The role of *Clostridium perfringens* in intestinal diseases

Raymond Kiu Kwong Ong

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

Quadram Institute Bioscience

United Kingdom

November 2018

Abstract

Clostridium perfringens, a spore-forming bacterium that produces >20 toxins, is known to cause both human and animal intestinal diseases including; foodborne diarrhoea, preterm necrotising enterocolitis (NEC) and necrotic enteritis (NE) in chickens. Currently, there is very limited information on genomic virulence determinants, and the phylogenic and epidemiology landscape of this enteric pathogen, thus in this thesis I sought to comprehensively explore intestinal-associated *C. perfringens* strains using both genomic and phenotypic methodologies.

Whole Genome Sequencing (WGS) and bioinformatic approaches was used to examine a novel collection of *C. perfringens* isolates and publically available genomes (n=552, including 109 public genomes) from a diverse range of hosts and disease states including; NEC-associated preterm infants, foodborne diarrhoea patients, NE-associated broilers, and healthy humans and animals. These genomes were analysed in combination or as discreet disease subsets to determine infection-linked virulence features, genomic epidemiology and hyper-virulent genotypes, and indicated a highly diverse pangenome (7.4% core genome), toxin-specific and novel virulence factors, widespread distribution of conjugative plasmids, long-term persistence, and inter-regional transmission events during disease outbreaks. This work represents the largest WGS-based phylogenetic and comparative genomics on *C. perfringens* to date.

Furthermore, a sub-set of *C. perfringens* isolates were characterised phenotypically using microbiological assays to determine hyper-virulent phenotypes, which linked to genomic analysis. These hyper-virulent strains were then used to establish a novel *C. perfringens* oral-challenge mouse model. This enteric infection model will allow further mechanistic work in understanding the role of *C. perfringens* in relevant intestinal diseases and may be used to facilitate therapy or vaccine development.

Overall, this multidisciplinary work provides important novel insights into the intestinal pathogen *C. perfringens* at both a genomic and phenotypic level and provides a platform for subsequent translational studies into efforts to reduce *C. perfringens*-associated disease burden in both humans and animals.

Declaration

This thesis is the result of my own work during my PhD and therefore no material presented has ever been submitted for the award of any degree at any university. All the materials and work done in collaboration presented in this thesis has been appropriately acknowledged and accredited in the text.

Assistance and guidance in bioinformatics analysis were provided by Mr S Caim and Dr A Mather (Quadram Institute Bioscience; QIB). Majority of the Whole Genome Sequencing (WGS) service was coordinated by Dr D Pickard at Wellcome Trust Sanger Institute (WTSI) unless otherwise stated. 16S rRNA sequencing library preparation on caecal samples was processed by Ms C Leclaire (QIB) and sequenced at WTSI.

Majority of the clinical bacterial isolates in the analyses obtained via collaborations with Prof J S Kroll, Dr A Shaw and Dr K Sim (Imperial College London; ICL), Dr C Amar (Public Health England), Dr R Dixon and Dr J Brown (University of Lincoln), Prof P O'Toole (University College Cork, Ireland), and Dr A Walker (University of Aberdeen). Chicken caecal samples were kindly provided by Dr R Dixon and Dr J Brown.

Mr H Bedwell (University of East Anglia; UEA) provided assistance in DNA extraction procedures of pure cultures under my supervision for WGS. Mrs K Cross (QIB) generated all electron microscopy images used in this thesis.

All major works including bioinformatics analyses, DNA extractions of bacterial cultures, bacterial isolation work, microbiological assays, *in vivo* tissue collections, tissue processing and imaging were performed by myself.

Raymond Kiu Kwong Ong

November 2018

Norwich, UK

Acknowledgements

Firstly, I wholeheartedly thank my primary supervisor Dr Lindsay J Hall (QIB) for the amazing PhD opportunity she offered, her extraordinary mentorship, brilliant research ideas and continuous guidance to complete my PhD. She has been truly supportive and enthusiastic in her supervisory role, teaching myself how to write good research papers, set up collaborations and present my own work to the public. I would also like to thank my secondary supervisor Prof Simon Carding (QIB) for his constructive criticisms on my work which was essential for my PhD.

Thanks to all my present and former colleagues in the Hall lab for making the lab such a conducive environment for learning and they have indeed enriched my postgraduate experience at QIB. A number of individuals I would like to specially mention here: Mr Shabhonam Caim (QIB) who taught myself bioinformatics skills and the morning coffees; Dr Ian O'Neill (QIB) and Dr Lukas Harnisch (QIB) for the help in setting up cell culture experiments; Dr Zoe Schofield (QIB) for her mentorship and assistance in *in vivo* work, and Ms Jennifer Ketskemety and Dr Melissa Lawon (QIB) who trained myself microbiology skills in the early stage of my PhD.

I am indebted to many collaborators who have directly or indirectly contributed to the studies: Prof J Simon Kroll, Dr Alex Shaw and Dr Kathleen Sim (ICL); Dr Kathie Grant, Dr Corinne Amar and Ms Anais Painset (PHE); Dr Ron Dixon and Dr Joseph Brown (University of Lincoln); Prof Gordon Dougan and Dr Derek Pickard (WTSI); Dr Paul Clarke (Norfolk and Norwich University Hospital); Dr Alan Walker (University of Aberdeen); Prof Paul O'Toole (University College Cork, Ireland); Dr Gusztav Belteki (Rosie Hospital); Ms Sara Goatcher (Banham Zoo and Africa Alive); Dr Alison Mather (QIB) and Dr Duncan Gaskin.

I would also like to express my gratitude to individuals who have offered their time, knowledge and resources, either via training, assistance or informal chat, for my PhD: Dr Fred Warren, Dr Benjamin Kirkup, Dr Stephen Robinson, Dr Fran Mulholland, Mrs Kathryn Cross, Dr Aimee Parker, Dr Stephanie Schuller, Dr Gemma Langridge, Dr Emma Manners, Prof John Wain, Dr Justin O'Grady, Dr Rob Kingsley, Dr Nathalie Juge, Dr Emmanuelle Crost, Dr Dimitrios Latousakis, Dr Louise Tailford, Dr Sandra Stringer, Dr Andrew Page, and Dr Melinda Mayer (QIB); Mr Gary Wortley, Dr

Govind Chandra and Dr Jacob Malone (JIC); Mr Harley Bedwell, Dr Benjamin Evans, Mr Rich Croft, Mr Simon Deakin, Mr Seth Thomas and Mr Andrew Hind (UEA); Ms Emma Cornwell (ICL); Dr Arnoud van Vliet (University of Surrey); Prof Julian Rood (Monash University, Australia) and Ms Stephanie Roy.

I also acknowledge the international bursary and research funding awarded by University of East Anglia (UEA) which is essential for the completion of my PhD. I am also grateful for the research funds awarded to my primary supervisor Dr Lindsay J Hall, her Wellcome Trust New Investigator Award, that supported my PhD research in both wet lab and dry lab work.

I would like to thank my parents who supported me financially, to the Ingram and Kim families for the dinners, to my second family the members of the church in Norwich whom I spent most time with outside work, and to my wonderful wife Jacinth who unreservedly supported and stood by me throughout my PhD, thank you for your boundless love and continuous encouragement.

Finally, I thank my Lord Jesus Christ - Who is the wisdom of God (1 Corinthians 1:24; Holy Bible Recovery Version), He is the wonderful One Who has always been the source of my joy, and the purpose and meaning of my life. To Him ‘Who is over all and through all and in all’ (Ephesians 4:6b) and may ‘…the earth… be filled with the knowledge of Jehovah, as water covers the sea’ (Isaiah 11:9)!

Table of contents

List of abbreviations

AAD	Antibiotic Associated Diarrhoea
AMR	Antimicrobial Resistance
ANI	Average Nucleotide Identity
BAMBI	Baby Associated MicroBiota of the Intestine study
BHI	Brain Heart Infusion
BLAST	Basic Local Alignment Search Tool
BLASTN	BLAST search using Nucleotide query
BLASTP	BLAST search using Protein query
CARD	Comprehensive Antibiotic Resistance Database
CFU	Colony Forming Units
CH	Care Home
COG	Clusters of Orthologous Group
CPB	Beta-toxin
CPB2	Beta2-toxin
CPE	Food-Poisoning Enterotoxin
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DMEM	Dulbecco's Modified Eagle Medium
DMU	Disease Modelling Unit
DNA	Deoxyribonucleic acid
EDTA	Ethylene Diamine Tetra-acetic Acid
EN	Enteritis Necroticans
ENA	European Nucleotide Archive
ETX	Epsilon-toxin
EYA	Egg Yolk Agar
fAFLP	Fluorescent Amplified Fragment Length Polymorphism
FP	Food Poisoning
GFP	Green Fluorescent Protein

GWAS	Genome-Wide Association Study
HGT	Horizontal Gene Transfer
IAP	Iota-toxin component a
IBP	Iota-toxin component b
IS	Insertion Sequence
ITX	Iota-toxin
JIC	John Innes Centre
LDA	Linear Discriminant Analysis
LOD	Limit of Detection
MALDI	Matrix-Associated Laser Desorption Ionisation
MIC	Minimum Inhibitory Concentration
MLST	Multilocus Sequence Typing
NCBI	National Center for Biotechnology Information
ND	Not Detectable
NE	Necrotic Enteritis
NEC	Necrotising Enterocolitis
NEOM	The Neonatal Microbiota Study
NGS	Next Generation Sequencing
NICU	Neonatal Intensive Care Unit
NNUH	Norfolk and Norwich Hospital
OD	Optical Density
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
PFO	Perfringolysin O
PHE	Public Health England
PLC	Alpha-toxin/ Phospholipase C
QCCH	Queen Charlotte's and Chelsea Hospital
QIB	Quadram Institute Bioscience
RH	Rosie Hospital

rRNA	Ribosomal Ribonucleic Acid
SD	Standard Deviation
SEM	Standard Error of the Mean
SMH	St Mary's Hospital
Tn	Transposon
TOF	Time-of-flight
TSC	Tryptose Sulfite Cycloserine
UEA	University of East Anglia
WGS	Whole Genome Sequencing
WTSI	Wellcome Trust Sanger Institute

List of figures

List of tables

List of peer-reviewed publications

Peer-reviewed journal articles published during my PhD (2015-2018), these articles were reproduced in Appendix 7.

- Kiu, R., Caim, S., Alcon-Giner, C., Belteki, G., Clarke, P., Pickard, D., Dougan, G., Hall, L J. Preterm infant-associated *Clostridium tertium*, *Clostridium cadaveris*, and *Clostridium paraputrificum* strains: genomic and evolutionary insights. *Genome Biology & Evolution* 9, 2707-2714, doi:10.1093/gbe/evx210 (2017).
- 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. *Frontiers in Microbiology* 8, 2485, doi:10.3389/fmicb.2017.02485 (2017).
- Kiu, R. & Hall, L. J. An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerging Microbes & Infections* 7, 141, doi:10.1038/s41426-018-0144-8 (2018).
- Kiu, R. & Hall, L. J. Response: Commentary: Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant pangenome diversity, and a diverse array of virulence factors. *Frontiers in Microbiology* 9, doi:10.3389/fmicb.2018.01857 (2018).

Chapter 1 General introduction

This chapter is a general introduction that forms the background of this thesis. The contents of this chapter are primarily based on (verbatim) the peer-reviewed review article published in *Emerging Microbes & Infections* that I am first author [1]. The published review article is reproduced in Appendix 7.

1.1 Introduction

Clostridium perfringens (formerly known as *Bacillus aerogenes capsulatus, Bacillus perfringens, Bacillus welchii* or *Clostridium welchii*) is a Gram positive, spore-forming, anaerobic, rod-shaped bacterium (Figure 1.1) [2]. It was first isolated and identified as a novel bacterium in 1891 by William H. Welch from the autopsy of a 38-year-old man, where gas bubbles were observed within infected blood vessels. This gas-forming trait (the original name *Bacillus aerogenes capsulatus*, 'aerogenes' literally means 'air-producing' in Latin) was later linked to gas gangrene symptoms seen in British and French soldiers during World War I [3]. This organism was later named as *Bacillus perfringens* (in Latin *perfringens* means 'breaking through' which mirrors its tissue-invading nature) by Veillon and Zuber [4], and designated as *Bacillus welchii* in 1900 by Walter Migula in Germany [5]. In the 1940s it was given its current name *Clostridium perfringens*, in the *Bergey's Manual of Determinative Bacteriology* [6].

C. perfringens is associated with diverse environments including soils, food, sewage, and as a member of the gastrointestinal (GI) tract microbial community (i.e. microbiota) of both diseased, and non-diseased humans and animals. Notably it has been associated with humans for thousands of years as evidenced by the recent identification of *C. perfringens* using Next Generation Sequencing (NGS) technology in the mummified GI tract of a >5000-year-old Neolithic 'Tyrolean Iceman' (also known as Ötzi), found in an Alpine glacier in 1991 [7].

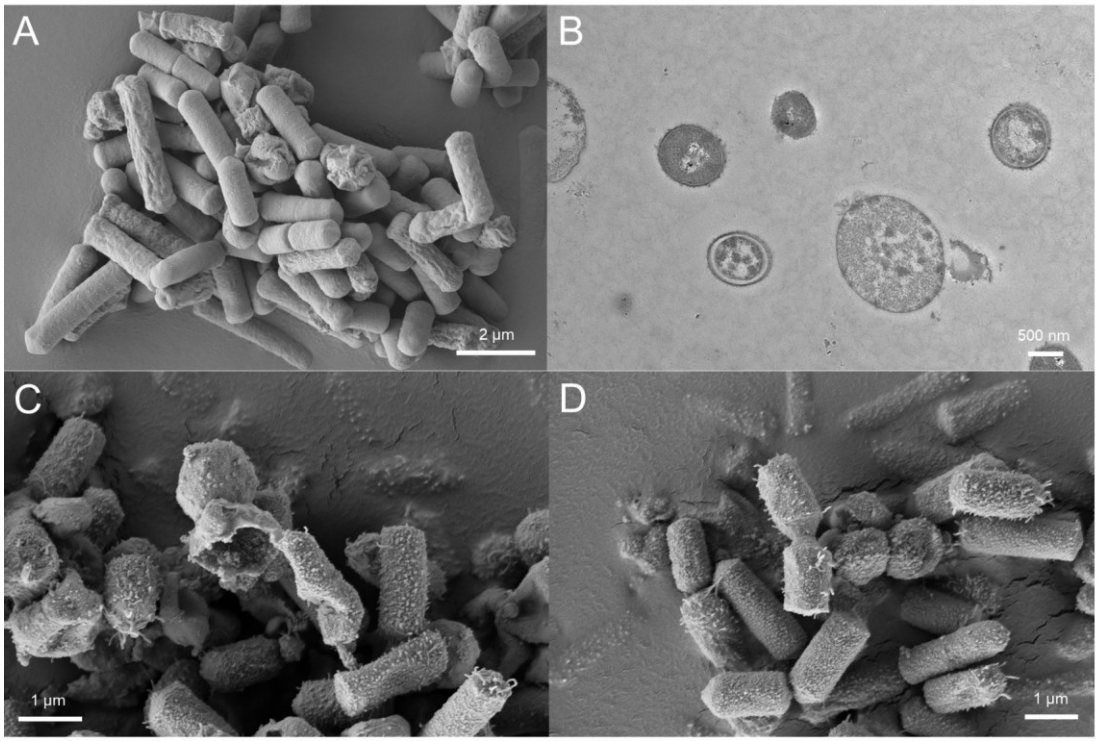


Figure 1.1: Electron microscopy images of *C. perfringens* vegetative cells and spores

(A) Scanning electron microscopy images of *C. perfringens* LH019. **(B)** Transmission electron microscopy of *C. perfringens* Q143 in spore forms. **(C)** Scanning electron microscopy image of the germination of *C. perfringens* Q143. **(D)** Binary fission of *C. perfringens* Q143 under scanning electron microscope. Images processed by Mrs Kathryn Cross.

Clinically, *C. perfringens* has been constantly associated with various significant systemic and enteric diseases, in both humans and animals, including gas gangrene (Clostridial myonecrosis), food poisoning (FP) and non-foodborne diarrhoea, and enterocolitis [8,9]. Importantly, *C. perfringens* strains are known to secrete >20 identified toxins or hydrolytic enzymes that could potentially be the principal virulence factors involved in pathophysiology [10].

This introductory chapter will explore phenotypic and genomic features of this important and re-emerging pathogen, including virulence factors, and antimicrobial resistance (AMR) profiles, and the clinical impact of this bacterium in relation to several medically important intestinal diseases, across several host species. Finally, some of the important unresolved questions in relation to *C. perfringens*-associated infections, and implications for future research directions are highlighted towards the end of this chapter.

1.2 Isolation, identification and typing methods

There are numerous methods for isolation of *C. perfringens*, and biochemical and molecular methods for identification as detailed in Table 1.1.

	Method	Method details in brief	Refs
Isolation	Direct plating	Direct plating on TSC-EYA + 18-24 h anaerobic incubation at 37ºC. Pitch black colonies with opaque halos are presumptively *C. perfringens.*	11
	Ethanol pre-treatment	Ethanol pre-treatment (50% ethanol) for 30 mins + plating on Fastidious Anaerobe Agar (sometimes supplemented with 0.1% taurocholate). Colonies that exhibit beta-haemolysis are preliminarily identified as *C. perfringens*.	12
Biochemical identification	Nagler's reaction	Known as lecithinase (alpha-toxin) test. Egg-yolk agar plates are split into two halves, where one half contains anti-alpha-toxin, and following anaerobic incubation, positivity is defined as turbidity around colonies on the anti-alpha-toxin-free side which confirms *C. perfringens*.	13
	Reverse CAMP test	*Streptococcus agalactiae* is streaked on blood agar, and *C. perfringens* streaked perpendicular to *S. agalactiae*. After 24-48 h of anaerobic incubation, a 'bow-tie' zone will form due to the synergistic haemolysis, this confirms *C. perfringens*.	14
Molecular identification	16S ribosomal RNA PCR	Appropriating full-length 16S rRNA universal primers/smaller region of 16S rRNA gene in PCR to amplify 16S rRNA gene + sequencing and identify using informatics approach (based on sequence similarity >97% to assign taxonomy).	15
	MALDI-TOF Mass Spectrometry	Rapid and inexpensive identification method based upon the mass spectrum analysis of highly-conserved ribosomal proteins from whole bacterial cells – apply bacterial colonies straight onto the MALDI-TOF metal target and followed by appropriate treatments and analysis.	8
Typing	Multiplex PCR	Multiplex PCR approach is commonly used to amplify key toxins genes to classify *C. perfringens* into 7 (A to G) different toxinotypes according to the toxin genes combination.	16
	Genome-wide sequence search	Genome-wide search on relevant toxin genes using sequence similarity search program e.g. BLAST for toxinotyping on *C. perfringens* genomes.	17
	Fluorescent amplified fragment length polymorphism (fAFLP)	This technique is based on selective PCR amplification of restriction fragments from genomic DNA digestion. It is a rapid and highly reproducible PCR method to discriminate sub-types in epidemiological study.	18

Table 1.1: Common isolation, identification and typing methods in *C. perfringens* research

Table adapted from Kiu and Hall (2018) under the Creative Commons BY licence [1]. PCR: Polymerase Chain Reaction.

Historically, bacterial typing systems were developed to distinguish distinct bacterial strains from a single species, and more importantly, to recognise epidemic clones with specific pathogenicity. *C. perfringens* strains, clinically well-known for toxin production, are typed (recently updated for 7 toxinotypes: A-G; previously only types A-E [19]) according to the combination of typing toxins (Figure 1.2) they produce, i.e. α-toxin (PLC), β-toxin (CPB), ε-toxin (ETX) and ι-toxin (ITX), enterotoxin (CPE) and NetB [20].

This toxinotyping method indicated that certain toxins were associated with certain hosts and diseases e.g. type B (particularly CPB it harbours), is exclusively linked to dysentery in sheep, and rarely seen in other hosts [21]. FP associated CPE was genotyped in type F strains (previously named as CPE-positive type A; not to be confused with heat-resistant type C strains) although CPE can also be produced by certain type C, D and E strains, whereas β2-toxin (CPB2) and θ-toxin (perfringolysin O; PFO) could be found in any toxinotypes [2,17,22]. However, it is important to note that no single strain is known to produce the entire panoply of toxins [17].

Since the advent of the Polymerase Chain Reaction (PCR) technique, multiplex PCR toxinotyping has now replaced the classical antisera methodology and is the routine rapid typing technique in clinical laboratories [16,20,23,24]. However, this recently-updated typing system is still largely limited to the identification of the 6 ‘typing’ toxins (or previously 4; although some do pick up accessory toxins) [22,25]. Moreover, linking toxinotype to disease phenotype is also highly debated as detection of typing PLC has little diagnostic value, as it can also be detected in healthy individuals [26]. Importantly, type A strains (most strains are type A, as this toxinotype is universal in the environment) were also associated with enteric diseases, these strains were detected to harbour other additional ‘minor’ toxins (linkage with diseases were suggested) that are not addressed in this current toxinotyping method [17,27].

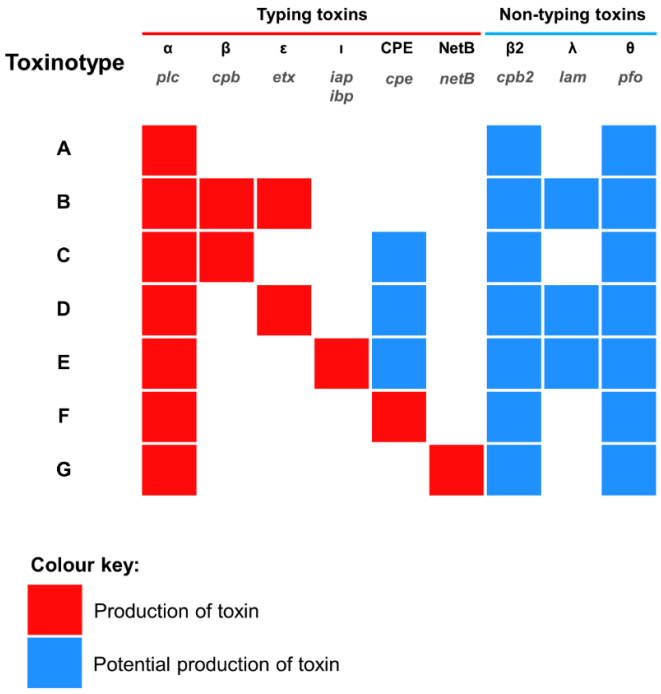


Figure 1.2: Current toxinotyping scheme of *C. perfringens*

Figure reproduced from Kiu and Hall (2018) under the Creative Commons BY licence [1].

Most recently, Whole Genome Sequencing (WGS) data of *C. perfringens* has revealed substantial genetic divergence in the sequenced genomes, when comparing similar toxinotypes [2,17,28]. Potentially, with the aid of bioinformatics tools, more novel toxins or virulence genes will be discovered and identified for investigation of *C. perfringens* pathogenicity, and disease transmission tracking in hospital and environmental settings. Genome-based research in a related *Clostridium* gut pathogen, *Clostridium difficile* (it has since been suggested to be taxonomically moved to *Clostridioides*), has highlighted the utility of WGS as an appropriate tool to understand bacterial pathogenicity, including identifying AMR genes, and monitoring *C. difficile* spread [29-32]. Importantly, WGS has also identified 'hypervirulent' clades of *C. difficile* that appear to be linked with more severe disease outcomes, which includes at least two different ribotypes [33,34], and was used in monitoring bacterial spread, and tracking sources of infection, at an international level [35,36]. These studies indicate that WGS will be key to understand, classify and characterise pathogens, including *C. perfringens,* in this genomic era.

1.3 Genomic characteristics

C. perfringens has a relatively low GC content, between 27-28%, and genome sizes range from 3.0-4.1 Mb, with 2,500 to 3,600 predicted genes present in each individual genome [17]. The first ever complete genome sequence of *C. perfringens* (strain 13) was published in early 2002; a historical gas gangrene-associated strain from 1939 (this also represents the first Gram-positive anaerobic pathogen to be sequenced) [37]. Virulence-linked genes within the sequence data include phospholipase C (PLC; α-toxin), hyaluronidase (μ-toxin), collagenase (κ-toxin), PFO (θ-toxin), and CPB2. Moreover, 61 sporulation- and germination-related genes were encoded, which supports the fact that *C. perfringens* is a spore-former. Notably, most virulence genes in this type A strain were not found on genomic islands, or near insertion/transposase sequences, and GC content of those genes were similar to the average GC percentage, implying that horizontal gene transfer (HGT) of these toxin genes was an ancient evolutionary event. This study also highlighted differences in environmental adaption genes, including glycolytic enzymes, and sporulation-linked genes, whereas recent studies indicated that phylogenetic relationships between *C. perfringens* (based on core genome analysis) does not relate to clone origin or toxinotypes [2,17].

Two more *C. perfringens* genomes were sequenced and published in 2006 [28]; strain ATCC13124 (a gas gangrene strain isolated 1941 by Mollby, et al. [38]), and strain SM101 (a FP strain isolated in 1953 [39]). Key findings indicated that ATCC13124 contains a putative restriction-modification system that is obstinate to genetic transformation, a range of glycolytic-related genes that are strain-specific (highlighting differences in environmental adaptation capabilities), and additionally that sporulation capacity is strain-specific, based on the comparative studies of sporulation-linked genes. Importantly, this study revealed an unexpectedly high degree of genomic diversity, with no two strains identical with respect to genomic islands.

Following the reduced cost of WGS, another 9 *C. perfringens* isolates were sequenced, and analysed alongside the pre-existing 3 strains in 2015 [2]. Core-genome regions were first reported to be highly syntenous (similar in nucleotide sequence arrangement), GC content (approximately 28.3%, similar to what was previously described), and importantly the core-genome phylogeny constructed did not show any potential relatedness/ clonal origins amongst different toxinotypes.

A recent large-scale genomic study (that I co-first authored) that investigated 56 *C. perfringens* strains, representing the largest *C. perfringens* genomic study to date, revealed a diverse pangenome (a repertoire of genes in a defined number of genomes), with only 12.6% core genes (common genes that are present in each genome; Figure 1.3) [17]. A wide array of toxin genes was profiled in this study: *plc*, α-clostripain (*ccp*) and microbial collagenase (*colA*) were conserved in all examined genomes, the pore-forming toxin PFO gene *pfo* was found to be encoded in most genomes (>90%), whereas the FP causative toxin CPE gene *cpe* was consistently detected in historical FP isolates (isolated before 1950s), and was also linked with enteritis in dogs and foals [40]. Importantly, prevalence of aminoglycoside-resistant/anti-defensin gene *mprF* (100% detection), and tetracycline-resistant efflux protein genes *tetA(P)* (>75% detection) encoded within the genomes underlined the potential antimicrobial resistant threat of *C. perfringens*-associated infection. This study also suggested that these diverse genetic variations may have been driven by HGT, especially prophage insertion, within the 'CRISPR-free' genomes (Clustered Regularly Interspaced Short Palindromic Repeats [CRISPR]; no single CRISPR prophage defence system detected in >70% of the genomes).

The substantial genetic divergence in the *C. perfringens* pangenome suggests there may be additional novel virulence-related genes encoded within the 'accessory genome', in addition to the plasmid-borne toxins, known to be primarily responsible for specific disease pathologies. Thus, it is important to analyse plasmid-encoded genes/toxins, in tandem to chromosomally-encoded genes in the accessory genome, to provide a more complete genomic picture of *C. perfringens*. Potentially, with the aid of bioinformatics tools, more novel toxins or virulence genes will be discovered and identified for investigation in relation to *C. perfringens* pathogenicity, and disease transmission tracking in hospital and environmental settings, as has been performed for the related gut pathogen *C. difficile* [30]. To fully determine and research the pangenome, and discover potentially novel virulence genes, more isolate sequencing will be required, including addition of isolates from a range of hosts and environmental conditions, which may pave the way for translational research into new treatment strategies against *C. perfringens* infections.

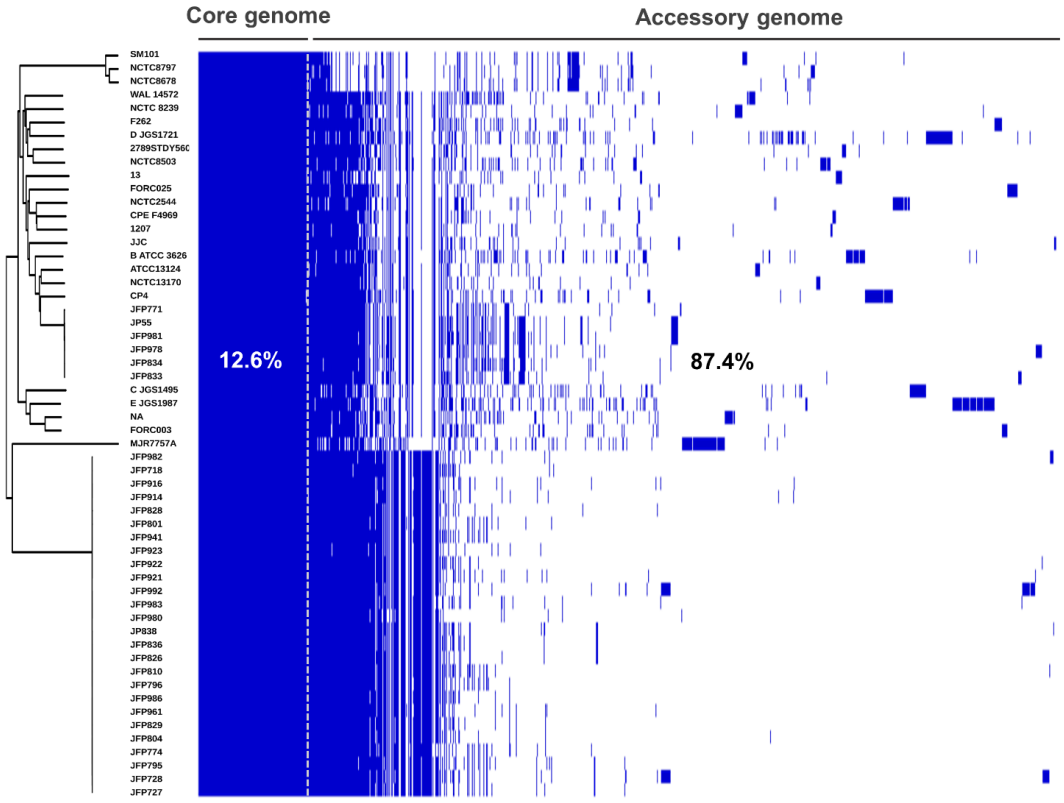


Figure 1.3: Linearised pangenome of 56 *C. perfringens* strains

Each coloured block represents synteny (or, homolog: identical gene) in the pangenome. Figure reproduced from Kiu et al. under the Creative Commons BY licence [17].

1.4 Plasmids

Importantly, *C. perfringens* strains are known to carry plasmids, which often encode virulence-associated proteins including toxins and antimicrobials [41]. Disease-associated toxins including; CPE, ETX, ITX, NetB, CPB2 and Binary Enterotoxin (BEC) have all been detected on *C. perfringens* plasmids [17,42]. HGT of plasmids between strains is reported to be via the *tcp* conjugation locus that exists in most known *C. perfringens* plasmids [43]. Important toxin-carrying plasmids will be briefly reviewed in order of toxinotypes. For comprehensive reviews on *C. perfringens* plasmids, and other related insertion sequences see Freedman et al. (2015) and Li et al. (2013) [41,42].

CPE is known to be involved in FP and non-foodborne gastroenteritis. It was reported that up to 70% of FP cases, causative *C. perfringens* strains (mainly type F, previously known as CPE-positive type A) were shown to encode chromosomal-*cpe,* rather than

plasmid-*cpe*, while the latter was associated with non-foodborne gastroenteritis (responsible for 5-15% of all cases) [44]. Some *cpe*-associated plasmids have also been shown to encode *cpb2* genes or *iab/ibp* (ITX) genes (e.g. pCPF5603) [41]. Type B strains carry plasmids which encode one of the deadliest toxins on earth, ETX, and CPB. Plasmid pCP8533etx encodes *etx* and *cpb2* genes, but not *cpb* [45]. Most plasmids in type B strains also possess other virulence genes encoding λ-toxin and urease [46]. Similarly, plasmids in type D strains also carry *cpe* and/or *cpb2* genes, and plasmids range in size from 45 kb to 110 kb [47].

C. perfringens type C strains possess plasmid-borne CPB gene *cpb*, and other plasmids reported to carry other toxin genes including *cpb2*, *tpeL*, or *cpe* [48]. In type E strains, two major families of plasmids have been identified: (1) plasmids carry *iap/ibp* (ITX) genes, λ-toxin gene and *cpe* gene, however, *cpe* contains nonsense and missense mutations and is thus non-functional [49]; (2) These plasmids, including pCPPB-1, encode both *iap/ibp* (ITX) genes and *cpe* gene but not the λ-toxin gene [50]. NetB is a chicken NE-associated toxin, which is plasmid encoded. These *netB*-encoded plasmids can potentially carry other virulence genes including tetracycline resistance genes (e.g. pJIR3537 and pCW3) and *cpb2* gene (e.g. pJIR3844) [51].

Notably, many toxin plasmids of *C. perfringens* are highly similar, sharing up to 35-40 kb of identical sequences [42]. In addition, some *C. perfringens* isolates can carry multiple near-identical plasmids that encode different toxin and AMR genes [51]. However, no one isolate has ever been found to possess plasmids that encode all key toxins; only certain plasmid combinations could can be maintained within single strain, which suggests plasmid incompatibilities due to presence of distinct plasmid segregation systems, and warrants further research [41,52,53]. Furthermore, the universal existence of *tcp* locus indicates that conjugative plasmid transfer could be a key HGT event for increasing virulence of *C. perfringens* strains.

1.5 Virulence-related factors

C. perfringens can generate a complement of extracellular toxins and hydrolytic enzymes (>20), survive in aerobic environments (i.e. oxygen tolerance), produce toxic gases, and rapidly grow. Therefore, it possesses the capacity to effect various histotoxic (i.e. toxigenic to tissues) infections in humans, including gas gangrene in

contaminated wounds, gastroenteritis (including foodborne and non-foodborne diarrhoea) in human adults, necrotic enteritis (NE) in animals, and recent links to necrotising enterocolitis (NEC) in preterm infants [54]. Relevant virulence factors of *C. perfringens* are represented and summarised in Figure 1.4.

1.5.1 Transmission and colonisation

1.5.1.1 Oxygen sensitivity

C. perfringens is commonly known as a strict anaerobe; however, this bacterium can also survive in the presence of oxygen, and/or low concentrations of superoxide or hydroxyl-radical-generating compounds [55]. It was shown to survive in aerobic conditions for >72h, which was attributed to its generating capacity of Superoxide Dismutase (SOD) [56]. Genomic analysis indicates there are putative NADPH dehydrogenase genes (*yadD* and *ycdF*), which generate enzymes that help maintain intercellular redox balance [55]. Notably, as an aero-tolerant anaerobe, *C. perfringens* could potentially survive through aerobic environments (such as on surfaces in hospital wards [57]) and trigger disease development in aerophillic environments (i.e. adult/preterm infant gut), and additionally oxygen-exposed tissues (such as gas gangrene), which may facilitate bacterial host-to-host transmission [15].

1.5.1.2 Sporulation

C. perfringens is a spore-former, which enables this bacterium to survive in extreme or nutrient-depleted conditions. This spore-forming characteristic plays a vital role in the transmission of this Gram-positive bacterium from diverse environments to hosts, leading to infection, including FP in human adults [58]. Some spores of *C. perfringens* (especially FP associated strains) had been demonstrated to resist extreme temperature conditions, which may contribute to *C. perfringens* survival and subsequent disease pathology [59,60]. Although, capable of sporulation, *C. perfringens* appears to be a less-prolific (*in vitro)* spore-former than its *Clostridium* relatives such as *C. difficile* [15,61]. For lab based assays, sporulation can be achieved using the well-experimented raffinose-supplemented Duncan-Strong sporulation medium [61].

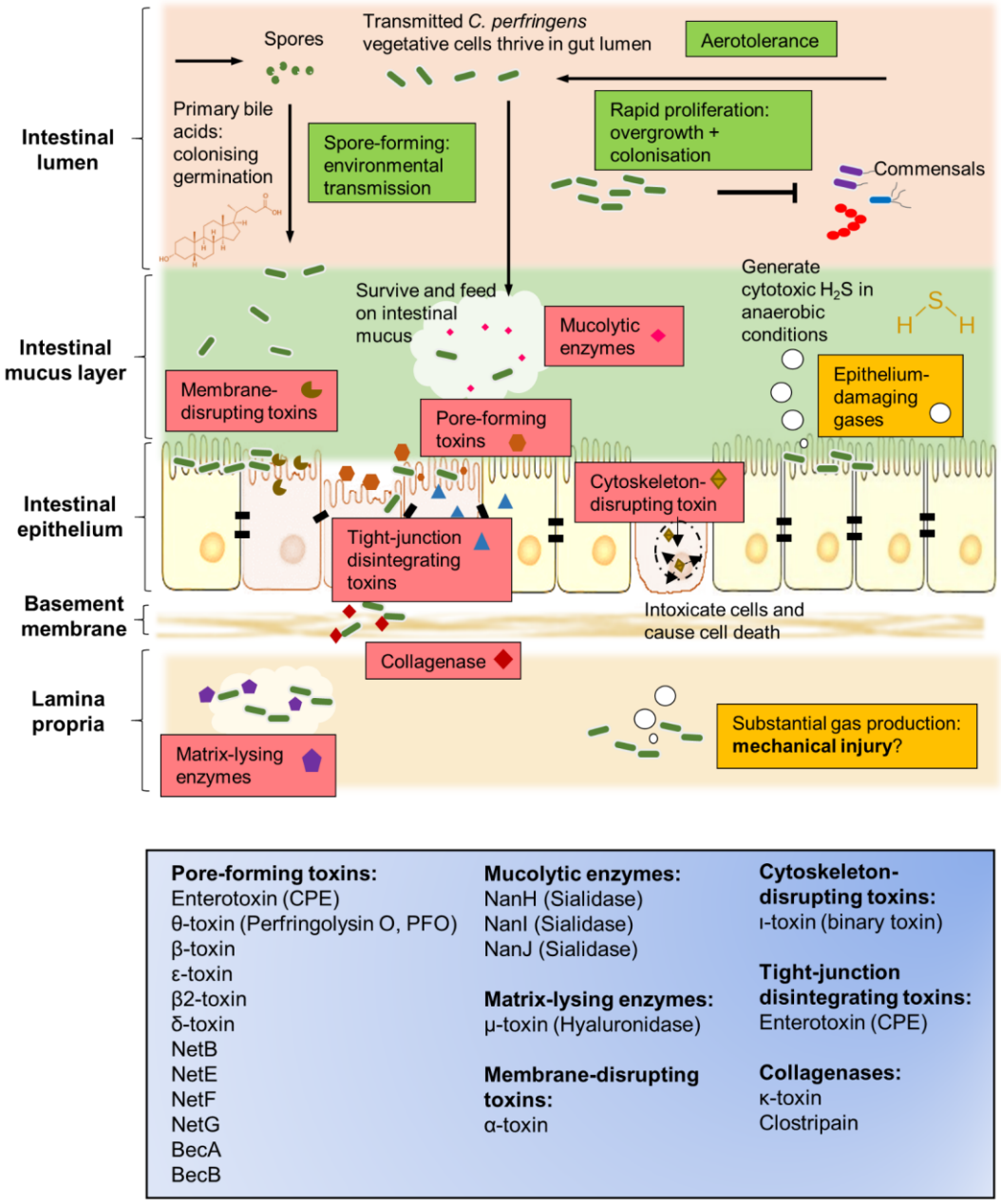


Figure 1.4: Disease-linked virulence factors of *C. perfringens* in intestinal infections

Figure reproduced from Kiu and Hall (2018) under the Creative Commons BY licence [1].

