently and replaced with phenol-red-free DMEM supplemented with 1% FBS. Diluted sterile-filtered (0.22- $\mu$ m filter) bacterial supernatants (selected *C. perfringens* strains were cultured from 1:100 confluent overnight cultures for 10 h before collection of supernatants; all strains reached a density of  $\sim 10^9$  colony-forming units

(c.f.u.) ml<sup>-1</sup> at the end of 10 h) were added neat into each well at 12.5% total volume followed by a 2-h 5% CO<sub>2</sub> incubation at 37 °C. Cell death measurement (fluorescence signal of lactate dehydrogenase released from cells with a damaged cell membrane) was performed and estimated using CytoTox-ONE Homogeneous Membrane Integrity Assay according to the manufacturer's instructions (Promega) using a black 96-well plate to reduce background readouts (Corning).

### Haemolysis assay

Columbian blood agar (CBA; 5% sheep blood; Oxoid) was used to identify the haemolytic trait of *C. perfringens* isolates linked to pore-forming toxins produced. Briefly, *C. perfringens* was streaked on the blood agar and incubated anaerobically overnight (<20 h). Isolated colonies with clear transparent halos formed in and around the colonies indicate complete haemolysis. Assay outcomes were observed manually.

### Oxygen tolerance assay

Pure cultures of 31 *C. perfringens* strains (including 30 infant-associated isolates and type strain ATCC13124) were grown anaerobically to confluency for 24 h in BHI. Subsequently, cultures were spotted in a dilution series on to BHI agar supplemented with 0.1% sodium taurocholate (potent germinant for spores). Spotted Petri dishes were dried and incubated under ambient (aerobic) conditions at room temperature (22 °C) for specified time periods before being returned to an anaerobic chamber for determination of the colony-forming unit count (maximum 14 d). All colony-forming units were counted after 15–20 h of anaerobic incubation at 37 °C. Cultures (spotted plates) that were not exposed to ambient conditions acted as controls for each strain. A viable percentage was shown after comparing with control cultures.

### Sporulation assay

Cultures of 31 *C. perfringens* strains (including 30 infant-associated isolates and type strain ATCC13124) were induced to sporulate using the modified Duncan-strong medium (proteose peptone 16 g l<sup>-1</sup>, yeast extract 4 g l<sup>-1</sup>, cysteine hydrochloride 0.5 g l<sup>-1</sup>, disodium phosphate 10 g l<sup>-1</sup> and raffinose 4 g l<sup>-1</sup>, with a final pH of 7.8) for 24-h anaerobic incubation using 1:100 overnight confluent cultures<sup>92</sup>. Cultures were pre-treated with sterile-filtered 70% ethanol for 4 h to eliminate vegetative cells and spot plated on BHI agar supplemented with 0.1% taurocholate (potent germinant) for spores. Agar plates were incubated anaerobically for 24 h before counting the colony-forming units.

### In vivo studies

**Ethics and licence.** All animal experiments and related protocols described were performed under the Animals (Scientific Procedures) Act 1986 (ASP) under project licence no. PP8873233 and approved by the Home Office and University of East Anglia (UEA) FMH Research Ethics Committee. Animals are monitored and assessed frequently during studies for physical condition and behaviour. Mice determined to have suffered from distress would be euthanized via the ASPA schedule 1 protocol (CO<sub>2</sub> and cervical dislocation). Trained and qualified animal technicians carried out animal husbandry at UEA Disease Modelling Unit (DMU).

### Animals, housing and animal study design

Specific pathogen-free C57BL/6 female mice ( $n = 40$ ; 3-week-old *Mus musculus*), obtained from Charles River (France) were used in the animal study. Animals were transported to and housed in the UEA DMU acclimatization barn under specific pathogen-free conditions for 2 weeks before moving to the infection suite. During the infection study, mice were housed with autoclaved bedding (and cage), food (stock pellets) and water with a 12 h light cycle (12 h:12 h light:darkness). Cages were changed in a laminar flow cabinet. The rooms were maintained at  $21 \pm 2$  °C and  $55 \pm 10\%$  humidity. Air change rates are 12–15 per h.

Mice were caged into eight groups (five mice in each group), which include control group, ABX group (only treated with antibiotics) and six experimental groups (*pfoA*<sup>+</sup> versus *pfoA*<sup>-</sup>). Group *pfoA*<sup>+</sup> ( $n = 3$ ) only was challenged with *C. perfringens* strains IQ146, IQ129 and LH115, which encode a *pfoA* toxin gene, whereas group *pfoA*<sup>-</sup> ( $n = 3$ ) was given only *C. perfringens* strains that do not encode the *pfoA* toxin gene, IQ147, IQ133 and LH043, respectively.

Experimental groups ( $n = 6$ ) and the ABX group were treated with a five-antibiotic cocktail comprising kanamycin (0.4 mg ml<sup>-1</sup>), gentamicin (0.035 mg ml<sup>-1</sup>), colistin (850 U ml<sup>-1</sup>), metronidazole (0.215 mg ml<sup>-1</sup>) and vancomycin (0.045 mg ml<sup>-1</sup>) in autoclaved drinking water for 3 d administered freely (Fig. 6a). Drinking water was then switched to sterile water and, 48 h later, mice in the experimental group and the ABX group were orally gavaged with 150 mg kg<sup>-1</sup> of clindamycin. After 24 h, mice were gavaged with 10<sup>9</sup> c.f.u. of *C. perfringens* in 100 µl and all mice were closely monitored for signs of disease symptoms, including significant weight loss (>20%). All animals were euthanized and the organs harvested at day 7 (Fig. 6a).

### Bacterial strains and growth

Bacterial stocks were recovered on BHI agar for purity checks each time and subsequently cultured in BHI broth overnight to reach confluency (10<sup>9</sup> c.f.u. ml<sup>-1</sup>, approximately 14–17 h). Bacterial pellets were washed and resuspended in sterile phosphate-buffered saline before feeding the mice via oral gavaging using a 20G plastic sterile feeding tube.

### Faecal sample collection and colony-forming unit enumeration

Faecal samples were collected in sterilized tubes daily and stored at -80 °C until further analysis. Serially diluted faecal mixtures were plated on fresh TSC agar and the colony-forming units enumerated after <24 h anaerobic incubation. Pitch-black colonies were counted as *C. perfringens* colonies.

### Tissue processing and H&E staining

Distal colons were fixed in 10% neutral buffered formalin (~4% formaldehyde; Merck) for <24 h and followed by 70% ethanol. Tissues were subsequently processed using an automated Leica Tissue Processor ASP-300-S and embedded in paraffin manually. Next, sectioning of tissues was performed using a Leica microtome (5-µm-thick sections) and left overnight for samples to air dry before staining and further analysis. Haematoxylin and eosin (H&E) staining was then performed using Leica Autostainer for structural imaging of intestinal samples.