Molecular regulation of sporulation in *C. perfringens* is less well studied compared to the model bacterium *Bacillus subtilis*. Previous studies have indicated that the transcriptional regulator CcpA (encoded by *ccpA*) is required for effectual sporulation [62] and Spo0A (*spo0A*), a transcriptional factor known to be present in *Clostridium* species including *C. perfringens*, is associated with sporulation initiation [28]. Sigma factors, including SigF, SigE, SigG and SigK (genes *sigF*, *sigE*, *sigG* and *sigK*

respectively) are known to regulate the sporulation process of *C. perfringens* [63,64]. In addition to this, it was also reported that *C. perfringens* sporulation processes could be regulated by an Agr-like quorum-sensing system, potentially promoted by the *codY* gene in type A FP strains [65,66]. Importantly, CPE is known to be produced by *C. perfringens* during sporulation, which correlates with the pathogenesis of FP-associated diarrhoea [58].

1.5.1.3 Germination

Germination of *C. perfringens* can be triggered by small molecules termed germinants – commonly known as sugars, nucleosides, amino acids, salts and purines [67]. Importantly, primary bile acids (including glycocholate, cholate and taurocholate) in the human GI tract are also known to act as potent germinants in *Clostridium* species, whereas secondary bile acids, such as deoxycholate (derivatives generated by certain gut microbes), can inhibit the *in vitro* proliferation of *C. perfringens* [68]. This proposed 'colonising germination' survival modus operandi of spore-formers, including *C. perfringens,* potentially acts as a signal to initiate growth, and to persist in the intestinal environment [15]. This persistence, and long-term gut colonisation may correlate with the ongoing symptoms (up to several weeks) of nonfoodborne diarrhoea in patients [69]. From a colonisation resistance perspective, inhibition of germination, by secondary bile products produced by resident commensal bacteria, implies a long-standing competition for limited nutrients and niches within the gut.

1.5.1.4 Rapid proliferation

Notably, *C. perfringens* represents one of the fastest-growing organisms currently known and is reported to have a very short 8-12 min generation time when cultured at 43°C in optimal media, and 12-17 min at 37°C [70]. Rapid growth of this bacterium potentially predisposes the host to tissue infection, as in the cases of avian NE, bovine necro-haemorrhagic enteritis (could be <5h), and gas gangrene. Also, as *C. perfringens* has a two-fold quicker generation time, when compared to intestinal commensals like wild-type *Escherichia coli* (typically 20-30 min in Luria-Bertani broth), this could represent a potent mechanism for outgrowing other resident bacteria, leading to efficient gut colonisation.

1.5.2 Disease initiations

1.5.2.1 Histotoxic gas production

Clostridial myonecrosis (better known as gas gangrene), is accompanied by profuse gas production, which could in theory be the mechanical determinant of tissue injury, alongside rapid cell proliferation of *C. perfringens*. Importantly, *C. perfringens* is known to generate hydrogen sulphide [71] (using ubiquitous sulfuric sources from the environment), a readily-permeable toxic substance to human cells, which when produced in excess is associated with other intestinal inflammatory diseases (e.g. Ulcerative Colitis) [72]. Thus, *C. perfringens*-associated gas production represents an understudied virulence trait in the pathogenesis of gut infection, such as preterm-infant NEC (discussed in more detail below), which symptoms significantly mimic those of gas gangrene; necrosis of tissue accompanied by abundant gas production [73,74].

1.5.2.2 Toxins and virulent enzymes

Presently, 23 virulence genes that encode toxins and virulent enzymes have been identified in *C. perfringens* (summarised in Table 1.2) as the most prolific toxin-producing pathogen currently known. Some key toxins will be briefly discussed and summarised in this sub-section.

1.5.2.2.1 Alpha-toxin

PLC (encoded by gene *plc* or *cpa*) was identified in 1941 as enzyme phospholipase (i.e. phospholipase C), which was the first bacterial toxin shown to possess enzymatic activity [75]. This toxin is present chromosomally in all strains of *C. perfringens* and acts on egg-yolk lipoprotein (lecithin), which results in increased turbidity- basis of a medical diagnostic test for *C. perfringens*, such as Nagler reaction or egg-yolk agar [11]. Biochemically, PLC has been shown to hydrolyse cell membrane phospholipids, which eventually leads to cell necrosis; a key characteristic in gas gangrene [38,76]. Indeed, this toxin is essential for gas gangrene pathology, as shown in an *in vivo* myonecrosis model [77]. Mechanistically, PLC may play three major roles in gas gangrene pathology; firstly, it is able to impact on the movement of immune cells such as neutrophils to infected tissues (mechanism currently unknown), hence potentially reducing pathogen clearance at infected sites. Secondly, it can cause constriction of blood vessels, which may reduce the blood supply to tissues, thus creating a micro-

aerophillic environment conducive to *C. perfringens* overgrowth. Thirdly, this toxin can activate inflammation cascades in host cell metabolism (arachidonic acid and protein kinase C), which may lead to direct immune-mediated pathology of tissues [78].

1.5.2.2.2 Beta-toxin

CPB, encoded by gene *cpb*, is a plasmid-encoded pore-forming toxin that is thought to be intestinal-necrotic, and also important for systemic enterotoxaemia, in humans and neonatal animals, as shown in *in vivo* studies [79,80]. *C. perfringens* type C strains that possess the *cpb* gene were associated with historical Clostridial gut infections; Darmbrand (post-world war II in Germany), and Pig Bel (Papua New Guinea) [81]. Previous intestinal loop studies have indicated a synergistic effect with CPE (encoded by *cpe*) [82]. Furthermore, this toxin shares sequence similarity with several pore-forming toxins produced by *Staphylococcus aureus* – ɣ toxin (A: 22% and B:28%), α-toxin (28%) and leucocidin (S:17% and F:28%) [83].

1.5.2.2.3 Perfringolysin O

PFO (also θ-toxin, encoded by gene *pfoA* or *pfo*), is also a pore-forming toxin that acts on cholesterol-comprising cell membranes. This toxin has been shown to be involved in the pathogenesis of gas gangrene, and haemorrhagic enteritis in calves [84,85]. Notably, PFO shares structural homology with similar pore-forming toxins identified in *Streptococcus*, *Bacillus*, *Listeria* and many other genera [25]. PFO is also known for its capacity to induce tumour necrosis factor alpha (TNF-α) and interleukin 6 (IL-6) expression in the host, and it could activate apoptosis through p38 MAPK (mitogen-activated protein kinase) pathway as demonstrated in *in vitro* models [86]. Interestingly, it has synergistic cytotoxic effects with PLC on bovine epithelial cells, highlighting the significant role of this toxin in disease development [85,87]. Please refer to the work by Verherstraeten et al. (2015) for a comprehensive review of PFO [25].

1.5.2.2.4 Beta2-toxin

CPB2 (encoded by *cpb2* plasmid gene), a pore-forming cytolytic toxin shares <15% sequence homology with CPB, distinguished itself as a novel toxin produced by *C. perfringens* [83]. CPB2-producing *C. perfringens* strains have been associated with gut diseases such as necrotising enteritis in piglets, and enterocolitis in foals [88]. Importantly, this pore-forming toxin has been suggested as an accessory toxin in *C.*

perfringens-associated non-foodborne diarrhoea [89], and proposed to play an important role in preterm-NEC in potential synergistic effects with antibiotic gentamicin [90].

		Gene	Toxin name	Abbreviation in this thesis	Alternative name	Mechanism of pathogenicity	Refs
1	Typing toxins [19]	*plc/cpa*	Phospholipase	PLC	α-toxin	Disruption of cell membrane	76
2		*cpb*	β-toxin	CPB	-	Pore-formation	91
3		*etx*	ε-toxin	ETX	-	Pore-formation	92
4		*iap*	ι-toxin component Ia	IAP (ITX)	-	Cytoskeleton disruption	50
5		*ibp*	ι-toxin component Ib	IBP (ITX)	-	Cytoskeleton disruption	50
6		*cpe*	Enterotoxin (CPE)	CPE	-	Pore-formation and tight-junction disintegration	50
7		*netB*	NetB	NetB	-	Pore-formation	93
8	Non-typing toxins	*cpb2*	β2 toxin	CPB2	-	Pore-formation	88
9		*lam*	λ-toxin	LAM	-	Potent protease	94
10		*pfo/pfo A*	Perfringolysin O	PFO	θ-toxin	Pore-formation	95
11		*cpd*	δ-toxin	CPD	-	Pore-formation	96
12		*ccp*	Clostripain	CCP	-	Digestion of collagen	37
13		*colA*	Microbial collagenase	ColA	κ-toxin	Digestion of collagen	97
14		*nanI*	Sialidase	NanI	-	Mucolysis	37
15		*nanJ*	Exo-α-sialidase	NanJ	-	Mucolysis	37
16		*nanH*	Neuraminidase	NanH	-	Mucolysis	98
17		*nagH*	Hyaluronidase	NagH	μ-toxin	Digestion of connective tissue	37
18		*tpeL*	Glucosylating toxin	TpeL	-	Induction of apoptosis	99
19		*becA*	Binary Enterotoxin Component A	BECA	-	Pore-formation	100
20		*becB*	Binary Enterotoxin Component B	BECB	-	Pore-formation	100
21		*netE*	NetE	NetE	-	Pore-formation	101
22		*netF*	NetF	NetF	-	Pore-formation	102
23		*netG*	NetG	NetG	-	Pore-formation	101

Table 1.2: Pathogenicity of the currently identified *C. perfringens* toxins

Table adapted from Kiu and Hall (2018) under the Creative Commons BY licence [1].

1.5.2.2.5 Epsilon-toxin

This deadly toxin ETX, is plasmid-encoded by *etx* gene *C. perfringens* type B and D strains and is essential for pathogenesis. Epsilon toxin is involved in animal (goat and sheep) enterotoxaemia [103]. ETX is currently thought to be the most potent of all toxins produced by *C. perfringens*, evidenced by the LD_{50} of 70ng/kg body weight, ranked only behind *C. botulinum* and *C. tetani* neurotoxins. It has been demonstrated to affect various organs, such as kidneys and liver, through unknown mechanisms, that allow this toxin to enter systemic circulation [104]. Due to its potential use as a biological

weapon, ETX-producing *C. perfringens* strains are on the export control list in a number of countries including USA and the UK [105,106].

1.5.2.2.6 Iota-toxin

This cytoskeleton-damaging toxin ITX (also considered as a typing toxin produced by *C. perfringens*), consists of two cell-binding monomers Ia (enzymatic component) and Ib (binding component; encoded on plasmids by *iap* and *ibp* gene respectively), which act synergistically to first translocate the toxin into host cells, then exert enzymatic activity on ADP-ribosylating actin to disassemble the cytoskeleton, which eventually leads to apoptosis and cell death [107]. Furthermore, ITX shares a certain degree of virulence similarity with oedema and lethal binary toxins from *Bacillus anthracis* [108,109].

1.5.2.2.7 Microbial collagenase

Microbial collagenase (also known as κ-toxin), encoded by chromosomal gene *colA*, is a key toxin produced by *C. perfringens* that degrades collagen. This enzyme might contribute to intestinal infection, as collagen, the substrate for collagenase, is a primary component of intestinal connective tissues/basal membrane of human and animal hosts, thus disruption may damage basal integrity, which may lead to eventual tissue necrosis [85,110]. However, Awad et al. indicated that κ-toxin is not a major determinant in a Clostridial myonecrosis mouse model (gas gangrene), despite its capacity to hydrolyse collagen [111].

1.5.2.2.8 Enterotoxin (CPE)

CPE (encoding gene *cpe*) is recognised as the key toxin to cause FP and non-foodborne diarrhoea, it has also been demonstrated to disrupt the intercellular claudin tight junctions in gut epithelial cells [112,113]. Importantly, this pore-forming toxin was demonstrated to bind and necrotise human ileal and colonic epithelium *in vitro*, and can induce cell death i.e. apoptosis, via the caspase-3 pathway [114]. Hence, the potential pathogenesis mechanisms underlying FP could result from CPE-induced tight junction rearrangements or pore-formation.

1.5.2.2.9 Sialidase

Three sialidases, encoded by genes *nanH, nanI and nanJ* in *C. perfringens*, are also named as neuraminidases or exo-sialidases (NanI and NanJ). This group of enzymes represent important virulence factors during *C. perfringens*-mediated tissue infection;

they catalyse hydrolysis of terminal sialic acids from glycoprotein, glycolipids and polysaccharides of cell membranes that aids in bacterial attachment to host cells [115]. This mucolytic potential suggests that *C. perfringens* may utilise intestinal mucus as a nutrient source, and thereby potentiates intestinal colonisation, which has been modelled using the *in vitro* Caco-2 cell line [116]. Importantly, *in vitro* studies have also demonstrated that α-toxin associated with NanI (exo-alpha-sialidase) increased the virulence of *C. perfringens* [117]. Furthermore, NanI was shown to potentiate the virulence of ETX, CPB and CPE, via binding-enhancing (ETX) and proteolytic activation (CPB and CPE) mechanisms, potentially enhancing *C. perfringens* pathogenesis [118]. Notably, in a gas gangrene mouse model, NanI and NanJ are not essential for virulence [119]. For a comprehensive review of this toxin, please refer to work by Li et al. (2016) [120].

1.5.2.2.10 NetB

NetB, a pore-forming toxin (encoded by *netB*), shares limited amino acid sequence similarity with CPB (38% sequence identity), PLC from *S. aureus* (31% sequence identity) and δ-toxin (40% sequence identity) from *C. perfringens* [121]. Less than a decade ago, this chicken-NE associated toxin (plasmid-borne), was discovered and shown to be essential (instead of α-toxin) for lesion formation both in cell line models, and avian *in vivo* models via genetic mutant strains, and thus fulfilled molecular Koch's postulates as a disease determinant [122]. This toxin is important in avian agriculture as *netB*-positive *C. perfringens* strains recovered from broiler chickens (healthy birds) can be as high as 61%, however expression of NetB was shown to be lower in healthy birds, when compared with NE-associated chickens (92% vs 29%) [123]. In addition, *netB* was later identified as part of a plasmid-encoded pathogenicity locus named NELoc-1 capable of being transferred via conjugation, indicating the potential spread in relation to chicken NE epidemiology [51,124]. Please refer to Review by Rood et al. for a comprehensive description on NetB and poultry NE [125].

These extracellular toxins represent potent virulence factors central to intestinal disease development. However, not all toxins are secreted by all strains of *C. perfringens* (excluding PLC), hence, *in silico* identification of virulence genes using NGS techniques and bioinformatics tools are essential for rapid and comprehensive geno-toxinotyping of *C. perfringens,* compared with conventional molecular tools.

1.5.3 Antimicrobial resistance (AMR)

AMR traits in *C. perfringens* pose a serious clinical treatment concern, due to its capacity to generate an array of lethal toxins. Presently there are several phenotyping studies published (based on *in vitro* susceptibility testing of Minimal Inhibitory Concentration; MIC) on the AMR profile of *C. perfringens*, however, there is currently limited genetic/WGS-based AMR genotyping studies (Table 1.3).

Tetracycline resistance in *C. perfringens* was first described in 1968, when 11 strains of *C. perfringens* were tested, and found to possess 'some degree' of resistance to tetracycline, thus penicillin G (or, benzylpenicillin) was recommended as the drug of choice for treating Clostridial infection [126]. Plasmid-carrying tetracycline resistance (*tet* components) in *C. perfringens* was genetically shown for the first time in porcine samples (3 strains) [127]. Tetracycline resistance elements *tetA*(P), was then detected in 81 tetracycline-resistance *C. perfringens* strains (100%), with 93% of these strains carrying a secondary resistance gene, either *tet*B(P) or *tet*(M), notably, no single strain possesses all *tet* genes [128]. Multidrug-resistant *C. perfringens* strains were first reported back in 1977, multiple strains (n=7) isolated from porcine faeces were demonstrated to be resistant against tetracycline (MIC > 32μg/ml), erythromycin (MIC > 128μg/ml), clindamycin (MIC > 64μg/ml), and lincomycin (MIC > 128μg/ml) [129]. A PCR-based AMR study on 160 environmental strains (water, soil and sewage) revealed encoded macrolide resistance genes *erm*(B)(26%), *erm*(Q)(1%), and *mef*(A)(18%), in addition to the commonly found *tetA*(P)(53%), *tetB*(P)(22%), and *tet*(M)(8%) [130].

Macrolide and lincosamide resistance (mainly erythromycin and lincomycin) appears widespread [131], and therefore is considered ineffective in treating *C. perfringens* infections. This is supported by a recent multidrug resistance study of 260 strains of *C. perfringens* isolated from diarrheal neonatal piglets in Thailand, where higher resistance was observed for erythromycin, lincomycin and tylosin [132].

Antibiotic class	Inhibitory mechanism	High susceptibility	Moderate resistance	High resistance	Refs
Penicillin	Cell wall synthesis	Penicillin G Amoxicillin	Ampicillin		129,133-137
Glycopeptides	Cell wall synthesis	Vancomycin Daptomycin			134,138
Imidazoles	Nucleic acid synthesis	Metronidazole			134,135
Cephalosporins	Cell wall synthesis	Cefotaxime	Ceftazidime Ceftriaxone Ceftiofur		134,136,137
Amphenicols	Protein synthesis	Chloramphenicol			129,134,135
Carbapenems	Cell wall synthesis		Imipenem		135
Tetracyclines	Protein synthesis			Tetracycline* Minocycline Doxycycline	17,129-131,136,139
Lincosamides	Protein synthesis			Lincomycin* Clindamycin*	131,132,135
Macrolides	Protein synthesis			Erythromycin* Tylosin	129-133
Quinolones	Nucleic acid synthesis			Norfloxacin Enrofloxacin Nalidixic acid	134
Aminoglycosides	Protein synthesis			Gentamicin Streptomycin	133,137
Polypeptides	Cell wall synthesis			Bacitracin Colistin	133,137

Table 1.3: A summary of *C. perfringens* antimicrobial resistance

*High-prevalence (both pheno- and genotypic) antibiotic resistance reported. Table reproduced from Kiu and Hall (2018) under the Creative Commons BY licence [1].

Recent WGS studies on AMR genes have detected *mepA* (multidrug-resistance gene), using various public AMR databases on *C. perfringens* strains [139], and also tetracycline resistant genes *tetA*(P), *tetB*(P) and *tet38*. The genes, *mprF* and *rpoB* (rifampin-resistant) have also been reported to be encoded [139]. Anti-defensin gene *mprF* (possibly involved in multidrug-resistant, including resistance against gentamicin) was recently reported in a large-scale genomic study of *C. perfringens* (n=56 strains) to be present in 100% of the genomes [17]. In this latest genomic study, *tetA*(P) was detected in 75% of the 56 strains, which is more prevalent than *tetB*(P)(42%). Interestingly, *ANT(6)-Ib*, an aminoglycoside resistance gene, was also reported to be encoded in a *C. perfringens* toxinotype C strain. Although mainly anaerobic bacteria like *C. perfringens* may have reduced transport of aminoglycosides intracellularly,

there are also findings that *C. perfringens* are sensitive to aminoglycosides like gentamicin at higher concentration, which indicates that *C. perfringens* might also have acquired additional resistance to aminoglycosides [140,141].

WGS will be a key tool to tackle *C. perfringens* AMR, particularly in rapid diagnostics, however gold-standard phenotypic MIC tests will also be required to clinically determine the antimicrobial susceptibilities and facilitate antibiotic management.

1.6 Clinical associations in humans and animals

C. perfringens has been constantly associated with gut diseases across both animal and human hosts (summarised in Table 1.4) as described in this section.

1.6.1 Animal hosts

1.6.1.1 Poultry NE

Poultry NE was first documented in England in 1961 [142]. Since then, NE has been consistently reported in every continent around the globe. Importantly, *C. perfringens* is unequivocally identified as the key aetiological organism of NE in broiler chickens [143]. Global financial impact of NE is significant, with an estimated economic loss of 6 billion US$ in 2015, projecting profit loss per bird at > US$0.062 [144].

Classic NE pathologies are characterised by gaseous lesions and mucosa necrosis in gas-filled distended small intestine, however, these can also involve kidney and liver as secondary infection sites [145]. Proposed key biological factors that contribute to NE are hydrolytic enzymes (e.g. collagenase) and toxin production (traditionally understood as PLC, and more recently NetB and TpeL, both are pore-forming toxins, were linked to the onset of NE) [146,147]. Other NE-predisposing factors include: a high-protein diet that favours the growth of *C. perfringens,* and environmental stress (e.g. high stocking density), which alters gut microbiota/host immunity that eventually increases risk of infection [148]. For a more detailed introduction to poultry NE see section 5.2.

Host species	Disease	Affected groups	Clinical symptoms	Associated toxinotypes/ toxins	Refs
Cow	Bovine necrotic enteritis	Neonatal calves	Distended colon; mucosal necrosis	Type A/ CPB2, PFO	87,149
Chicken	Poultry necrotic enteritis	Neonatal chicks 2-5 weeks post-hatching	Gaseous lesions; mucosal necrosis; distended intestines	Type A, C, G/ CPB, NetB, TpeL	150
Pig	Swine enterocolitis	1-7 days old neonatal piglets	Severe diarrhoea, necrotic mucosa, villus atrophy	Type A, C/CPB2	151
Dog	Canine gastroenteritis	n/a	Haemorrhagic/ necrotic intestines	Type A, F/enterotoxin CPE, CPB2	102
Horse	Equine necrotising enterocolitis	1-14 days old neonatal foals	Bloody diarrhoea; haemorrhagic and necrotic intestines	Type A, F/ CPE, CPB2, NetE, NetF & NetG	102
Human	Acute watery diarrhoea (food poisoning)	Both children and adults	8-14h after food ingestion: intestinal cramp, watery diarrhoea without vomiting (self-limiting: 12-24h)	Type F/ CPE	152
Human	Non-foodborne diarrhoea (antibiotic-associated diarrhoea and sporadic diarrhoea)	Older adults (>60 years old)	Abdominal pain and diarrhoea (prolonged duration >3 days to several weeks), often accompanied by blood in the faeces.	Type F/ CPE	153
Human	Enteritis necroticans (Pigbel)	Children/ malnourished adults	Intestinal gangrene (small intestine as key infection site)	Type C/ CPB	81
Human	Preterm necrotising enterocolitis	1-14 days old neonatal preterm infants	Distended abdomen; pneumatosis intestinalis (gas cysts formation); intestinal necrosis	Type A/ CPB2	8,9

Table 1.4: *C. perfringens*-associated intestinal diseases in both animals and humans

Table reproduced from Kiu and Hall (2018) under the Creative Commons BY licence [1].

1.6.1.2 Equine Acute Necrotising Enterocolitis

Acute necrotising enterocolitis (ANEC) in foals/horses (caused by *C. perfringens*) is a severe intestinal disease that resembles the classic clinical signs of *C. perfringens*-infections; rapid disease development and necrotic intestines (mainly the colon) [154]. ANEC symptoms are characterised by bloody diarrhoea, followed by a haemorrhagic and necrotic gut [155]. Type F *C. perfringens* strains (previously type A) harbouring both CPE and CPB2 are often associated with this deadly condition [154]. More recently, NetE, NetF and NetG toxins (pore-forming toxins) were proposed to underlie the pathogenesis of foal NEC [40].

1.6.1.3 Swine enterocolitis

C. perfringens type A strains (also less frequently, CPE-harbouring type F strains) are widely considered as the invasive agent of non-haemorrhagic enterocolitis in piglets, although the actual pathogenesis remains undefined. Similar to other *C. perfringens*-intestinal infection, this disease commonly affects one-week-old piglets, suspected to inherit (i.e. via vertical transmission during birth) the bacterium from the microbiota of mother sows [156]. Symptoms involve severe diarrhoea (non-haemorrhagic), accompanied by necrotic mucosa and atrophy of intestinal villi. CPB2 was initially believed to drive the development of this disease (backed by epidemiological studies in 1990s), and as such was used as a diagnostic biomarker, although this has become controversial in recent years [88]. Type C strains that carry CPB are associated with haemorrhagic enterocolitis in piglets and largely affect 1 to 4 days old neonates [156]. In contrast to type A-infection (to a lesser degree, type F-infection), type C-infection is characterised by haemorrhagic enterocolitis, which is proposed to be driven by the presence of type C strains and low trypsin secretion (trypsin can inactivate CPB) in the immature host gut [157].

1.6.2 Human hosts

1.6.2.1 Darmbrand

Darmbrand, which literally means ‘burning (fire) bowels’ in German, was used to describe a particular type of necrotic inflammatory gut disease (also known as enteritis necroticans, EN) associated with *C. perfringens*, that occurred epidemically post-second World War (1944-1949) in north-west Germany [158]. Darmbrand causative strains were later classified as type C strains that carry CPB, and these strains are highly resistant to heat [159]. Many of these type C strains also harbour CPE, which has been shown to act synergistically with CPB in an intestinal loop model and therefore CPE was suggested to play a role in some cases of EN [82]. Notably, the heat-resistance of this strain might be attributed to the production of a small acid-soluble protein (Ssp4) that could potentially play a central role in dissemination of *C. perfringens* strains [60]. Darmbrand was believed to be facilitated by poor post-war sanitary (bacterial contamination), and malnutrition (protein-malnourished) conditions, as this disease disappeared within a few years after its first recognition [160].

1.6.2.2 Pigbel

Enteritis necroticans (EN), or commonly known as pig-bel in the highlands of Papua New Guinea (PNG), is a form of inflammatory gut disease that was associated with pork-feasting activities that took place among PNG Highlanders in outbreaks first documented back in 1966 [161]. The classical description of the symptoms is 'spontaneous gangrene of small intestine, without obvious vascular or mechanical cause' which resembles Darmbrand, and occurred particularly in children [162]. The aetiology of this fatal infection was thought to be large co-consumption of *C. perfringens* type C-contaminated pork (due to poor hygiene), and large amount of sweet potatoes which contain trypsin inhibitor (trypsin secreted in the gut could break down CPB secreted by *C. perfringens*) [163]. Similar cases of EN were also reported in Uganda [164], Indonesia [165], Thailand [166], and Nepal [167]. Notably, EN-like cases were also reported (rare cases) in developed countries, including USA, in exclusively diabetic subjects, who also have attenuated trypsin production [168].

1.6.2.3 Acute Watery Diarrhoea (FP)

C. perfringens has been associated with FP since it was first documented in both the UK and the USA in the 1940s [169]. *C. perfringens*-type A FP, is ranked the second most prevalent bacterial FP in the US, estimated at 1 million cases per annum, after *Salmonella* [170]. In the European Union (EU) member countries, *C. perfringens*-linked FP outbreaks were projected at approximately 5 million cases per year (2011) [171]. *C. perfringens*-linked foodborne cases are suggested to be under-reported due to its self-limiting symptoms, thus the statistics published based on laboratory-confirmed cases may be lower than the actual numbers, suggesting a higher actual epidemiological impact.

Hallmark symptoms include, rapid appearance within 8-14 hrs after food ingestion, intestinal cramp, watery diarrhoea without fever or vomiting [172]. Currently, FP *C. perfringens* cases are thought to be caused by CPE (encoded by *cpe* gene). This secreted pore-forming toxin is also demonstrated to disrupt intestinal tight junction barriers and initiate disease development [112]. For a more detailed description on *C. perfringens*-associated acute watery diarrhoea (or, foodborne diarrhoea), see section 4.2.

1.6.2.4 Non-foodborne diarrhoea

C. perfringens has also been associated with non-foodborne diarrhoea (a distinct disease entity from FP, characterised primarily by more severe symptoms and longer duration), which includes antibiotic-associated diarrhoea (AAD), and sporadic diarrhoea [173,174]. AAD typically occurs in 5–25% of patients after administration of broad-spectrum antibiotics [175]. Non-foodborne diarrhoea typically affects older adults that are >60 years of age (although sporadic diarrhoea is also associated with younger age groups) [176]. Clinical symptoms are abdominal pain, prolonged diarrhoea (>3 days to several weeks), and bloody stools [153,177]. Patients suffering from non-foodborne diarrhoea, particularly AAD, can become seriously ill due to diarrhoea-induced dehydration, and may progress to develop colitis, although full recovery is common [178].

Although other pathogenic bacteria including *C. difficile* (most common AAD pathogen; ~25% of cases [179]) and *S. aureus* have also been implicated in AAD, enterotoxigenic *C. perfringens* type F strains (producing CPE) are estimated to be responsible for up to 15% of all AAD cases [173,180,181]. CPE, produced by *C. perfringens* type F strains (encoded by plasmid-borne *cpe* gene), has been reported to be the aetiological agent for *C. perfringens*-associated non-foodborne diarrhoea, as evidenced by the high prevalence in diarrhoea patients, but not healthy individuals. Importantly, type F strains have also been linked to sporadic diarrhoea, although to a lesser extent [173]. AAD-associated *C. perfringens* strains are also described to be more adherent to Caco-2 intestinal cells, when compared to other FP strains, which is attributed to the production of sialidase NanI [182]. The spore-forming nature of *C. perfringens* could also potentially contribute to the commonly observed disease persistence and relapse [183]. Molecular detection of faecal CPE or PCR confirmation of *cpe* gene represents the current clinical diagnostic method for *C. perfringens*-AAD [24,184].

1.6.2.5 Preterm infant NEC

Preterm-NEC has been clinically linked with *C. perfringens* since the 1970s, as type A *C. perfringens* were isolated from necrotic tissues in many NEC cases. Notably, it was even called ‘gas gangrene of the bowel’ owing to its reflection of the highly similar diseased histology with the infamous tissue myonecrosis [74].

NEC has ~14% prevalence in preterm infants less than 1000 g birth weight (i.e. extremely low birth weight, ELBW), a high mortality of 30% with surgical NEC, and overall mortality up to 50%. NEC-related deaths account for 10% of all mortality causes in neonatal intensive care unit (NICU) according to a UK-nationwide study [185], and is the most severe and lethal neonatal GI emergency worldwide [186]. The additional annual cost for treating NEC in preterm infants is conservatively estimated at £13.5 million in the UK (excluding long-term post-surgical treatment), whilst in the US $1 billion is estimated to be spent annually by neonatal departments on NEC [187].

The involvement of bacteria in the pathogenesis of NEC has been strongly implicated since the disease was first described [188]. Both *Klebsiella* and *C. perfringens* have been linked to clinical NEC in recent years [8]. Most recently, the emergence of two extensive preterm infant NEC cohort studies that profiled metagenomic faecal samples, have both indicated that *C. perfringens* is significantly abundant in the infant gut microbiota prior to NEC development [8,9]. These studies are supported by the substantial number of NEC studies (14 studies to date since 1970s, more than any individual bacterial agents, as summarised in Table 1.5) that have reported *C. perfringens* as a potential pathogenic agent [189]. Thus, *C. perfringens* appears to be a pathogenic agent of microbial-NEC, i.e. *C. perfringens*-associated NEC, and may link to specific disease symptoms [190]. For a more detailed description on *C. perfringens*-associated NEC, see section 3.2.

Reporting year	Major findings/ associations with *C. perfringens*	Disease presentation	Identification technique/ strains identified	Country	Refs
1976	*C. perfringens* isolated from 6 out of 7 NEC patients	Pneumatosis intestinalis	Culturing/ *C. perfringens* type A	USA	74
1978	*C. perfringens* detected in 3 of 4 deceased NEC patients	Severe pneumatosis intestinalis	Culturing/ *C. perfringens*	USA	191
1984	*C. perfringens* isolated from 5 NEC patients, 2 died; *C. perfringens* being the most found pathogen in 50 NEC patients in a 4-year cohort of 375 babies.	Pneumatosis intestinalis, portal venous gas, sepsis	Culturing/ *C. perfringens*	Australia	192
1984	*C. perfringens* detected in 2 NEC patients	n/a	Culturing/ *C. perfringens*	USA	193
1985	*C. perfringens* detected in 10 NEC patients with 78% fatal outcome	n/a	Culturing/ *C. perfringens*	Australia	194
1985	*C. perfringens* detected in 9 NEC patients with 78% fatal outcome	Intestinal gangrene and extensive pneumatosis intestinalis	Culturing/ *C. perfringens*	USA	195
1990	*C. perfringens* was more abundant in 1 out of 4 NEC infants	n/a	Culturing/ *C. perfringens*	England	196
2004	Case-control study of 12 preterm infants. 3 *Clostridium-perfringens*-positive infants developed NEC later (100%) and eventually died (100%)	n/a	16S rRNA PCR/ *C. perfringens* (95% identity)	France	197

Table 1.5: Studies reporting clinical cases of *C. perfringens*-associated NEC

Table reproduced from Kiu and Hall (2018) under the Creative Commons BY licence [1].

Continued on next page

Reporting year	Major findings/ associations with *C. perfringens*	Disease presentation	Identification technique/ strains identified	Country	Refs
2008	Case-control study of 9 NEC patients associated with *C. perfringens*. Outcomes compared with non-*C. perfringens*-NEC (n=32): higher mortality (44% vs 7%) and high percentage of portal venous gas (78% vs 25%)	Abdominal distention; bloody stools; pneumatosis intestinalis; portal venous gas	Culturing; multiplex-PCR toxinotyping/ *C. perfringens* type A	Germany	198
2009	*C. perfringens* isolated from necrotic colon wall of 1 NEC preterm patient	Abdominal distention; necrotic intestine; pneumatosis intestinalis; multiple perforation; haemolysis	Culturing; Nagler reaction on egg yolk agar	Germany	190
2010	*C. perfringens* isolated from 3 NEC patients (faecal samples)	n/a	Culturing; 16S rRNA PCR; multiplex-PCR toxinotyping/ *C. perfringens* type A	Australia	11
2010	*C. perfringens* detected from 3 NEC patients (peritoneal fluid samples) who eventually died from multisystem organ failure	Abdominal distention; pneumatosis intestinalis; extensive bowel necrosis	Intra-hospital culturing	Switzerland	199
2015	A 2-year cohort study of 369 preterm infants. Overabundant *C. perfringens* were detected in 4 NEC (Bell 2/3) patients before disease onset (in a total of 12 NEC patients) using NGS technology	n/a	16S rRNA metagenomics (V3-V5 regions); culturing and MALDI-TOF; multiplex-PCR toxinotyping/ *C. perfringens* type A, 3 isolates harboured beta2-toxin	England	8
2016	A case-control study of 3 NEC patients. *C. perfringens* were found to be significantly more abundant from meconium to NEC onset.	Pneumatosis intestinalis and/or portal venous gas	16S rRNA metagenomics (V3-V4 regions)	The Netherlands	9

Table 1.5: Studies reporting clinical cases of *C. perfringens*-associated NEC

Table reproduced from Kiu and Hall (2018) under the Creative Commons BY licence [1].

Continued from previous page

Recent advancement of WGS and genomic analyses, particularly those focussing on other gut pathogens including *C. difficile*, *Klebsiella pneumoniae* and *Salmonella enterica* has proved successful in understanding the pathogenic role of these bacteria in disease development and tracking bacterial transmission [200,201]. These enteric pathogens have all been sequenced extensively with genomic information made public on NCBI (National Center for Biotechnology Information) genome database (as of November 2018; bracketed digits indicate number of genomes): *C. difficile* (1404), *E. coli* (13475), *K. pneumoniae* (6503) and *S. enterica* (9798). Yet, including two batch-additions of 30 genomes towards the end of 2016 and >55 genomes in mid-2018, there are only 112 publicly available genomes of *C. perfringens*, highlighting the lack of in-depth understanding of this pathogen with respect to its genomic information. WGS-based investigation may allow identification of putative/novel disease-associated genotypes and virulence factors, that may facilitate intervention strategies and/or therapy development.

1.7 Unanswered questions

The clinical link between intestinal diseases and *C. perfringens* is clear and defined, however the underlying factors responsible for specific aspects of pathology remains uncertain (Table 1.6).

Deciphering the genes that are involved in the sporulation, germination, enzyme/toxin production, oxygen tolerance, AMR and other novel virulence factors could lead to more targeted clinical preventions in *C. perfringens*-associated adult and neonatal intestinal diseases, whether it be humans or animals. Important questions may be answered utilising WGS and *in silico* tools, to delve into the genetic makeup of this notorious, yet under-studied pathogen. In tandem further *in vitro* and *in vivo* studies should be carried out to confirm the importance of suspected novel virulence traits, i.e. to fulfil Molecular Koch's postulates.

What are the key knowledge gaps in understanding *C. perfringens* as a pathogen, and what are the potential approaches to investigate/mitigate the virulence of *C. perfringens* in clinical diseases?

Questions	Approaches
• Compared to other pathogens (e.g. *C. difficile*), there is limited WGS data for *C. perfringens,* what is the genomic diversity of this important pathogen? • Can we use WGS to 'type' *C. perfringens* isolates, and can WGS be used in diagnostics (including AMR profiling), and to track spread in outbreaks settings? • *C. perfringens* causes diverse diseases in a wide host range, but what are the microbial virulence factors responsible for pathology, and is this host/strain specific? • What are the host immune cells/ signalling cascades involved in *C. perfringens*-associated clearance and/or pathology? • *C. perfringens* can reside as a 'normal' member of the gut microbiota, what are the external factors that may lead to overgrowth and disease? • Can *C. perfringens* be directly linked to NEC pathogenesis? • What are the therapeutic approaches that could be used to prevent *C. perfringens* infection in multi-host species?	• Isolate diverse strains from various environments, i.e. from both case and control samples • Perform comprehensive bioinformatics analyses to understand evolution e.g. SNP based phylogeny, and investigate variants and spread in clinical settings • Carry out pan-GWAS/comparative genomics to identify/predict specific virulence genes related to disease development. • Develop/optimise clinical diagnostics from faecal samples via e.g. ultra-long read real-time sequencing method such as Oxford Nanopore • Develop and characterise *in vivo* infection models for different disease types (e.g. NEC and gastroenteritis) • Pinpoint specific virulence genes involved in infection via bacterial transcriptomics (RNAseq) and knock-out bacterial strains, to fulfil molecular Koch's postulates. • Perform microbiota profiling (e.g. 16S rRNA metagenomics or shotgun metagenomics) to investigate the impact of *C. perfringens* to microbiota members, and the role of microbiome in *C. perfringens* infections • Understand host defence mechanisms using immunological approaches, and knock out *in vivo* models • Develop therapeutics for *C. perfringens* infections including phage therapy, vaccines, microbiota treatment i.e. probiotics

Table 1.6: Knowledge gaps in understanding *C. perfringens* and proposed research approaches
Table reproduced from Kiu and Hall (2018) under the Creative Commons BY licence [1].

These approaches may help establish future platforms for disease prevention strategies such as vaccines, phage therapy, microbiota-based therapeutics, or implementation of specific antibiotic administration policies. Furthermore, understanding the genomes will potentially enable epidemiological tracking (as is taking place for other

pathogenic bacteria), allowing us to pinpoint the origin and route of spread of isolates in the hospital settings, which is vital within clinical contexts.

1.8 Hypotheses

1.8.1 Overarching hypothesis

C. perfringens plays a causative role in intestinal diseases, including preterm-NEC, foodborne diarrhoea and poultry-NE, via an array of toxin and accessory virulence genes, which are widely disseminated via clonal expansion.

1.8.2 Study-specific hypotheses

Chapter 3 - Specific nosocomial virulent genotypes of *C. perfringens* are linked to preterm-NEC disease progression

Chapter 4 - Specific epidemic strains of *C. perfringens* are responsible for gastrointestinal-associated outbreaks

Chapter 5 - Broiler caecal contents exhibit microbiome signatures for poultry NE, and specific genotypes of chicken-associated *C. perfringens* are linked with poultry NE

Chapter 6 - *C. perfringens* is a genetically diverse, zoonotic and widely-disseminated pathogen

Chapter 7 - *C. perfringens* isolates exhibit strain-specific virulence properties, and certain disease-linked strains are 'fitter' than other non-pathogenic strains

Chapter 8 – Hyper-virulent *C. perfringens* strains can induce disease symptoms in mice

1.9 Primary research aims

This research had 5 primary research aims:

1. To isolate and sequence (WGS) gut-associated *C. perfringens* from clinical samples, cohorts including preterm infants, FP patients, care-home non-foodborne diarrhoea patients and broiler chickens.
2. To genotype *C. perfringens*: using bioinformatics to determine full virulence profiles and genetic variations of *C. perfringens* isolates, and link disease outcomes with pathogenic genotypes
3. To determine genomic epidemiological spread of disease-associated *C. perfringens* genotypes and understand global phylogenetics of *C. perfringens*.
4. To phenotype *C. perfringens* virulence traits and characterise *C. perfringens* pathogenesis using *in vitro* models for the selection of phenotypically 'hyper-virulent' strains to be employed in animal models.
5. To set up and characterise a novel oral-challenged mouse model of *C. perfringens* intestinal infections.

1.10 Structure of the thesis

The thesis is organised into 9 chapters.

Chapter 1 is an introductory chapter that presents a comprehensive literature review on the research background and summarises the hypotheses and aims of the thesis.

Chapter 2 comprises the details of materials and research methods, including various bioinformatics approaches used in the studies.

Chapter 3 presents a study on the genomic virulence and epidemiology of preterm-associated *C. perfringens*. This chapter provides insights into the full virulence profiles and pangenome of hospital-linked *C. perfringens*, and a novel understanding of the nosocomial spread and persistence of *C. perfringens* in a multicentre neonatal study.

Chapter 4 describes phylogenomic investigation on food-poisoning and care home-associated *C. perfringens* isolates across England and Wales. This study determines

specific virulence plasmid disseminations and identified potential clonal expansion of enterotoxigenic *C. perfringens* strains.

Chapter 5 contains a microbiome study on chicken-associated caecum contents and a genomic investigation on NE-related isolates. The findings identify specific NE-diagnostic microbiome marker and combinatory role of NE-linked virulence genes.

Chapter 6 is a population genomic study on the complete set of isolates available to date including public genomes and newly sequenced genomes. This study presents the pangenome of *C. perfringens*, and identifies potential zoonotic relationship, genetically distinct lineages and intercontinental disseminations.

Chapter 7 describes the phenotypical characterisation of a sub-set of preterm-associated *C. perfringens* isolates. This chapter identifies hyper-virulent *C. perfringens* strains that were selected for use in a novel mouse infection model study described in Chapter 8.

Chapter 8 presents a novel oral-challenge *C. perfringens* infection model that reveals unexpected immune-mediated microscopic pathologies. This reproducible infection model has the potential for use in future therapy development against *C. perfringens* intestinal infections.

Chapter 9 is the final discussion chapter, summarises the impact of the new findings, future perspectives and conclusions of the thesis.

Chapter 2 Materials and methods

2.1 Materials

2.1.1 Reagents, kits and equipment

Primary reagents, kits, equipment and tools used/mentioned in this chapter are listed in Table 2.1 to Table 2.3.

Materials	Components and conditions	Supplier
Phosphate Buffer Saline (PBS)	PBS, pH7.2	Sigma-Aldrich, Dorset, UK
25% sucrose in TE buffer	TE buffer: 10mM Tris and 1mM EDTA pH8.0	Sigma-Aldrich, Dorset, UK
Lysozyme	Lysozyme re-suspended in 0.25M Tris pH8.0	Roche, West Sussex, UK
Proteinase K	~15-20mg/ml proteinase K	Roche, West Sussex, UK
RNase A	10mg/ml RNase A (boiled before use)	Roche, West Sussex, UK
EDTA	0.5M EDTA, pH8.0 (pH adjusted)	Sigma-Aldrich, Dorset, UK
Sarkosyl NL30	30% Sarkosyl	Sigma-Aldrich, Dorset, UK
Phenol-Chloroform-Isoamyl alcohol (IAA) solution	25:24:1 (Phenol-Chloroform-Isoamyl Alcohol)	Sigma-Aldrich, Dorset, UK
E buffer	10mM Tris pH8.0	Sigma-Aldrich, Dorset, UK
Chloroform:IAA	24:1 Chloroform:IAA	Sigma-Aldrich, Dorset, UK
Ethanol	Absolute ethanol (>99.5%)	Sigma-Aldrich, Dorset, UK
Phase-lock gel tubes for DNA extraction using Phenol-Chloroform	MaXtract High Density tubes	Qiagen, Manchester, UK

Table 2.1: Materials used in Phenol-Chloroform DNA extraction for WGS samples

Equipment	Model	Supplier
Anaerobic chamber	Ruskinn Concept Plus	Baker Ruskinn, Bridgend, UK
Autoclave (portable)	Classic 2100 Standard (9-litre)	Prestige Medical, Blackburn, UK
Centrifuge	Eppendorf 5810R	Eppendorf, Stevenage, UK
Microscope (fluorescence and brightfield)	Carl Zeiss Axio Imager 2	Carl Zeiss, Cambridge, UK
Gel electrophoresis unit	ENDURO Gel XL	Labnet, New Jersey, USA
Homogeniser for DNA extraction samples	FastPrep-24	MP Biomedicals, Santa Ana, USA
Incubator	Cooling incubator	Binder, Tuttlingen, Germnay
Micro-centrifuge	Prism	Labnet, New Jersey, USA
Micro-centrifuge (refrigerated)	Prism R	Labnet, New Jersey, USA
Microplate reader	FLUOstar Omega	BMG Labtech, Ortenberg, Germany
Microplate reader (anaerobic condition)	Tecan infinite F50	Tecan, Männedorf, Switzerland
Nanodrop spectrophotometer	Nanodrop 2000	Thermo Fisher Scientific, Loughborough, UK
PCR station	UV HEPA PCR Systems	UVP, Upland, USA
PCR thermal cycler	Applied Biosystems Veriti 96-well thermal cycler	Thermo Fisher Scientific, Loughborough, UK
pH meter	Martini MI 151	Rocky Mount, USA
Pressure transducer (USB connection)	PX409-015GUSBH	Omega Engineering Limited, Manchester, UK
Qubit meter	Qubit 2.0	Thermo Fisher Scientific, Loughborough, UK
Safety cabinet	Walker Class II MSC	Walker Safety Cabinets, Glossop, UK

Table 2.2: Details of equipment used in the laboratory

Item	Supplier
Agar	Oxoid (Thermo Fisher Scientific), Loughborough, UK
Agarose for Gel Electrophoresis	Sigma-Aldrich, Dorset, UK
Brain Heart Infusion (BHI)	Oxoid (Thermo Fisher Scientific), Loughborough, UK
Caspase-Glo 3/7 Assay	Promega, Southampton, UK
CytoTox-One Homogeneous Membrane Integrity Assay	Promega, Southampton, UK
D-cycloserine	Oxoid (Thermo Fisher Scientific), Loughborough, UK
DNA ladder (1kb and 100bp)	Promega, Southampton, UK
DNA lo-bind tubes (1.5 and 2.0ml)	Eppendorf, Stevenage, UK
DPX mountant for histology	Sigma-Aldrich, Dorset, UK
Egg yolk emulsion	Oxoid (Thermo Fisher Scientific), Loughborough, UK
FastDNA SPIN kit for Soil	MP Biomedicals, Santa Ana, USA
Feeding tube 20G	Instech, Plymouth, USA
Inoculation sterile loops (disposable)	Microspec, Wirral, UK
KAPA 2G Robust PCR kit	KAPA Biosystems, Wilmington, USA
Lead acetate test strips (detection of H_2S)	Sigma-Aldrich, Dorset, UK
Molecular water (water for molecular biology)	Merck Millipore, Billerica, USA
Perfringens Agar Base	Oxoid (Thermo Fisher Scientific), Loughborough, UK
QIAGEN Plasmid Midi Kit	Qiagen, Manchester, UK
QIAquick PCR Purification Kit	Qiagen, Manchester, UK
Qubit dsDNA BR assay	Invitrogen (Thermo Fisher Scientific), Loughborough, UK
Reinforced Clostridial Medium (RCM)	Oxoid (Thermo Fisher Scientific), Loughborough, UK
Spreader	Microspec, Wirral, UK
Staurosporine	Abcam, Cambridge, UK

Table 2.3: Materials and kits

2.1.2 Ethics approval and consent for participation in human studies

2.1.2.1 BAMBI study

BAMBI study (Baby Associated MicroBiota of the Intestine study) carried out at Norfolk and Norwich University Hospital (NNUH) was approved by the University of East Anglia (UEA) Faculty of Medical and Health Sciences (FMH) Ethics Committee. Sample collection followed the protocols outlined by NRES approved UEA Biorepository (Licence no.: 11208). Written consent was given by the parents for their infants for participation in this study [202].

2.1.2.2 NEOM study

NEOM study (The Neonatal Microbiota Study) was approved by West London Research Ethics Committee Two (ClinicalTrials.gov identifier NCT01102738; Licence no.: 10/H0711/39). Parents gave written informed consent for their infant to participate in the study [8].

2.1.3 Faecal samples, bacterial isolates and isolate DNA extracts

Samples (faecal samples/isolate/isolate DNA) were obtained through collaborations with various organisations as described (Table 2.4).

Collaborators	Organisations	Sample cohort/sample type	Location
Dr Paul Clarke	NNUH	Preterm infants (BAMBI)/ faecal samples	Norwich, UK
Dr Gusztav Belteki	Rosie Hospital	Preterm infants (BAMBI)/ faecal samples	Cambridge, UK
Prof J Simon Kroll Dr Alex Shaw Dr Kathleen Sim	Imperial College London	Preterm infants in SMH and QCCH (NEOM and two-year NEOM)/bacterial isolates	London, UK
Dr Corinne Amar	Public Health England (PHE)	Outbreak patients in England and Wales/isolate DNA	Colindale, UK
Sara Goatcher	Banham Zoo and Africa Alive	Captive animals/faecal samples	Suffolk, UK
Dr Ron Dixon Dr Joseph Brown	Lincoln University	Poultry and farm/bacterial isolates	Lincoln, UK
Prof Paul O'Toole	University College Cork	Older adults in care homes (ELDERMET)/ bacterial isolates	Cork, Ireland
Dr Alan Walker	University of Aberdeen	Healthy adults/isolate DNA	Aberdeen, UK

Table 2.4: Brief description of sample sources and relevant collaborators

NNUH: Norfolk and Norwich University Hospital; QCCH: Queen Charlotte's and Chelsea Hospital; SMH: St Mary's Hospital.

2.1.3.1 Faecal sample collection in BAMBI study

Faecal samples were routinely collected from infant nappies in the NICUs into sterile stool containers and stored at 4 °C and stored at -80 °C within 5 days.

2.1.3.2 Faecal sample collection in NEOM study

Samples were collected by nursing staff from nappies into sterile Eppendorf tubes and stored in a -20 °C freezer on the NICU within 2h of collection and stored at -80 °C within 5 days.

2.2 DNA extraction

2.2.1 Using FastDNA SPIN kit

This method was used on bacterial/faecal samples for 16S rRNA metataxonomics, 16S preliminary identification and qPCR bacterial load estimation.

Briefly, DNA extraction (~100mg for faecal sample or ~100µl for bacterial sample) was performed using FastDNA SPIN kit for Soil, according to the protocol described by the manufacturers, with a modification to the homogenisation of samples - 3 min at speed setting of 6.0, using FastPrep-24 instrument. Samples were suspended in 1ml of Sodium Phosphate Buffer. DNA extraction protocol was then performed accordingly, with a final elution volume of 65µl. Extracted DNA was stored in sterile DNase-free and RNase-free DNA lo-bind tubes at -20°C after quantification until further analysis.

2.2.2 Phenol-Chloroform extraction

This optimised 2-day protocol was used to generate ultra-high quality pure bacterial genomic DNA for WGS, originally obtained from Dr Derek Pickard (Wellcome Trust Sanger Institute; WTSI) [202].

2.2.2.1 First-day protocol

10ml pure cultures in Brain Heart Infusion (BHI) were harvested (5-7h) for phenol-chloroform DNA extraction (washed with PBS). Briefly, bacterial pellets were re-suspended in 2ml 25% sucrose in 10mM Tris and 1mM EDTA at pH 8.0. Cells were lysed using 50µl 100mg/ml lysozyme. 100µl 20mg/ml Proteinase K, 30µl 10mg/ml

RNase A, 400µl 0.5M EDTA and 250µl 10% Sarkosyl NL30 were added into the lysed bacterial suspension after 1h-incubation at 37°C. This is followed by 60-120 mins incubation on ice (until complete lysis is observed) and 50°C overnight water bath.

2.2.2.2 Second-day protocol

Two to three rounds of equal volume (5ml each) phenol-chloroform-isoamyl alcohol extraction using 15ml gel-lock tubes to separate DNA from bacterial waste after reconstituting the overnight samples in E buffer to 5ml. Chloroform-isoamyl alcohol extraction was then performed to remove residual phenol, followed by ethanol precipitation and 10ml 70% ethanol wash by centrifugation (at least twice). Air-dried DNA pellets were finally re-suspended in 200-300µl of E buffer and stored in -20°C after quantitation until further analysis. Refer to Table 2.1 for details of materials and conditions. This adapted and optimised extraction protocol was originally provided by Dr Derek Pickard (WTSI).

2.2.3 Pre-sequencing analysis of genomic DNA quality and quantity

DNA quality was assessed by Nanodrop spectrophotometer (acceptable thresholds: 260/280 ratio ~1.8-2.0, 260/230 ratio > 1.8) and DNA concentration was quantified using Qubit 2.0 dsDNA BR assay kit.

2.2.4 Plasmid DNA extraction

Plasmid DNA was extracted using QIAGEN Plasmid Mini/Midi kit and performed according to manufacturer's instructions.

2.3 DNA Sequencing

2.3.1 Sanger Sequencing on full-length 16S rRNA amplicons (bacteria)

2.3.1.1 PCR primers, conditions, product confirmation and purification

Full-length 16S PCR was performed as a culture-independent method to identify bacterial genera. Briefly, KAPA2G Robust PCR reagents were used in preparing the PCR master mix. DNase- and RNase-free sterile water was used in preparing dilution and master mix for PCR. 5µl of each diluted bacterial DNA extract (optimal DNA

concentration for PCR: 10-30 ng/µl) was amplified with PCR master mix comprising: 10 µM primers (1µl/sample)[203], 10mM of each dNTP mix (1 µl/sample), GC buffer with $MgCl_2$ (10µl/sample), Millipore H_2O (31.6 µl/sample) and Kapa2G Robust DNA polymerase (0.4µl/sample), master mix was prepared according to careful calculation.

Primers	Sequences: 5' -3'
fD1	AGA GTT TGA TCC TGG CTC AG
fD2	AGA GTT TGA TCA TGG CTC AG
rP1	ACG GTT ACC TTG TTA CGA CTT

Table 2.5: Sequences of universal primers used for full-length 16S rRNA PCR

Primer abbreviations: f, forward; r, reverse; D, distal; P, proximal [203].

PCR Step	Temperature	Duration	Cycles
Initial denaturation	94	5 min	1
Denaturation	94	1 min	35
Annealing	43	1	
Extension	72	2 min	
Final extension	72	2 min	1

Table 2.6: PCR amplification conditions for full-length 16S rRNA PCR

PCR was performed in thermal cycler according to the conditions detailed in Table 2.6. PCR products (~1,500bp) were separated on 1% agarose gel and DNA bands were stained with ethidium bromide for x-ray visualisation, negative control (molecular water) was included to detect any possible contamination of DNA. 1kb ladder was used to estimate the size of the PCR amplicon bands. Primers and reagents were removed from the PCR products using a QIAquick PCR purification kit according to manufacturers' protocol before performing Sanger Sequencing.

2.3.1.2 Sequencing

PCR amplicons were then prepared according to sample submission guide (Eurofins, Luxembourg) for nucleotide reading service at Eurofins (Sanger sequencing method). Most final PCR products ranged between 900-1000bp (after quality trimming). Illumina short-read sequencing

2.3.1.3 WGS

This work was performed at the WTSI (Hinxton, UK) facilitated by Dr Derek Pickard and in collaboration with Prof Gordon Dougan. DNA libraries were subjected to WGS

pipeline on Illumina HiSeq 2500 platform to generate 101bp/125bp paired-end reads as described previously [34]. The sequencing library preparation process was detailed in Table 2.7 and Table 2.8.

Step	Step description	Instruments/ details
1	Quantitation of DNA	Using Biotium Accuclear Ultra high sensitivity dsDNA Quantitative kit
2	DNA shearing	mechanical shearing using Covaris LE220 instrument
3	Purification of sheared samples	Agencourt AMPure XP SPRI beads on Agilent Bravo WS
4	Library construction (end-repair, A-tailing and ligation)	using 'NEB Ultra II custom kit' on an Agilent Bravo WS automation system
5	PCR	using KapaHiFi Hot start mix and IDT 96 iPCR tag barcodes on Agilent Bravo WS automation system
6	Post-PCR purification	Agencourt AMPure XP SPRI beads on Beckman BioMek NX96 liquid handling platform
7	Quantitation of libraries	Biotium Accuclear Ultra high sensitivity dsDNA Quantitative kit
8	Pooling libraries	Pool in equimolar amounts on a Beckman BioMek NX-8 liquid handling platform
9	Normalisation	libraries normalised to 2.8nM
10	Loading to sequencer	loading on requested Illumina sequencing platform

Table 2.7: WGS library process with details

PCR Step	Temperature	Duration	Cycles
Initial denaturation	95	5 min	1
Denaturation	98	30 s	5
Annealing	65	30 s	
Extension	72	1 min	
Final extension	72	10 min	1

Table 2.8: PCR conditions for WGS libraries

2.3.1.4 Metataxonomics using 16S rRNA

This work was performed at the WTSI (Hinxton, UK). Library preparation was carried out by Charlotte Leclaire. Bacterial DNA concentration was quantified using Qubit and normalised to 5 ng/µl for all samples. Extracted DNA was used as a template for PCR amplification (Table 2.9) of the V1+V2 regions of the 16S rRNA gene as detailed

in Table 2.10 for the primer sequences. 16S rRNA gene libraries were sequenced on the Illumina MiSeq platform with 250bp paired-end reads.

PCR Step	Temperature	Duration	Cycles
Initial denaturation	94	3 min	1
Denaturation	94	45 s	25
Annealing	55	15 s	
Extension	72	30 s	

Table 2.9: Amplification conditions of PCR for 16S metataxonomics (V1+V2 regions)

Primers	Sequences: 5'-3'
V1FW_SD501	AATGATACGGCGACCACCGAGATCTACACAAGCAGCATATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V1FW_SD502	AATGATACGGCGACCACCGAGATCTACACACGCGTGATATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V1FW_SD503	AATGATACGGCGACCACCGAGATCTACACCGATCTACTATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V1FW_SD504	AATGATACGGCGACCACCGAGATCTACACTGCGTCACTATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V1FW_SD505	AATGATACGGCGACCACCGAGATCTACACGTCTAGTGTATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V1FW_SD506	AATGATACGGCGACCACCGAGATCTACACCTAGTATGTATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V1FW_SD507	AATGATACGGCGACCACCGAGATCTACACGATAGCGTTATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V1FW_SD508	AATGATACGGCGACCACCGAGATCTACACTCTACACTTATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V1FW_SA501	AATGATACGGCGACCACCGAGATCTACACATCGTACGTATGGTAATTGTAGMGTTYGATYMTGGCTCAG
V2RV_SD701	CAAGCAGAAGACGGCATACGAGATACCTAGTAAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD702	CAAGCAGAAGACGGCATACGAGATACGTACGTAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD703	CAAGCAGAAGACGGCATACGAGATATATCGCGAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD704	CAAGCAGAAGACGGCATACGAGATCACGATAGAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD705	CAAGCAGAAGACGGCATACGAGATCGTATCGCAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD706	CAAGCAGAAGACGGCATACGAGATCTGCGACTAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD707	CAAGCAGAAGACGGCATACGAGATGCTGTAACAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD708	CAAGCAGAAGACGGCATACGAGATGGACGTTAAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD709	CAAGCAGAAGACGGCATACGAGATGGTCGTAGAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD710	CAAGCAGAAGACGGCATACGAGATTAAGTCTCAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD711	CAAGCAGAAGACGGCATACGAGATTACACAGTAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT
V2RV_SD712	CAAGCAGAAGACGGCATACGAGATTTGACGCAAGTCAGTCAGCCGCTGCCTCCCGTAGGAGT

Table 2.10: Primer sequences for amplifying V1+V2 regions of 16S rRNA gene

2.3.2 454 pyrosequencing

This work was performed by Dr Kathleen Sim and Dr Alex Shaw at Imperial College London. The V3-V5 regions of the 16S rRNA gene were amplified using forward primer 357F and reverse primer 926Rb (Table 2.11) as detailed in section 2.7 in Sim (2015) [12]. Briefly, 16S rRNA amplicons were sequenced by Imperial College London in-house 454 Life Sciences GS Junior Titanium machine, or the Department of Biochemistry of University of Cambridge, using 454 Life Sciences GS FLX Titanium machine following the Roche Amplicon Lib-L protocol. The workflow and protocols of the 454 sequencing are detailed on p.64 in Sim (2015) [12].

Primers	Sequences: 5' – 3'
357F	CTATCCCCTGTGTGCCTTGGCAGTCTCAGCCTACGGGAGGC AGCAG
926Rb	CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNNNNNNNN CCGTCAATTY MTTTRAGT

Table 2.11: Sequences of primers used in 454 pyrosequencing

This information is taken from Sim (2015) [204].

2.4 Bioinformatics

Bioinformatic tools/pipelines applied in this thesis are described in this section. Most of the written scripts developed for use during my PhD can be found in my GitHub repositories[1].

2.4.1 Computing environment and resources

All bioinformatics command-line computing tools were run on Norwich Bioscience Institute (NBI) High-Performing Computing Cluster (HPC) on CentOS Linux operating system. Relevant tool and dependencies (modules required by another tool) are loaded (using command *source*) on to the shell environment before usage of a software/tool. SLURM workload manager[2] (v16.05.8) was routinely used to submit computational jobs to HPC.

[1] https://github.com/raymondkiu/informatics-tools

[2] https://slurm.schedmd.com/

2.4.2 16S rRNA full-length amplicon sequence analysis

Primarily for bacterial species identification, 16S rRNA nucleotide sequences for bacterial isolates were mapped with NCBI nucleotide database (16S ribosomal RNA sequences: Archaea and Bacteria) using NCBI BLASTN, reference bacterial species with highest identity hits/similarity scores (sequence identity must be ≥97% for species level threshold) were assigned for preliminary identification[3].

2.4.3 16S rRNA microbiome analysis via OTU (Operational Taxonomic Unit) assignment

Software 'Quantitative Insights Into Microbial Ecology' (QIIME) v1.9.1 was used for analysing the 16S amplicon data (454 pyrosequencing and Illumina paired-end sequencing data) [205].

2.4.3.1 454 long-read data

454 pyrosequencing Standard Flowgram Format (SFF) files were converted to FASTA files (process_sff.py) followed by demultiplexing and quality-filtering step at quality score 50 (split_libraries.py).

2.4.3.2 Illumina paired-end data

FASTQ paired-end reads were assembled/merged using PEAR [206], at Phred quality score 33 and followed by quality filtering: --phred_quality_threshold 19 and --phred_offset 33 (split_libraries_fastq.py).

2.4.3.3 OTU assignment and clustering

Chimeras were removed (identify_chimeric_seqs.py) using USEARCH v6.1 [207] prior to open-reference OTU assignment (pick_open_reference_otus.py) to representative clustered sequences based on SILVA rRNA database (SSU_Ref_128; released in September 2016) [208] using UCLUST clustering algorithm at 97% sequence identity [207]. Taxa read counts <5 and samples with <100 total reads were omitted for clarity of data representation. BIOM files were subsequently generated for data visualisation in software MEGAN v6.10.4 [209]. Reads were normalised as Relative Abundance (%) before visualising using R programming environment [210].

---

[3] http://blast.ncbi.nlm.nih.gov/

2.4.4 Bacterial whole genome *de novo* assembly

2.4.4.1 Short-read genome assembly

An optimised and standardised approach was employed to generate *de novo* assembly for data consistency. Bacterial genomes sequenced at WTSI were assembled using the presented pipeline (Figure 2.1) [211]. Briefly, *de novo* assemblers Velvet v1.2.10 or SPAdes v3.8.1 are used to assemble short-read sequencing data, initial assembly was further improved by Sanger assembly improvement pipeline v1.7.0 which comprises SSPACE as scaffolder to scaffold contigs and GapFiller to fill up gaps in between contigs by picking up unmapped reads from FASTQ reads (Figure 2.1) [211-214].

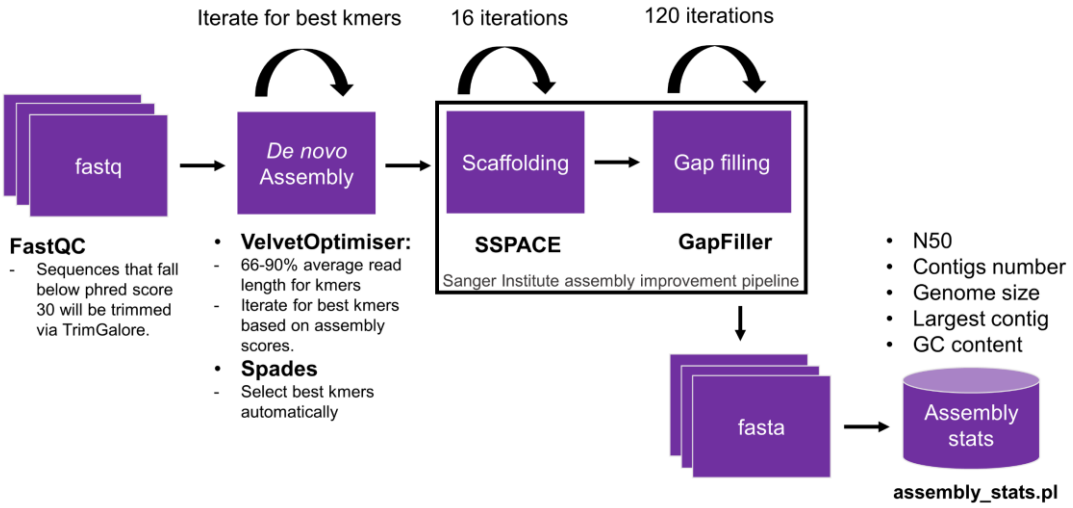


Figure 2.1: *De novo* genome assembly workflow used for short-read sequencing data

2.4.4.2 Long-read genome assembly

Canu v1.5, successor of long-read sequence-assembler Celera Assembler, was used to assemble long-read sequencing data in this study [215]. Canu assembly mainly carries out 4 major tasks: noise detection, sequence correction, sequence trimming and sequence assembly. Workflow was demonstrated in Figure 2.2. Standard workflow was carried out to obtain long-read assembly for analysis.

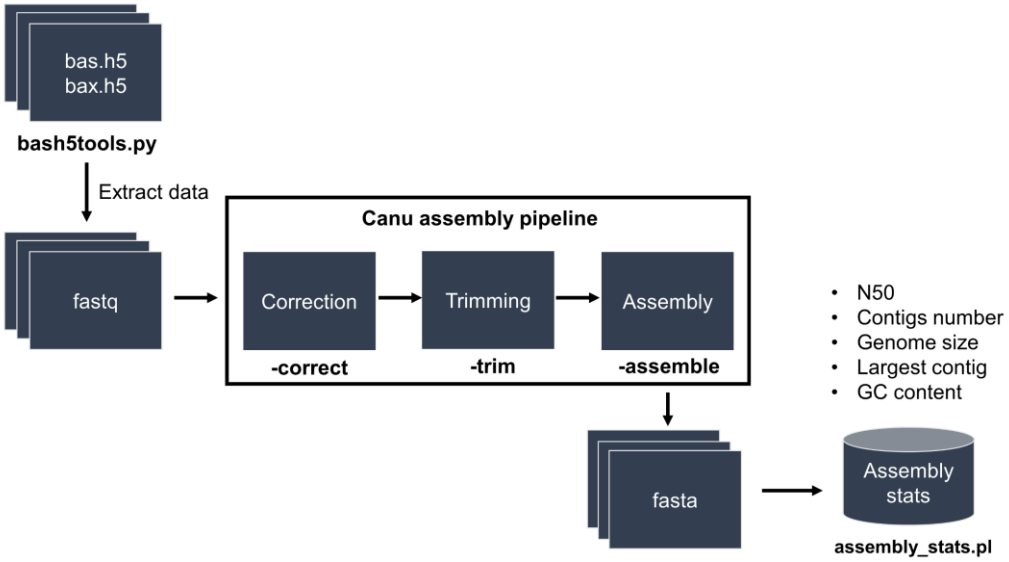


Figure 2.2: *De novo* genome assembly workflow used for long-read sequencing

2.4.4.3 Genome contamination checks

Genome assemblies were checked for potential sequence contamination prior to downstream analysis. Genome contamination checks (based on assemblies) were firstly performed using Kraken (based on MiniKraken database [216]) to detect sequences of other bacterial species. Any genome assemblies found to be contaminated with >5% of other bacterial species were excluded from analyses. Genomes were subjected to Average Nucleotide Identity (ANI) method (using mummer-based pyani[4]) to ensure genome quality. ANI >95% (mapped against reference genome or pair-wise comparison) confirms species assignment [217]. Sequences with <95% ANI were excluded from analysis. Assemblies >500 contigs (considered as highly fragmented assemblies) were also excluded from studies.

2.4.5 Genome annotation

Prokka v1.13 was used for genome annotation [218]. Specifically, an in-house genus-specific database (genus *Clostridium*) was generated (Table 2.12) and compiled using CD-HIT v4.6.6 as BLAST database based on NCBI Reference Sequence Database (RefSeq) high quality genome annotations including 35 *Clostridium* species. This

[4] https://github.com/widdowquinn/pyani

genus-specific database was used by --usegenus command with priority over default protein databases.

No	RefSeq Genome	Strain	NCBI Accession
1	*C. perfringens*	1207_CPER	SAMN03197169
2	*C. perfringens*	ATCC13124	SAMN02604008
3	*C. perfringens*	B_ATCC3626	SAMN02436295
4	*C. perfringens*	CBA7123	SAMD00057672
5	*C. perfringens*	C_JGS1495	SAMN02436294
6	*C. perfringens*	CP15	SAMN06252080
7	*C. perfringens*	CPE_F4969	SAMN02436168
8	*C. perfringens*	D_JGS1721	SAMN02436277
9	*C. perfringens*	E_JGS1987	SAMN02436167
10	*C. perfringens*	FORC003	SAMN03140316
11	*C. perfringens*	FORC025	SAMN04209542
12	*C. perfringens*	JFP718	SAMN05323879
13	*C. perfringens*	JFP836	SAMN05323894
14	*C. perfringens*	JJC	SAMN02317206
15	*C. perfringens*	JP55	SAMN03372134
16	*C. perfringens*	JP838	SAMN03377063
17	*C. perfringens*	LLY_N11	SAMN07624709
18	*C. perfringens*	NCTC8239	SAMN02436239
19	*C. perfringens*	13	SAMD00061119
20	*C. perfringens*	2789STDY5608889	SAMEA3545297
21	*C. acetobutylicum*	ATCC824	SAMN02603243
22	*C. baratii*	Sullivan	SAMN03222823
23	*C. botulinum*	ATCC3502	SAMEA1705919
24	*C. butyricum*	KNU-L09	SAMN04293668
25	*C. chauvoei*	JF4335	SAMEA102059668
26	*C. difficile**	630	SAMEA1705932
27	*C. difficile**	BR81	SAMN06286529
28	*C. difficile**	E15	SAMEA3138880
29	*C. difficile**	M120	SAMEA3138368
30	*C. novyi*	NT	SAMN02603464
31	*C. paraputrificum*	LH025	SAMEA104287982
32	*C. pasteurianum*	DSM525	SAMN02990135
33	*C. sordellii*[#]	AM370	SAMN04263499
34	*C. tertium*	LH009	SAMEA104287969
35	*C. tetani*	E88	SAMN02603289

Table 2.12: Genomes used to generate *Clostridium* genus-specific annotation database

*Recently suggested to be moved to genus *Clostridioides* [219]. [#]Recently suggested to be classified as genus *Paeniclostridium* [220].

2.4.6 Pangenome analysis

Pangenome analysis software Roary v3.12 was run with annotated GFF files (from Prokka annotation) as the inputs at BLASTP identity 90% (-i 90) to obtain the pangenome statistics (also with -s option, thus paralogs were not split in the gene clustering) and core gene alignments (-e -n; via MAFFT v7.305) [221]. Core genes were defined as genes present in 99% of the genomes (-cd 99; or default setting). Similarly, accessory genes were defined as genes not present in 99% of the computed genomes. Pangenome statistical plots were generated using R scripts from Roary package and visualised in Phandango [222].

2.4.7 Phylogenetic and SNP analysis

Core genome trees were constructed based on core gene (genes present in 100% of the computed genomes) alignments generated from Roary pipeline by concatenating all single-copy core-gene sequences and aligned via MAFFT v7.305 [221,223]. SNPs within core-genome (multi-FASTA alignment format) were extracted using SNP-sites v2.3.3 [224]. SNP distances were computed using snp-dists v0.2 based on multi-FASTA alignment files[5]. Snippy v4.0 was used for reference-based (read-mapping) SNP analysis. FastTree v2.1.9 was used to construct Maximum-likelihood tree using generalised time-reversible (GTR) model on 1000 bootstrap replicates [225]. Trees were visualised, rooted and manually annotated using iTOL v3 [226]. Phylogenetic clustering was assigned via hierBAPS analysis (as described in section 2.10) [227]. Binary trees were generated based on gene-presence-absence matrix outputs from Roary and parsed with in-house Python scripts to construct binary nucleotide alignments. Phylogenies were constructed as described.

2.4.8 Genome-wide sequence search

Genome-wide screening of genes was performed using BLASTN-based command-line tool ABRicate v0.5[6], with 90% identity and 80% coverage minimum cutoffs to infer identical genes (best-match approach). Gene sequences of interest were retrieved from public databases, primarily NCBI Nucleotide database and UniProt database, to build the screening BLASTN database using *makeblastdb* command. ARIBA v2.8.1

[5] https://github.com/tseemann/snp-dists

[6] https://github.com/tseemann/abricate

(assembly-based) was also used as a secondary approach to confirm detections of genes in raw sequencing FASTQ files [228].

Databases used for gene screening include: toxin genes (custom, n=22; Table 2.13), AMR genes (CARD database v2.0.0, n=2239 [229]), putative virulence genes including collagen binding protein and enterotoxin genes (custom, n=8; Table 2.14), Tcp conjugative loci and insertion sequence (IS) elements (custom, n=14; Table 2.15). Gene *cnaD* was recently associated with chicken NE and putatively identified to encode collagen binding protein [230].

	Gene name	Product	Alternative name	NCBI Accession
1	*plc/cpa*	Phospholipase	Alpha-toxin	D63911.1
2	*cpe*	Enterotoxin CPE	-	M98037.1
3	*ccp*	Clostripain	-	NC_003366.1
4	*colA*	Microbial collagenase	Kappa-toxin	D13791.1
5	*nanI*	Sialidase NanI	-	NC_003366.1
6	*nanJ*	Exo-alpha-sialidase NanJ	-	NC_003366.1
7	*nanH*	Neuraminidase nanH	-	Y00963.1
8	*cpb*	Beta-toxin	-	KP064410.1
9	*cpb2* variant 1	Beta-2 toxin variant 1	-	AP003515.1
10	*cpb2* variant 2	Beta-2 toxin variant 2	-	CP009558.1
11	*pfo*	Perfringolysin O	Theta-toxin	BA000016.3
12	*nagH*	Hyaluronidase NagH	Mu-toxin	NC_003366.1
13	*etx*	Epsilon-toxin	-	M95206.1
14	*iap*	Iota-toxin component Ia	-	NC_015712.1
15	*ibp*	Iota-toxin component Ib	-	NC_015712.1
16	*netB*	Pore-forming toxin NetB	-	FJ189503.1
17	*becA*	Binary enterotoxin component a	-	NC_023918.1
18	*becB*	Binary enterotoxin component b	-	NC_023918.1
19	*tpeL*	Peptidase	-	EU848493.1
20	*cpd*	Delta-toxin	-	EU652406.1
21	*netF*	NE Toxin NetF	-	KJ606986
22	*lam*	Lambda toxin	-	AJ439340

Table 2.13: Accessions of *C. perfringens* toxin gene sequences

	Gene name	Product	NCBI Accession
1	*cna*	Collagen adhesion protein	BAB62495.1
2	*cnaA*	Collagen adhesion protein A	ALJ54440.1
3	*cnaC*	Collagen adhesion protein C	DQ366035.1
4	*cnaD*	Collagen adhesion protein D	CP000246.1
5	*entA*	Putative enterotoxin A	ABG85708.1
6	*entB*	Putative enterotoxin B	ABG85644.1
7	*entC*	Putative enterotoxin C	BAB80858.1
8	*entD*	Putative enterotoxin D	ABG83751.1

Table 2.14: Accessions of additional virulence-associated genes

	Gene name	Product	Protein function	NCBI Accession
1	*tcpA*	TcpA	Conjugative locus	ABF47308.1
2	*tcpB*	TcpB	Conjugative locus	ABF47311.1
3	*tcpC*	TcpC	Conjugative locus	ABF47313.1
4	*tcpD*	TcpD	Conjugative locus	ABF47314.1
5	*tcpE*	TcpE	Conjugative locus	ABF47315.1
6	*tcpF*	TcpF	Conjugative locus	ABF47316.1
7	*tcpG*	TcpG	Conjugative locus	ABF47318.1
8	*tcpH*	TcpH	Conjugative locus	ABF47319.1
9	*tcpI*	TcpI	Conjugative locus	ABF47320.1
10	*tcpJ*	TcpJ	Conjugative locus	ABF47321.1
11	IS*1470*	IS1470 transposase	Insertion sequence element	ABG85734.1
12	IS*1470*-like	IS1470-like transposase	Insertion sequence element	BAE79022.1
13	IS*1469*	IS1469 transposase	Insertion sequence element	YP_473486.1
14	IS*1151*	IS1151 transposase	Insertion sequence element	YP_473484.1

Table 2.15: Accessions of conjugative loci and insertion sequence elements

2.4.9 Bacterial pan-GWAS: gene-trait association analysis

Scoary v1.6 was used to infer specific gene-trait statistical association based on the Roary output gene-presence-absence matrix [231]. Script scoary.py was run with default parameters with an additional 10 permutations.

2.4.10 Plasmid *in silico* analysis

Sequencing reads were utilised for computational plasmid prediction via software PlasmidSeeker v1.0 [232]. Plasmid prediction was based on 8,514 plasmid sequences available on NCBI RefSeq databases (including 35 *C. perfringens* plasmids,

downloaded in April 2017; Table 2.16) [7]. All sequencing reads (single FASTQ) were searched for matching k-mers at k-mer length of 20 and screening cutoff at P-value 0.05 based on FASTQ reads. Top predicted plasmids appeared in each 'cluster' (with highest k-mer identity; k-mer percentage ≥80% is the minimum cutoff) was extracted as predicted plasmids.

Potential plasmid sequences from high-sequencing-coverage assemblies (unassembled single contig) were extracted using in-house Perl scripts. Plasmid annotation (GenBank format) was generated via Prokka v1.13 and Artemis v17 [233]. Plasmid comparison and visualisation were performed using Python tool Easyfig v2.2.2 [234].