### Microscopy imaging

Bright-field microscopy was performed using Olympus BX60 with a microscope camera Jenoptik C10 and ProgRes CapturePro software v.2.10 (Supplementary Figs. 1–8).

### Histopathological scoring

All distal colonic section images (two to three distal colonic images per animal) were examined and graded single-blinded by L.J.H. The histological severity of intestinal inflammation was graded using a histopathological scoring system (0–14; denoting increasing severity) based on three general pathological features: inflammatory cell infiltrate (score value: 0–4), epithelial changes that include epithelial hyperplasia and goblet cell loss (score value: 0–5) and mucosal architecture—villous blunting (score value: 0–5) as described by Erben et al.<sup>93</sup>. The overall additive score (0–14) was the sum of each component score value. An overall pathological score of 1–4 suggests minimal colitis, 5–8 mild colitis, 9–11 moderate colitis and >12 marked/severe colitis.

### Morphometry

Crypt length of distal colons was examined digitally via ImageJ2 v.2.3.0 based on 20 crypts per representative image of distal colonic microscopy slides.

### Cytokine analysis

Proximal murine colons were homogenized in lysis buffer (150 mmol l<sup>-1</sup> of NaCl, 20 mmol l<sup>-1</sup> of Tris, 1 mmol l<sup>-1</sup> of EDTA, 1 mmol l<sup>-1</sup> of EGTA and 1% Triton X-100 at pH 7.5) using a FastPrep-24 bead-beating grinder (MP Biomedicals) at speed 4.0 for 40 s followed by speed 6.0 for 40 s. Samples were then centrifuged for 12 min at 4 °C. Supernatants were transferred to clean tubes and further quantified (protein concentration) before cytokine analysis.

Protein concentration was quantified by BCA (bicinchoninic acid) assay (BioRad) and cytokines were analysed using a customized Mesoscale Discovery U-Plex assay (MSD) according to the manufacturer's instructions. Cytokine concentrations were normalized to total protein input and samples were run in duplicates. The plate was read using an MSD QuickPlex SQ 120 imager and data analysed on the MSD Discovery Workbench software.

### Gut microbiome analysis

Genomic DNA of murine faecal samples was extracted, sequenced and analysed following the protocol described previously<sup>6</sup>. Briefly, DNA was extracted from samples using FastDNA Spin Kit for Soil (MP Biomedicals) following the manufacturer's instruction, while extending the bead-beating (in FastPrep-24 bead-beating grinder) step to 80 s at speed 6.0 (ref. 59). Next, genomic DNA extracts were subjected to the 16S rRNA gene V1 + V2 region library preparation before sequencing on an Illumina NovaSeq sequencing platform at 2 × 250 bp (paired-end sequencing)<sup>59</sup>. Sequencing raw reads (FASTQ) were first merged using PEAR v.0.9.6 (ref. 94), then underwent quality and chimera filtering via QIIME v.1.9.1 (ref. 95), subsequently with operational taxonomic unit (OTU) assignment using a SILVA\_132 database<sup>96</sup>. BIOM output of OTU tables was read using MEGAN6 (ref. 97).

### Data visualization

Phylogenetic trees (mid-point rooted and unrooted) were graphed using iTOL v.6.5.8 (ref. 86). Various statistical graphs including pie charts, line charts, dot plots, bar plots and box plots were drawn in R v.4.1.2 (ref. 98), using R libraries tidyverse v.1.3.1 (ref. 99), ggplot2 v.3.3.5 (ref. 100) and ggpubr v.0.4.0 (ref. 101). R library vegan v.2.6.2 (ref. 102) was used to construct non-metric multidimensional scaling (NMDS) plot for microbiome data and estimation of the inverse Simpson's index.

### Sample size, randomization and blinding

No statistical methods were used to predetermine sample size for experiments. Sample size for the in vivo study was selected following the 3Rs principles—reduction, replacement and refinement—ensuring sufficient sample size for meaningful statistical analysis.

Randomization was not attempted during the in vivo study to avoid cross-contamination because treatments were performed according to the individual cages; however, animals were initially allocated to each cage randomly. For in vitro experiments, randomization is not relevant due to the way bacterial isolates were selected with an aim of formally characterizing *pfoA*<sup>+</sup> isolates ( $n = 15$ ) and *pfoA*<sup>-</sup> isolates ( $n = 15$ ).

Single blinding was used in the in vivo colon tissue histopathological scoring analysis to prevent bias. For in vivo sample collection and analysis, the investigators were blinded to group allocation. For other in vitro experiments, data collection and analysis were not performed blind to the conditions of the experiments.

### Data exclusion

Two *C. perfringens* isolate genomes (sequenced in the present study) were excluded from genome analysis due to detected sequence contamination by CheckM v.1.1.3 (>10%). No animals or data points were excluded from analyses.

### Statistics and reproducibility

Statistical tests were performed via R base packages stats v.4.1.2 (ref. 98), including Fisher's exact test (two sided), Kruskal–Wallis test, Wilcoxon's test, analysis of variance (ANOVA), post-hoc Tukey's test and point-biserial correlation test. The Shapiro–Wilk normality test was used to formally test for data normality where appropriate. R library rstatix v.0.7.0 (ref. 103) was used for post-hoc Dunn's test of multiple comparisons (two sided), whereas rcompanion v.2.4.16 (ref. 104) was utilized to conduct pairwise post-hoc tests (Fisher's exact test).

For both oxygen tolerance and sporulation assays, experiments were performed on three biologically independent replicates per sample (bacterial strain;  $n = 3$ ). In cell toxicity assay, cell death measurements were performed in two independent experiments, each in three biological replicates per sample (sterile supernatants of bacterial strains;  $n = 6$ ). The animal study was conducted once with five C57BL/6 mice each group or cage ( $n = 40$ ). For murine distal colonic samples, two to three subsections of each processed distal colon (cross-sections) were photographed using a digital bright-field microscope and used for further phenotypic examinations.