No	Plasmid	No	Plasmid
1	pCP-TS1	19	pJFP838C
2	pCP8533etx	20	pJFP838A
3	pCP-OS1	21	pCP13 DNA
4	pCPPB-1	22	pCPF4969
5	pCW3	23	pJIR3844
6	pIP404	24	pJIR3843
7	pJIR3536	25	pJIR3537
8	pCPF5603	26	pCpb2-CP1
9	pBCNF5603	27	pNetB-NE10
10	pCP8533S12	28	pDel1_1
11	pFORC3	29	pDel1_2
12	pJFP55F	30	pDle1_3
13	pJFP55G	31	pDel1_4
14	pJFP55H	32	pCP15_1
15	pJFP55J	33	pCP15_2
16	pJFP55K	34	pCP15_3
17	pJFP838E	35	pCP15_4
18	pJFP838D		

Table 2.16: Plasmid sequences of *C. perfringens* included in the *PlasmidSeeker* database

[7] http://bioinfo.ut.ee/plasmidseeker/

2.4.11 Bacteriophage *in silico* detection

Web tool PHASTER is utilised for detection of bacteriophage species existing in bacterial genomes. Annotated GenBank files were submitted to the PHASTER web server (either manual submission or via API server)[8] and annotated data were parsed with in-house scripts, visualised with heatmaps generated within R programming environment. The detection of phage species is based on the PHASTER scoring method and classification as follows: (1) intact: >90; (2) questionable: 70-90; (3) incomplete: <70 [235]. Prophage with scores >70 will be considered as complete prophage, whereas 'incomplete' prophage will be specified wherever mentioned. Sequences of prophages were extracted by default using PHASTER, manually annotated and colour-coded using Prokka v1.13 and Artemis v17 for visualisation in EasyFig v2.2.2. Phage-related genes in the pangenome were obtained using string-search based on 'phage', 'head', 'tail', 'integrase', 'recombinase', 'terminase' and 'capsid'.

2.4.12 CRISPR array mining

MinCED[9] (v0.2.0), a command-line software derived from CRISPR Recognition Tool (CRT) [236], was used to perform rapid CRISPR-search in genome assemblies at default parameters. Outputs were parsed using BASH text-processing commands to generate CRISPR counts for each genome. CRISPR counts were visualized in bar charts generated using R package ggplot2 and function geom_bar.

2.4.13 Pangenome-wide functional annotation

Functional categories (Clusters of Orthologous Groups [COG] categories) were assigned to genes for biological interpretation via eggNOG-mapper v0.99.3 based on EggNog database (bacteria) on default parameters [237,238].

2.4.14 Linear discriminant analysis (LDA)

LDA was performed using LEfSe (Linear discriminant analysis Effect Size) via Galaxy server[10]. Alpha value for non-parametric Kruskal-Wallis test was set at 0.05

[8] http://phaster.ca/

[9] https://github.com/ctSkennerton/minced

[10] http://huttenhower.sph.harvard.edu/galaxy/

and threshold on LDA score at 2.5 for statistical significance. Graph was illustrated using the LEfSe plotting module [239].

2.4.15 Nucleotide sequence accessions

All genome assemblies were submitted to ENA (European Nucleotide Archive) as in Table 2.17. Project accession details for individual isolate was summarised in Appendix 3.

Study	Primary Accession	Secondary Accession	Genome no.
NEOM isolate assemblies (Chapter 3)	PRJEB25762	ERP107717	185
PHE isolate assemblies (Chapter 4)	PRJEB25764	ERP107719	109
BAMBI (Chapter 6) and Lincoln (Chapter 5)	PRJEB25766	ERP107721	83
NCTC long-read assemblies (Chapter 6)	PRJEB28310	ERP110497	25
Zoo microbiome study (Chapter 6)	PRJEB28310	ERP110496	17
ELDERMET study (Chapter 6)	PRJEB28308	ERP110495	17
2-year NEOM study (Chapter 6)	PRJEB28307	ERP110494	25
		Total	**461**

Table 2.17: Accessions (ENA) for all newly-sequenced genomes described in this thesis
NCTC (PacBio-sequenced; NCTC3000 project) genomes were newly assembled in my work.

2.5 *In vitro* methods

2.5.1 Bacterial isolation

2.5.1.1 Direct plating

Direct plating method for isolating *C. perfringens* is an effective yet reliable approach [11]. Firstly, faecal samples were weighed in sterile Eppendorf tubes (approx. 100mg/sample) and then serially diluted in sterile Phosphate Buffer Saline (PBS) - neat, 10-fold, 100-fold, 1000-fold and 10000-fold. The mixture contents were homogenised using a horizontal mixer and vortexed, and plated (100µl/ sample/ plate) on selective media Tryptose-Sulfite-Cycloserine Egg Yolk Agar (TSC-EYA; Table 2.18) [11,240].

Composition	Concentration
Perfringens Agar Base	46g/L
D-cycloserine	400mg/L
Egg yolk emulsion	100ml/L

Table 2.18: Composition of TSC-EYA agar used for primary isolation of *C. perfringens*

After 18-24h anaerobic incubation at 37°C, pitch-black colonies with opaque halos (halos represent lecithinase activity of *C. perfringens*, which morphology was presumptively identified as *C. perfringens*) were re-streaked with sterile culture loops on fresh TSC-EYA agar plates to obtain pure colonies and stocks (see Figure 2.3). After 2-3 rounds of re-streaking of isolates, pure colonies were obtained.

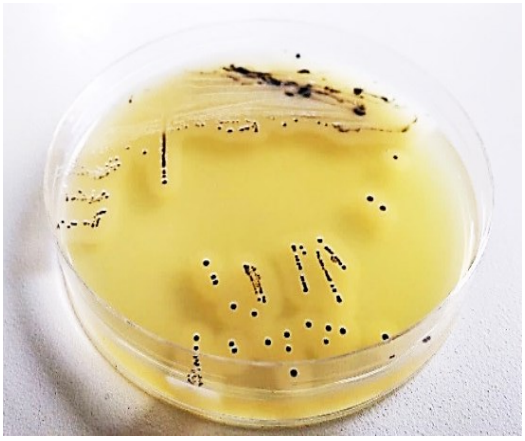


Figure 2.3: Distinctive *C. perfringens* colonies with opaque halos on TSC-EYA.

2.5.1.2 Ethanol-shock method

30 min-ethanol treatment (50% ethanol in Robertson's Cooked Meat Media broth) to the collected faecal samples at room temperature, aimed to eliminate non-spore formers, followed by plating on autoclaved Fastidious Anaerobic Agar (FAA) supplemented with defibrinated sheep blood and 0.1% sodium taurocholate. After 48h anaerobic incubation (in anaerobic jar), colonies with haemolytic characteristic growth on agar were presumptively identified as *C. perfringens*, which were subjected to further identification.

Composition	Concentration
Fastidious Anaerobic Agar	46g/500ml
Defibrinated sheep blood (0.5%)	25ml/500ml
Sodium taurocholate (0.1%)	0.5g/500ml

Table 2.19: Composition of FAA for isolation of *C. perfringens*.

2.5.2 Molecular identification of *C. perfringens*

2.5.2.1 16S rRNA PCR and sequencing

Genomic DNA of isolates were extracted (section 2.2.1) and 16S rRNA PCR (full 16S gene) was performed. This method is described in section 2.3.1.

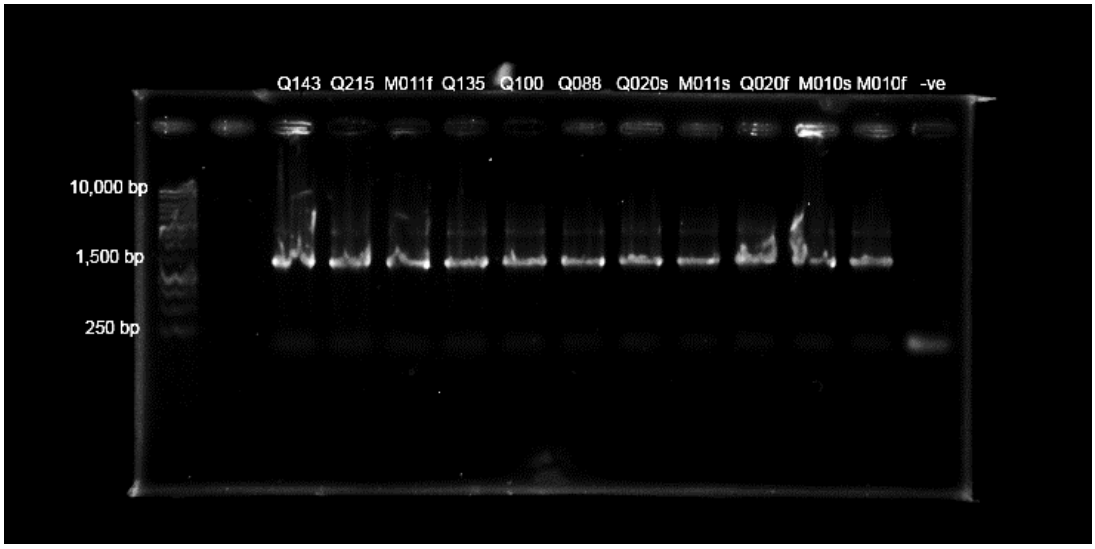


Figure 2.4: Agarose gel showing full-length 16S rRNA gene amplicons
16S rRNA full-length amplicons were approximately 1,500 bp as shown.

2.5.2.2 Matrix-Assisted Laser Desorption Ionisation Time-of-Flight (MALDI-TOF)

All samples provided by Prof J S Kroll were identified by MALDI-TOF, performed according to manufacturer's instructions using Bruker Microflex LT by the Microbiology Department at Imperial College Healthcare NHS Trust. For details please refer to p.66 Sim (2015) [12]. Isolates used for phenotyping studies were re-identified (for reaffirmation) using Bruker Microflex at UEA Norwich Medical School with the assistance of Dr Emma Manners.

2.5.2.3 Multiplex-PCR toxinotyping

A sub-set of *C. perfringens* isolates were toxinotyped by Dr Joseph Brown at Lincoln University using multiplex-PCR technique to screen for toxin genes including *plc*, *cpb*, *cpb2*, *etx*, *iap/ibp* and *pfo* [241].

2.5.3 Culturing

2.5.3.1 Pure cultures

All isolates were cultured in sterile BHI for further analysis unless otherwise indicated, including genomic DNA extraction for WGS and phenotyping. Briefly, stocks are streaked on BHI agar and incubated in anaerobic cabinet at 37°C overnight (approximately 15-18h). Single colonies were then picked for further culturing according to specific experimental protocols.

2.5.3.2 BHI Agar

BHI agar was made using BHI supplemented with 1.5% pure agar. BHI agar was the primary agar used in phenotyping studies.

2.5.3.3 TSC-EYA Agar

Isolates were streaked on TSC-EYA for brief identification (<18h overnight incubation, Figure 2.3).

2.5.3.4 Bacterial culture stocks

All bacterial culture stocks were preserved using autoclaved RCM with glycerol (30%, V/V) in sterile cryogenic storage vials and stored long-term at -80°C until further analysis.

2.5.4 Colony Forming Unit (CFU) enumeration

Calculation of CFU is performed according to the formula below:

$$\text{CFU/ml} = \text{number of colonies (CFU)} \times \text{dilution factor} \times 50\ (\text{ml}^{-1})$$

For faecal sample,

$$\textbf{CFU/g = number of colonies (CFU) x dilution factor x 50 (g}^{-1}\textbf{)}$$

For calculation of Limit of Detection (LOD),

$$\textbf{CFU/ml = 1 (CFU) x lowest dilution factor x 50 (ml}^{-1}\textbf{)}$$

2.5.5 Cell lines and maintenance

Rat (*Rattus* norvegicus) small intestinal cell line IEC-6 (ATCC CRL-1592™) and human colonic cell line Caco-2 (ATCC HTB-37™) were employed for cell culture experiments.

2.5.5.1 IEC-6

Frozen stocks were resuscitated/resuspended in pre-warmed (37°C) media Dulbecco's Modified Eagle Medium (DMEM, Thermo Fisher Scientific), supplemented with 10% Foetal Bovine Serum (FBS; Thermo Fisher Scientific), 4mM L-glutamine (Thermo Fisher Scientific) and 0.1 U/ml bovine insulin (Sigma-Aldrich). Cells were immediately transferred to T25 sterile culture flasks (25cm^2; Corning) and incubated at 37°C in 5% CO_2 incubator for 24h. Spent media was removed after 24-48h, replaced with fresh warm media as described and incubated until confluency was reached. For making cell line stocks, cells were grown to confluency (70-80%) before trypsinisation (1% trypsin for 10 min at 37°C) and cell density adjusted to approximately 1x10^6 cells/ml. 5% of DMSO (cryoprotectant) was added to 1ml aliquots before transferring to -80°C freezer for 24h. After 24h, stocks were stored to -196°C liquid nitrogen for long-term storage (cell banking). Cell line passages 6 – 20 were used, cells were split at ratios 1:10-20 for each passage. Cells were counted on haemocytometer (Neubauer Chamber).

2.5.5.2 Caco-2

Cells were resuscitated and maintained as described above except the working medium – DMEM supplemented with 20% FBS. Cells were split at ratio 1:5-10. Cell line passages 30-45 were used in experiments.

2.5.6 Growth kinetics assay

2.5.6.1 CFU kinetics: culture setup and drop-plating

Bacterial stocks were checked for purity on BHI agar prior to experiments. Single colonies were picked and cultured in BHI broth anaerobically overnight (approximately 14-16h) to reach stationery phase (approximately 1 x 10^9 c.f.u). Confluent cultures were diluted 1-million-fold (10^6) at the start of the 24-h experiment, estimating 10^3 starting CFU/ml/ well. Growing cultures were homogenised by inverting 5 times before taking 100µl of total liquid for plating.

Serial dilutions were performed with sterile PBS in sterile Eppendorfs before drop-plating using Miles and Misra method on BHI agar [242]. For each time point and sample, 3 technical replicates were drop-plated for statistical accuracy. This experiment was performed in 3 independent cultures (n=3) for each isolate. Plates were incubated at 37°C in anaerobic cabinet for 15-20h before CFU counting. Plates of highest dilutions with CFU were selected for maximum accuracy.

2.5.6.2 Optical density kinetics measurement

200µl of each sample (triplicates) was transferred into sterile 96-well plate anaerobically. Culture setup is described in section 2.5.6.1. Plate reader TECAN was used to monitor the growth kinetics for a continuous 24h. Plate reader was set to shake plates every 15mins for 1min to avoid cell clogging.

2.5.7 Gas production assay

Autoclaved gas vials filled with 30ml sterile BHI were pre-reduced overnight in anaerobic chamber prior to experiment. Overnight cultures were diluted to approximately 10^3 CFU/ml each vial and air-sealed with butyl septum and aluminium crimp, then moved to 37ºC incubator.

Briefly, gas produced by each culture was measured by USB pressure transducer connected to a 3-way stopcock – inserting a 23G hypodermic syringe needle (attached to a 10-ml syringe) into the sealed vial headspace, the gas pressure readouts were indicated on the display (using the recommended OMEGA software on a laptop). By withdrawing the syringe plunger, the headspace pressure returned to ambient pressure as indicated by a zero reading. Gas volume was indicated on the syringe barrel after

the headspace pressure returns to zero [243]. Each reading was then recorded every hour for an extended period of 10h. Measurements were carried out in a fume hood. This experiment was performed in 3 independent cultures (n=3) for each isolate.

2.5.8 Hydrogen sulphide production assay

Overnight pure cultures were inoculated in 20ml fresh sterile BHI in 50ml Falcon tubes at 1% inoculum and followed by 8h anaerobic incubation (negative control *Bifidobacterium longum* was inoculated at 5% and incubated for 20h). H_2S test strips (also known as lead acetate test strips: H_2S detection range 5-400ppm) were attached in each tube to visually determine the amount of H_2S produced by each strain using blackening score (1-10). NaHS was added to BHI to serve as positive control as it produces H_2S when dissolved in liquid (which gives a blackening score of 10). This experiment was performed in 3 independent cultures (n=3) for each isolate.

2.5.9 Oxygen tolerance assay

Pure cultures (12 strains) were grown anaerobically to confluency for 24h in BHI. Cultures were spotted in a dilution series onto BHI agar supplemented with 0.1% sodium taurocholate. Spotted plates were dried and incubated under ambient (aerobic) conditions at room temperature (21ºC) for specified time periods (24h-336h) before being returned to anaerobic chamber for determination of CFU count. All CFU were counted after 12-15h anaerobic incubation. Cultures (spotted plates) that were not exposed to oxygen acted as controls for each strain. Viable percentage was shown after comparing with control cultures. This experiment was performed in 3 independent cultures (n=3) for each isolate.

2.5.10 Sporulation assay

Pure cultures were induced to sporulate using the modified Duncan-Strong Medium (at pH 7.8) as detailed in Table 2.20 for 24h anaerobic incubation unless otherwise indicated [61].

Sporulated cultures were treated with sterile-filtered 70% ethanol for 4h to eliminate all remaining vegetative cells and spotted on BHI agar supplemented with 0.1% taurocholate (a potent germinant) for spores.

Composition	Weight/litre
Yeast extract	4g
Sodium thioglycolate	1g
Disodium phosphate (Na2HPO4.7H2O)	10g
Raffinose	4g
Proteose peptone	15g

Table 2.20: Composition of the modified Duncan-Strong Medium [61].

2.5.11 Bile salt assays

2.5.11.1 Germination response of spores

Ethanol-treated sporulated pure cultures were serially diluted and plated on BHI agar and BHI agar supplemented with 0.1% bile salts (taurocholate, cholate, chenodeoxycholate and deoxycholate), followed by anaerobic incubation for >24h. CFU were counted and fold-change was calculated compared with CFU on BHI plates without bile salt supplementation.

2.5.11.2 Responses of vegetative cells

Pure cultures were incubated anaerobically overnight before serially diluted and plated on BHI agar (control), BHI agar supplemented with 0.1% bile salts (taurocholate, cholate and deoxycholate).

2.5.11.3 Mixed-bile assay

Pure cultures were incubated to confluency overnight before serially diluted and plated BHI agar supplemented with 0.1% taurocholate + 0.1% cholate + 0.1% deoxycholate. CFU were checked after >24 h incubation.

2.5.11.4 Deoxycholate assay

Pure cultures were incubated anaerobically overnight before being reinoculated into BHI broth containing different concentration of deoxycholate at 0.01%, 0.05% and 0.1%.

2.5.12 Minimum Inhibitory Concentration (MIC) assay

The MIC of each antibiotic against *C. perfringens* strains was determined based on the modified method of microdilution technique for antimicrobial susceptibility testing [244]. MIC is defined as the lowest antibiotic concentration that inhibits the visible growth of *C. perfringens* after 24h-incubation.

Sterile-filtered antibiotics were pre-made in stocks at desirable concentration prior to storage at -20°C. Bacterial culturing was described previously. 96-well plate was utilised for this assay – approximately 10^4 CFUs of each strain was added into each well (sterile BHI) containing a different antibiotic concentration (double dilution) accordingly. BHI without antibiotic served as a positive control for growth and supplemented with thiamphenicol at 15μg/ml served as negative control. FLUOstar Omega microplate reader was used to measure liquid turbidity at 595nm (absorbance) to determine whether there is microbial growth in each well after 24h anaerobic incubation. Absorbance >0.5 is considered as significant growth.

2.5.13 Cell line toxicity assay

2.5.13.1 Necrosis assay

IEC-6 cells (passage 6-20) were seeded at 20,000 cells/well in 96-well plate (tissue-treated). After 72h 5% CO_2 incubation, approximately 56,000 cells per well (firm monolayer) were subjected to experiments. Briefly, phenol-red DMEM was removed, cells were washed twice gently and replaced with sterile phenol-red free DMEM supplemented with 1% FBS. Diluted sterile-filtered (0.22μm) bacterial supernatants (in ratio 1:2, 1:4 and 1:8 as necessary) were added into each well at 5% total volume followed by 2h 5% CO_2 incubation. Necrosis measurement was performed using CytoTox-One Homogeneous Membrane Integrity Assay according to manufacturer's instructions.

For Caco-2 cells (passage 30-45), cells were seeded at 20,000 cells/well. After 72h 5% CO_2 incubation, a confluent monolayer was formed. Protocol follows as described in previous paragraph, with an exception that supernatants were added in 25% of total volume.

2.5.13.2 Apoptosis assay

Cell seeding follows the above description. After 3h 5% CO_2 incubation, caspase activity was measured using Caspase-Glo 3/7 Assay according to manufacturer's instructions. Staurosporine (1μM) was used as positive control.

2.5.13.3 Hydrogen sulphide toxicity assay

Sodium hydrosulfide (NaHS) was used as a sulfide donor as an equivalence of H_2S in this study [245]. Concentration of soluble NaHS was prepared ranged from 0.0001% to 1% in PBS, sterile-filtered (0.22μm) and stored at 4ºC prior to experiment. Cell seeding was performed as described, with all NaHS solution being added to working volume (200μl) at 5% each well. Supernatants were harvest after 4h 5% CO_2 incubation. CytoTox-One Homogeneous Membrane Integrity Assay was used to measure potential cell toxicity according to manufacturer's instructions.

2.5.14 PFO expression assay

BHI agar supplied with 5% sheep blood was used to identify PFO expression of *C. perfringens* isolates. Briefly, *C. perfringens* was streaked on the blood agar and incubated anaerobically for overnight (<20h). As PFO is known produce beta-haemolysis in blood-supplemented media (complete lysis of blood cells), therefore isolates with clear transparent halos formed around the colonies (indicating lysis of blood cells) visualised were defined as PFO-positive isolates [246].

2.5.15 Bacterial transformation

Plasmid pMTL8315fdx-gfp+ was the target plasmid (see Figure 2.5 for plasmid map) to be inserted into *C. perfringens* isolates. This modified version of plasmid pMTL8135 with green fluorescent protein (GFP) gene was a gift from Dr Duncan Gaskin [247]. This shuttle plasmid (stored in *E. coli* TOP10 strain as stocks at -80ºC) comprises *catP* gene for selection which confers chloramphenicol resistance in Gram-negative bacteria and thiamphenicol resistance in Gram-positive bacteria. Optimised

transformation protocol [11] (electroporation) was obtained online from Clostridia Research Group at the University of Nottingham and performed accordingly [248].

2.5.16 Plasmid stability test

Plasmid stability test was performed to determine the stability of the specific plasmid in the transformants up to 60 generations. Briefly, transformed isolates were cultured in BHI without antibiotic supplementation (thiamphenicol) to reach confluency every 12h (starting cultures at 1000-fold dilution, 1000-fold increase equates approximately 10 generations). Sub-culturing was carried out every 12h (10 generations) for 6 cycles (60 generations). Cultures were serially diluted every cycle (12h) and plated on both selective (antibiotic-supplemented) and non-selective (without antibiotic) BHI plates. CFU were counted after 15-20h incubation to compare plasmid retention percentage.

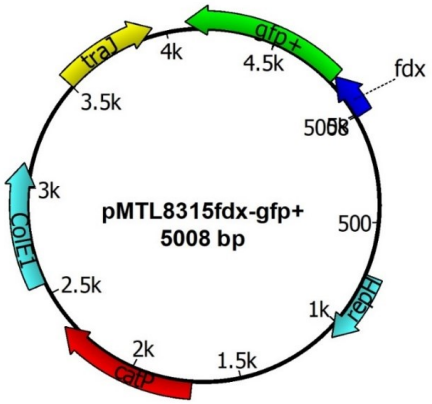


Figure 2.5: Plasmid map of pMTL8315fdx-gfp+

This annotated circular plasmid map was generated using Unipro UGENE v1.23.

2.5.17 Real-time PCR (RT-PCR)

RT-PCR was performed with help from Dr Lukas Harnisch to semi-quantify the relative total microbial load in mouse faecal samples as described previously [249]. Briefly, DNA of faecal samples (similar weight) were extracted and amplified using Quantifast SYBR green mastermix (Qiagen) and full-length 16S rRNA universal primers (Table 2.5). Cycling was performed on a Roche LightCycler 480 using the PCR amplification conditions in Table 2.6. Cycle threshold (CT) value of each sample

[11] http://www.clostron.com/CacElectroporationProtocol.pdf

was compared directly as a relative comparison of total microbial load in faecal samples.

2.6 *In vivo* methods

2.6.1 Ethics and licence

All animal experiments and related protocols described were performed under the Animals (Scientific Procedures) Act 1986 (ASPA) under project licence (PPL: 80/2545) and personal licence (PIL: I7382F677) and approved by Home Office and 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 were euthanised via ASPA Schedule 1 protocol (CO_2 and cervical dislocation). Trained and qualified animal technicians carried out animal husbandry at UEA Disease Modelling Unit (DMU).

2.6.2 Experimental animals and housing conditions

C57BL/6 wild-type female mice (juvenile mice: 3-5 weeks, adult mice: 7-9 weeks), obtained from UEA DMU were used in animal experiments. Animals were bred and housed in DMU barn under specific pathogen-free conditions and moved to DMU infection suite prior to the study. During infection study, mice were housed with autoclaved bedding (and cage), food (stock pellets) and water with 12h light cycle (12h of light and 12h of darkness). Cages were changed in laminar flow cabinet.

2.6.3 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 (i.e. 10^9 CFU/ml, approximately 14-17h). Bacterial pellets were washed and re-suspended in sterile PBS before feeding the mice via oral gavaging using 20G plastic sterile feeding tube.

2.6.4 Faecal sample collection and CFU enumeration

Faecal samples were collected in sterilized tubes daily and stored at -80ºC until further analysis. Serially diluted faecal mixtures were plated on fresh TSC agar and CFU enumerated <24h anaerobic incubation. Pitch black colonies were counted as *C. perfringens* colonies. Black colonies were re-streaked on new BHI plates to confirm morphologies of experimental strains.

2.6.5 Tissue processing: fixation, embedding and sectioning

Intestinal sections were fixed in 10% neutral buffered formalin for <24h and followed by 70% ethanol. Tissues were processed <5 days in 70% ethanol using an automated Leica Tissue Processor ASP-300-S (Quadram Institute Bioscience; QIB) and embedded in paraffin manually. Sectioning was performed using a microtome (5-µm-thick sections; QIB) and left overnight for samples to air-dry prior to staining and further analysis.

2.6.6 H&E staining

H&E (haematoxylin and eosin) staining was performed for structural imaging of intestinal samples using the protocol provided by Dr Zoe Schofield (Table 2.21).

Step no.	Reagent	Time
1	Xylene	2mins
2	Xylene	2mins
3	Xylene	2mins
4	100% ethanol	2mins
5	100% ethanol	2mins
6	90% ethanol	2mins
7	70% ethanol	2mins
8	Running water	2mins
9	Distilled water	30s
10	Haematoxylin	3mins
11	Running water	2mins
12	70% ethanol	2mins
13	Eosin	6mins
14	95% ethanol	2mins
15	100% ethanol	2mins
16	100% ethanol	2mins
17	Xylene	2mins
18	Xylene	2mins
19	Mount slides with DPX	

Table 2.21: H&E staining procedures

2.6.7 Mouse infection model

In-house wild-type C57BL/6 3-4 weeks old female mice (after weaning) were treated with a five-antibiotic cocktail that comprises kanamycin (0.4mg ml^{-1}), gentamicin (0.035mh ml^{-1}), colistin (850U ml^{-1}), metronidazole (0.215mg ml^{-1}) and vancomycin (0.045mg ml^{-1}) in autoclaved drinking water for 3-4 days administered *ad libitum* [34].

Drinking water was then switched to sterile water, and 48h later mice were orally gavaged with 150mg/kg of clindamycin. After 24h, mice were challenged with 10^9 CFU of *C. perfringens* in 100µl and all mice were closely monitored for signs of disease symptoms including significant weight loss (>20%) that will require euthanisation. Tetracycline was supplemented (0.001mg/ml) in drinking water post oral-challenge in tetracycline model.

2.6.8 Pathological scoring method

All colonic section images were examined and graded single-blinded by Dr Lindsay J Hall. The histological severity of infectious colitis was graded using a pathological scoring system (0-14; denoting increasing severity) based on three general pathological features: (1) inflammatory infiltration (0-4), (2) epithelial hyperplasia and goblet cell loss (0-5), and (3) mucosal architecture (0-5) [250]. The overall score was the sum of each component score. An overall pathological score of 1-4 indicates minimal colitis, 5-8 as mild colitis, 9-11 as moderate colitis, above 12 as marked/severe colitis.

2.7 Bright-field and fluorescence microscopy

Bright-field microscopy was performed at QIB Analytical Science Unit, using Olympus BX60 with a microscope camera Jenoptik C10 with ProgRes CapturePro software v2.10. Zeiss Axio Imager 2 (QIB) was used for fluorescence microscopy.

2.8 Electron microscopy

Scanning Electron Microscopy and Transmission Electron Microscopy images were produced by Mrs Kathryn Cross (QIB). Overnight bacterial cultures in BHI were provided and further processed for the EM.

2.9 Graphing and illustrations

Figures were generated using GraphPad Prism v6.0.4 for Windows, GraphPad Software, La Jolla California USA (www.graphpad.com) unless otherwise indicated. R packages (including key graphing packages *gplots* and *ggplot2*) were often used to generate various graphs and figures, and primary R-coding scripts used in this thesis

were made available on my GitHub repository (github.com/raymondkiu/Thesis). RStudio v1.1 for Windows and for Mac were utilised for R programming [210,251]. Several plots created using Perl and Python were indicated in the legends. Microsoft PowerPoint 2016 was frequently used to edit figures and graphs.

2.10 Statistical analyses

All statistics were performed using GraphPad Prism version 6.04 for Windows, GraphPad Software, La Jolla California USA (www.graphpad.com) unless otherwise indicated in figure legends. P values <0.05 are considered as statistically significant. Data expression and statistical tests were indicated in figure legends where appropriate. Principal Component Analysis (PCA) was performed in R using package *ggplot2* and function *autoplot*. Hierarchical population clustering of DNA sequencing data was assigned in R using packages *rhierbaps*, *ggtree,* and *phytools*, and function *hierBAPS*.

Chapter 3 Genomic virulence and phylogenetic analysis of preterm-associated *C. perfringens* identifies NEC-linked extra-virulent genotypes, intra- and inter-hospital transmissibility

3.1 Introduction

This chapter will focus on discussing genomic and phylogenetic analyses of *C. perfringens* isolated from preterm infants, including NEC infants and non-NEC control infants, delving into virulence genes to identify extra-virulent genotypes and tracing genomic epidemiology of *C. perfringens* within and across two hospitals (Figure 3.1). Initial faecal sample collection and *C. perfringens* isolation work were carried out by Dr A Shaw and Dr K Sim (Prof J S Kroll's group at Imperial College London) under NEOM study. I obtained the isolates via collaboration, extracted the genomic DNA (with support from Mr H Bedwell), whole-genome sequenced at WTSI (handled by Dr D Pickard) and further analysed the genomes.

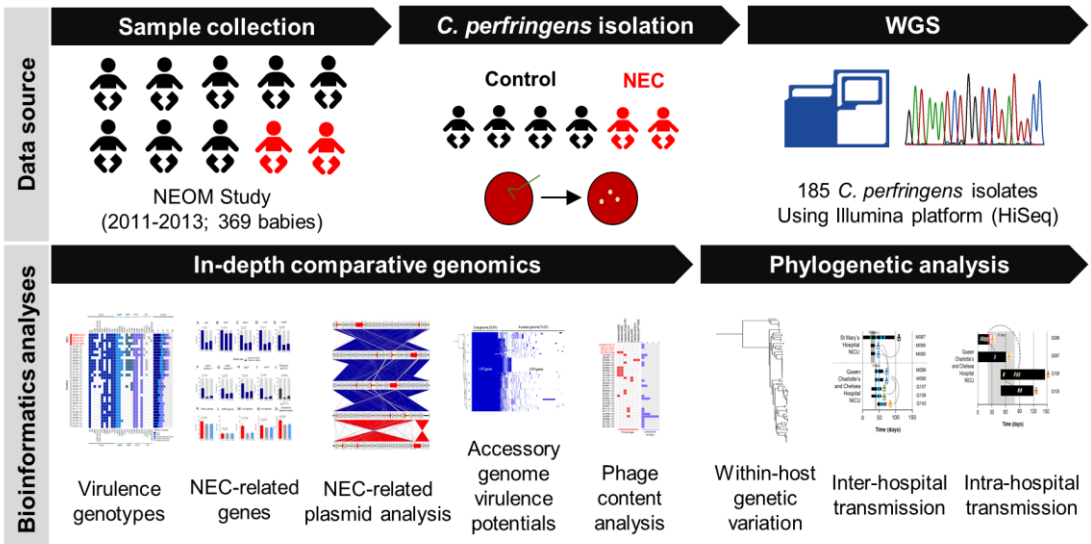


Figure 3.1: Schematic representation of data source and analyses described in Chapter 3

3.2 Background

3.2.1 Preterm NEC

By definition, NEC is an acquired preterm inflammatory gut disease that manifests as tissue necrosis anywhere along the GI tract [73,252]. NEC was first reported in premature infants in the 1820-50s (in Paris and Vienna). After World War II, the emergence of special-care baby units (SCBU) in developed countries facilitated the increasing survival of preterms, which also corresponded to increasing incidence of NEC-like cases and NEC symptoms (first clinically characterised in 1952) [252,253].

Still, it was not until the arrival of modern NICU in the 1960s, that the rapid rise in NEC incidence began to be widely recognised. However, at this time each NEC symptom was categorised under different disease entities. NEC, was first coined in 1964 to address similar disease symptoms under diverse designations including neonatal appendicitis, peritonitis, and pneumatosis intestinalis in infants [188]. After half a century, NEC remains the most severe and lethal neonatal GI emergency worldwide [186,254].

3.2.2 NEC pathogenesis and risk factors

Importantly, the treatment for NEC has been non-specific since the introduction of disease staging criteria i.e. Bell stage. This is perhaps due to our limited understanding of NEC pathogenesis and the view of NEC as a single disease. Gordon et al, recently proposed that NEC is as a disease spectrum – an outcome of several aetiological pathways, instead of a single disease. Preterm-NEC as a clinical endpoint of multiple-pathway disease is represented with associated risk factors in Figure 3.2. Primary risk factors include feeds, antibiotics and microbial infection.

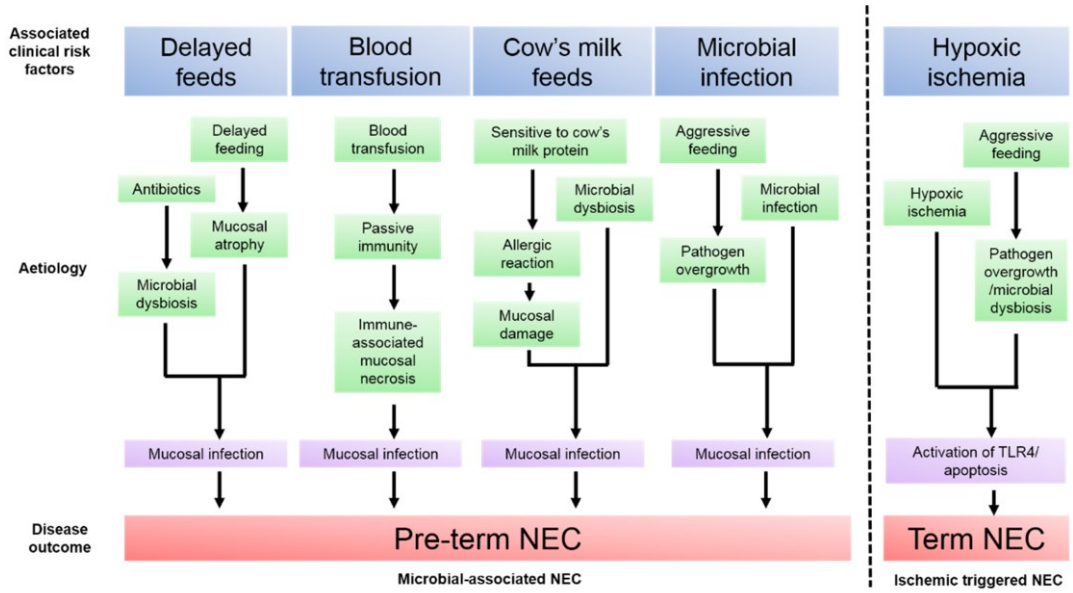


Figure 3.2: Proposed aetiological pathways of both preterm and term NEC development

NEC development was proposed as a disease spectrum. This figure was adapted from Gordon et al. (2012) [255].

3.2.3 Clinical features

NEC is commonly characterised by intestinal necrosis primarily affecting the ileum, though it can affect any part of the GI tract. There are several classic symptoms of NEC: feeding intolerance, abdominal distention, pneumatosis intestinalis (formation of gas cysts in the gut wall), portal venous gas and gas-filled bowel necrosis. Preterm infants with NEC often deteriorate rapidly, frequently within 24 h from initial signs to severe conditions that require surgical operation or cessation of vital signs [73].

3.2.4 Staging criteria and treatment strategy

Presentation of NEC typically occurs in the first 1-4 weeks after premature birth [192]. A classification system, Bell's staging criteria (or more recently, modified Bell's staging criteria as in Table 3.1), has been developed to diagnose, stage and treat NEC in NICUs since the 1970s [73,256,257]. These staging criteria encompass three major stages (stages I, II and III). Clinically, as Bell stage I covers general/non-specific signs, only Bell stage II and III are considered as definite NEC.

3.2.5 The roles of gut microbiota

The role of bacteria in the pathogenesis of NEC has been strongly implicated since the disease was first described [188]. Notably, NEC does not occur in germ-free (i.e.

microbial-free) environments including germ-free *in vivo* models [258], thus NEC has been postulated to be driven by a microbial infectious component. Ultra-rapid progression of NEC from non-specific signs to systemic shock also implies bacterial overgrowth especially quick-growing pathogens including *Klebsiella* spp. and *C. perfringens*, and both have been linked to clinical NEC in recent years [8,190,192,259].

Importantly, opportunistic gut pathogens like *C. perfringens* have been isolated from neonates (70%) in the first few days of life, indicating that NEC may have a microbial infection component [260]. Hallmark NEC symptoms, including pneumatosis intestinalis, also suggests gas-producing bacteria including *Clostridium* spp. as one of the key causative agents [261-264]. Gram-negative bacteria (e.g. *K. pneumoniae*) have also been associated with NEC cases. This has been linked to the fact that Toll-like receptor 4 (TLR4, sensor for the cell wall component lipopolysaccharide [LPS]) is elevated in preterm infants, which may account for the inflammatory cascade induced in the premature intestine prior to NEC onset [265].

Other pathogens such as *Cronobacter sakazakii* have also been isolated from certain NEC 'outbreaks', which again emphasises the potential bacterial role in NEC onset [266]. Importantly, antibiotic-associated disruptions in the preterm gut microbiota may negatively impact on colonisation resistance mechanisms, potentially allowing pathogen invasion and favouring adherence and proliferation of resilient pathogens [267,268]. Preterm infants harbour more potentially pathogenic microbes, with significantly reduced levels of beneficial groups of microbes such as *Bifidobacterium*, which are present at very high levels in vaginally delivery full term breast-fed infants. Indeed, previous studies, and ongoing studies in the Hall lab, have indicated that supplementation of preterm infants with *Bifidobacterium* spp., reduce NEC incidence [269-271].

Stage	Classification	Systemic signs	Intestinal signs	Radiologic signs
IA	Suspected NEC	Temperature instability, apnoea, bradycardia, lethargy	Increased pre-gavage residuals, mild abdominal distention, emesis, guaiac-positive stool	Normal or intestinal dilation, mild ileus
IB	Suspected NEC	Same as above	Bright red blood from rectum	Same as above
IIA	Definite NEC – mildly ill	Same as above	Same as above, plus absent bowel sounds, with or without abdominal tenderness	Intestinal dilation, ileus, pneumatosis intestinalis
IIB	Definite NEC – moderately ill	Same as above, plus mild metabolic acidosis, mild thrombocytopenia	Same as above, plus absent bowel sounds, definite abdominal tenderness, with or without abdominal cellulitis or right lower quadrant mass	Same as IIA, plus portal venous gas, with or without ascites
IIIA	Advanced NEC – severely ill, bowel intact	Same as IIB, plus hypotension, bradycardia, severe apnoea, combined respiratory and metabolic acidosis, disseminated intravascular coagulation, and neutropenia	Same as above, plus signs of generalized peritonitis, marked tenderness, and distention of abdomen	Same as IIB, plus definite ascites
IIIB	Advanced NEC – severely ill, bowel perforated	Same as IIIA	Same as IIIA	Same as IIB, plus pneumoperitoneum

Table 3.1: Modified Bell's staging criteria for preterm NEC

This table was adapted from Lee and Polin [272].

Despite the strong evidence linking bacterial infection in the pathogenesis of NEC, no single common pathogen has been consistently found/isolated in the past 50 years [273]. Instead, more than 12 species of potentially NEC-linked pathogens were summarised in Coggins, et al. [270] (Table 3.2). This indicates that microbial-NEC should not be considered as a single-pathogen-cause disease, but as a disease with a multi-pathogen aetiology (which may also link to specific disease symptoms) [262].

	Genus	Species	Refs
Gram-positive	*Clostridium*	*Clostridium butyricum*	274
		Clostridium difficile	275
		Clostridium perfringens	8,9
		Clostridium neonatale	276,277
	Enterococcus		278
	Staphylococcus	*Staphylococcus aureus*	279
		Staphylococcus epidermidis	280
Gram-negative	*Cronobacter*	*Cronobacter sakazakii*	266,281
	Escherichia	*Escherichia coli*	282
	Bacteroides	*Bacteroides dorei*	9
	Klebsiella		8
	Pseudomonas	*Pseudomonas aeruginosa*	283
	Salmonella	*Salmonella enteritidis*	284

Table 3.2: Pathogenic bacteria associated with preterm NEC

This table was adapted from Coggins et al. [270]

3.2.6 *C. perfringens*-associated NEC

Most recently, several extensive preterm infant NEC cohort studies have again implicated that pathogenic bacteria are involved in the onset of NEC based upon analysis of the faecal gut microbiota using NGS techniques [8,9]. Importantly, *C. perfringens*, has been shown to significantly increase in the infant gut microbiota prior to NEC development in these metagenomics studies. These studies are supported by the substantial number of NEC studies [270] (>14 to date since 1970s, more than any individual bacterial agents; see Table 1.5) that have reported *C. perfringens* as a potential pathogenic agent, in addition to its established pathogenic biology in many other neonatal animal gut diseases, and further supported by *in vivo* studies [189,285]. Thus, *C. perfringens* appears to be a strong candidate as a direct causative agent of microbial-NEC, which may be termed as *C. perfringens*-associated NEC [190].

Being universal and resilient in habitable niches, spore former *C. perfringens* may readily pass on to in-hospital neonates through either environmental or oral transmission, and then initiate diseases in the intestine [260]. Proposed underlying mechanisms for *C. perfringens*-associated NEC include reduced bowel peristalsis in preterm infants (increased pathogen retention time), universal administration of broad-

spectrum antibiotics [286,287] (that leads to reduced microbiota diversity, and the potential rapid overgrowth of resistant spores), reduced gut barrier integrity (non-existent mucus layer), lack of protective bacteria [8,9,186] (including *Bifidobacterium* species, which may antagonise the growth of *C. perfringens*), and immature/underdeveloped immune system that is ineffective at fighting off pathogenic bacteria [73] (Figure 3.3). Bile salt factors may also be linked to disease onset. Reduced diversity of the microbiota in preterm infants, including bacteria that can de-conjugate bile salts to secondary bile acids, may allow germination of *C. perfringens,* which has also been directly indicated in the pathogenesis of the well-studied *C. difficile* colitis (both species form spores and secrete colitis-related toxins) [288]. Finally, the presence of 'hyper-virulent' strains of *C. perfringens* may be implicated.

Presently no particular toxin(s) secreted by *C. perfringens* has been specifically linked with preterm-NEC, although CPB2 has been suggested [8]. Notably, *C. perfringens* can also be part of the 'healthy' human microbiota, thus suggesting other factors play a role [194].

Importantly, no NEC-associated preterm infant isolates are known to be sequenced at present, and genomic investigation may allow identification of NEC-associated virulence factors, that may facilitate intervention strategies and/or therapy development.

Using various computational pipelines and approaches (see section 2.4), I assembled *C. perfringens* genomes and performed downstream analyses including constructing pangenome, annotating gene functions, screening genes of interest (e.g. toxin and AMR genes), analysing 16S rRNA metagenomic taxa abundance, phylogenetic analysis to track bacterial transmission and phage analysis. The work in this chapter will aim to fill up the knowledge gap in understanding the role of *C. perfringens* in preterm and NEC diseases via in-depth genomic analysis.

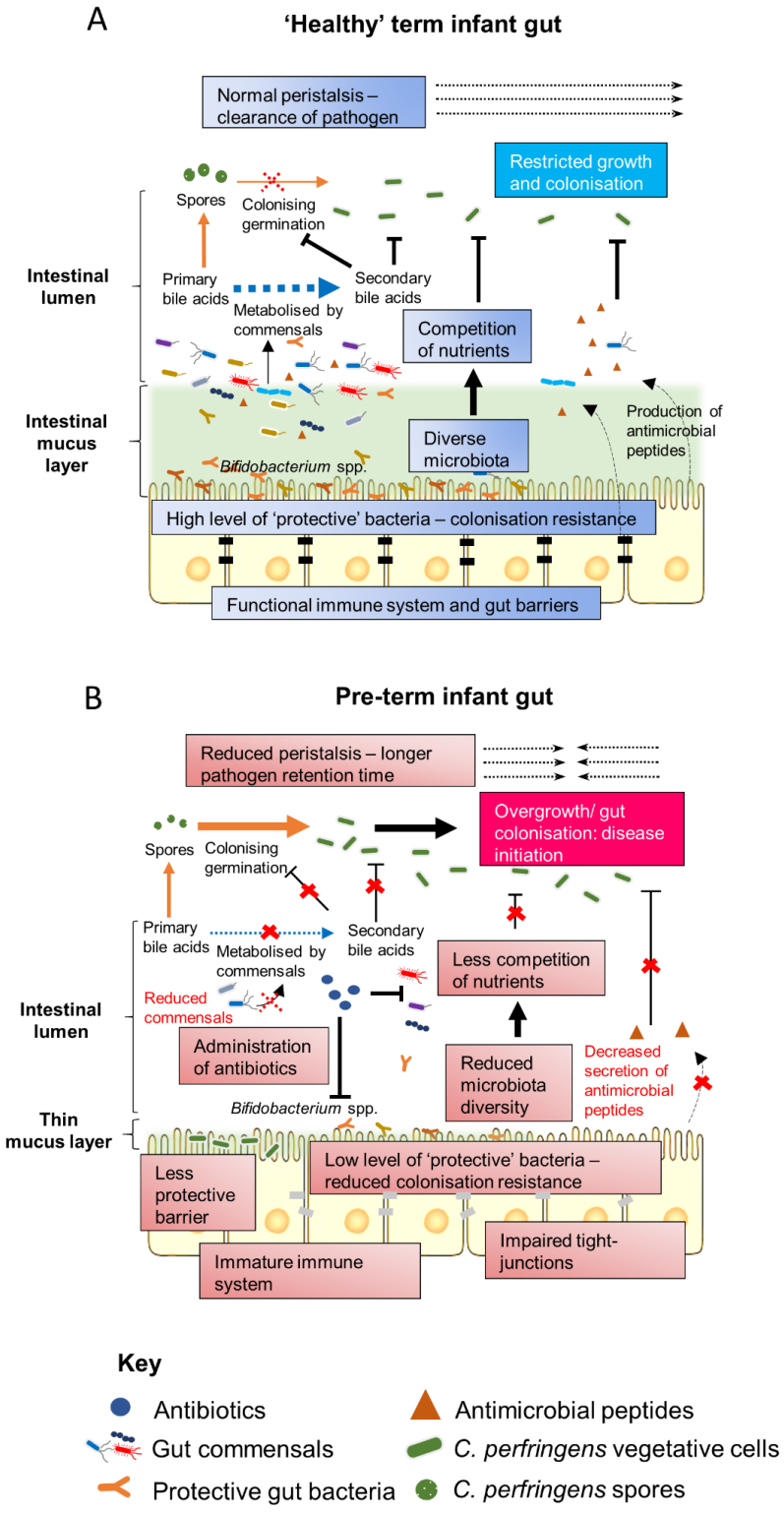


Figure 3.3: Proposed infection mechanisms underlie *C. perfringens*-associated NEC

(A) In non-NEC 'healthy' term infant gut **(B)** In preterm infant gut that potentially leads to *C. perfringens*-associated NEC. Figure reproduced from Kiu and Hall (2018) under the Creative Commons BY licence [1].

3.3 Hypothesis and aims

Specific nosocomial virulent genotypes of *C. perfringens* are linked to preterm-NEC disease progression, specifically to determine:

(1) Genomic virulence features in NEC and control *C. perfringens* isolates obtained from preterm population, including toxin genes, AMR genes, plasmids and phage contents using comparative genomics.

(2) Genomic epidemiology of preterm-associated *C. perfringens* using phylogenetic approaches.

3.4 Study population

The work described in this chapter is based on the preterm gut microbiome data (16S rRNA metagenomics, sequenced by Prof J S Kroll's group at Imperial College London) and *C. perfringens* isolates collected during a 2-year preterm cohort study NEOM [12], a large-scale multi-centre study which was carried out between 2011-2013 in primarily Queen Charlotte's and Chelsea Hospital (QCCH) and St Mary's Hospital (SMH). These isolates were then obtained and prepared at QIB for sequencing at WTSI. Samples include 6 NEC Bell 2/3 infants, 3 NEC Bell 1 infants and 29 non-NEC infants at longitudinal time points.

These 88 newly-sequenced preterm-associated *C. perfringens* isolates were obtained from 38 individual preterm infants at multiple-time points (most individuals). Two public genome assemblies were used as reference genomes in the analysis (ATCC13124 and NCTC8239).

Among these 88 isolates, 60 isolates were obtained from QCCH (24 in 2011 and 36 in 2012), where important surgeries were carried out while 28 isolates were obtained from SMH (6 in 2011 and 22 in 2012) as shown in Figure 3.4A. All genome assemblies analysed were of high quality at <60 contigs (range: 11-59; mean contig number: 25), with 95.5% of the assemblies <40 contigs (Figure 3.4B). All isolates have a mean sequencing coverage (or, depth) of 142X (depth range: 82-206X), mean genome size of 3.2 Mb (range: 3.0-4.1 Mb), 3,009 genes per genome (range: 2,803-3,998 genes) and ANI> 96.0% (pair-wise comparison for all genomes; Appendix 1 Table S3.2). For

statistics on NEC and control individuals, 9 infants were diagnosed with NEC during their hospitalisation period (6 of Bell 2/3; 3 of Bell 1) and 29 were control infants, or non-NEC infants. In terms of isolate number, 24 isolates including multiple-time-point isolations, were obtained from NEC-infants whereas 64 isolates were isolated from control infants (Figure 3.4C-D).

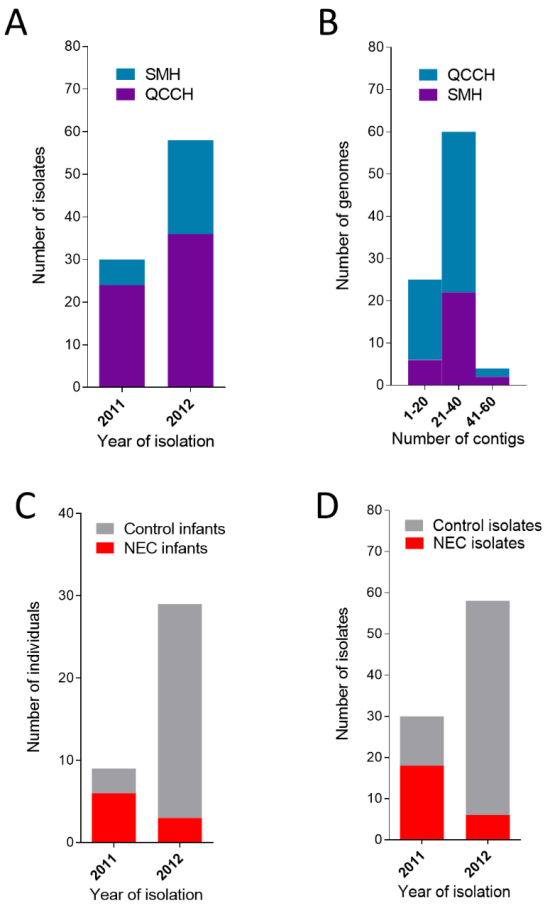


Figure 3.4: Statistics of preterm-study cohorts and isolates

(A) Temporal distribution of isolates from two hospitals SMH and QCCH in 2011 and 2012. A total collection of 88 isolates (multiple time-points) from 38 individual *C. perfringens*-positive preterm infants. **(B)** Quality of genome assemblies based on contig number. **(C)** Distribution of NEC case and control individuals in this study. **(D)** Distribution of NEC case and control isolates.

I also obtained *C. perfringens* isolates from 10 twin-pairs and 1 triplets (remainder singleton) for the analysis (Table 3.3). All relevant clinical metadata of preterm individuals where isolates were obtained from, including admission hospital and delivery mode, are shown in Table 3.3. Clinical metadata (Table S3.1) and genome assembly statistics (Table S3.2) of each isolate are detailed in Appendix 1.

	Infant ID	NEC/ Control	Delivery mode	Admission Hospital	Births	Matching twin/triplet
1	M009	Control	CS	SMH	Singleton	
2	M010	NEC (Bell I)	CS	SMH	Twin	M011
3	M011	NEC (Bell I)	CS	SMH	Twin	M010
4	M030	NEC (Bell I)	CS	SMH	Twin	M029
5	M087	Control	CS	SMH	Twin	M088
6	M088	Control	CS	SMH	Twin	M087
7	M089	Control	CS	SMH	Twin	M090
8	M090	Control	CS	SMH	Twin	M089
9	M091	Control	NVD	SMH	Triplet	
10	M092	Control	NVD	SMH	Singleton	
11	M094	Control	NVD	SMH	Singleton	
12	M096	Control	NVD	SMH	Singleton	
13	M111	Control	CS	SMH	Twin	M112
14	M112	Control	CS	SMH	Twin	M111
15	M124	Control	CS	SMH	Twin	M125
16	M125	Control	CS	SMH	Twin	M124
17	Q020	NEC (Bell II/III)	NVD	QCCH	Singleton	NA
18	Q087	Control	CS	QCCH	Twin	Q088
19	Q088	NEC (Bell II/III)	CS	QCCH	Twin	Q087
20	Q090	Control	NVD	QCCH	Singleton	
21	Q096	Control	CS	QCCH	Twin	Q097
22	Q097	Control	CS	QCCH	Twin	Q096
23	Q100	NEC (Bell II/III)	NVD	QCCH	Singleton	
24	Q101	Control	CS	QCCH	Singleton	
25	Q105	Control	CS	QCCH	Triplet	Q106, Q107
26	Q106	Control	CS	QCCH	Triplet	Q105, Q107
27	Q107	Control	CS	QCCH	Triplet	Q105, Q016
28	Q118	Control	CS	QCCH	Singleton	
29	Q126	Control	CS	QCCH	Singleton	
30	Q135	NEC (Bell II/III)	CS	QCCH	Singleton	NA
31	Q137	Control	NVD	QCCH	Twin	Q138
32	Q138	Control	NVD	QCCH	Twin	Q137
33	Q142	Control	NVD	QCCH	Twin	Q143
34	Q143	NEC (Bell II/III)	NVD	QCCH	Twin	Q142
35	Q164	Control	NVD	QCCH	Singleton	
36	Q167	Control	NVD	QCCH	Twin	Q168
37	Q168	Control	NVD	QCCH	Twin	Q167
38	Q215	NEC (Bell II/III)	CS	QCCH	Twin	Q216

Table 3.3: Clinical metadata of 38 individual preterm infants

CS: Caesarean section; NVD: Normal vaginal delivery; SMH: St Mary's Hospital; QCCH: Queen Charlotte's and Chelsea Hospital.

3.5 Results

3.5.1 NEC microbiome signatures

During the two-year NEOM study (2011-2013), 12/369 individuals were diagnosed with definite NEC. Further 16S rRNA sequencing analysis of the NEC-associated microbiomes revealed 2 distinct microbial signatures: *Clostridium* was statistically

overrepresented in 4 samples whereas *Klebsiella* genus was enriched in the remaining 8 samples prior to definite-NEC onsets [8,12]. Of note, clinical syndromes between these two NEC-signature groups were undistinguished.

Importantly, *C. perfringens* isolation work had been routinely carried out in the NEOM study [12]. With those *Clostridium*-NEC infants, *C. perfringens* was confirmed as the sole *Clostridium* species detected by 16S rRNA sequencing technique via Clostridial isolation method (culturing) in the laboratory [12]. NEC-associated *C. perfringens* were then briefly characterised by toxinotyping schemes and all determined as distinctive-type A strains (only encode typing toxin gene *plc*; different patterns via fAFLP typing), three isolates were found to encode additional *cpb2* gene. As *C. perfringens* was found to be the predominating species in these NEC cases prior to disease onset, it was proposed as the microbial causative agent of NEC [8,12].

I obtained the previously generated raw 16S rRNA microbiome data from Dr A Shaw and re-analysed with the 16S rRNA SILVA database (Release 128; September 2016) using similar OTU taxa assignment approach as in Sim et al. to determine and visualise the longitudinal microbiome dynamics of these *C. perfringens*-positive NEC neonates (Figure 3.5) [8]. The microbiome dynamics 4 *C. perfringens*-associated NEC infants (Q215, Q088, Q143 and Q020) indicate an evident and significant increase of *Clostridium* taxa prior to NEC onset (Figure 3.5A). Two other *Klebsiella*-NEC (with *C. perfringens* isolated during the NEOM study) showed a distinct microbiome profile characterised by increased abundance of *Klebsiella/Citrobacter* genus (they are indistinguishable at 97% nucleotide identity) rather than *Clostridium* (Figure 3.5B). Three suspected-NEC neonates (M010, M011 and M030, data not published previously; Bell I neonates) display diverse gut microbiome dynamics, although M010 appears have increased abundance of *Clostridium* before suspected-NEC diagnosis (Figure 3.5C).

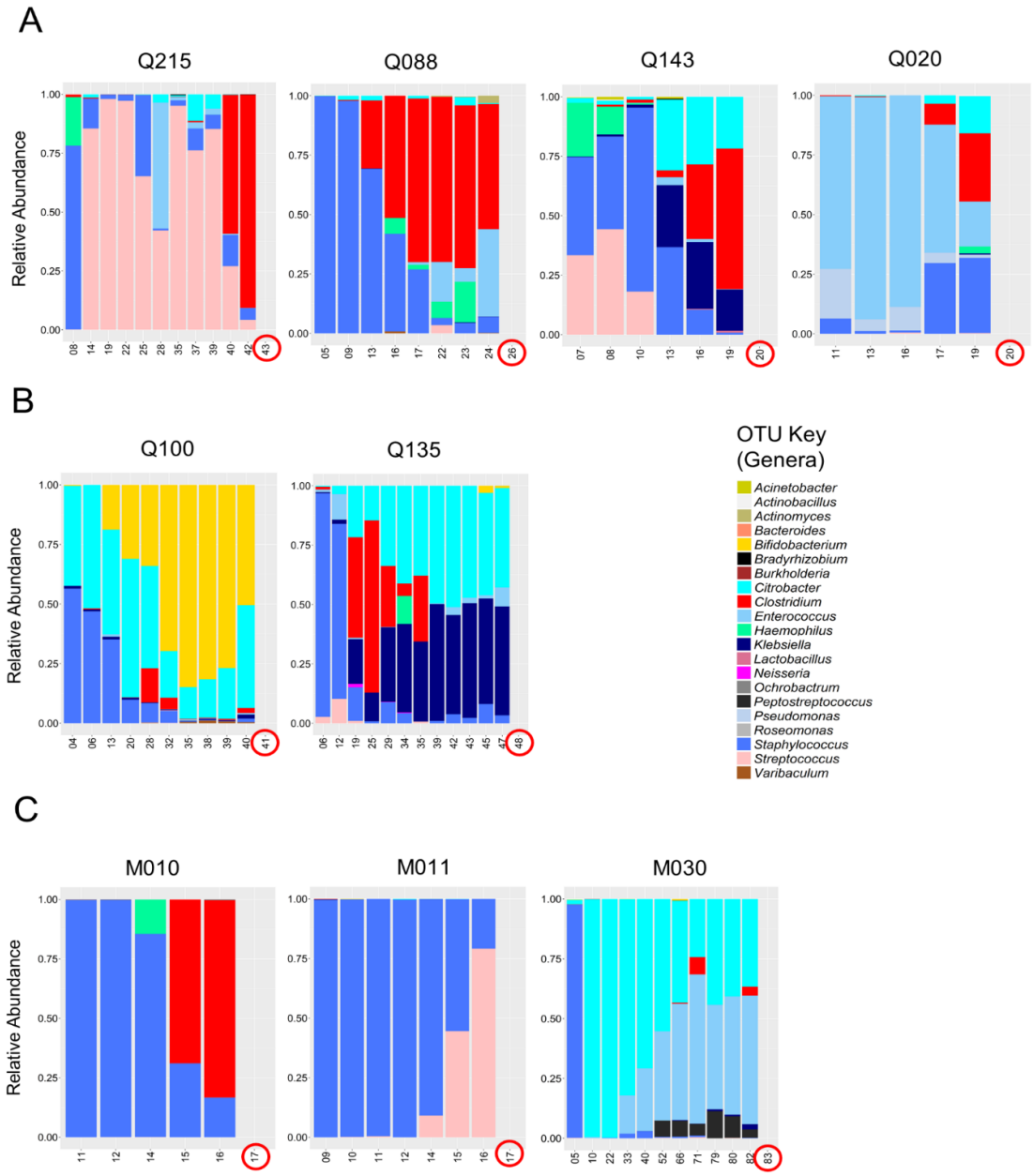


Figure 3.5: Faecal microbiome dynamics in infants with microbial associated NEC of different Bell stages

(A) 4 infants with '*C. perfringens*-associated' NEC Bell stages 2/3 as determined in Sim et al. (2015). *Clostridium* genus was detected in abundance prior to NEC onset. **(B)** 2 infants with '*Klebsiella*-associated' NEC Bell stages 2/3. **(C)** 3 infants with NEC Bell stage 1 (generic diagnosis, not clinically considered as medical NEC). The Day of Life (DOL) of each preterm infant when the samples were collected is shown on the x-axis and the day of NEC diagnosis is circled in red. Bar charts were produced in R using package *ggplot2*.

Microbiome profiling also indicated that the beneficial *Bifidobacterium* genus (Figure 3.6) was present at high levels in three control neonates (M087, Q087 and Q101), and potential pathogens including *Staphylococcus*, *Clostridium* and *Citrobacter* were

found at lower levels indicating a negative relationship between these microbes. This suggests enhanced colonisation resistance capacity of *Bifidobacterium* in the neonatal gut, which agrees with recent microbiome studies [289].

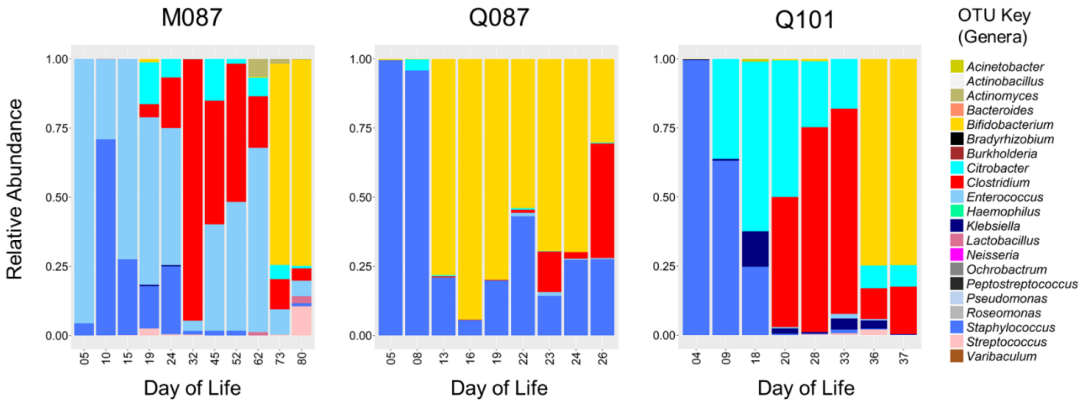


Figure 3.6: Potential role of *Bifidobacterium* as shown in non-NEC preterm microbiomes

Microbiome dynamics in three non-NEC neonates M087, Q087 and Q101 that harboured high abundance of *C. perfringens* (confirmed by culturing). These data suggest the potential colonisation resistance impact of *Bifidobacterium* genus against pathogenic bacteria in infant gut including *C. perfringens* as it appears the bloom of *Bifidobacterium* (yellow bar) directly antagonised the intestinal abundance of *C. perfringens* (red bar), potentially contributed to the prevention of *C. perfringens*-associated NEC. Bar charts were produced in R using package *ggplot2*.

In twin-paired gut microbiome analysis (one developed NEC in each pair), it was reported that NEC is an acquired microbial gut infection rather than genetic/hereditary in Sim et al. (Figure 3.7) [12]. There is an obvious increase of *Clostridium* taxa in neonate Q088 prior to NEC onset, however, although the matching twin sibling (Q087) also appeared to be colonised with *Clostridium* (low levels) it also had a *Bifidobacterium*-dominated gut microbiota in later time points (Figure 3.7A). Q143 infant was a *C. perfringens*-associated NEC case, whereas the twin sibling Q142 (control) did not acquire any *Clostridium* in this period and did not develop NEC (Figure 3.7B). This further suggests the impact of *C. perfringens* overgrowth in NEC development, with a strong influence in the wider microbiota in disease pathogenesis.

3.5.2 Comparative genomics of preterm-associated *C. perfringens*

In this section I will describe the results from comparative genomics of case and control *C. perfringens* isolates including linking pathogenicity associations with

'extra-virulent' genotypes, virulence plasmids, accessory virulence genes and bacteriophage contents in the *C. perfringens* genomes.

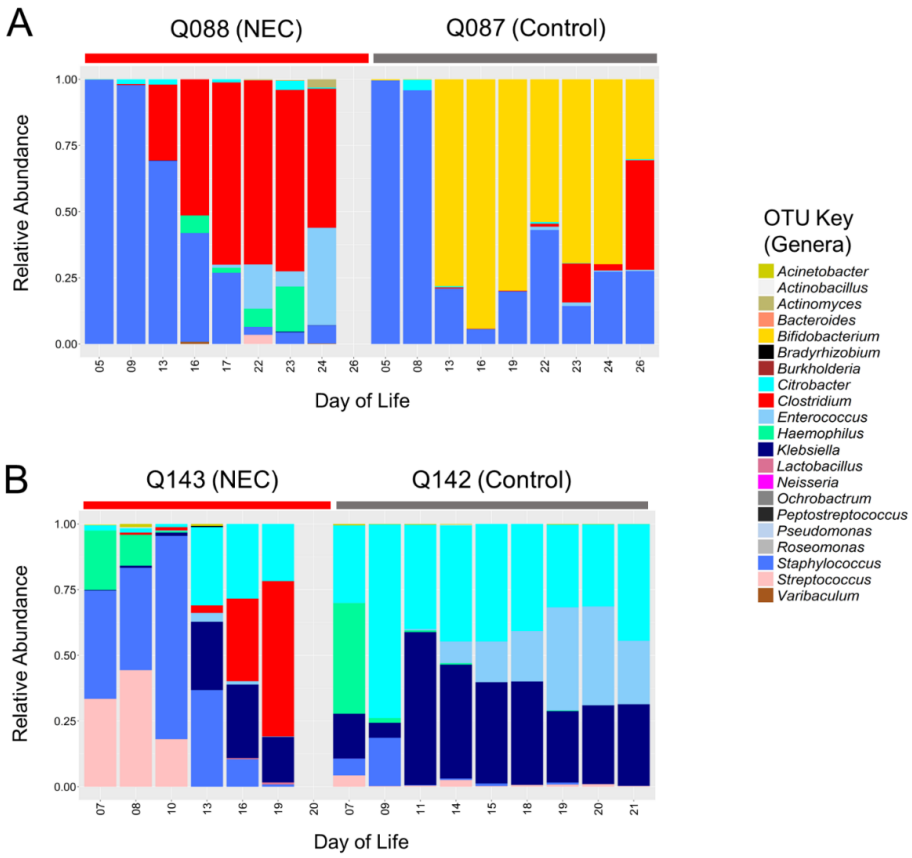


Figure 3.7: Twin-paired microbiota profiles

(A) *Bifidobacterium*-dominated microbiome in Q087 at later time points suggests protective effects of *Bifidobacterium* against NEC onset. **(B)** Infant Q142 did not acquire *Clostridium* throughout hospitalisation period. These data support the overgrowth of *C. perfringens* in the development of microbial NEC and the potential protective effects of genus *Bifidobacterium*. Bar charts were produced in R using package *ggplot2*.

3.5.2.1 Pathogenicity associations

Plasmid-carried toxin genes including *cpe* and *netB* have been associated with FP diarrhoea and chicken NE respectively, while key typing toxins ι-toxin, ε-toxin and β-toxin are mainly associated with diseases in ruminant animals including calves and sheep [1,19,290]. However, to date, there are no in-depth genomic studies describing *C. perfringens* isolates from preterm-NEC, and therefore the importance of toxins in disease pathology is unknown. In addition, accessory/secondary toxins (hydrolytic enzymes), AMR capacity, colonisation related binding proteins, enterotoxins and

virulence plasmids could be important in defining hyper-virulent *C. perfringens* strains that are distinct from 'commensal' *C. perfringens* [230,291]. I therefore performed comparative genomics on 88 *C. perfringens* isolates to explore virulence potentials encoded within the genomes.

Phylogenetics was performed to construct a SNP phylogeny (in core gene alignment) based on 90 isolates (including 2 reference genomes). Isolates from 4 neonates – Q215, Q088, Q143 and Q020 were defined as *C. perfringens*-associated NEC isolates (case isolates in this section) and the rest defined as control isolates. Genes encoding toxins (n=21), antimicrobials (based on CARD AMR database), collagen-binding proteins (n=4), enterotoxins (n=4) were individually profiled (virulence plasmids in the genomes were also determined) to compare between case and control isolates (Figure 3.8).

Via hierBAPS analysis, 90 genomes (as type A except reference genome type F NCTC8239) clustered into 6 distinct lineages (lineages II-IV; Figure 3.8), split into two major distinct clades (lineage I vs II-VI). Lineage I, a separate distant-away genetic cluster (>30,000 SNPs away from remaining clusters), comprises a number of the isolates from Q088 (other Q088 isolates clustered in lineage II), suggesting multiple strains can be present within one individual. NEC isolates are exclusively nested within lineages II and III (Figure 3.8), while lineages IV-VI encompass only control isolates, suggesting a potential genetic distinction between extra-virulent strains and commensal strains. Importantly, lineage IIIc appears to comprise isogenic strains (denoted by the apparent tight clustering), suggesting probable intra- and inter-hospital transmission as highly-similar strains were isolated from infants are two hospitals (QCCH and SMH respectively; further analysed and discussed in section 3.5.3).

To further analyse the gene associations with disease representations (case vs control), single-timepoint isolates were selected according to the following three criteria: (1) isolates must be either case (*C. perfringens* associated-NEC) or control (non-NEC); (2) first-time point control isolates (one isolate for each individual); (3) last-time point (closest to NEC) case isolates. Isolates Q100, Q135, M010, M030 and M011 were excluded from gene-association analysis as they did not fit the above criteria, with 33 isolates included for further analyses (Figure 3.9).

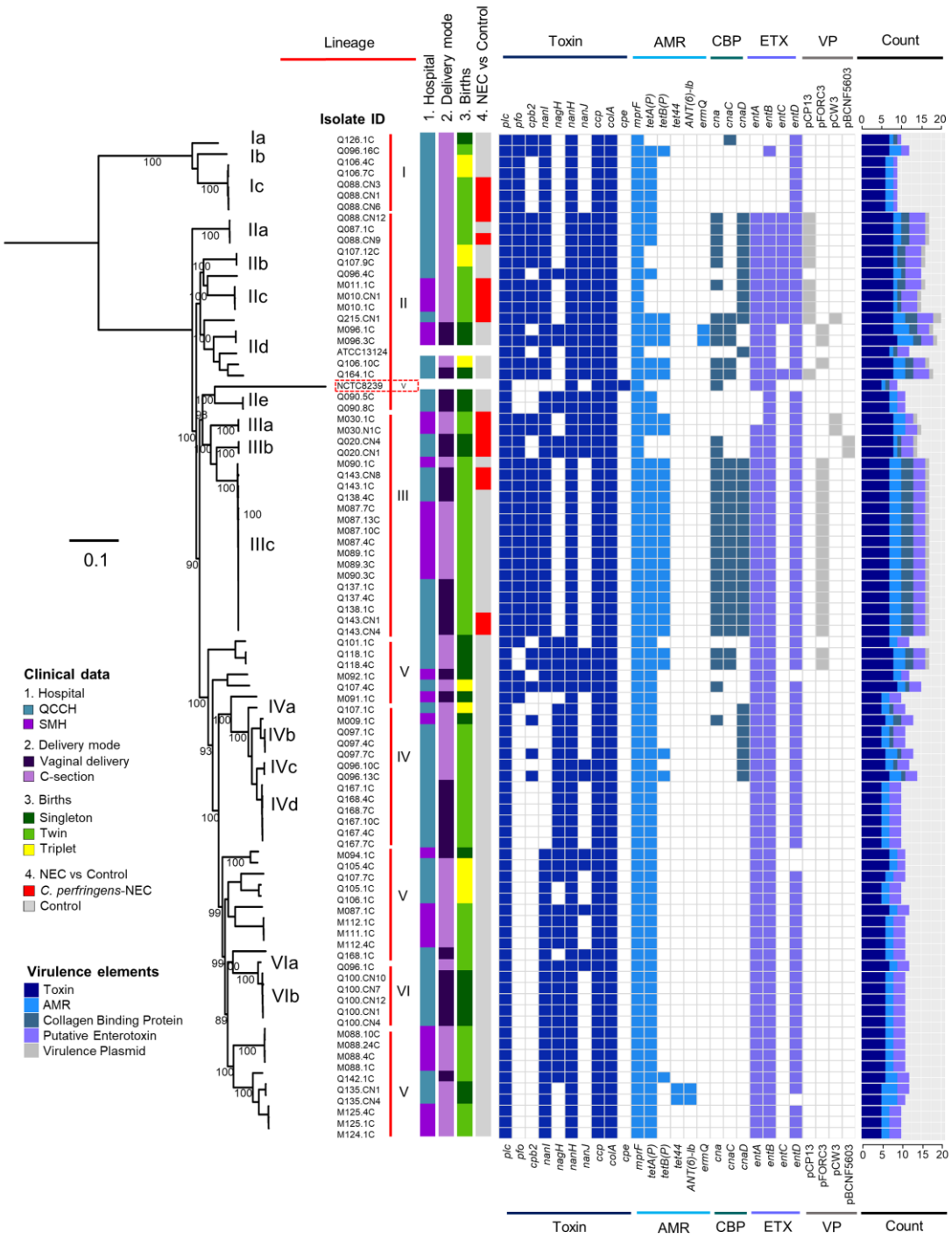


Figure 3.8: Phylogeny of preterm-associated isolates aligned with full virulence profiles

A mid-point rooted maximum-likelihood tree constructed from 90 *C. perfringens* isolates aligned with clinical metadata of host neonates and full virulence profiles of each isolate. This phylogenetic tree (including reference genome ATCC13124 and NCTC8239) is constructed from 125,930 core SNPs (an alignment of 1,973 core genes). Genetic clustering assigned via hierBAPS analysis. Bootstrap values are shown on the tree. AMR: Antimicrobial Resistance; CBP: Collagen Binding Protein; ETX: Probable Enterotoxin; VP: Virulence Plasmids.

Statistical testing (Fisher's exact test), indicates that overall toxin gene, *cna*-family gene and overall virulence elements (adding up all virulence-related genes and virulence plasmid count) were statistically more abundant in the *C. perfringens*-associated NEC isolates, indicating these NEC isolates are 'extra-virulent' strains encoding significantly more virulence genes compared with 'commensal'/control isolates (P<0.05; Figure 3.10). Individual virulence genes including *pfo* (encodes PFO) and *cna* (encodes collagen-binding proteins) were significantly enriched in NEC isolates (P<0.05). Genes *cnaD*, *cpb2* (variant 1), *entC* (enterotoxin) and plasmid pCP13 were associated with NEC isolates (trending towards statistical significance). Only 4 AMR genes were detected including the intrinsic resistance gene *mprF* (100%), tetracycline resistance genes *tetA(P)*(~94%) and *tetB(P)*(~30%), and *ermQ* (macrolide-resistance, encoded in one isolate).

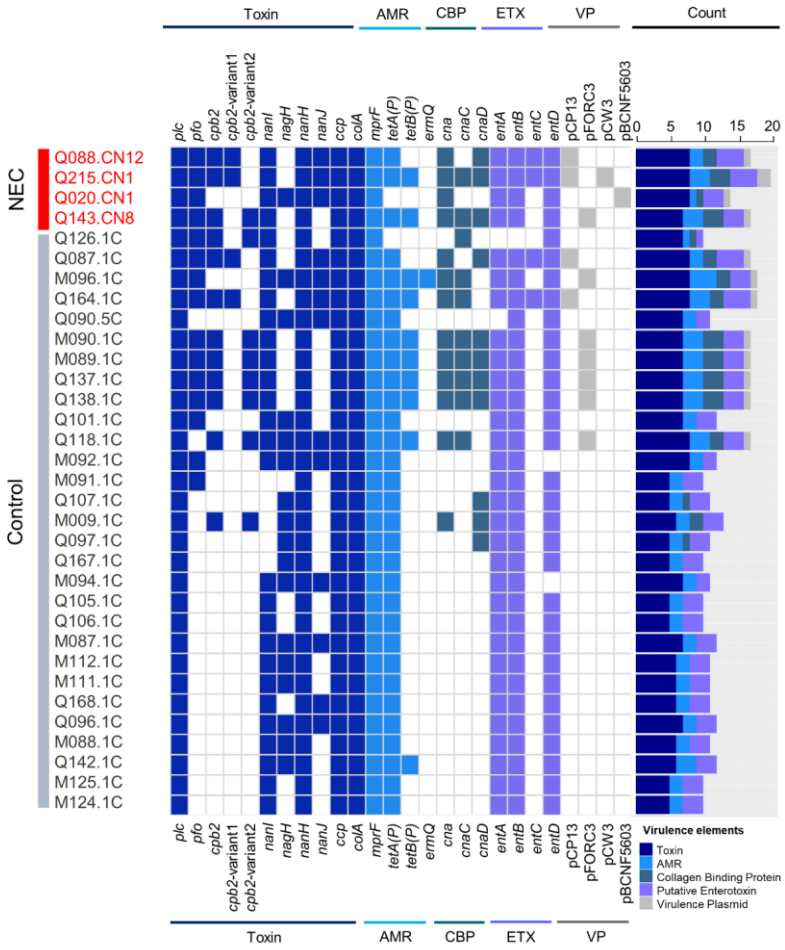


Figure 3.9: Virulence profiles for selected *C. perfringens* isolates used in association study

Isolates qualified for selection criteria for statistical analyses (n=33), aligned with full virulence profiles. *C. perfringens*-associated definite-NEC isolates were colour-coded in red (n=4).