### Ethics statement

Faecal collection from Norfolk and Norwich University Hospital (NNUH) and Rosie Hospital (BAMBI study) was approved by the Faculty of Medical and Health Sciences Ethics Committee at the UEA and followed protocols laid out by the UEA Biorepository (licence no. 11208). Faecal collection from Imperial Healthcare NICUs (NeoM and NeoM2 studies) was approved by West London Research Ethics Committee (REC) under the REC approval reference no. 10/H0711/39. In all cases, doctors and nurses recruited infants after parents had given written consent. Ethical approval for the SERVIS study was approved by the North East and N Tyneside committee 10/H0908/39, and signed parental consent was obtained from every parent. We have anonymized identifiers of patients, hospitals and associated clinical data.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Genome assemblies of 171 *C. perfringens* isolates were retrieved from the National Center for Biotechnology Information's (NCBI) GenBank (downloaded on 2 April 2020), while 96 *C. perfringens* food-poisoning-associated, isolated genomes were downloaded (May 2022) from the European Nucleotide Archive under accession no. PRJEB25764 (Supplementary Table 3). Another 117 *C. perfringens* metagenome-assembled genomes and 17 *C. perfringens* isolated genomes were retrieved from the UHGG v.2.0 collection (May 2022; Supplementary Table 3). Sequencing raw reads and draft genome assemblies for 272 *C. perfringens* isolates generated in the present study are publicly available in the NCBI Sequence Read Archive (SRA) and GenBank respectively, under accession no. PRJNA755973 (Supplementary Table 3). The 16S rRNA gene amplicon sequence reads (in vivo microbiome study) are publicly available in SRA under accession no. PRJNA755974 (Supplementary Table 4). Accessible links to sequence databases including ResFinder v.4.0, TOX1per v.1.1, *tcp* loci, *pcp* loci and *C. perfringens*-associated virulence genes (Supplementary Table 5) used in the present study are available in Supplementary Table 6. Anonymized clinical metadata have been made available in Supplementary Table 2. Computationally extracted plasmid sequences, histology images and data used for the figures (Source data files) are listed in the inventory ('Inventory\_of\_supplementary\_info.xlsx') which is openly shared via the GitHub repository: <https://github.com/ramondkiu/Infant-Clostridium-perfringens-Paper> (ref. 105). Source data are provided with this paper.

### Code availability

R scripts associated with data visualization and statistical analysis are available and shared via GitHub (<https://github.com/ramondkiu/Infant-Clostridium-perfringens-Paper>)<sup>105</sup>.

### References

- Kiu, R. & Hall, L. J. An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerg. Microbes Infect.* **7**, 141 (2018).
- Li, J., Paredes-Sabja, D., Sarker, M. R. & McClane, B. A. *Clostridium perfringens* sporulation and sporulation-associated toxin production. *Microbiol. Spectr.* <https://doi.org/10.1128/microbiolspec.TBS-0022-2015> (2016).
- Rood, J. I. et al. Expansion of the *Clostridium perfringens* toxin-based typing scheme. *Anaerobe* **53**, 5–10 (2018).
- Ronco, T. et al. Genome analysis of *Clostridium perfringens* isolates from healthy and necrotic enteritis infected chickens and turkeys. *BMC Res. Notes* **10**, 270 (2017).
- Lacey, J. A. et al. Whole genome analysis reveals the diversity and evolutionary relationships between necrotic enteritis-causing strains of *Clostridium perfringens*. *BMC Genom.* **19**, 379 (2018).
- Kiu, R. et al. Genomic analysis on broiler-associated *Clostridium perfringens* strains and exploratory caecal microbiome investigation reveals key factors linked to poultry necrotic enteritis. *Anim. Microbiome* **1**, 12 (2019).
- Mahamat Abdelrahim, A. et al. Large-scale genomic analyses and toxinotyping of *Clostridium perfringens* implicated in foodborne outbreaks in France. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2019.00777> (2019).
- Kiu, R. et al. 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. *Micro. Genom.* <https://doi.org/10.1099/mgen.0.000297> (2019).
- Camargo, A. et al. Intra-species diversity of *Clostridium perfringens*: a diverse genetic repertoire reveals its pathogenic potential. *Front. Microbiol.* **13**, 952081 (2022).
- Neu, J. & Walker, A. W. Necrotizing enterocolitis. *N. Engl. J. Med.* **364**, 255–264 (2011).
- Lim, J. C., Golden, J. M. & Ford, H. R. Pathogenesis of neonatal necrotizing enterocolitis. *Pediatr. Surg. Int.* **31**, 509–518 (2015).
- Lin, P. W. & Stoll, B. J. Necrotizing enterocolitis. *Lancet* **368**, 1271–1283 (2006).
- Rees, C. M., Eaton, S. & Pierro, A. National prospective surveillance study of necrotizing enterocolitis in neonatal intensive care units. *J. Pediatr. Surg.* **45**, 1391–1397 (2010).
- Shah, T. A. et al. Hospital and neurodevelopmental outcomes of extremely low-birth-weight infants with necrotizing enterocolitis and spontaneous intestinal perforation. *J. Perinatol.* **32**, 552–558 (2012).
- Bazaciu, C. & Neu, J. Necrotizing enterocolitis: long term complications. *Curr. Pediatr. Rev.* **15**, 115–124 (2019).
- Coggins, S. A., Wynn, J. L. & Weitkamp, J. H. Infectious causes of necrotizing enterocolitis. *Clin. Perinatol.* **42**, 133–154 (2015).
- Carlisle, E. M. & Morowitz, M. J. The intestinal microbiome and necrotizing enterocolitis. *Curr. Opin. Pediatr.* **25**, 382–387 (2013).
- Morowitz, M. J., Poroyko, V., Caplan, M., Alverdy, J. & Liu, D. C. Redefining the role of intestinal microbes in the pathogenesis of necrotizing enterocolitis. *Pediatrics* **125**, 777–785 (2010).
- Hanke, C. A. et al. *Clostridium perfringens* intestinal gas gangrene in a preterm newborn. *Eur. J. Pediatr. Surg.* **19**, 257–259 (2009).