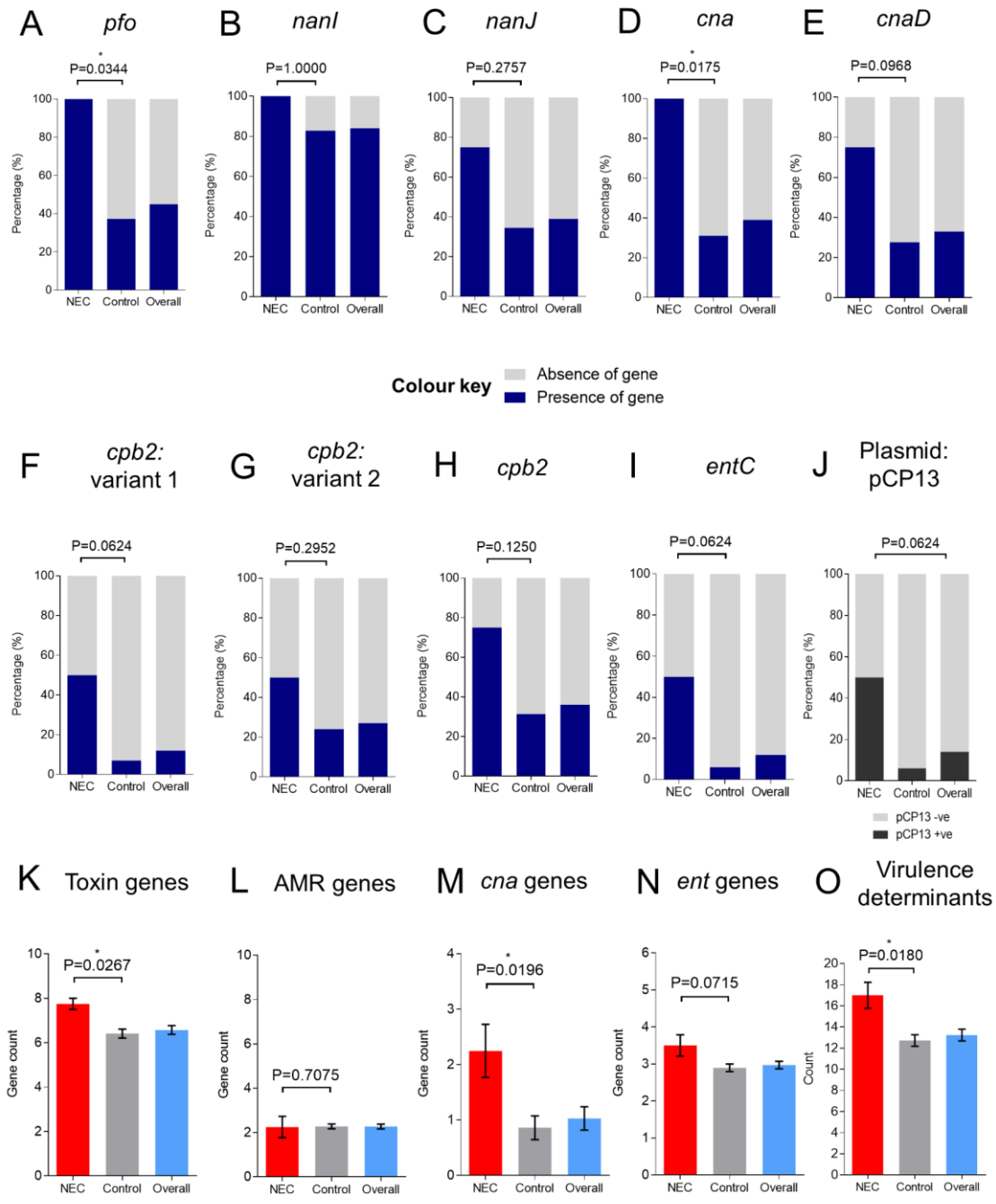


Figure 3.10: Enrichment analysis of key virulence elements in representative strains

(A) PFO gene *pfo*. **(B)** and **(C)**: sialidase genes *nanI* and *nanJ*. **(D)** to **(F)**: CPB2 toxin gene *cpb2* in both variants and overall association. **(G)** Enrichment of plasmid pCP13 in the isolates. **(H)** Toxin gene count comparison. **(I)** Antimicrobial gene count comparison. **(J)** Comparison of all virulence element counts (toxin + AMR + virulence plasmid count). **Data:** (A) to (G) fraction percentage (Fisher's exact test, two-tailed, *P<0.05, NEC: n=4, control: n=29). (H) to (J): mean count ± SEM (non-parametric Mann Whitney U test, two-tailed, *P<0.05, NEC: n=4, control: n=29). All estimated P values are printed in the figure. P<0.10 indicates weak association, P<0.05 indicates statistically significant difference.

3.5.2.2 *In silico* virulence plasmid analysis

Plasmids are known to be associated with intestinal diseases in both human and animals, carrying an array of virulence factors including toxin and AMR genes [1,41,42,291]. Importantly, suspected virulence plasmid pCP13 was identified in two NEC isolates Q215.CN1 and Q088.CN12, showing strong correlation with NEC disease representation (Figure 3.10). There are several known virulence factors encoded on this 54 kb-plasmid (first isolated from *C. perfringens* strain 13, a human gas gangrene isolate) [37]. Collagen binding protein *cna* and β2-toxin gene *cpb2* (variant 1) are encoded on this plasmid. Plasmid pFORC3 was detected in NEC isolate Q143.CN8, and all isolates in lineage IIIc (potentially isogenic strains). This 56 kb-plasmid encodes virulence genes including *cpb2* (variant 2), collagen binding protein *cnaC* and AMR elements *tetA(P)* and *tetB(P)*.

To further confirm whether these detected 'plasmids' are 'intact plasmids', which is important in understanding its potential transfer and spread, the sequences of plasmids were individually extracted from these high-sequencing-coverage (depth range: 102 X-206 X) genome assemblies using *in silico*-approaches and compared with reference plasmids.

Isolates Q215.CN1, Q088.CN12, M010.1C and Q164.1C were computationally detected (raw read-based) to carry plasmid pCP13 and hence sequences were extracted (from assemblies) and compared in Figure 3.11A. Genomic mapping shows that these plasmids are near-identical in nucleotide identity (>97%), plasmid size (~54 kb), GC content (~25.4%) and gene content (62-65 CDS encoded), suggesting this plasmid has been circulating intact (Figure 3.11B; Table 3.4).

A total of 9 plasmid pFORC3-carrying isolates were selected for comparative analysis due to different k-mer coverage (Table 3.4), which initially suggested these are not intact pFORC3-family plasmids (Figure 3.12A-B). Eight '*in-silico* plasmids' (not experimentally extracted) including pQ143.CN8 were found to be 14 kb (~20%) larger than the expected reference plasmid pFORC3, while pM096.3C is more similar to reference plasmid pF262A (from bovine isolate F262) at 48 kb length.

Reference plasmid	Predicted plasmid	Genome features			K-mer identity (%) (Read-mapping)	Similarity by BLASTN (Assembly-based)	
		Size (bp)	GC (%)	CDS		Identity (%)	Coverage (%)
pCP13	**pCP13**	**54,310**	**25.50**	**63**	n/a	n/a	n/a
	pQ215.CN1	54,407	25.47	62	99.52	99.98	100.00
	pQ164.1C	54,647	25.44	65	94.82	97.08	100.00
	pQ088.CN12	54,421	25.50	62	99.41	99.95	100.00
	pM010.1C	54,421	25.50	64	99.40	99.97	100.00
pFORC3	**pFORC3**	**56,577**	**27.28**	**79**	n/a	n/a	n/a
	pM087.7C	70,121	26.05	80	91.13	99.56	71.05
	pM089.1C	70,119	26.06	80	91.12	99.56	71.04
	pM090.1C	70,120	26.06	80	91.08	99.64	71.24
	pQ137.1C	70,119	26.05	78	91.12	99.72	67.82
	pQ143.CN8	70,111	26.05	79	91.12	99.56	67.46
	pQ138.1C	70,119	26.04	79	91.12	99.55	76.91
	pQ106.10C	70,123	26.05	79	91.11	99.56	71.05
	pQ118.4C	70,126	26.04	78	90.00	99.56	71.05
	pM096.3C	48,245	27.27	57	98.64	83.85	99.97
pF262A	**pF262A**	**48,604**	**27.15**	**55**	n/a	n/a	n/a
	pM096.3C	48,245	27.27	57	as pFORC3	97.42	97.14

Table 3.4: Similarity comparison of predicted plasmids with reference plasmids

Importantly, pQ143.CN8 which is 14 kb larger than pFORC3, was found to encode a series of extra potential virulence factors including multiple ABC transporters that are linked to multidrug resistance (Figure 3.12C). In brief, no intact pFORC3 plasmid, a plasmid sourced from aquarium water in South Korea, was encoded within these preterm-associated *C. perfringens* isolates. In fact, these plasmids could potentially be unidentified novel plasmids (linked to additional virulence in drug resistance) derived from pFORC3 plasmid backbone that warrants further experimental validations. Of note, these pFORC3-derived plasmids encode conjugative systems (Tcp loci were identified via sequence search) and are expected to be capable to strain to strain transmission [51,292].

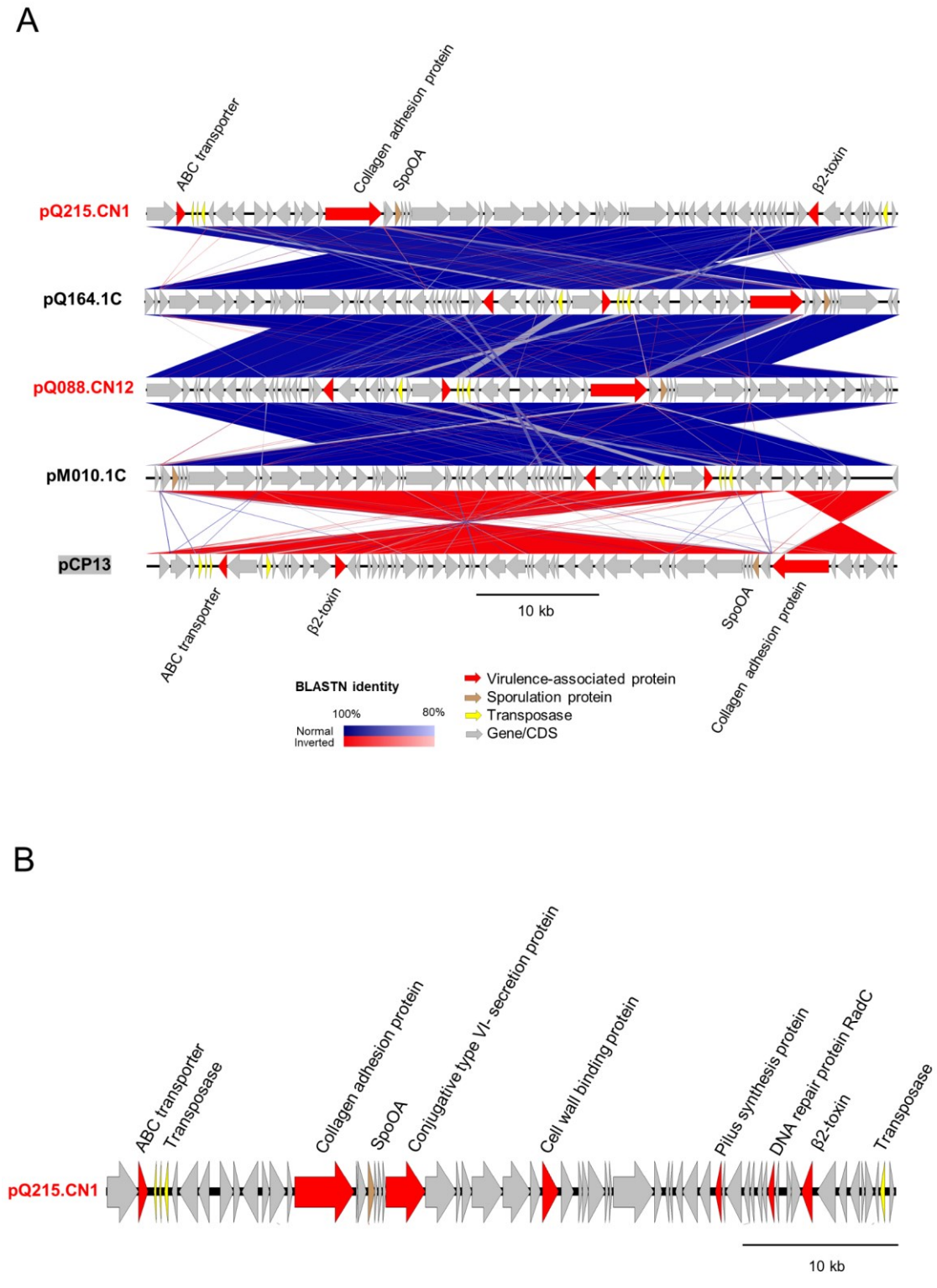


Figure 3.11: Comparisons of pCP13-family plasmids extracted from preterm isolates

(A) Genomic mapping of pCP13 plasmids (shaded) from isolates Q215.CN1, Q088.CN12, M010.1C and Q164.1C. **(B)** Closing in on potential virulence factors encoded on pQ215.CN1. Graphs generated via Easyfig.

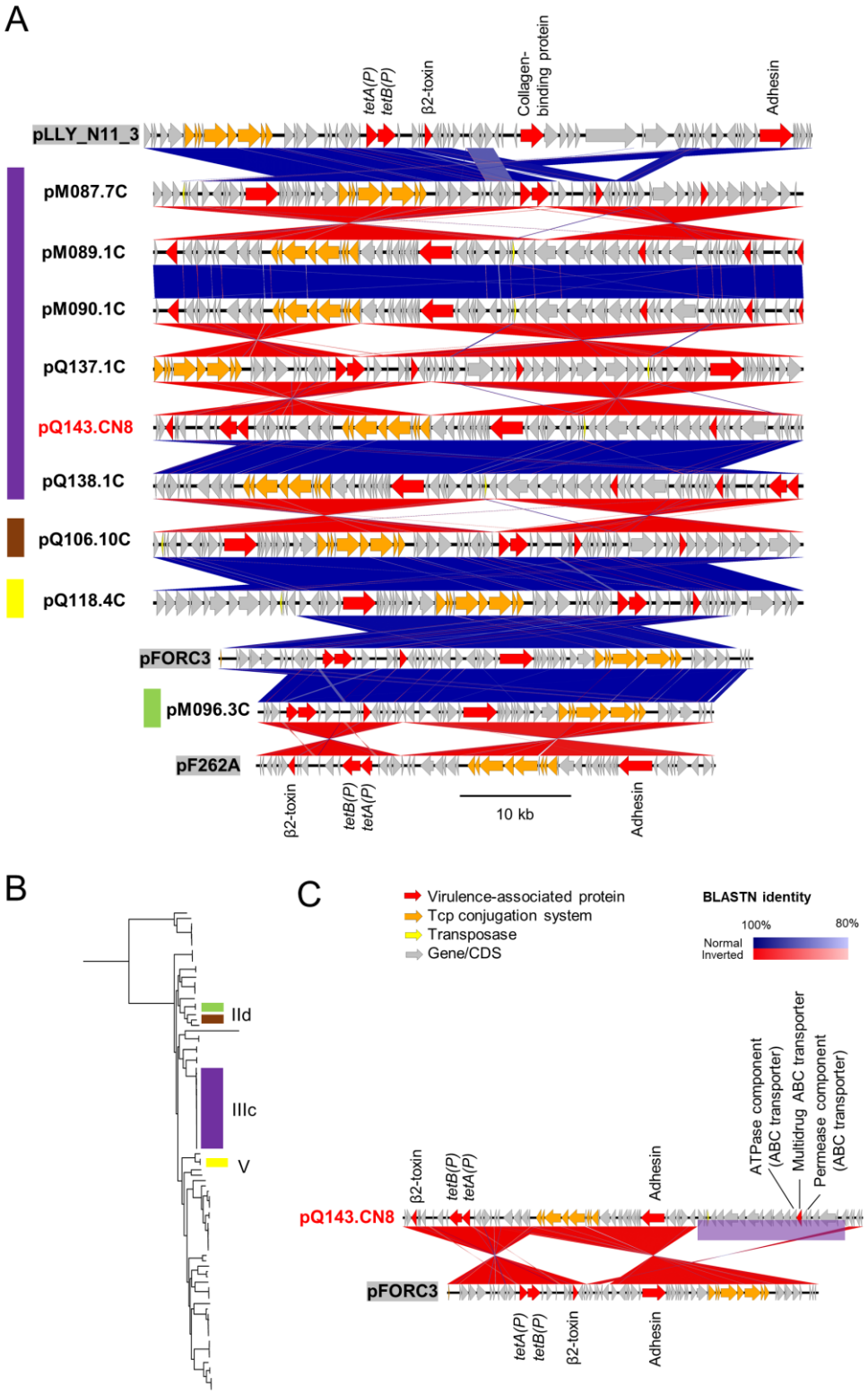


Figure 3.12: Comparisons of pFORC3-family plasmids extracted from preterm isolates

(A) Genomic mapping of pFORC3 plasmids from 9 isolates where pFORC3 plasmid was computationally detected at different k-mer identity with reference plasmids pLLY_N11_3, pFORC3 and pF262A. **(B)** Colour-coded phylogenetic positions of the isolates compared in (A). **(C)** Potential additional genetic regions (light purple region) found on plasmid pQ143.CN8 that encoded extra virulence genes compared to reference plasmid pFORC3. Graphs generated via Easyfig and iTOL.

3.5.2.3 Lineage-specific virulence and evolution observed in accessory genomes

The pangenome of the 90 isolates (including two reference genomes) comprises 1,973 (25.8%) core genes and 5,674 (74.2%) accessory genes (Figure 3.13). As a *C. perfringens* genome is approximately 3.2 Mb in size and carries 3,009 genes (Appendix 1 Table S3.2), *C. perfringens* is predicted to harbour ~34.5% (~1,038) accessory genes, which is hypothesised to be important in virulence contribution.

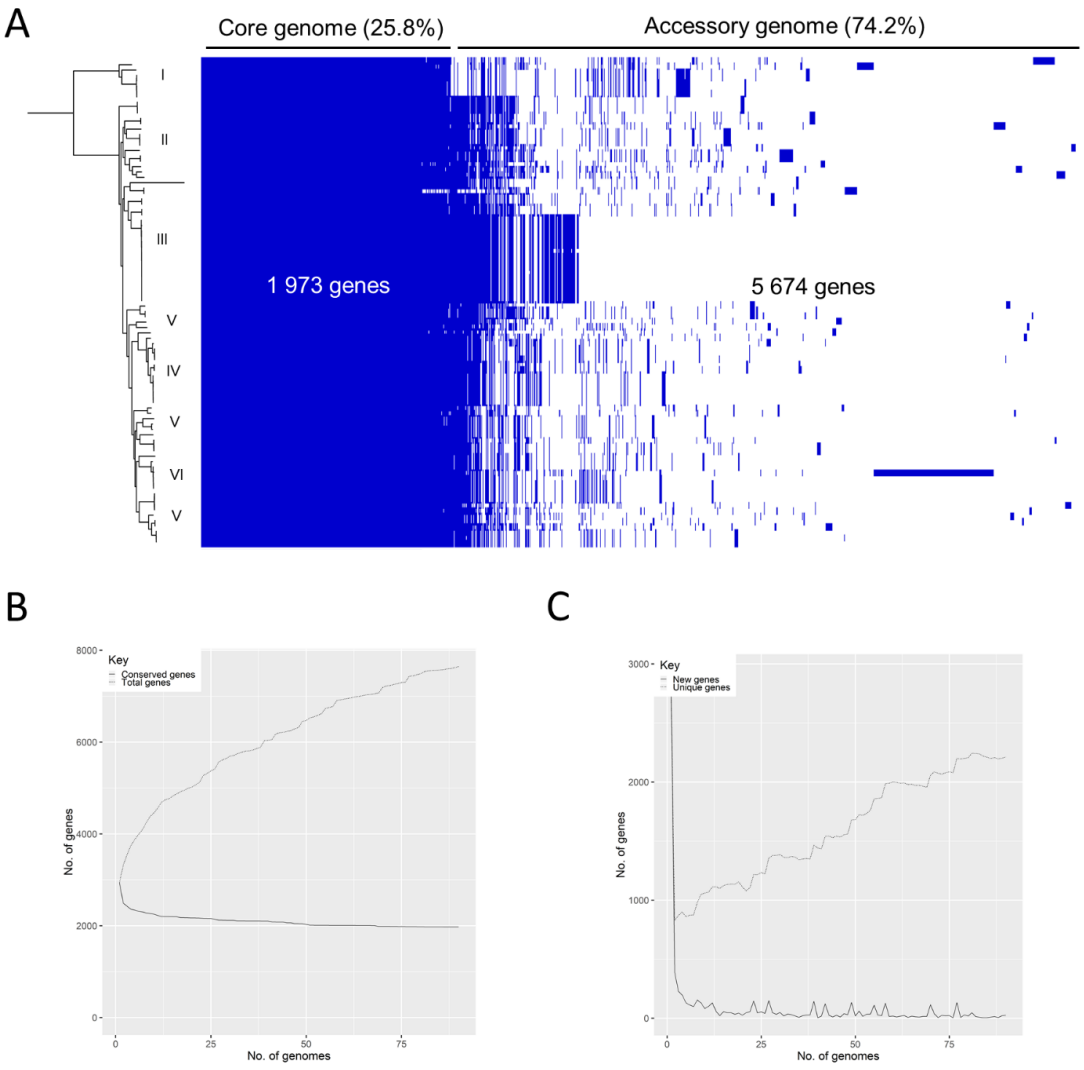


Figure 3.13: Pangenome of 90 preterm-associated isolates

(A) Pangenome visualised in heatmap using Phandango. Blue cells indicate orthologous groups (syntenies). **(B)** Pangenome stats: core vs total genes. **(C)** new vs unique genes. Stats graphs generated using R scripts in Roary package.

By delving into accessory genome and trait-association analysis (Appendix 1 Table S3.3 and S3.4), I observed several NEC-isolate specific genes (n=56; sensitivity cut-off=75%, specificity cut-off=60%) which might contribute additional fitness/virulence to the strains. These genes include thioester-forming surface-anchored protein (*cnaD*), pilus biosynthesis protein HicB, cell-wall binding protein (*cna*; identified in 9/12 [75%] NEC-associated isolates, while not identified in 50/78 [65%] in non-NEC isolates), and carbohydrate-related ABC transporters (n=3; identified in 9/12 [75%] NEC-associated isolates, while not identified in 49/78 [63%] in non-NEC isolates).

Interestingly, lineages II and III (hypervirulent clusters) that comprise all the NEC isolates encode certain virulence-related specific genes (n=82; sensitivity cut-off=80%, specificity cut-off=75%) including PFO (*pfo*; sensitivity: 35/38 [92.1%], specificity: 41/52 isolates in other lineages do not encode *pfo* [78.4%]), thioester-forming surface-anchored protein (*cnaD*; sensitivity: 32/38 [84.2%], specificity: 44/52 isolates in other lineages do not encode *cnaD* [90.3%]), pilus biosynthesis protein HicB (sensitivity: 32/38 [84.2%], specificity: 44/52 [90.3%]), cell-wall binding protein (*cna*; sensitivity: 32/38 [84.2%], specificity: 44/52 [90.3%]), S-layer homology domain-containing protein (sensitivity: 35/38 [92.1%], specificity: 45/52 [86.5%]), germination-associated cotH protein (group_654; sensitivity: 37/38 [97.3%], specificity: 42/52 [80.7%]) and carbohydrate-related ABC transporter (n=5; sensitivity: 92-97%, specificity: 86-94%), thus revealing new and important accessory virulence potentials. Specific orthologous groups of the pangenome (multi-FASTA format) can be obtained from my GitHub repository[12].

These data support findings via virulence association analysis discussed in previous section with regard to *pfo* and *cna* virulence genes. Gene enrichment analysis also showed that *pfo* is highly specific to lineage II and III isolates (92.1%; 35/38 isolates; P<0.0001) compared to lineage IV-VI isolates, collagen-binding protein encoding genes *cna* (78.9%; 30/38 isolates; P<0.0001) and *cnaC* (71.1%; 27/38 isolates; P<0.0001) are significantly enriched in lineages II and III. These three important virulence genes were rarely encoded in lineage IV-VI (n=44; *pfo:* 6.8%; 3/44 isolates; *cna:* 9%; 4/44 isolates; *cnaC*: 18.9%). Hence, it is proposed that lineage II-III

[12] https://github.com/raymondkiu/Thesis/blob/master/Chapter3_pangenome

comprises extra-virulent strains that harbour significantly more virulence genes than isolates in lineage IV-VI, which may be regarded as commensal strains (phenotypes need to be experimentally addressed).

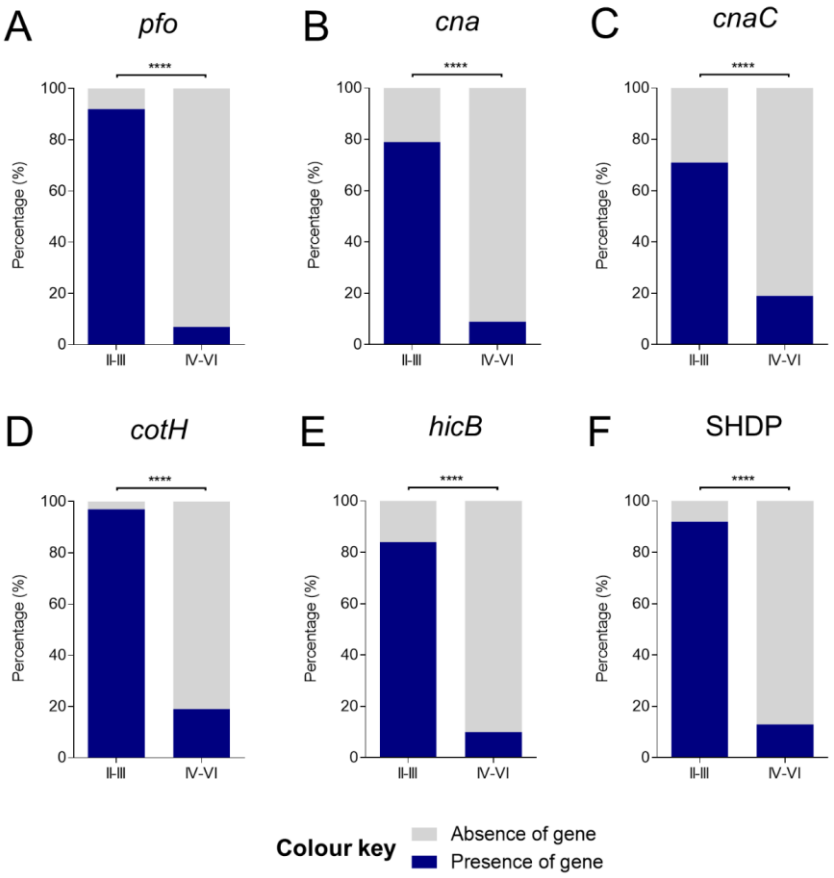


Figure 3.14: Enrichment of virulence genes in lineages II-III and lineages IV-VI isolates

(A) PFO gene *pfo.* **(B)** Collagen binding protein gene *cna* **(C)** *cnaC* **(D)** spore germination-related *cotH* **(E)** pilus-synthesis gene *hicB* **(F)** SHDP: S-layer homology domain-containing protein. **Data:** fraction percentage (Fisher's exact test, two-tailed, ****P<0.0001, lineages II-III: n=38, lineages IV-VI: n=44). P<0.05 indicates statistically significant difference.

3.5.2.4 Bacteriophage contents

Bacteriophage, or simply phages, are viruses that infect bacteria. Temperate phages are known to reside within bacterial genomes and are known to contribute virulence factors to their hosts, including toxins [293]. To date, prophage representation within *C. perfringens* is not well understood. Hence, here I determined the phage contents and virulence-related elements carried by prophages residing within these preterm-associated *C. perfringens* strains to give exploratory insights and understanding towards this increasingly important aspect of microbiology research via bioinformatics approaches.

Using pangenome-based analysis, prophage-related genes were determined to contribute 4% to the *C. perfringens* pangenome (part of accessory genome) (Figure 3.15A). No significant differences of prophage contents between NEC (n=0.5) and control (n= 0.72) were observed (Figure 3.15B-C; P=0.5663).

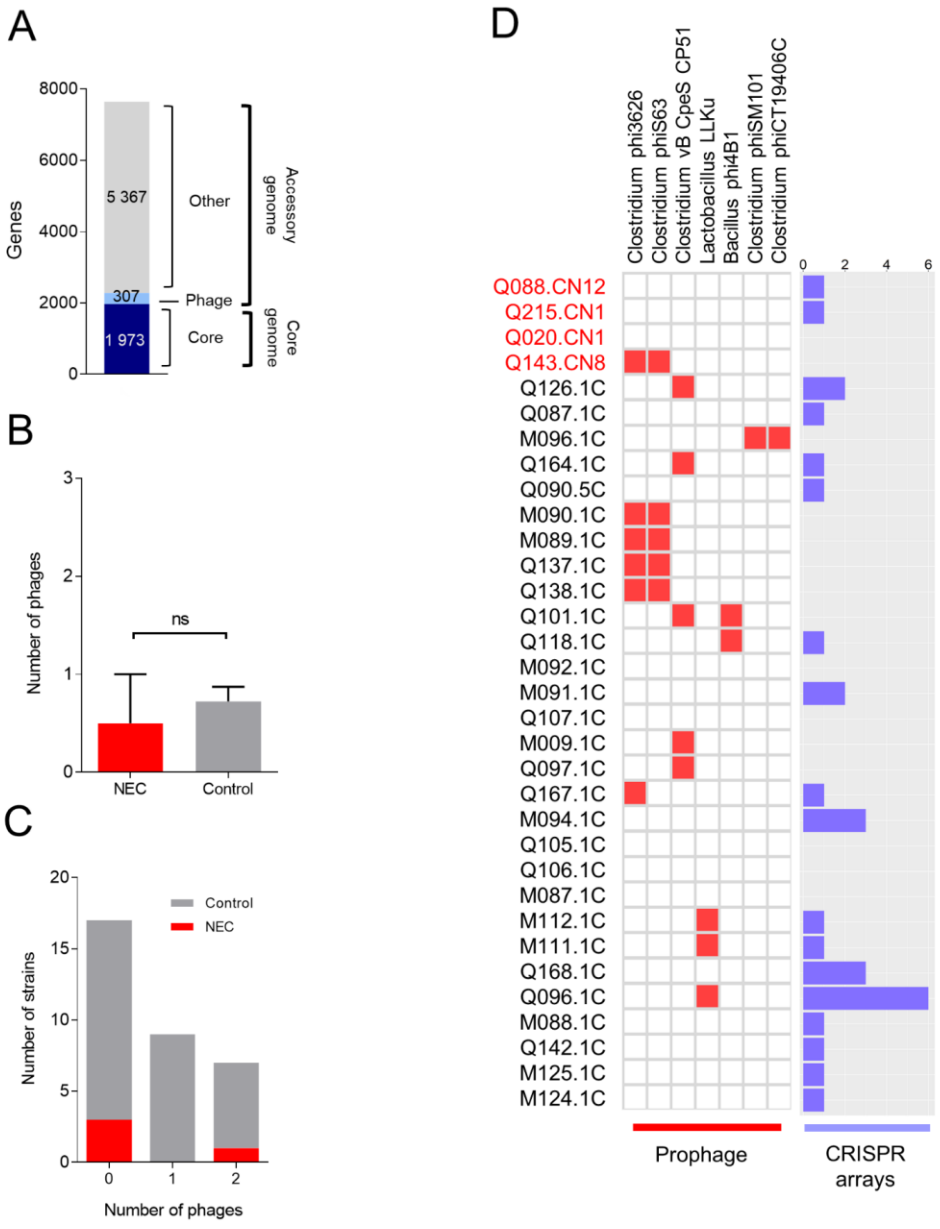


Figure 3.15: Prophages detected in preterm-associated *C. perfringens* genomes

(A) Contribution of prophage-related genes in *C. perfringens* pangenome (90 genomes). Core genome represents genes present in all (100%) computed genomes. Gene numbers are printed within the corresponding bars**. (B)** Statistics of integrated prophage in *C. perfringens* strains categorised by NEC and control groups. **(C)** Frequencies of integrated phage(s) in *C. perfringens* strains. **(D)** Heatmaps showing detected prophages (red-colour cells) corresponding to *C. perfringens* strains and CRISPR arrays present in each individual isolate. Heatmap and bar chart were generated in R.

A total of 7 prophage species were identified in 33 representative *C. perfringens* strains from different individual neonates including genera associated with *Clostridium*, *Lactobacillus* and *Bacillus* (Figure 3.15D). Phages identified in these *C. perfringens* strains including *Clostridium* phage ϕ63 has been reported to potentially influence sporulation of its host, which is critical for host persistence and fitness [294]. Another key *Clostridium* phage ϕ3626 was suggested to affect sporulation-onset mechanisms [295]. *Clostridium* phage vB_CpeS-CP51 detected in 5 *C. perfringens* genomes (15%; 5/33) has previously been reported to exist in a foal case isolate, JFP55 (or, JP55) [102]. Overall, prophages were not identified in 17/33 of the (51%) *C. perfringens* strains.

Importantly, CRISPR-cas systems in bacteria are used to prevent invasion of lytic phages. Thus, CRISPRs were identified in these 33 *C. perfringens* strains, with the highest number of 6 CRISPRs present in one control strain (Figure 3.15D). No differences were found between NEC (n=0.5) and control isolates (n=0.93; P=0.7258) in terms of CRISPR arrays they encode, although it appears that control strains have more CRISPRs (Figure 3.16).

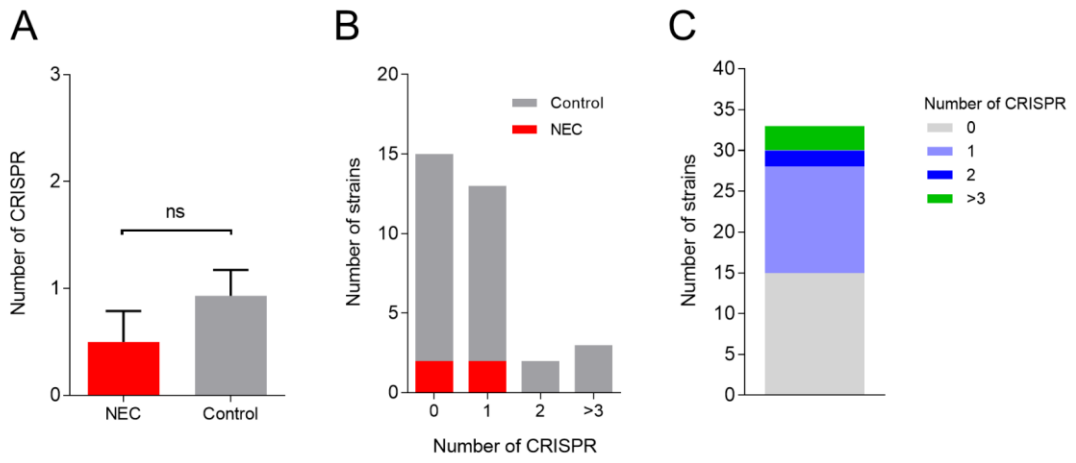


Figure 3.16: CRISPR systems in *C. perfringens* strains

(A) Number of CRISPRs present in *C. perfringens* strains categorised by NEC and control groups. **(B)** Frequency of CRISPR counts in *C. perfringens* strains. **(C)** An overview of CRISPR proportions in isolates.

Prophage has been reported to have a role in carrying virulence-related genes, such as *Siphoviridae* bacteriophages which carry enterotoxin A that contributes to *S. aureus*

virulence [296]. Subsequently I investigated the virulence elements potentially carried by the prophages of preterm-associated *C. perfringens* strains. Several virulence-related genes were detected via sequence search and/or annotation-based screening including genes that encode sialidase NanH, probable enterotoxin EntB and EntD, and ABC transporter (Figure 3.17A). Sporulation-linked genes were also encoded within prophages (Figure 3.17B), which may be associated with sporulation modulation in *C. perfringens* strains [297].

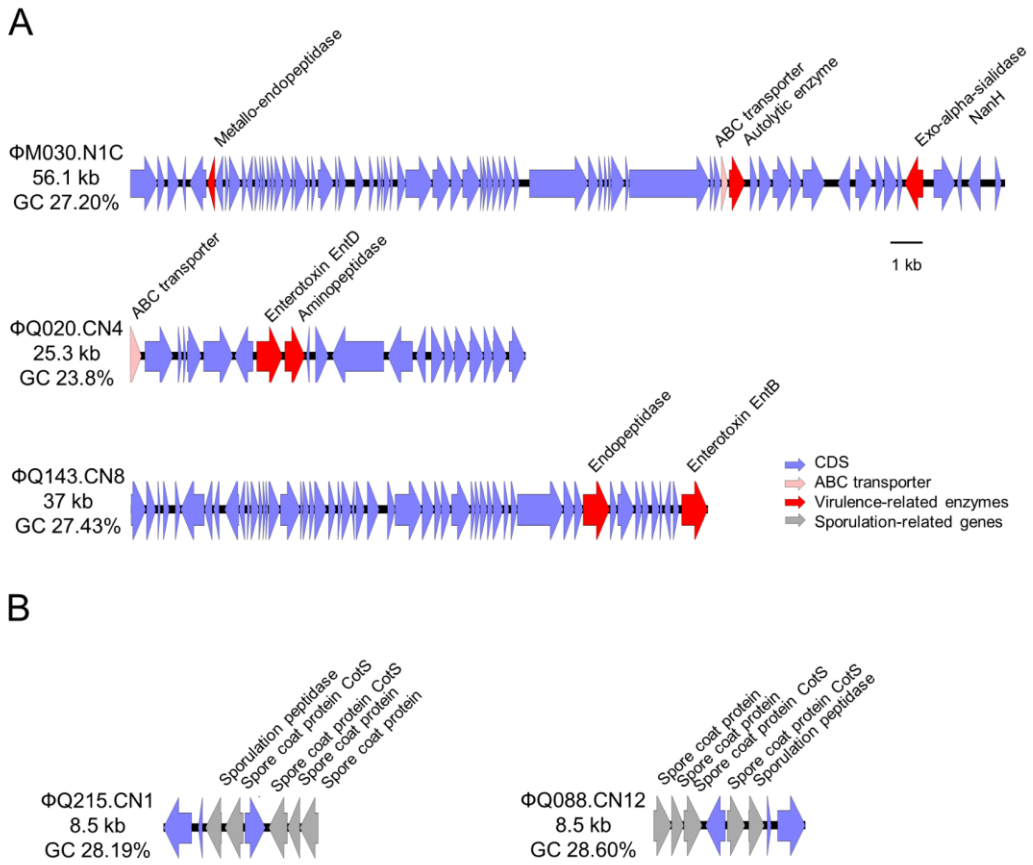


Figure 3.17: Virulence- and sporulation-related genes encoded within prophage genomes

(A) Annotated genetic maps showing putative virulence-associated genes within extracted prophages (from complete prophages). **(B)** Sporulation-related genes were enriched in these phage-encoded regions (from incompletely predicted prophage sequences).

Importantly, enterotoxin gene *entB* has been consistently detected in 19 phage-carrying isolates from 8 individuals (Table 3.5), suggesting this enterotoxin is prophage-specific which has previously not been described. This is evidenced by a dissimilarity in *entB* sequence identities between these prophage-carrying *entB* (~94.4%; Table 3.5) and *entB* in other isolates (<93%), indicating prophage-*entB*

differs in nucleotide sequence. This exploratory *C. perfringens* prophage investigation indicates an important virulence contribution which requires further research into this understudied area.