20. Claud, E. C. Neonatal necrotizing enterocolitis: inflammation and intestinal immaturity. *Antiinflamm. Antiallergy Agents Med. Chem.* **8**, 248–259 (2009).
21. Heida, F. H. et al. A necrotizing enterocolitis-associated gut microbiota is present in the meconium: results of a prospective study. *Clin. Infect. Dis.* **62**, 863–870 (2016).
22. Sim, K. et al. Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clin. Infect. Dis.* **60**, 389–397 (2015).
23. Butel, M. J. et al. Clostridial pathogenicity in experimental necrotising enterocolitis in gnotobiotic quails and protective role of bifidobacteria. *J. Med. Microbiol.* **47**, 391–399 (1998).
24. Waligora-Dupriet, A. J., Dugay, A., Auzeil, N., Huerre, M. & Butel, M. J. Evidence of clostridial implication in necrotizing enterocolitis through bacterial fermentation in a gnotobiotic quail model. *Pediatr. Res.* **58**, 629–635 (2005).
25. Kosloske, A. M., Ball, W. S. Jr, Umland, E. & Skipper, B. Clostridial necrotizing enterocolitis. *J. Pediatr. Surg.* **20**, 155–159 (1985).
26. Kosloske, A. M., Ulrich, J. A. & Hoffman, H. Fulminant necrotising enterocolitis associated with clostridia. *Lancet* **312**, 1014–1016 (1978).
27. Revitt-Mills, S. A., Watts, T. D., Lyras, D., Adams, V. & Rood, J. I. The ever-expanding tcp conjugation locus of pCW3 from *Clostridium perfringens*. *Plasmid* **113**, 102516 (2020).
28. Watts, T. D. et al. pCP13, a representative of a new family of conjugative toxin plasmids in *Clostridium perfringens*. *Plasmid* **102**, 37–45 (2019).
29. Lacey, J. A., Johanesen, P. A., Lyras, D. & Moore, R. J. In silico identification of novel toxin homologs and associated mobile genetic elements in *Clostridium perfringens*. *Pathogens* **8**, 16 (2019).
30. Kiu, R. et al. Genomic analysis of *Clostridium perfringens* BEC/CP1E-positive, toxinotype D and E strains isolated from healthy children. *Toxins* **11**, 543 (2019).
31. Feng, Y. et al. Phylogenetic and genomic analysis reveals high genomic openness and genetic diversity of *Clostridium perfringens*. *Microb. Genom.* <https://doi.org/10.1099/mgen.0.000441> (2020).
32. Kiu, R., Caim, S., Alexander, S., Pachori, P. & Hall, L. J. 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 (2017).
33. Lakshminarayanan, B. et al. Prevalence and characterization of *Clostridium perfringens* from the faecal microbiota of elderly Irish subjects. *J. Med. Microbiol.* **62**, 457–466 (2013).
34. Hamza, D., Dorgham, S. M., Elhariri, M., Elhelw, R. & Ismael, E. New insight of apparently healthy animals as a potential reservoir for *Clostridium perfringens*: a public health implication. *J. Vet. Res.* **62**, 457–462 (2018).
35. Savinov, S. N. & Heuck, A. P. Interaction of cholesterol with perfringolysin O: what have we learned from functional analysis? *Toxins* **9**, 381 (2017).
36. Stevens, D. L. & Bryant, A. E. The role of clostridial toxins in the pathogenesis of gas gangrene. *Clin. Infect. Dis.* **35**, S93–S100 (2002).
37. Verherstraeten, S. et al. The synergistic necrohemorrhagic action of *Clostridium perfringens* perfringolysin and alpha toxin in the bovine intestine and against bovine endothelial cells. *Vet. Res.* **44**, 45 (2013).
38. Zhang, J. et al. Verbascoside protects mice from clostridial gas gangrene by inhibiting the activity of alpha toxin and perfringolysin O. *Front. Microbiol.* **11**, 1504 (2020).
39. Li, Z., Yan, C., Gong, X. & Wang, J. Severe intravascular hemolysis from *Clostridium perfringens* septicemia in a neonate with necrotizing enterocolitis in China: a case report. *Infect. Drug Resist.* **15**, 1461–1465 (2022).
40. Suzaki, A. et al. Pathogenic characterization of *Clostridium perfringens* strains isolated from patients with massive intravascular hemolysis. *Front. Microbiol.* **12**, 713509 (2021).
41. Awad, M. M., Ellemor, D. M., Boyd, R. L., Emmins, J. J. & Rood, J. I. Synergistic effects of alpha-toxin and perfringolysin O in *Clostridium perfringens*-mediated gas gangrene. *Infect. Immun.* **69**, 7904–7910 (2001).
42. Dittmar, E. et al. Necrotizing enterocolitis of the neonate with *Clostridium perfringens*: diagnosis, clinical course, and role of alpha toxin. *Eur. J. Pediatr.* **167**, 891–895 (2008).
43. Park, J. M., Ng, V. H., Maeda, S., Rest, R. F. & Karin, M. Anthrolysin O and other gram-positive cytolysins are toll-like receptor 4 agonists. *J. Exp. Med.* **200**, 1647–1655 (2004).
44. Yamamura, K. et al. Inflammasome activation induced by perfringolysin O of *Clostridium perfringens* and its involvement in the progression of gas gangrene. *Front. Microbiol.* **10**, 2406 (2019).
45. Yuan, J., Najafav, A. & Py, B. F. Roles of caspases in necrotic cell death. *Cell* **167**, 1693–1704 (2016).
46. Rood, J. I. & Cole, S. T. Molecular genetics and pathogenesis of *Clostridium perfringens*. *Microbiol. Rev.* **55**, 621–648 (1991).
47. Verherstraeten, S. et al. Perfringolysin O: the underrated *Clostridium perfringens* toxin? *Toxins* **7**, 1702–1721 (2015).
48. Kumar, N. et al. Genome-based infection tracking reveals dynamics of *Clostridium difficile* transmission and disease recurrence. *Clin. Infect. Dis.* **62**, 746–752 (2015).
49. Garcia-Fernandez, S. et al. Whole-genome sequencing reveals nosocomial *Clostridioides difficile* transmission and a previously unsuspected epidemic scenario. *Sci. Rep.* **9**, 6959 (2019).
50. Moffatt, C. R., Howard, P. J. & Burns, T. A mild outbreak of gastroenteritis in long-term care facility residents due to *Clostridium perfringens*, Australia 2009. *Foodborne Pathog. Dis.* **8**, 791–796 (2011).
51. Ellis, C. L., Rutledge, J. C. & Underwood, M. A. Intestinal microbiota and blue baby syndrome: probiotic therapy for term neonates with cyanotic congenital heart disease. *Gut Microbes* **1**, 359–366 (2010).
52. Malagon, I. et al. Gut permeability in paediatric cardiac surgery. *Br. J. Anaesth.* **94**, 181–185 (2005).
53. McElhinney, D. B. et al. Necrotizing enterocolitis in neonates with congenital heart disease: risk factors and outcomes. *Pediatrics* **106**, 1080–1087 (2000).
54. Browne, H. P. et al. Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543–546 (2016).
55. Rolfe, R. D., Hentges, D. J., Campbell, B. J. & Barrett, J. T. Factors related to the oxygen tolerance of anaerobic bacteria. *Appl. Environ. Microbiol.* **36**, 306–313 (1978).
56. Lemyre, B. et al. A decrease in the number of cases of necrotizing enterocolitis associated with the enhancement of infection prevention and control measures during a *Staphylococcus aureus* outbreak in a neonatal intensive care unit. *Infect. Control Hosp. Epidemiol.* **33**, 29–33 (2012).
57. Dufrene, Y. F. & Viljoen, A. Binding strength of gram-positive bacterial adhesins. *Front. Microbiol.* **11**, 1457 (2020).
58. Yonogi, S. et al. BEC, a novel enterotoxin of *Clostridium perfringens* found in human clinical isolates from acute gastroenteritis outbreaks. *Infect. Immun.* **82**, 2390–2399 (2014).
59. Alcon-Giner, C. et al. Microbiota supplementation with *Bifidobacterium* and *Lactobacillus* modifies the preterm infant gut microbiota and metabolome: an observational study. *Cell Rep. Med.* **1**, 100077 (2020).
60. Robertson, C. et al. Incidence of necrotising enterocolitis before and after introducing routine prophylactic *Lactobacillus* and *Bifidobacterium* probiotics. *Arch. Dis. Child Fetal Neonatal Ed.* **0**, F1–F7 (2019).