Isolate	Host	Group	Gene	BLASTN Query coverage	Coverage (%)	Identity (%)
M087.10C	M087	Control	*entB*	1-1650/1650	100	94.44
M087.13C	M087	Control	*entB*	1-1650/1650	100	94.44
M087.4C	M087	Control	*entB*	1-1650/1650	100	94.44
M087.7C	M087	Control	*entB*	1-1650/1650	100	94.44
M089.1C	M089	Control	*entB*	1-1650/1650	100	94.44
M089.3C	M089	Control	*entB*	1-1650/1650	100	94.44
M090.1C	M090	Control	*entB*	1-1650/1650	100	94.44
M090.3C	M090	Control	*entB*	1-1650/1650	100	94.44
Q096.4C	Q096	Control	*entB*	1-1650/1650	100	94.38
Q107.12C	Q107	Control	*entB*	1-1650/1650	100	94.44
Q107.7C	Q107	Control	*entB*	1-1650/1650	100	94.86
Q107.9C	Q107	Control	*entB*	1-1650/1650	100	94.44
Q137.1C	Q137	Control	*entB*	1-1650/1650	100	94.44
Q137.4C	Q137	Control	*entB*	1-1650/1650	100	94.44
Q138.1C	Q138	Control	*entB*	1-1650/1650	100	94.44
Q138.4C	Q138	Control	*entB*	1-1650/1650	100	94.44
Q143.1C	Q143	NEC	*entB*	1-1650/1650	100	94.44
Q143.CN1	Q143	NEC	*entB*	1-1650/1650	100	94.44
Q143.CN4	Q143	NEC	*entB*	1-1650/1650	100	94.44
Q143.CN8	Q143	NEC	*entB*	1-1650/1650	100	94.44

Table 3.5: Enterotoxin sequences detected in predicted prophage genomes

3.5.3 Genomic epidemiological analysis reveals inter- and intra-hospital transmission

3.5.3.1 Frequent hospital transfer may facilitate bacterial transmission

In this section, potential *C. perfringens* transmissibility inferred by phylogenomic data will be demonstrated. *C. perfringens* isolates (n=88) were obtained during the 2-year NEOM study at both QCCH and SMH (Figure 3.18). As inferred by phylogenetic analysis and clinical data, two clusters (Ic and IIIc; assigned via hierBAPS analysis; overlapping hospitalisation period) were identified to be associated with potential intra-hospital and inter-hospital transmission as highlighted in grey boxes in Figure 3.18.

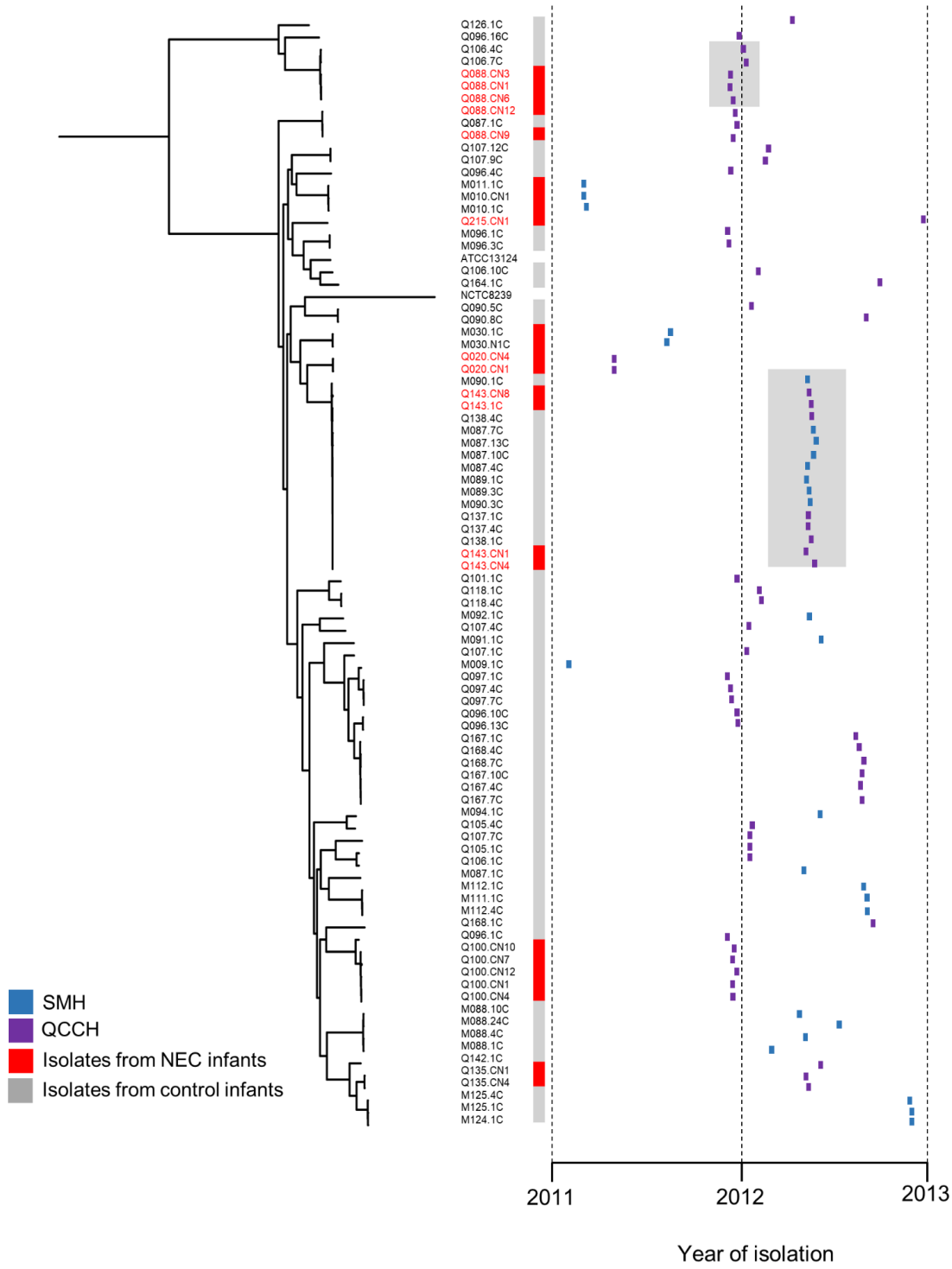


Figure 3.18: Isolation time-points for all 88 *C. perfringens* isolates in this study

Mid-point rooted maximum likelihood tree for 90 *C. perfringens* genomes (including 2 reference genomes ATCC13124 and NCTC8239) from preterm babies isolated in two hospitals in between 2011-2013. Timepoints of isolation for each isolate were indicated in the figure. Two clusters of isolates associated with potential patient-to-patient transmission and inter-hospital transmission are highlighted by shaded boxes. Red-labelled isolates in the colour strip indicates NEC-associations. Isolates from *C. perfringens*-associated definite-NEC (Bell stages 2/3) were printed in red.

Transmissibility of *C. perfringens* has never been studied using WGS techniques to date. Using the available clinical metadata (NEOM) and phylogenetic analysis, it is possible to observe and trace the potential transmission routes of certain strains in between the hospitals QCCH and SMH. Babies are transferred frequently between these two sister hospitals in central London as QCCH handles more surgery-related cases in neonates (Figure 3.19). It is therefore hypothesised that this will facilitate the transmission of bacteria, especially spore-formers like *C. perfringens*, which can survive for long period of time in hospital environments.

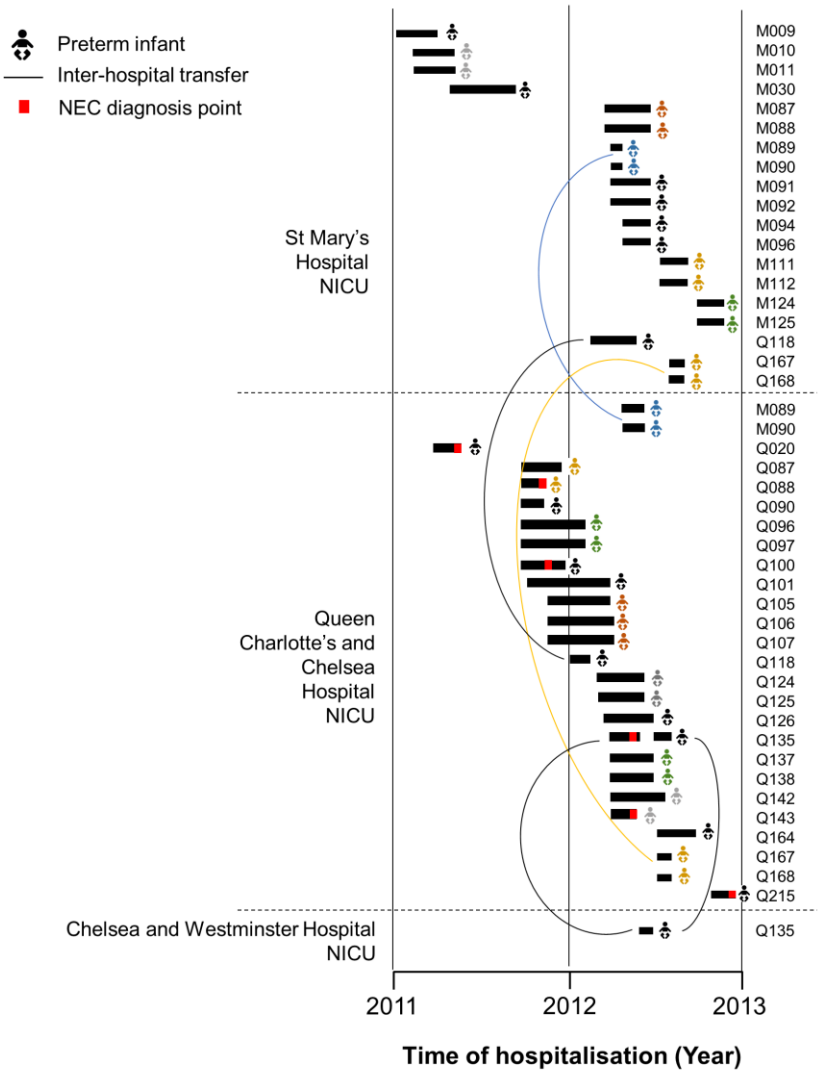


Figure 3.19: Inter-hospital transfer map of 38 *C. perfringens*-positive preterm infants

These include infants screened positive of *C. perfringens* (2011-2013) in between SMH, Chelsea and Westminster Hospital and QCCH NICUs. Colour-coded infant symbols indicate twins/triplets otherwise singleton.

3.5.3.2 Within-host SNP variations

To define *C. perfringens* at a strain level, within-host SNP variation analysis was performed using two approaches: reference-based and assembly-based. A total of 16 pairs of strains (two identical *C. perfringens* colonies isolated from same sample and time point were sequenced) from 13 infants were selected to pre-determine the SNP count range of an identical strain (normally assume 0 SNP). A narrow r ange of 1-12 (reference-based, or, 0-6 via assembly-based method) SNPs were found to be the SNP variations at strain level (confirmed by core genome ANI>99.98%; Table 3.6).

	Host ID	Isolate	ANI (%)	Day of life	SNP count (reference-based)	SNP count (assembly-based)
1	M010	M010.CN1	100.00	15	5	1
		M010.CN2	99.99			
2	M030	M030.N1C	100.00	71	1	0
		M030.N2C	99.99			
3	M087	M087.1C	100.00	31	3	1
		M087.2C	99.99			
		M087.10C	100.00	62	2	0
		M087.11C	100.00			
4	M088	M088.1C	100.00	17	4	2
		M088.2C	99.98			
		M088.10C	100.00	45	5	2
		M088.11C	99.99			
5	M089	M089.1C	100.00	22	5	1
		M089.2C	99.99			
6	M090	M090.1C	100.00	15	10	1
		M090.2C	99.99			
7	Q020	Q020.CN1	100.00	19.5	12	6
		Q020.CN2	99.99			
		Q020.CN4	99.99	19	2	1
		Q020.CN5	99.99			
8	Q087	Q087.1C	100.00	37	1	1
		Q087.2C	99.99			
9	Q088	Q088.CN1	100.00	13	1	1
		Q088.CN2	99.99			
10	Q090	Q090.5C	100.00	25	4	2
		Q090.6C	99.99			
11	Q100	Q100.CN1	100.00	29	5	1
		Q100.CN2	99.99			
12	Q168	Q168.1C	100.00	30	5	1
		Q168.2C	100.00			
13	Q215	Q215.1C	100.00	40	6	1
		Q215.2C	99.99			
				Minimum	1	0
			SNP Statistics	Median	4.5	1
				Maximum	12	6

Table 3.6: Within-host SNP variation analysis

Two colonies were selected from the same sample (same time-point) and sequenced to analyse *C. perfringens* within-host SNP variations. Pair-wise ANI was performed via mummer alignment. Reference-based (read-mapping) SNP method was performed using raw-read FASTQ via direct reference-mapping. Assembly-based SNP approach was performed using genome assemblies FASTA via Roary pangenome pipeline, based on alignment of 117,731 core-SNPs (1,741 core genes).

3.5.3.3 Inter-hospital strain transmission indicated by sequencing data

Although most twin-pairs harbour identical strains, indicating the role of maternal/vertical transmission (Figure 3.18), there are two important clusters suggestive of intra- and inter-hospital *C. perfringens* transmission events that took place during the 2-year clinical study. This section will discuss cluster IIIc that involved 6 neonates residing in both QCCH and SMH, carrying identical *C. perfringens* strains (0-6 SNPs) over an overlapping period of approximately 35 days (Figure 3.20A-B).

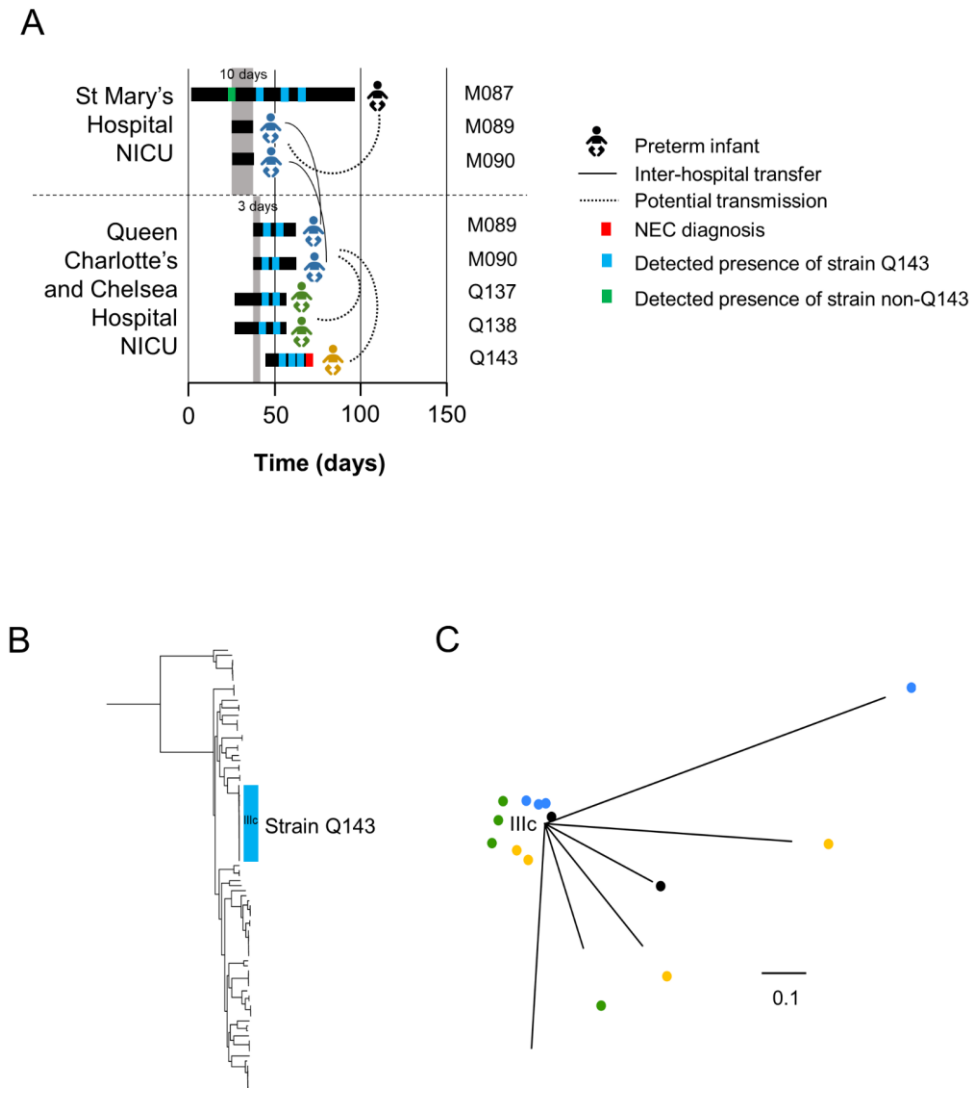


Figure 3.20: Potential inter-hospital transmission of *C. perfringens* strain Q143

(A) Epidemiological map of 6 preterm infants in both SMH and QCCH NICUs carrying NEC-associated *C. perfringens* strains. Same coloured infant symbols indicate twins. Black bars indicate period of hospitalisation. **(B)** Phylogenetic position of strains Q143. **(C)** Unrooted tree of 16 isolates from 6 infants constructed from 9 SNPs (strains Q143; lineage IIIc). Coloured symbols are indicative of the infants as in previous figures.

Lineage IIIc comprises 16 genetically identical isolates (longitudinal samples) from 6 neonates, having only a range of 0-6 SNPs between them (Appendix 1 Table S4; Figure 3.20C), fulfilling the criterion of being isogenic or identical at strain level, confirmed by ANI>99.99% (Table 3.6). This strain was designated Q143 after the ID of the neonate who carried this strain and who developed NEC. Q143 was genotyped as an extra-virulent strain associated with NEC representations (section 3.5.2), possessing significantly more virulence genes ($P<0.05$) than the proposed commensal *C. perfringens*.

With limited sampling of *C. perfringens*, potential transmission routes were likely to be sourced from SMH when SMH-residing neonates M089, M090 (twin pair) and M087 carried strain Q143 over time (Figure 3.20). M089 and M090 were transferred to QCCH after the first 10 days in SMH ward due to medical reasons while M087, having the same strain Q143, was never transferred to any other hospital, suggesting within-hospital transmission of strain Q143 from M087 to M089-M090 or vice-versa. Importantly, twin-pair Q137 and Q138 (within QCCH NICU) carried Q143 strain 3 days after the arrival of M089 and M090. Neonate Q143, a baby residing with QCCH NICU, was colonised with strain Q143 at 3 different time-points prior to NEC development (with significant overgrowth of *C. perfringens* in the gut microbiome as indicated by the 16S rRNA gut microbiome analysis), eventually this neonate succumbed to NEC, diagnosed as Bell II/III, and ultimately died.

This data demonstrates the transmissibility of *C. perfringens* extra-virulent-genotypes between neonates in two hospitals, highlighting the importance of high-resolution WGS techniques for tracking bacteria at the strain level within health care settings.

3.5.3.4 Within-ward transmission of *C. perfringens*

Based on phylogenetic analysis, within-ward transmission of multiple *C. perfringens* strains was also observed (Figure 3.21). Strain 1 first isolated from neonate Q088 (who eventually developed NEC), was detected in another neonate Q106 (control infant) 67 days after its first detection and close to 30 days after neonate Q088 passed away due to NEC, suggesting strain 1 is a persistent strain within the hospital ward. Strain 2 (extra-virulent genotype) was isolated both from Q088 (NEC) and the matching twin sibling Q087 (did not develop NEC), suggesting this strain could be due to maternal transmission as it was not detected in any other infants during the two-year period.

Neonate Q106 carried three different strains (from three different genetic clusters) during 100-day stay at QCCH NICU, demonstrating the underestimated diversity of *C. perfringens* strains present within an individual.

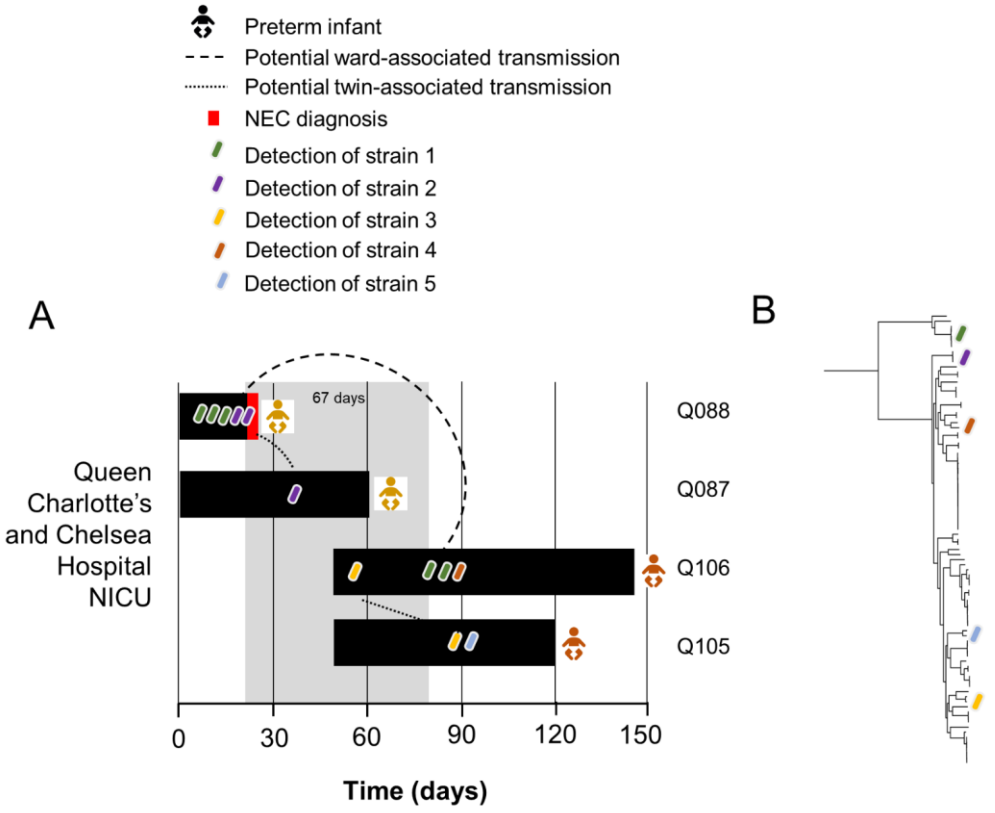


Figure 3.21: Inference of intra-hospital transmission of multiple *C. perfringens* strains

(A) Epidemiological map of 4 preterm infants in QCCH NICU showcasing the potential within-ward transmissions of NEC-associated *C. perfringens* strains based on WGS data. Strain identity is defined by SNP distance 0-6 SNPs. Black bars indicate hospitalisation periods of infants. Same-colour infant symbols are indicative of twins. **(B)** Phylogenetic positions of strains 1-5 in the mid-point rooted maximum likelihood tree.

3.6 Discussion

This WGS study represents the first of its kind to investigate preterm-associated *C. perfringens* isolates obtained from a preterm cohort of 369 infants collected over 2 years. These analyses identified hypervirulent NEC-associated genotypes, putative avirulent genotypes, and inferred potential intra- and inter- hospital transmission of *C. perfringens* hypervirulent strains via in-depth phylogenomic analyses.

Previous studies have indicated that a 'typical' preterm infant intestinal microbiota harbours a higher abundance of *Enterococci*, *Enterobacteriaceae* and *Staphylococci*,

and a lower abundance of beneficial bacterial genera *Bifidobacterium* and *Lactobacillus* [298,299]. Importantly, healthy term-infant gut microbiota is dominated by the 'probiotic' genus *Bifidobacterium* (particularly in vaginal-delivered breast-fed infants), which is linked with improved infant wellbeing [300], and several clinical studies have used probiotic supplementation to reduce NEC incidence [268,286,301]. From the metataxanomic profiling it appears that abundance of *Clostridium* species (including *C. perfringens* confirmed by culture) correlates negatively with *Bifidobacterium* levels, indicating an antagonistic relationship as shown in infant M087, who was also shown to harbour Q143 NEC-associated *C. perfringens* strains (Figure 3.6). *Bifidobacterium* has been demonstrated, both *in vitro* and *in vivo,* to antagonise specific pathogens including *C. perfringens*, although the mechanisms underpinning these findings is currently unclear and further experimental validation is required to identify the most effective strains/species of *Bifidobacterium* that may inhibit *C. perfringens*-related infections [189,302].

It is apparent that the resident microbiota plays a role in NEC, however WGS analaysis also indicated presence of 'hyper-virulent' *C. perfringens* strains in NEC infants. PFO gene *pfo* (or, *pfoA*) was shown to be more abundantly encoded in all 4 NEC isolates (100%) compared to control isolates (38%; $P<0.05$), indicating the potential virulence role of this accessory toxin, which is also known to act synergistically with α-toxin in the development of gas gangrene and necrohaemorrhagic enteritis in calves [84,87]. This pore-forming toxin is known to potentiate the effect of ε-toxin in a mouse model of type D enterotoxaemia that primarily occurs in sheep and goats, which disease is mainly characterised by sudden death with macroscopic identification of necrotic and hemorrhagic lesions in the small intestine, which is similar to clinical findings in NEC [303,304]. Importantly, current dogma suggests that PFO is almost universally encoded by *C. perfringens*, however in this study only 45% of the *C. perfringens* strains were identified to encode PFO. These data, which includes strains isolated from healthy infants, suggests a more 'pathogenic' nature for PFO-carrying strains, as most PFO-carrying *C. perfringens* currently reported have been isolated from diseased patients/individuals/animals [17,25]. Additionally, this underrated 'accessory toxin' could be included in the routine toxin typing scheme in reference laboratories, as this may indicate a risk marker for pathogenic *C. perfringens* strains in relation to preterm

communities. To define what a 'commensal' *C. perfringens* is, with respect to toxins, will require a larger-scale genomic study including more healthy control samples.

Alongside well studied virulence traits, it was notable that other less well studied factors were also identified to be enriched in the NEC isolates, including collagen adhesin gene *cna*. This collgen adhesin has been associated with *C. perfringens*-porcine enteritis, and chicken NE [230,305], which indicates that *cna* may facilitate adhesion of *C. perfringens* within the premature gut, potentially increasing risk of NEC-associated pathology.

HGT plays a key role in emergence of hyper-virulent strains that are associated with severe infection. Importantly, analysis of strains indicated that plasmids detected in NEC isolate Q143.CN8 and several other strains encoded virulence elements *cnaC, cpb2* (variant 2) and AMR elements *tetA(P)* and *tetB(P)*. Comparisons with a similar known plasmid pFORC3, indicated that these plasmids are larger, and appear to encode additional virulence determinants such as ABC transporters. Notably, CPB2-var2 encoded on the plasmid (linked to IS*1151* in plasmid pCPF5603 previously) also known as CPB2h1 (detected in human isolates), was demonstrated experimentally to be >10 times more potent than CPB2h2 (CPB2-var1; linked to IS*1470*-like; more similar to veterinary CPB2) in Caco-2 cells, which was previously explained as possible host-adaptation [89], and may link to the enhanced pathogenesis of these NEC isolates. Importantly, gentamicin (an antibiotic frequently administered to preterms residing in NICU), and streptomycin were also reported to induce the expression of CPB2 in *C. perfringens,* which supports the accessory role of this toxin [90].

As both NEC isolates Q215.CN1 and Q088.CN12 were predicted to encode intact plasmid pCP13, this suggests a role for factors encoded on this plasmid to be involved in NEC onset: CPB-var1 (more similar to veterinary CPB2, suggesting the veterinary source of these isolates), collagen adhesin and ABC transporter (important in AMR). Furthermore, the detection of plasmid pM096.3C as reference plasmid pF262A (from bovine case isolate F262) suggests another potential veterinary origin of this isolate, indicating a potential zoonotic reservoir of these *C. perfringens* strains [149].

Interestingly, previously unreported factors, such as surface layer, or S-layer, were also detected within these *C. perfringens* genomes. Previous studies have shown that the S-layer plays an important role in both the pathogenesis and intestinal colonisation

in pathogenic *C. difficile* [307,308]. However, there has been no S-layer reported in *C. perfringens,* although certain surface envelopes have been suggested [309]. S-layer homology domain-containing protein was inferred to be encoded exclusively in lineages II and III isolates (extra-virulent clusters), which may indicate that S-layers in *C. perfringens* could be associated with the enhanced virulence capacity of these clinical strains, however further studies would be required to confirm these findings.

Tracking of potentially hyper virulent strains is important, particularly within health-care settings. Indeed, inter-hospital transmission, traced via WGS techniques, has been used previously [310,311]. Spore-forming *C. perfringens* is expected to survive harsh environments, including on hospital surfaces, by forming endospores, as demonstrated in the related pathogen *C. difficile.* Indeed, this analysis indicates long-term persistence of *C. perfringens* strains in a sterilised hospital environment, as a particular strain re-appeared in the same ward, but in a different infant, after 67 days (Figure 3.21). This could indicate the circulating strain is either present in the environment, or the reservoir could as well be healthcare staff [312]. WGS surveillance could therefore be applied to track certain high-risk clones, as suggested by this work, i.e. extra-virulent clusters of *C. perfringens,* to minimise the risk of *C. perfringens*-associated NEC as one of the major factors in preterm NEC [15,30]. However, a larger-scale study is necessary to gain an in-depth understanding of *C. perfringens* transmission in different hospitals, tracing the routes of transmission, which is essential in infection control [312].

Interestingly, lineages IV-VI consist primarily of strains isolated from healthy preterm infants, and which appear less virulent based on the range of virulence factors encoded. Circulation of these strains was also apparent, which also indicates hospital ward-associated strains, or direct transfer from mothers (vertical transmission) to infants during birth, as exemplified by those strains existing in matching twin pairs. Further studies examining strains isolated from mothers pre- and post-birth would be required to confirm this hypothesis.

A key limitation of this study is the number of NEC strains isolated from *C. perfringens*-associated NEC patients (n=4) compared to control patients (n=33). As the prevalence of NEC is low (however fatality is high) among premature babies (5-15%), isolates from these case patients represent a realistic isolate number from a

relatively large longitudinal clinical study involving 369 individuals over a period of two years [8,12,73]. To date, no specific study has ever described NEC-associated *C. perfringens* from a genomic perspective even though this pathogen has been strongly implied in many clinical cases [1]. Although this remains an important caveat of this current investigation, these data provide novel insights into this unsolved question of *C. perfringens*' involvement in NEC via genomic approaches.

In conclusion, these data highlight that 'hypervirulent' genotypes of *C. perfringens* may be associated with preterm NEC, and that these strains may be transmitted within and between hospital wards. Identification of several important virulence genes including PFO may be vital for NEC pre-diagnostic purposes, which requires further laboratory validations. Notably, avirulent cluster of isolates were associated with 'healthy' preterms indicating the potential binary nature of this species, a pathogen and a commensal. Nevertheless, experimental confirmation will be necessary to determine the differential phenotypes of isolates in both clusters.

3.7 Future research directions

This work represents the first reported large-scale screening of *C. perfringens* isolates from both NEC and healthy preterms (n=369), however, to more fully explore *C. perfringens* populations in preterm infants, a larger-scale multi-site study is required, including more frequent clinical samplings and isolation. Alternatively, in-depth shotgun metagenomics could be implemented as culture-independent computational approach to acquire individual *C. perfringens* genomes at strain level [313]. This type of study would help define the potential role of e.g. PFO in preterm NEC, highlighting areas for experimental studies into underlying disease mechanisms.

Tracking pathogens is vital to infection control within hospitals. Real-time surveillance of *C. perfringens* could be proposed in NICUs using new rapid WGS approaches e.g. Nanopore sequencing, which allows bacterial genome sequencing in under 2 hours [314]. This real-time technique also allows rapid pathogen diagnosis via metagenomics in NICUs, which time-saving approach (compared to culturing-based method) may help identify presence of hypervirulent *C. perfringens* strains potentially <5 h (from clinical sampling to pathogen genotype identification) that could be critical in particularly NEC diagnosis [315]. Hyper-virulent *C. perfringens* strains could be

predicted based on full toxin profiles and virulence genes including *pfo* and *cna,* could act as pre-diagnosis biomarkers for potential development of microbial-NEC. For routine *C. perfringens* typing, NEC-associated virulence genes including *pfo* and *cna* should be included in PCR screening, in addition to the conventional toxinotyping based on 6 typing toxins (α-toxin, β-toxin, ε-toxin, ι-toxin, CPE and NetB) for early prediction of microbial-NEC.

To confirm the novel plasmid findings, long-read sequencing techniques e.g. PacBio and Nanopore should be implemented to validate the plasmid content. However, although there are several very good *in silico* tools currently used to explore plasmid content, this is not definitive and thus plasmid isolations should be performed.

Crucially, to probe gene-specific disease determinants in *C. perfringens*-associated NEC a robust and clinically relevant animal model is required, as current *in vivo* models do not replicate the infectious component. Therefore, to advance these pathogenesis studies, and expedite development of preventative therapies development of a suitable *in vivo* model is required, ideally an economical model such as a mouse model.

Chapter 4 Phylogenomic analysis of diarrhoea-associated *C. perfringens* identifies isogenic strains in multiple outbreaks, and novel virulence-related features

4.1 Introduction

This chapter described the phylogenetic and virulence analyses of 109 representative *C. perfringens* isolates recovered from non-foodborne diarrhoea (n=35), associated with care home (CH) residents, and FP patients (n=74), obtained between 2011-2017 in England and Wales from Public Health England (PHE). Isolates were whole-genome sequenced and in-depth bioinformatics analyses carried out to determine the epidemiology of these *C. perfringens* outbreak strains in tandem with full virulence profiling (toxin genes and AMR genes), virulence plasmids detection and comparison, and accessory virulence potentials on a strain-level (Figure 4.1).

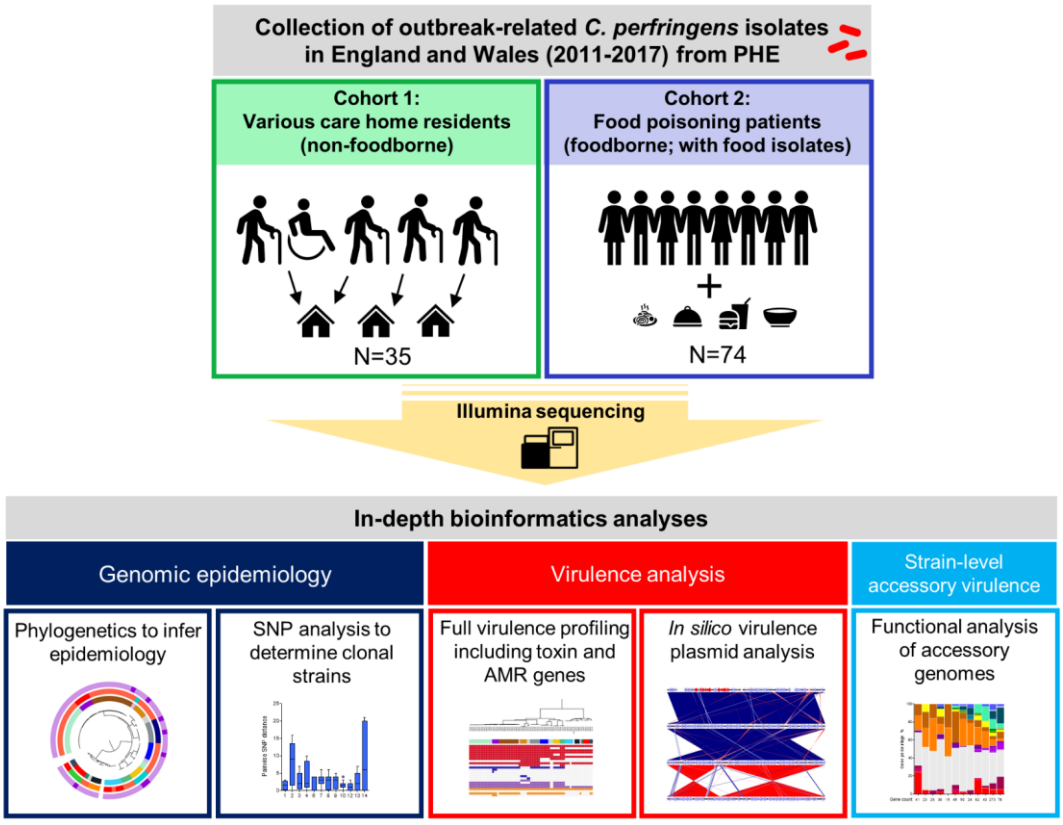


Figure 4.1: Schematic of dataset and phylogenomic analyses described in Chapter 4

4.2 Background

C. perfringens is an important pathogen known to cause disease in humans and animals [84]. This Gram-positive spore former has been associated with intestinal diseases in humans including foodborne and non-foodborne diarrhoea, and preterm NEC [8,24]. Notably, the pathogenesis of *C. perfringens*-associated infections is largely attributed to the wide array of toxins this species can produce, with >20 toxins identified [10,17] .

C. perfringens-associated FP, also termed acute watery diarrhoea, was first documented in the UK and USA in the 1940s [39]. First present-day record of *C. perfringens*-type A FP outbreaks dated back to 1943 when children ate school meals severely contaminated with rod-shaped bacteria including *C. perfingens* [169]. Typical symptoms appear within 8-14 hrs after food ingestion of live bacterial cells include intestinal cramp, watery diarrhoea without fever, and rare vomiting, which are normally resolved within 12-24 hrs [172]. Suspected *C. perfringens*-outbreaks are defined in the UK as 'two or more cases of illness with an aetiology suggestive of a toxin-mediated infection (point source distribution of cases with a predominance of diarrhoea and short duration of illness), and identification of a plausible source of *C. perfringens* exposure (based on analysis by Environmental or Health Protection teams), or culture of *C. perfringens* from one or more faecal specimens' [316]. Importantly, *C. perfringens* is the second most common foodborne pathogen in the UK after *Campylobacter*, however cases are under reported due to the mild and self-limiting nature of the illness, and current estimate suggest ~ 80,000 cases/annum [316,317].

In the UK, AAD and non-foodborne outbreaks of *C. perfringens* diarrhoea were frequently reported in the 1980s amongst the elderly, particularly in hospital settings [174]. With this type of illness, symptoms are more severe than foodborne diarrhoea, and are longer lasting (3 days to several weeks), often chronic, and infections are more likely to be spread amongst cases [153]. This type of *C. perfringens* infection has also been more commonly reported in elderly patients, especially those residing in CHs. In the North East of England between 2012-2014, these infections accounted for 83% of gastrointestinal CH outbreaks [316]. Although fatality is uncommon and hospitalisation rate of these illnesses is currently low, *C. perfringens* gastrointestinal infections cause

an estimated 55 deaths/year in England and Wales according to the UK Food Standards Agency [318].

The newly expanded and revised toxinotyping scheme classifies *C. perfringens* into 7 toxinotypes (type A-G accordingly; formerly 5 types A-E) according to the combination of typing toxins produced [19]. Human FP *C. perfringens* cases are primarily caused by type F strains, which produce CPE (encoded by *cpe* gene) [319]. This potent pore-forming toxin is also reported to disrupt intestinal tight junction barriers, which is associated with intestinal disease symptoms [112]. *C. perfringens,* and associated encoded toxins, have been extensively studied with respect to disease pathogenesis, with a strong focus on animal infections [87,135,138,230,320]. Recent studies analysing a range of diverse *C. perfringens* strains (from both animal and human-associated infections) has indicated a plastic and divergent pangenome, with a significant proportion of accessory genes predicted to be involved in virulence mechanisms, that may enhance host colonisation and disease initiation [17]. However, to date there are limited studies describing human outbreak-associated *C. perfringens* infections, and no known studies have utilised WGS data to probe the genomic epidemiology of associated strains [17].

To address this important knowledge gap, I applied in-depth genomics and phylogenetic analyses to whole genome sequences of 109 *C. perfringens* isolates associated with outbreaks or incidents of *C. perfringens* diarrhoea, either foodborne or non-foodborne-derived (mainly in CHs) collected between 2011-2017. Furthermore, I also identified the distribution of known virulence-related determinants including toxin and AMR genes, virulence-associated plasmid contents, within food as well as case isolates, and additionally probed putative functional capabilities of the accessory genomes. Importantly, I determined that isogenic strains were associated with 9 care-home (CH) outbreaks in North East England between 2013-2017, and additionally uncovered the significant involvement of virulence plasmid-carrying *C. perfringens* strains (prevalence of 82%) in these outbreaks.

4.3 Hypothesis and aims

Specific epidemic strains of *C. perfringens* are responsible for GI-associated outbreaks, specifically aiming to determine:

(1) Genomic epidemiology of FP and CH associated *C. perfringens* isolates.

(2) Genomic virulence features of these outbreak-linked strains, including plasmids and accessory genomes.