61. Hall, L. J. et al. Natural killer cells protect against mucosal and systemic infection with the enteric pathogen *Citrobacter rodentium*. *Infect. Immun.* **81**, 460–469 (2013).
62. Bouladoux, N., Harrison, O. J. & Belkaid, Y. The mouse model of infection with *Citrobacter rodentium*. *Curr. Protoc. Immunol.* **119**, 19.15.11–19.15.25 (2017).
63. Lundberg, R. Humanizing the gut microbiota of mice: opportunities and challenges. *Lab. Anim.* **53**, 244–251 (2019).
64. Tarracchini, C. et al. Unraveling the microbiome of necrotizing enterocolitis: insights in novel microbial and metabolomic biomarkers. *Microbiol. Spectr.* **9**, e0117621 (2021).
65. Alsaied, A., Islam, N. & Thalib, L. Global incidence of necrotizing enterocolitis: a systematic review and meta-analysis. *BMC Pediatr.* **20**, 344 (2020).
66. Haigh, E. Gas gangrene of the colon in a newborn infant; report of a case successfully treated by total colectomy. *Br. J. Surg.* **43**, 659–661 (1956).
67. Raskind, C. H., Dembry, L. M. & Gallagher, P. G. Vancomycin-resistant enterococcal bacteremia and necrotizing enterocolitis in a preterm neonate. *Pediatr. Infect. Dis. J.* **24**, 943–944 (2005).
68. Stewart, C. J., Estes, M. K. & Ramani, S. Establishing human intestinal enteroid/organoid lines from preterm infant and adult tissue. *Methods Mol. Biol.* **2121**, 185–198 (2020).
69. Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).
70. Kotsanas, D. et al. Novel use of tryptose sulfite cycloserine egg yolk agar for isolation of *Clostridium perfringens* during an outbreak of necrotizing enterocolitis in a neonatal unit. *J. Clin. Microbiol.* **48**, 4263–4265 (2010).
71. Kiu, R. et al. Preterm Infant-Associated *Clostridium tertium*, *Clostridium cadaveris*, and *Clostridium paraputrificum* Strains: genomic and Evolutionary Insights. *Genome Biol. Evol.* **9**, 2707–2714 (2017).
72. Collins, J. et al. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature* **553**, 291–294 (2018).
73. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
74. Kiu, R. sequence-stats: generate sequence statistics from FASTA and FASTQ files. *GitHub* <https://github.com/ramondkiu/sequence-stats> (2020).
75. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
76. Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
77. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
78. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
79. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
80. Tonkin-Hill, G. et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
81. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
82. Page, A. J. et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Micro. Genom.* **2**, e000056 (2016).
83. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
84. Tonkin-Hill, G., John, A. L., Stephen, D. B., Simon, D. W. F. & Jukka, C. R. R. BAPS: an R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res.* **3**, 93 (2018).
85. Katz, L. S. et al. MashTree: a rapid comparison of whole genome sequence files. *J. Open Source Softw.* **4**, 1762 (2019).
86. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
87. Kiu, R. TOXlper: rapid *Clostridium perfringens* toxinotype assignment using genome assemblies. *GitHub* <https://github.com/ramondkiu/TOXlper> (2020).
88. Bortolaia, V. et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**, 3491–3500 (2020).
89. Seemann, T., Klotzl, F. & Page, A. J. snp-dists: pairwise SNP distance matrix from a FASTA sequence alignment. *GitHub* <https://github.com/tseemann/snp-dists> (2018).
90. Seemann, T. ABRicate: mass screening of contigs for antimicrobial and virulence genes. *GitHub* <https://github.com/tseemann/abricate> (2018).
91. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
92. Duncan, C. L. & Strong, D. H. Improved medium for sporulation of *Clostridium perfringens*. *Appl. Microbiol.* **16**, 82–89 (1968).
93. Erben, U. et al. A guide to histomorphological evaluation of intestinal inflammation in mouse models. *Int. J. Clin. Exp. Pathol.* **7**, 4557–4576 (2014).
94. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
95. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
96. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
97. Huson, D. H. et al. MEGAN community edition—interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957 (2016).
98. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021).
99. Wickham, H. et al. Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
100. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
101. Kassambara, A. ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0. <https://CRAN.R-project.org/package=ggpubr> (2020).
102. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
103. Kassambara, A. rstatix: pipe-friendly framework for basic statistical tests. R package version 0.7.0. <https://CRAN.R-project.org/package=rstatix> (2021).
104. Mangiafico, S. S. rcompanion: functions to support extension education program evaluation. R package version 2.4.16. <https://CRAN.R-project.org/package=rcompanion> (2022).
105. Kiu, R. et al. Particular genomic and virulence traits associated with preterm-derived toxigenic *Clostridium perfringens* strains. *Zenodo* <https://doi.org/10.5281/zenodo.7756267> (2023).

## Acknowledgements

This research was supported in part by the NBI Computing infrastructure for Science (CiS) group through the provision of a high-performance computing cluster. L.J.H. is supported by: Wellcome Trust Investigator Awards (nos. 100974/C/13/Z and 220876/Z/20/Z); the Biotechnology and Biological Sciences Research Council (BBSRC), Institute Strategic Programme Gut Microbes and Health BB/R012490/1, and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356. D.P. was funded by the Wellcome Trust. C.J.S. is supported by the Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (no. 221745/Z/20/Z). We thank the sequencing team at both the Wellcome Trust Sanger Institute and the Quadram Institute Bioscience (QIB) for genome sequencing. We would like to express our appreciation to our colleague S. Carding, and former colleagues S. Caim, L. Harnisch and B. Kirkup at QIB for their discussion on the data and assistance in various training sessions associated with this work.

## Author contributions

R.K., A.G.S., K.S., G. Douce, J.S.K. and L.J.H. conceived the study. R.K., S.A.D., S.D.R. and L.J.H. provided the methodology. R.K. provided the software. K.S., A.G.S. and L.J.H. validated the study. R.K. and L.J.H. did the formal analysis. R.K., K. S., A.G.S., A.A.G., C.A.P., S.A.D., W.J.F., E.C., J.M., C.A., S.P., Z.S. and L.J.H. carried out the investigations. R.K., A.G.S., E.C., D.P., G.B., J.E.B., G.R.Y., C.J.S., G. Dougan, P.C. and J.S.K. provided the resources. R.K., A.G.S., K.S. and C.A. curated the data. R.K. and L.J.H. wrote the original draft of the MS. R.K., A.G.S., K.S., Z.S., J.E.B., C.J.S., J.S.K. and L.J.H. reviewed and edited the MS. R.K. carried out the visualization. J.S.K. and L.J.H. supervised the study. R.K. administered the project. J.S.K. and L.J.H. acquired funds.

## Competing interests

C.J.S. declares performing consultancy for Astarte Medical, receiving lecture honoraria from Nestlé Nutrition Institute and supervising a BBSRC collaborative training partnership PhD student for which Nestlé are involved (no salary or other personal payment is provided by Nestlé). The remaining authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41564-023-01385-z>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41564-023-01385-z>.

**Correspondence and requests for materials** should be addressed to Lindsay J. Hall.

**Peer review information** *Nature Microbiology* thanks Horst Posthaus and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

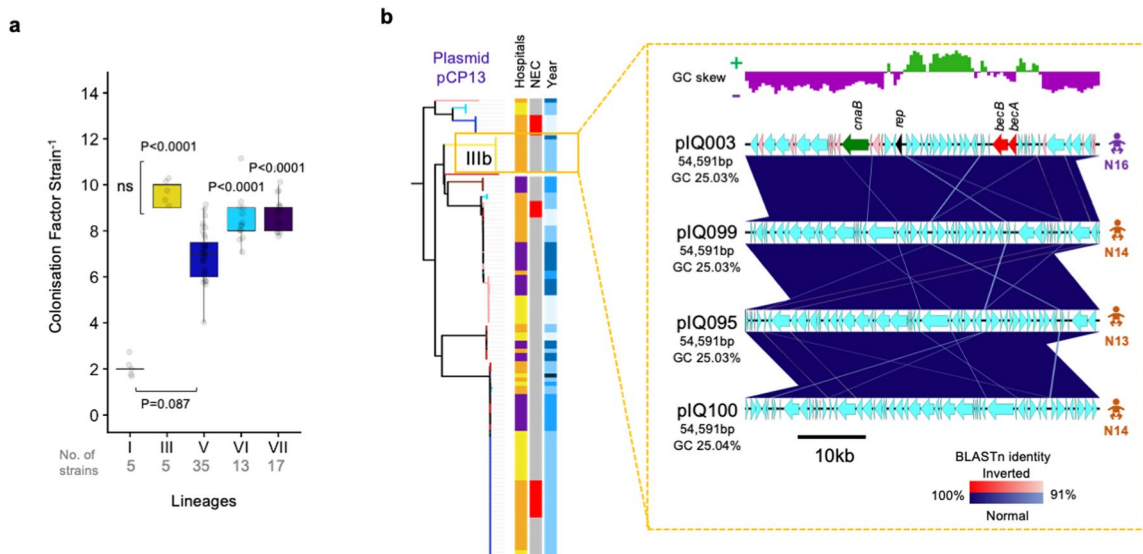
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>1</sup>Gut Microbes and Health, Quadram Institute Bioscience, Norwich, UK. <sup>2</sup>Faculty of Medicine, Imperial College London, London, UK. <sup>3</sup>Department of Medicine, University of Cambridge, Cambridge, UK. <sup>4</sup>Neonatal Intensive Care Unit, The Rosie Hospital, Cambridge, UK. <sup>5</sup>Hub for Biotechnology in the Built Environment, Northumbria University, Newcastle upon Tyne, UK. <sup>6</sup>Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK. <sup>7</sup>Newcastle Neonatal Services, Newcastle upon Tyne NHS Foundation Trust, Newcastle upon Tyne, UK. <sup>8</sup>Norfolk and Norwich University Hospital, Norwich, UK. <sup>9</sup>Norwich Medical School, University of East Anglia, Norwich, UK. <sup>10</sup>Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. <sup>11</sup>School of Biological Sciences, University of East Anglia, Norwich, UK. <sup>12</sup>Intestinal Microbiome, School of Life Sciences, ZIEL—Institute for Food & Health, Technical University of Munich, Freising, Germany. ✉e-mail: [lindsay.hall@quadram.ac.uk](mailto:lindsay.hall@quadram.ac.uk)

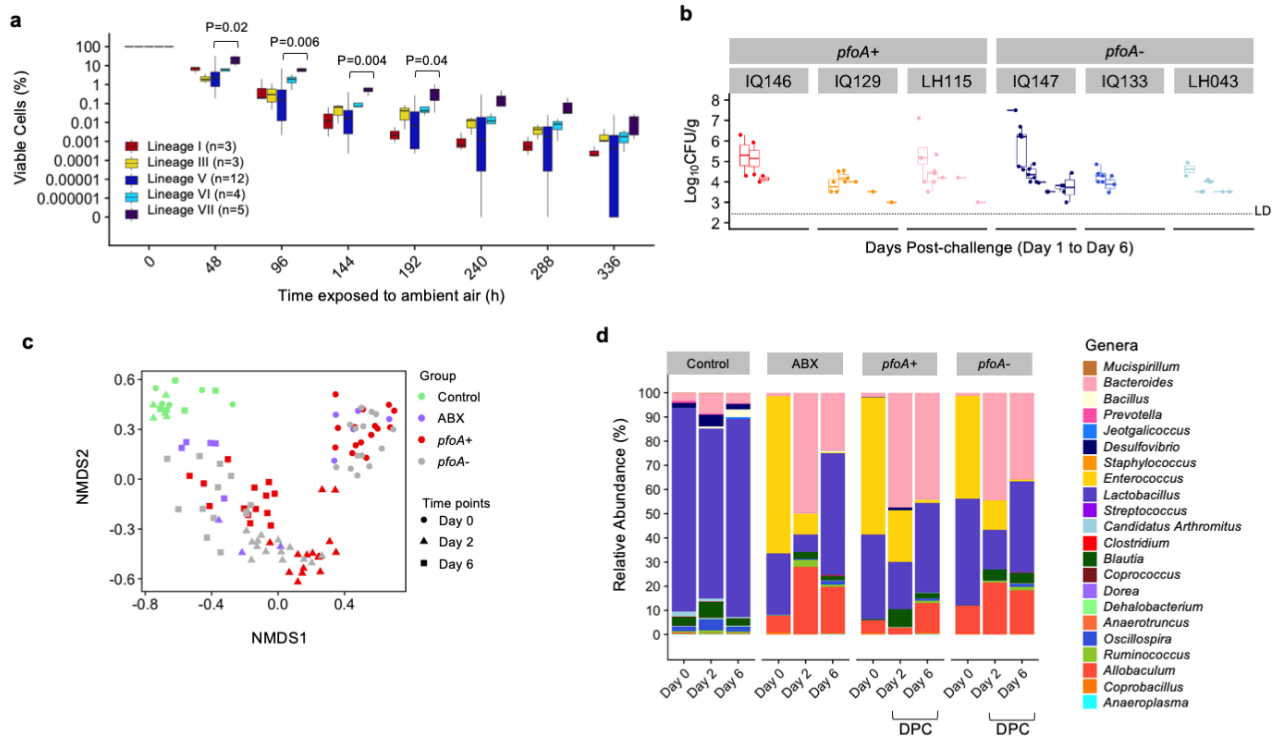




### Extended Data Fig. 1 | Genomic analysis of *C. perfringens* strains.

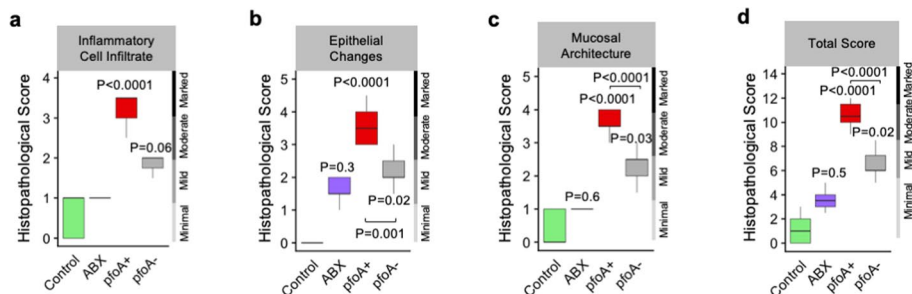
**a**, Colonisation factor comparison across all isolates by lineage. Colonisation factors are significantly more in lineages III, VI and VII than lineages I and V, while no significant difference between lineages III, VI and VII, similarly also lineages I and V. The box in boxplot represents 50% of the central data, in between lower

and upper quartile, with a central line representing the median while whiskers show the most extreme data points. Statistical analyses were performed with Kruskal-Wallis test and post-hoc Dunn's test of multiple comparison with P values adjusted by Benjamini-Hochberg method. **b**, Comparative genomic architecture analysis of identical plasmids from sub-lineage IIIb isolates.

**Extended Data Fig. 2 | Strain-level in vitro and in vivo assessment of**

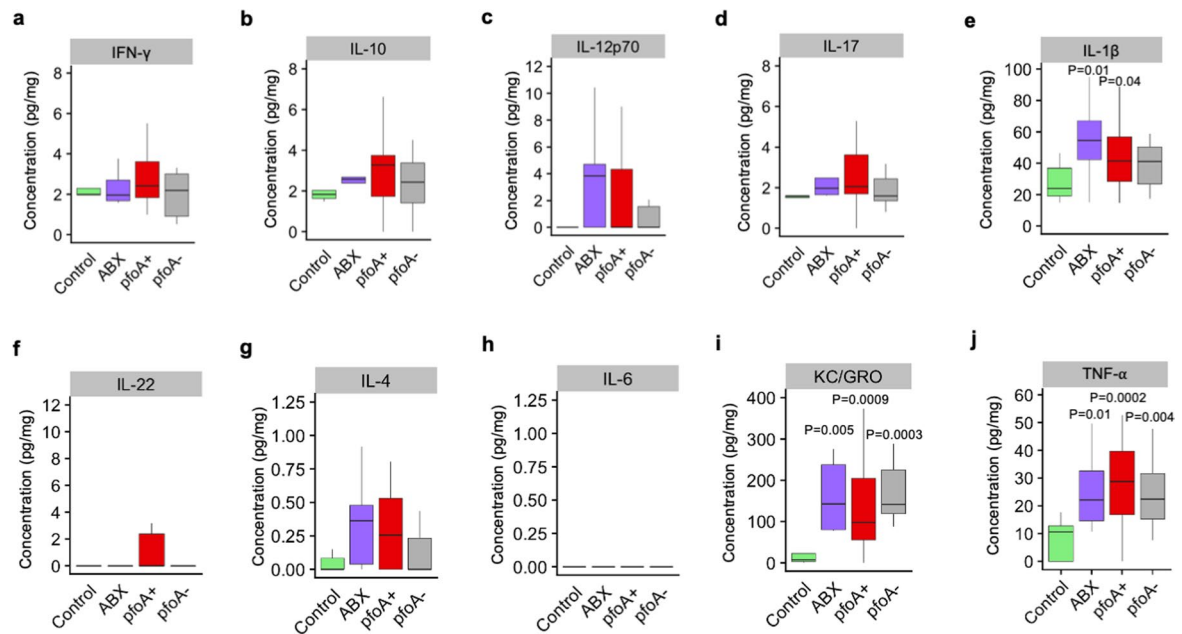
**C. perfringens** virulence potentials. **a**, Oxygen tolerance test. Comparison of viability of *C. perfringens* strains ( $n = 31$ ) from lineages I, III, V, VI and VII exposed to ambient air (aerobic) over 14 days in 3 biologically independent samples for each strain. Cell viability in percentage was shown from 48 h to 192 h between lineage VII and lineage V with statistical difference. The box in boxplot represents 50% of the central data, in between lower and upper quartile, with a central line representing the median while whiskers show the most extreme data points (also in **b**). **b**, Intestinal colonisation of *C. perfringens* monitored over 7 days

post-challenge. LD, limit of detection. **c**, NMDS plot of murine genus-level gut microbiome profiles on day 0, day 2 and day 6. **d**, Comparison of genus-level murine gut microbiome on day 0, day 2 and day 6. A total of 4 groups were compared, Control ( $n = 5$ ), ABX ( $n = 5$ ), *pfoA+* ( $n = 15$ ) and *pfoA-* ( $n = 15$ ). DPC: days post-challenge. (All statistical analyses were performed with Kruskal-Wallis test and post-hoc Dunn's test of multiple comparisons with P values adjusted by Benjamini-Hochberg method. The box in boxplots represents 50% of the central data, in between first and third quartile, with an inner line representing the median).



**Extended Data Fig. 3 | Histopathological analysis on murine distal colonic sections. a-d**, Histopathological scores for intestinal inflammation of murine distal colons comparison between groups in three main general criteria (features), with indicative severity scale (minimal, mild, moderate and marked). There were 8 groups of mice ( $n = 5$  each group) including control groups (Control and ABX) and 6 experimental groups (3 groups of *pfoA*<sup>+</sup> and 3 groups of *pfoA*<sup>-</sup>). The box in boxplot represents 50% of the central data, in between lower and upper quartile, with a central line representing the median while whiskers show the most extreme data points. All statistical analyses were performed

with Kruskal-Wallis test and post-hoc Dunn's test of multiple comparisons with P values adjusted by Benjamini-Hochberg method, unless otherwise stated. Significance was tested against untreated group, otherwise significance brackets will be used in pairwise comparison. **a**, Inflammatory infiltrate: immune cell infiltration (score value: 0–4). **b**, Epithelial changes: epithelial hyperplasia and goblet cell loss (score value: 0–5). **c**, Mucosal architecture: villous blunting (score value: 0–5). **d**, Total score (additive score value: 0–14). Increments in score value denotes increasing severity. Scoring was based on two H&E-stained images per distal colon.



**Extended Data Fig. 4 | Evaluation of pro-inflammatory cytokine profiles in proximal murine colons. a-j.** Concentration (pg/mg) of cytokines IFN- $\gamma$  (a), IL-10 (b), IL-12p70 (c), IL-17 (d), IL-1 $\beta$  (e), IL-22 (f), IL-4 (g), IL-6 (h), KC/GRO (i) and TNF- $\alpha$  (j) respectively, in murine proximal colonic samples, stratified into groups: Control ( $n = 5$ ), ABX ( $n = 5$ ), *pfoA*<sup>+</sup> ( $n = 15$ ), *pfoA*<sup>-</sup> ( $n = 15$ ). All individual samples were measured in technical duplicates ( $n = 2$ ). The box in boxplot represents

50% of the central data, in between lower and upper quartile, with a central line representing the median while whiskers show the most extreme data points. All statistical analyses were performed with Kruskal-Wallis test and post-hoc Dunn's test of multiple comparisons with P values adjusted by Benjamini-Hochberg method, unless otherwise stated. Significance was tested against untreated group, otherwise significance brackets will be used in pairwise comparison.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection	The dataset of 171 <i>Clostridium perfringens</i> public bacterial genome assemblies were downloaded from NCBI GenBank (on 2 April 2020) for comparison with novel isolate sequences (272) generated in this study. 96 <i>C. perfringens</i> isolate genomes were downloaded (May 2022) from ENA accession PRJEB25764. Another 117 <i>C. perfringens</i> metagenome assembled genomes and 17 <i>C. perfringens</i> isolate genomes were retrieved from the Unified Human Gastrointestinal Genome v2.0 (UHGG) collection. Genomes were all retrieved via web browser, no specific tools were used.
Data analysis	1) fastp v0.20.0 2) SPAdes v3.14.1 3) Sequence-stats v1.0 4) checkm v1.1.3 5) gtdb-tk v1.5.1 6) dRep v3.2.2 7) Prokka v1.14.6 8) Panaroo v1.2.8 9) snp-sites v2.3.3 10) IQ-TREE v.2.0.5 11) RhierBAPS v1.1.3 12) Mashtree v1.2.0 13) iTOL v6.5.8 14) ABRicate v1.0.1 15) snp-dists v0.7 16) MAFFT v7.305b 17) Easyfig v2.2.2 18) ProgRes CapturePro v2.10 19) ImageJ2 v2.3.0 20) PEAR v0.9.6 21) QIIME v1.9.1 22) MEGAN v6 23) R v4.1.2 24) R tidyverse v1.3.1 25) R ggplot2 v3.3.5 26) R ggpubr v0.4.0 27) R vegan v2.6.2 28) R stats v4.1.2 29) R rstatix v0.7.0 30) R rcompanion v2.4.16. R scripts and source data files associated with data visualisation and statistical analysis are available and shared via GitHub ( <a href="https://github.com/raymondkiu/Infant-Clostridium-perfringens-Paper">https://github.com/raymondkiu/Infant-Clostridium-perfringens-Paper</a> )

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Genome assemblies of 171 *C. perfringens* isolates were retrieved from NCBI GenBank (downloaded on 2 April 2020), while 96 *C. perfringens* food-poisoning associated isolate genomes were download (May 2022) from European Nucleotide Archive (ENA) under project accession PRJEB25764 (Supplementary Table 3). Another 117 *C. perfringens* metagenome assembled genomes and 17 *C. perfringens* isolate genomes were retrieved from the Unified Human Gastrointestinal Genome v2.0 (UHGG) collection (May 2022; Supplementary Table 3). Sequencing raw reads and draft genome assemblies for 272 *C. perfringens* isolates generated in this study are publicly available in the NCBI Sequence Read Archive (SRA) and GenBank respectively, under project PRJNA755973 (Supplementary Table 3). 16S rRNA gene amplicon sequence reads (in vivo microbiome study) are publicly available in SRA under project PRJNA755974 (Supplementary Table 4). Accessible links to sequence databases including ResFinder v4.0, TOXlper v1.1, tcp loci, pcp loci and *C. perfringens*-associated virulence genes (Supplementary Table 5) used in this study are available in Supplementary Table 6. Anonymised clinical metadata has been made available in Supplementary Table 2. Computationally extracted plasmid sequences, histology images, and data used for the figures (source data files) are listed in the inventory ("Inventory\_of\_supplementary\_info.xlsx") which is openly shared via GitHub repository: <https://github.com/ramondkiu/Infant-Clostridium-perfringens-Paper>

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

### Reporting on sex and gender

*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*

### Population characteristics

*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*

### Recruitment

*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*

### Ethics oversight

*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences     Behavioural & social sciences     Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Sample size

For genomic analysis, *Clostridium perfringens* strains were isolated from longitudinal stool samples of 70 infants residing in hospital NICU, analysed by Whole Genome Sequencing (n=272). An additional 171 *C. perfringens* genomes were obtained from NCBI GenBank, 96 *C. perfringens* isolate genomes were downloaded from ENA accession PRJEB25764, 117 *C. perfringens* metagenome assembled genomes and 17 *C. perfringens* isolate genomes were retrieved from the Unified Human Gastrointestinal Genome v2.0 (UHGG) collection. Sample size for in vivo study was selected following 3 R's principles - Reduction, Replacement and Refinement, ensuring sufficient sample size for meaningful statistical analysis. For in vitro assays, both sporulation and oxygen tolerance assays, experiments were performed in 3 biologically independent replicates per sample. For haemolysis assay, also 3 biological replicates were attempted. Specifically for cell toxicity (Caco2 cell line) assays, two separate experiments were performed with 3 biologically independent replicates each sample (in total 6 biological replicates; bacterial strain supernatants) to achieve biologically/statistically meaningful results.

### Data exclusions

2 novel isolate genomes were excluded from genome analysis due to detected sequence contamination by checkm software (>10%).

Replication	Cell toxicity assay was performed twice with 3 biological replicates each sample per experiment. For sporulation, oxygen tolerance and haemolysis assays, 3 biological replicates for each sample in single experiments. All attempts of replication were successful and no experimental data were excluded.
Randomization	Randomisation was not possible in in vivo study to avoid cross-contamination as treatments were performed according to the individual cages. Randomisation is not relevant to all other in vitro experiments due to the way isolates were selected with an aim to characterise pfoA+ isolates (n=15) and pfoA- isolates (n=15).
Blinding	Single-blinding was used in in vivo colon tissue histopathological scoring analysis to prevent bias. For in vivo sample collection and analysis, the investigators were blinded to group allocation. For other in vitro experiments, data collection and analysis were not performed blind to the conditions of the experiments.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Caco-2 cell line maintained in liquid nitrogen at Quadram Institute Bioscience (QIB). Commercial source: <a href="https://www.atcc.org/products/htb-37">https://www.atcc.org/products/htb-37</a>
Authentication	Cell lines were validated using microscopy to determine typical cell morphology and structure.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell line was used in the study

## Animals and other research organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Mus musculus C57BL/6 mice, female, 3-5 weeks old.
Wild animals	This study did not involve wild animals.
Reporting on sex	Sex was not considered in the study design.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All animal experiments and related protocols described were performed under the Animals (Scientific Procedures) Act 1986 (ASPAs) under project licence no. PP8873233 and approved by Home Office and University of East Anglia (UEA) FMH Research Ethics Committee. Animals are monitored and assessed frequently during studies for physical condition and behavior. Mice determined to have suffered from distress would be euthanised via ASPA Schedule 1 protocol (CO2 and cervical dislocation). Trained and qualified animal technicians carried out animal husbandry at UEA Disease Modelling Unit (DMU).

Note that full information on the approval of the study protocol must also be provided in the manuscript.