4.4 Study population

Two distinct groups of *C. perfringens* isolates were obtained via a collaboration with Dr Corinne Amar at the Gastrointestinal Bacteria Reference Unit (GBRU) of PHE. A total of 35 novel isolates from CH non-foodborne diarrhoea outbreaks mainly in North East of England [321], and 74 isolates from different individuals from FP outbreaks between 2011-2017 in England and Wales were used in this study, with samples sequenced using Illumina platforms at WTSI (isolates PH001-PH090) and PHE (isolates PH091-PH156). I performed all the genomic and phylogenetic analyses described in this chapter, with computing assistance from Mr Shabhonam Caim for the PacBio long-read assembly of reference genome NCTC8239 (from NCTC3000 project) used in this study. All the supporting materials are presented in Appendix 2. Clinical metadata and assembly statistics of isolates, and initial PCR identifications of *cpe* genes are detailed in Appendix 2 Table S4.1.

4.5 Results

4.5.1 Whole-genome based phylogenetic analysis reveals potential epidemiological clusters

Initially, I analysed the population structure of all isolates sequenced associated with FP (n=74), and cases of diarrhoea from CHs (n=35) (Figure 4.2). The majority of CH isolates nested within cluster VIII, which implied a regional-specific (North East England) relationship among strains. Quality of the genomic assemblies of draft genomes was also determined (Figure 4.2C); >70% of isolates assemblies were <200 contigs.

Separate analysis of CH isolates indicated two distinct phylogenetic clades (Figure 4.3A), with lineage I containing the reference genome NCTC8239 (from NCBI RefSeq database), a historical *cpe*-positive isolate (from salt beef) associated with an

outbreak of FP [39]. The remaining isolates were within another clade, residing within lineages II, III and IV (clusters assigned via hierBAPS analysis) indicating these CH isolates may be genetically distinct from typical FP isolates as in Figure 4.2A.

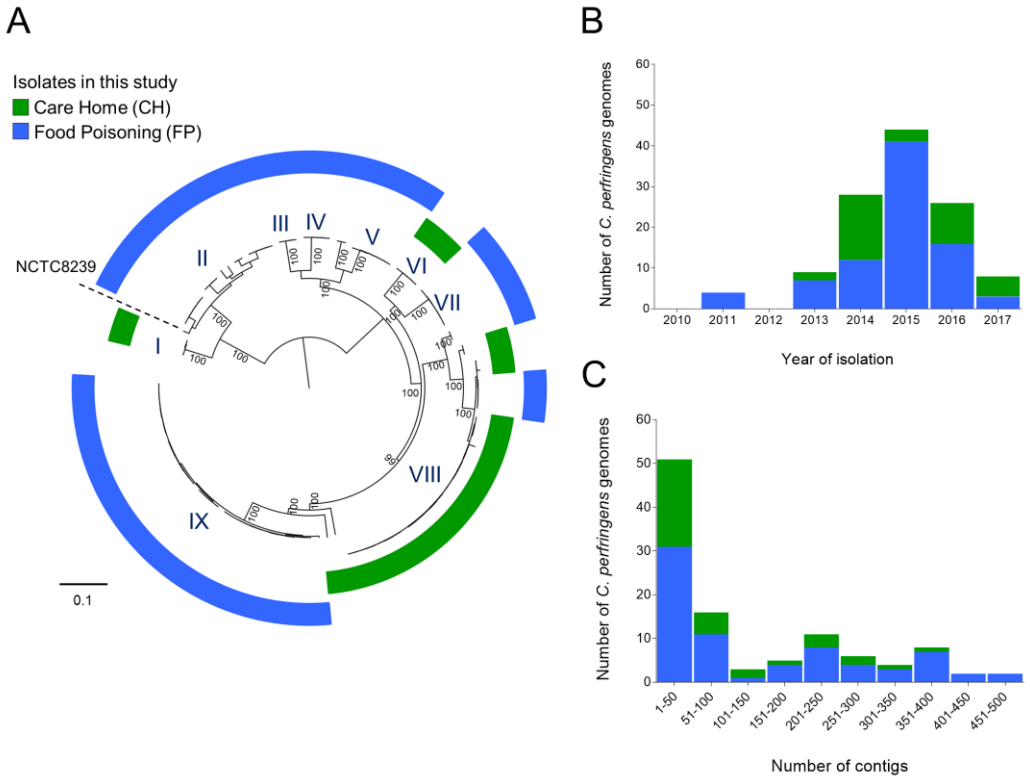


Figure 4.2: Population structure and sample distribution statistics for genome assemblies

(A) Population structure of 109 *C. perfringens* isolates analysed in this study. Mid-point rooted maximum-likelihood phylogeny inferred from 73,882 SNPs identified in 110 diarrhoea-associated *C. perfringens* isolates (including NCTC8239). The colour-coded rings indicated cohort-specific origins of isolates. Cluster VIII (green ring; clusters determined via hierBAPS clustering analysis) consists of primarily isolates obtained from multiple care home-associated outbreaks. Historical food poisoning isolate NCTC8239 was used as a public reference genome as indicated in the figure. Bootstrap values are represented in the tree. Branch lengths are indicative of the estimated nucleotide substitution per site (SNPs). **(B)** Temporal distribution of all 109 *C. perfringens* genomes included in this study. **(C)** Contig count distribution of *C. perfringens* genome assemblies in this study. More than 70% of the total assemblies are <200 contigs.

Interestingly, 18 closely-related strains from North East England were observed to nest within lineage IVc (Figure 4.3B), which were obtained from 9 different CH outbreaks between 2013-2017. This suggests a close epidemiological and geographical link within this lineage, as confirmed by core-gene SNP investigation, indicating all these isolates lie within pair-wise genetic distances of <80 SNPs (29.9 ± 16.6 SNPs; mean ± SD; Appendix 2 Table S4.2).

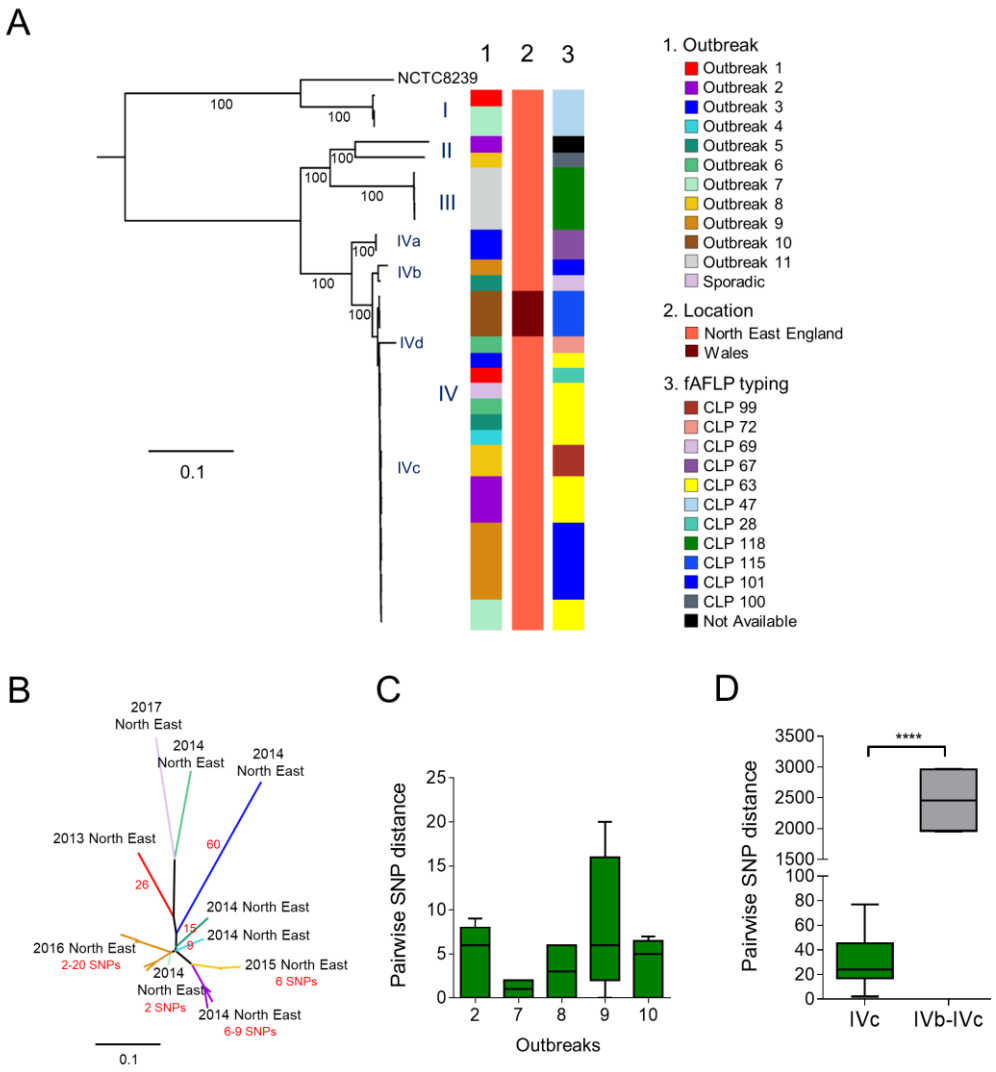


Figure 4.3: Phylogenomic analysis of care home-associated *C. perfringens* isolates

(A) Mid-point rooted maximum-likelihood phylogeny inferred from 64,560 SNPs (in core gene alignment) identified in 35 care home-associated *C. perfringens* isolates. The colour strips indicate diarrhoea outbreaks, location of outbreaks and fAFLP types respectively corresponding to the isolates. Branch lengths are indicative of the estimated SNP distance. Lineages and sub-lineages were determined via hierBAPS (level 1 & 2) clustering analysis. NCTC8239, a food poisoning isolate, was used as a public reference genome in this tree. Bootstrapping values are represented on the tree. **(B)** Unrooted maximum-likelihood tree (inferred from 191 SNPs in 18 genomes) of a sub-lineage IVc showing SNP distances in between 18 North-East England-derived isolates of individual outbreaks (labeled in locations and years, and SNP range in outbreaks; branches are colour-coded corresponding to individual outbreaks). SNP distances between branches are indicated in red numbers. **(C)** Pairwise within-outbreak core-SNP distance between isolates. **(D)** Pairwise outside-sub-lineage (IVb vs IVc) SNP comparison between isolates. Data: Mann-Whitney test. **** P<0.0001.

Notably, outbreak-associated CH isolates (within sub-lineage IVc) were all distinctly clustered within individual outbreaks, and pair-wise genetic distances were <20 SNPs (Figure 4.3C; 6.6 ± 6.6 SNPs; mean ± SD), compared to outside-sub-lineage (vs sub-lineage IVb) pair-wise SNP count of 2460 ± 503 SNPs, suggesting potential

involvement of an isogenic strain (genetically highly similar) within individual outbreaks (Figure 4.3D).

Importantly, the discriminatory power of WGS analysis was shown to be greater than the currently used fAFLP typing method, as fAFLP (type CLP 63, yellow-coded) failed to discriminate isolates from 6 different care-home outbreaks (CH outbreaks 2-7; Figure 4.3A), while SNP analysis clearly distinguished these strains (Figure 4.3B) [18].

The FP phylogenetic tree also demonstrated a clear separation of major clades (Figure 4.4A), where lineage I was a distinct clade from the remaining lineages II-VII. This may suggest two genetically distinct clusters of *C. perfringens* strains that are capable of causing diarrhoea in humans. Isolates from 3 foodborne-outbreaks in lineage VII appeared to be convergent in the phylogeny. Further analysis indicated that these isolates demonstrated geographical heterogeneity; isolates from 2 outbreaks in London (2013) were closely related, while isolates from North East (2015) were genetically more similar (Figure 4.4B). This may suggest geographical separation of a common ancestor at an earlier time point and may also indicate the potential widespread distribution of a closely related strain (Appendix 2 Table S4.3).

Isolates from the same outbreaks appeared to be 'clonal' and isogenic, as pairwise genetic distances were between 0-21 SNPs (mean genetic distance: 2.6 ± 2.7 SNPs; Figure 4.4C), when compared to same-lineage-between-outbreaks SNP distances of >1200 SNPs (Figure 4.4D). Importantly, food source isolates were not distinguishable from human clinical isolates (genetically similar, 0-16 SNPs) in 7 individual FP outbreaks (Figure 4.4A). This is consistent with the hypothesis that contaminated food was the source of these *C. perfringens* FP outbreaks, which included all meat-based food stuffs i.e. goat curry, cooked sliced beef, lamb, chicken curry, three-bird roast, cooked turkey and cooked meat (Appendix 2 Table S4.1). Notably, isolates from different geographical locations converged within the same lineages, suggesting circulation and spread of genetically-similar strains between cities and towns, or potentially a larger reservoir of *C. perfringens* within food chains or humans, and therefore warrants further large-scale investigation [322].

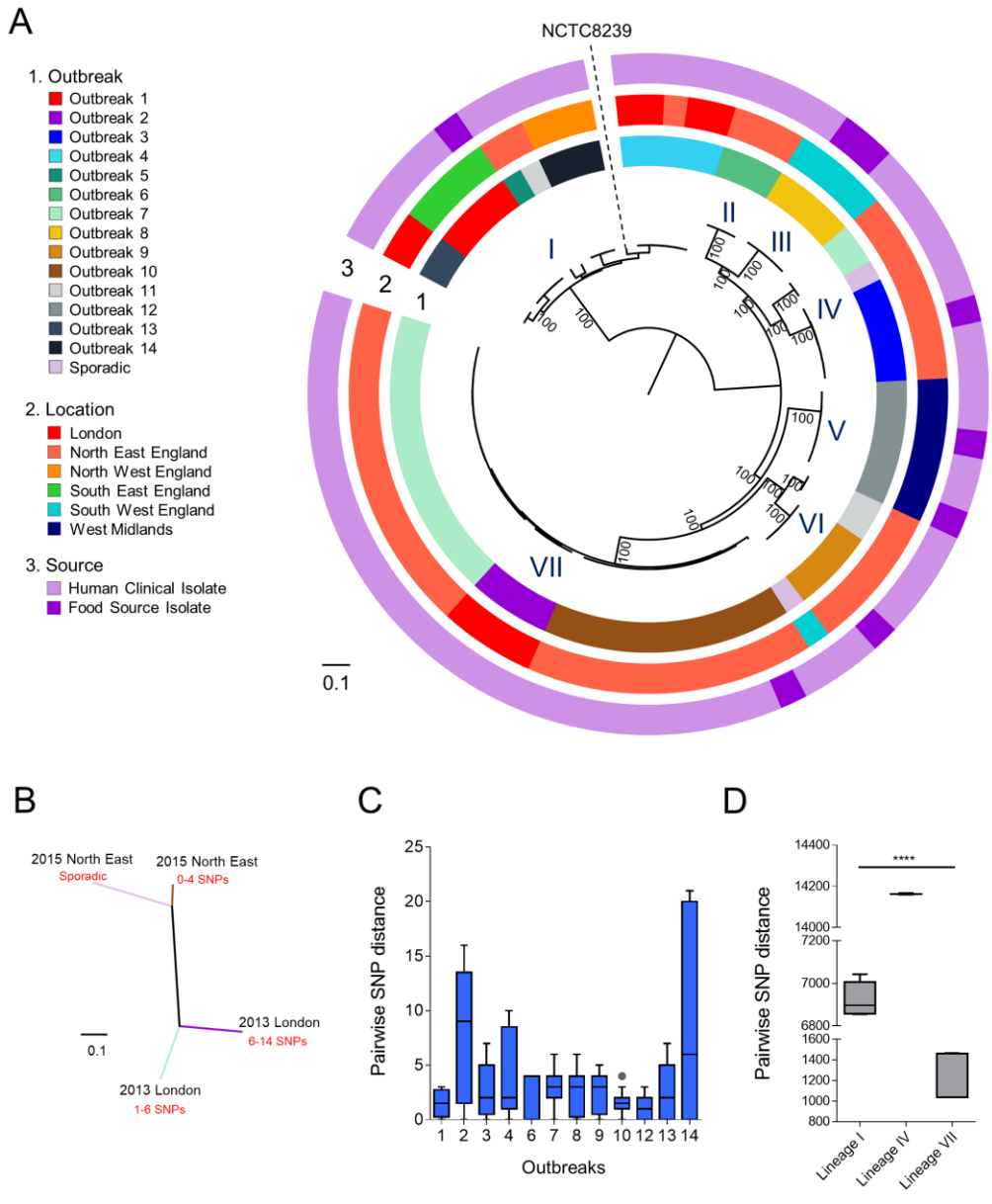


Figure 4.4: Phylogenomic analysis of 75 foodborne-associated *C. perfringens* isolates

(A) Mid-point rooted maximum-likelihood phylogeny of food-poisoning *C. perfringens* inferred from 70,613 SNPs (in core gene alignment) identified in 75 individual isolates. NCTC8239, a food poisoning strain isolated in 1949, (encodes *cpe* gene) is a RefSeq public reference genome. Lineages were determined via hierBAPS clustering analysis. Bootstrap values are represented in the tree. **(B)** Unrooted maximum-likelihood tree inferred from 2,505 SNPs in lineage VII (31 isolates). Four distinct clusters are identified in different outbreaks comprising genetically-similar strains (labeled in locations and years, and SNP range in outbreaks; branches are colour-coded corresponding to outbreak labels). **(C)** Pairwise core-SNP distance comparison in between isolates within outbreaks. **(D)** Pairwise core-SNP comparisons of within-major-lineage isolates in between individual outbreaks. Lineage I: outbreaks 1,4,13 and 14; Lineage IV: outbreaks 3 and 7; Lineage VII: outbreaks 2, 7 and 10. Data: Kruskal-Wallis test; **** P<0.0001.

4.5.2 Care home isolates possess higher virulence capacities

Diarrhoea symptoms associated with *C. perfringens* infection are due to the pore-forming toxin CPE, produced by *C. perfringens* type F strains [1,323]. Additional virulence determinants implicated include sialidase (NanI), which has been linked to enhanced intestinal attachment and plays a potential accessory role in enhancing CPE cell-toxicity [118,182], and also pore-forming toxin perfringolysin (PFO), a toxin known to act synergistically with PLC (phospholipase produced by all *C. perfringens* strains) to inflict damage on intestinal cells [87]. Moreover, drug-resistant *C. perfringens* were reported to be prevalent, particularly within poultry, thus AMR profiles of *C. perfringens* may be linked with prolonged *C. perfringens* associated-infections, and may hamper downstream treatments strategies, particularly in vulnerable or immune-compromised communities [131,183]. To probe these important virulence-associated traits isolates were screened for toxin and AMR genes, based on both genome assemblies and raw sequence reads.

Enterotoxin gene *cpe* was detected in all, but 4 isolates (PH017, PH029, PH045 and PH156 were *cpe*-negative), which was confirmed by PCR, with the exception of PH029, which was determined to be *cpe*-positive via PCR (96.4%; Appendix 2 Table S4.1). CH isolates (average 9.6 ± 1.0 toxin genes per isolate) encoded significantly more toxin genes ($P<0.001$; Mann-Whitney test) than FP (average 7.3 ± 1.9 toxin genes) isolates (Figure 4.5A). CH isolates in lineage II-IV generally possessed identical toxin profiles (Figure 4.6A), encoding significantly more toxin genes than FP isolates including colonisation-related sialidase-encoding genes *nanI*, *nanJ* and *nagH*, hemolysin PFO gene *pfo*, and *cpb2* (Figure 4.5C-G) that transcribes a vital accessory toxin CPB2, which is also associated with CPE-mediated pathogenesis [89]. CH isolates however, did not harbor many AMR genes; 6 isolates (out of 35; 17%) encoded tetracycline resistant genes *tet(P)*, one isolate encoded aminoglycoside resistant gene *APH(3')*, and remaining isolates did not encode any acquired AMR genes, other than the intrinsic AMR gene *mprF*.

In contrast, FP isolates had a more variable virulence gene profile (Figure 4.6B). Isolates in lineage I had identical toxin genes, including *cpe*, but these isolates did not encode toxins such as PFO, CPB2 and sialidase NanI, and only 3 isolates in this lineage carried tetracycline resistant genes (19%). Most isolates within lineages II-VII (92%; 53/58), which were genetically distinct from lineage I, encoded *tetA(P)*.

Notably, *tetA(P)* was significantly enriched in all FP isolates (74.3%; 55/74; P<0.0001; Fisher's exact test; Figure 4.5H) compared to CH isolates (17.1%; 6/35). Furthermore, most isolates in FP lineages II-VII also encoded toxin genes *cpe, nanI* and *pfo*, and 16 isolates (28%) possessed the accessory toxin gene *cpb2*, statistically these isolates (8.0 ± 1.5 toxin genes) encoded more toxin genes than those belonging to lineage I (4.9 ± 0.3 toxin genes; P<0.0001; Mann-Whitney test; Figure 4.5I) suggesting isolates in lineages II-VII were more virulent than those in lineage I. Importantly, plasmid detection indicated that virulence plasmid pCPF5603 was associated with CH isolates (97% in CH plasmid-carrying isolates; 34/35), while plasmid pCPF4969 was linked to FP isolates (92% in FP plasmid-carry isolates; 4/49; P<0.0001; Figure 4.5J).

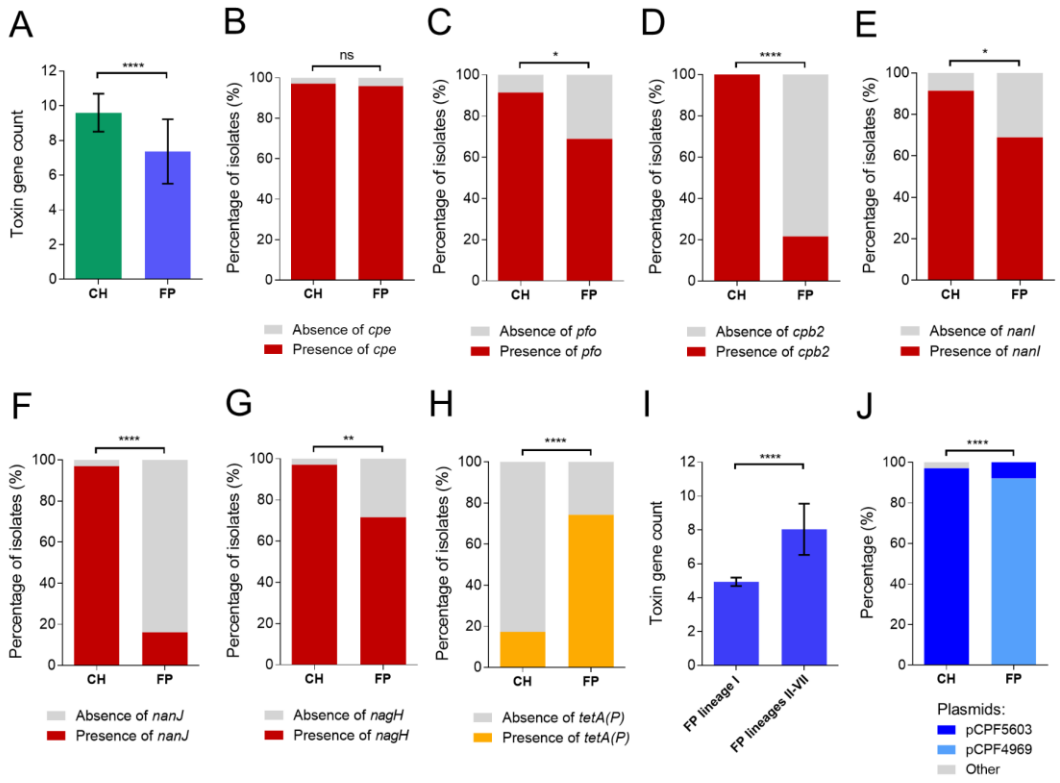


Figure 4.5: Comparisons of toxin gene count, enrichments of toxin, AMR gene and plasmids

(A) Toxin gene count comparison between CH and FP isolates. Data: mean ± SD. Mann-Whitney test. **** P<0.0001. Enrichments of variable toxin genes (B-G): **(B)** *cpe* **(C)** *pfo* **(D)** *cpb2* **(E)** *nanI* **(F)** *nanJ* **(G)** *nagH* **(H)** Percentage of AMR gene *tetA(P)*-carrying isolates in CH and FP groups. **(I)** Toxin gene count comparison of isolates in FP lineage I and lineages II-VII. Data: mean ± SD. Mann-Whitney test. **** P<0.0001. **(J)** Main virulence plasmids found in CH and FP isolates. Data: percentage of isolates that carry either pCPF5603 or pCPF4969 plasmids (excluding isolates not detected to carry any plasmids). Data from (B)-(H) and (J): Fisher's exact test (two-tailed). * P<0.05, ** P<0.01, **** P<0.0001, ns=non-significant.

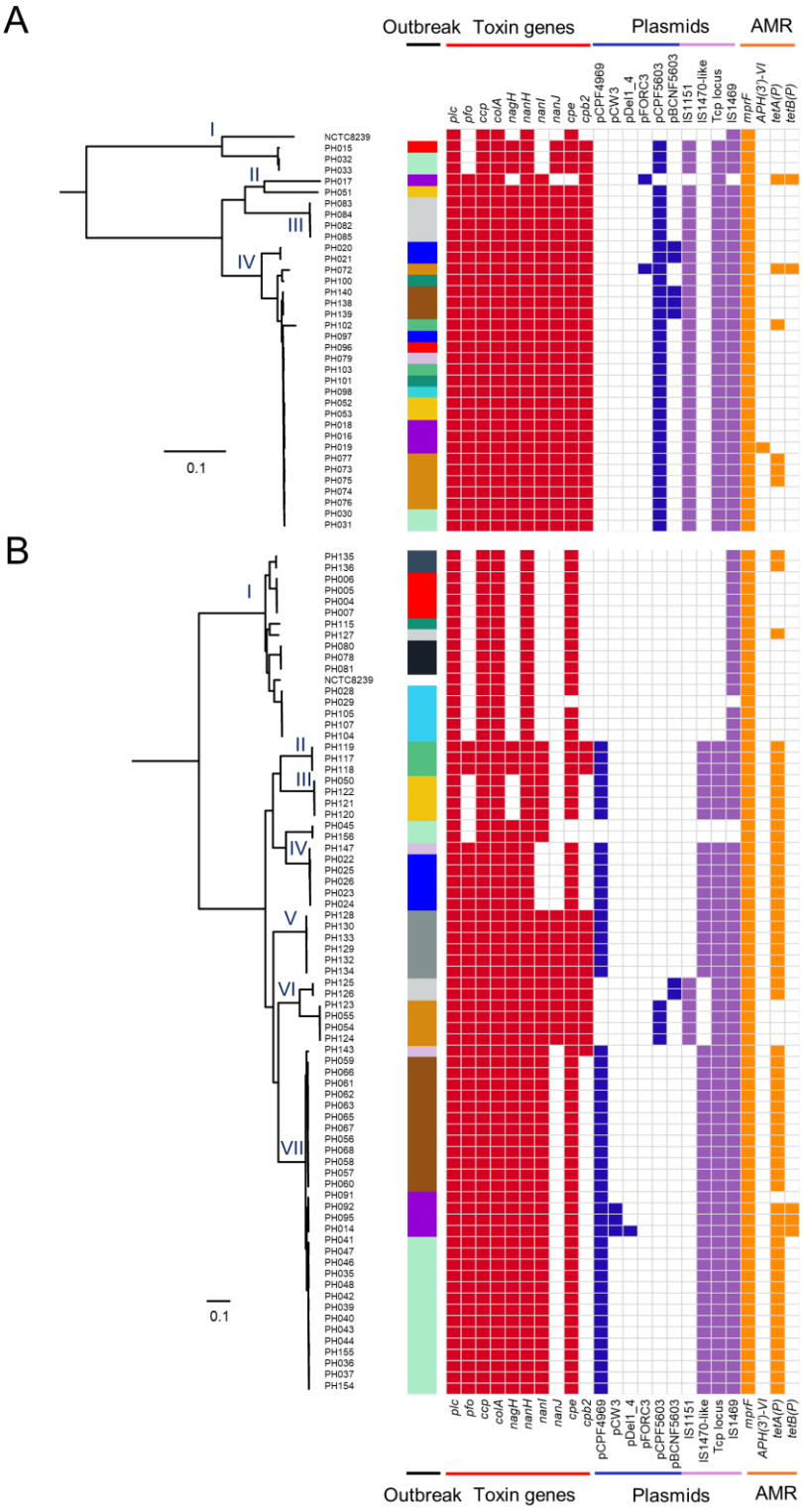


Figure 4.6: Full virulence profiles of *C. perfringens* isolates including virulence plasmids

Binary heatmaps displaying presence and absence of toxin genes, AMR genes, plasmids, plasmid-related sequences and *tcp* conjugative loci in corresponding isolates: **(A)** CH isolates **(B)** FP isolates. Outbreaks were colour-coded according to the colour system in previous figures in this chapter. Coloured cells represent presence of genes and white cells represent absence of genes. Heatmaps were generated in R.

4.5.3 Specific plasmid-associated lineages and potential CPE-plasmid transmission

Most disease-associated toxins secreted by *C. perfringens* are plasmid-borne, including CPE, although notably chromosomally encoded *cpe* strains are primarily linked to food-poisoning [70,324]. CPE has been associated not only with FP diarrhoea, but also a more severe and durable non-foodborne diarrhoea, which usually lasts >3 days to several weeks [1,325]. However, FP is thought to be caused by type F *C. perfringens* strains that chromosomally encode the *cpe* gene (e.g. NCTC8239), whilst non-foodborne diarrhoea is usually associated with plasmid-borne *cpe* strains (e.g. F5603) [173,322,325]. Data in previous section (Figure 4.5C) demonstrated that CH isolates (non-foodborne) predominantly harboured pCPF5603 plasmids (33/35 isolates; 97%; this plasmid encodes *cpb2* and *cpe*), whilst FP isolates carried primarily pCPF4969 plasmids (45/75 isolates, 60%; 45/49 plasmid-carrying isolates, 92%; this plasmid encodes *cpe* but not *cpb2*). Both reference plasmids pCPF4969 and pCPF5603 were originally extracted from non-foodborne sporadic-diarrhoea-associated *C. perfringens* type F strains F4969 and F5603 [45,326]. Findings were confirmed by genome-wide plasmid-specific sequence search (pCPF5603:IS*1151*, pCPF4969:IS*1470*-like and conjugative system *tcp* genes which are present in both plasmids, Figure 4.6) [43,327].

To further examine and confirm the predicted plasmids, plasmid contigs (complete single contig) were extracted from genome assemblies (3 isolates in both CH and FP groups), and compared with the reference plasmid sequence data (pCPF5603 and pCPF4969) from NCBI Nucleotide Database (Figure 4.7A-C). The extracted plasmid sequences from FP and CH isolates closely resembled the respective reference plasmids, with near-identical nucleotide identity (>99.0%), plasmid size and GC content (Table 4.1; Figure 4.7B-C), thus supporting the findings that these two near-identical plasmids (pCPF4969 and pCPF5603) were present in these isolates.

These data suggested potential circulation and transfer of plasmids in outbreak-associated *C. perfringens*, given that pCPF4969 has previously been shown experimentally to transfer among *C. perfringens* strains via conjugation using the encoded *tcp* loci (*cpe*-positive to *cpe*-negative strains) [292]. The data showed that plasmid-*cpe C. perfringens* strains (these strains did not encode another copy of *cpe* on the chromosome, confirmed by in-depth sequence search) were predominantly associated with FP (82.6%; 90/109), rather than previously reported chromosomal-*cpe*

strains (17.4%; 19/109 in this study). Importantly, plasmid transfer may have occurred in CH 'outbreak 7' isolates (n=4), as two isolates resided within lineage IV, whilst the remaining two isolates nested within the genetically distant CH lineage I (genetic distance >10,000 SNPs), however all 4 isolates harboured identical plasmid pCPF5603 (Figure 4.6). This analysis also indicated that multiple distinct strains, but carrying the same plasmid, may be implicated in these outbreaks.

Reference plasmid	Predicted Plasmid	Genome features			K-mer identity (%) (Read-based)	Similarity by BLASTN (Assembly-based)	
		Size (bp)	GC (%)	CDS		Identity (%)	Coverage (%)
pCPF5603	pPH077	75,398	25.54	87	99.64%	99.98	100.00
	pPH032	75,402	25.53	87	99.58%	99.98	100.00
	pPH031	75,401	25.54	87	99.63%	99.97	100.00
pCPF4969	pPH062	70,638	26.59	74	99.66%	99.98	100.00
	pPH041	70,811	26.57	75	99.53%	99.97	100.00
	pPH050	70,624	26.57	75	99.40%	99.98	100.00

Table 4.1: Similarity comparisons of predicted plasmids with reference plasmids

Predicted plasmids were compared with reference plasmids pCPF5603 and pCPF4969.K-mer identity (based on FASTQ) determined by PlasmidSeeker; BLASTN sequence similarity and coverage determined by ABRicate v0.5. Genome features of plasmids were extracted from Prokka annotations.

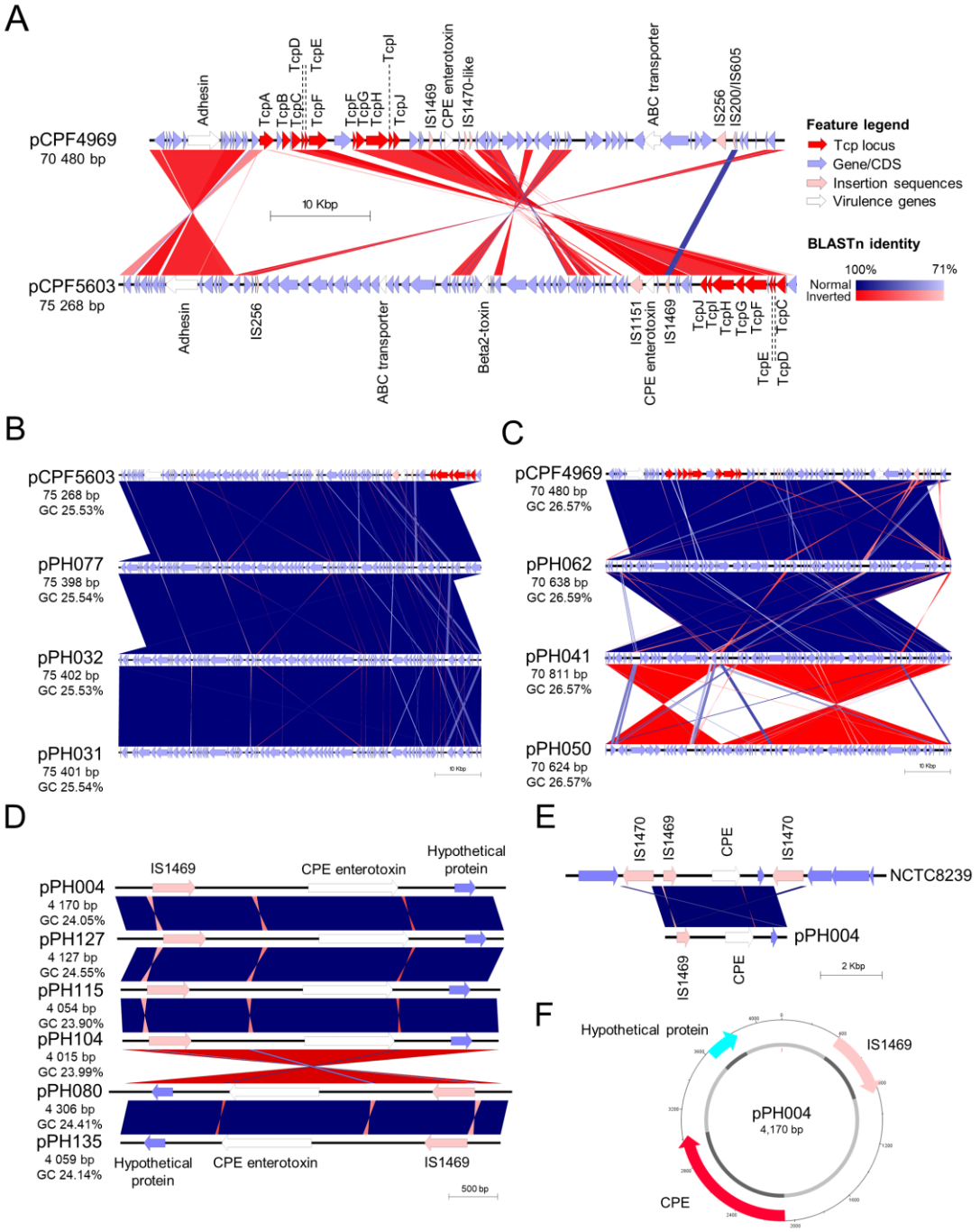


Figure 4.7: Investigations on predicted plasmids carried by CH and FP isolates

(A) Comparative genomic visualisation of reference plasmids pCPF4969 and pCPF5603 with annotated features. **(B)** Genomic comparison of pCPF5603 plasmid and pCPF5603-like plasmids computationally extracted from CH genomes. **(C)** Plasmid comparison in between pCPF4969 plasmid and FP-isolate predicted plasmids. **(D)** CPE-regions (Tn*5565*) extracted computationally from FP lineage I representative isolate assemblies (single contigs in the assemblies but without flanking IS1470 elements). **(E)** An extracted 11-kbp region of NCTC8239 that encodes *cpe* and flanking IS1470 elements compared to predicted plasmid pPH004. **(F)** Circular form Tn*5565* from PH004, illustrated using DNAPlotter v17.0.1. Figures A-E were produced using Easyfig v2.2.2.

Interestingly, within the FP phylogeny there was a distinct lineage of isolates (lineage I) that may encode *cpe* chromosomally, as no known plasmids, nor plasmid-specific sequences were detected. These data support previous studies which indicated that *C. perfringens* with chromosomally-encoded *cpe* are genetically divergent from plasmid-*cpe* carriers, and was further evidenced by historical chromosomal-*cpe* strain NCTC8239 (known to carry *cpe* on chromosomes) nesting within this lineage (FP lineage I; Figure 4.6) [322,326,328].

To further investigate this hypothesis using genomic approaches, the *cpe*-encoding region (complete single contig from high-coverage assemblies) was extracted from representative isolates in lineage I (n=6), and comparative genomics was performed (Figure 4.7D). These consistently smaller (~4.0-4.3 kb) contigs were almost identical in nucleotide identity (>99.9%) when compared with the *cpe*-encoding region of chromosomal-*cpe* strain NCTC8239, confirming that these isolates possessed the same genomic architecture in the *cpe* encoding region as NCTC8239 (Figure 4.7E). Nevertheless, all these regions lacked the typical flanking elements IS*1470* that is proposed to be present in chromosomal-*cpe* carriers (confirmed by genome-wide screening of IS*1470*, no single IS*1470* element was detected in genomes of these isolates), and was confirmed as transposable element Tn*5565* without flanking IS*1470* (Figure 4.7E-F) [173].

Furthermore, PH029 was the only outbreak isolate not detected to encode *cpe* within the lineage I outbreak cluster, despite having a clonal relationship with PH028, PH104, PH105 and PH107 (Figure 4.6B), and indicated Tn*5565* loss may have occurred due to extensive sub-culturing (this is supported by initial PCR results in the lab being *cpe*-positive; see Appendix 2 Table S4.1). The data also indicates that *cpe* was closely associated with IS*1469*, independent of where it was encoded, as this IS was detected exclusively in all *cpe*-encoding genomes (100% sensitivity and specificity; Figure 4.6A-B).

4.5.4 Accessory genome virulence potentials

The 109 *C. perfringens* isolates analysed in this study had a pangenome consisting of 6,219 genes (Figure 4.8); 1,965 core genes (31.5%), and 4,254 accessory genes (68.5%). Based on these statistics, approximately 30-40% genes in any individual *C.*

perfringens strain are encoded within the accessory genome, potentially driving evolution and genome restructuring. Mobile genetic elements (MGEs) including plasmids, genomic islands and bacteriophages could potentially contribute to virulence, given the plasticity of the genome. To explore this in more detail, I further analysed the accessory genomes, comparing different sub-sets of *C. perfringens* isolates.

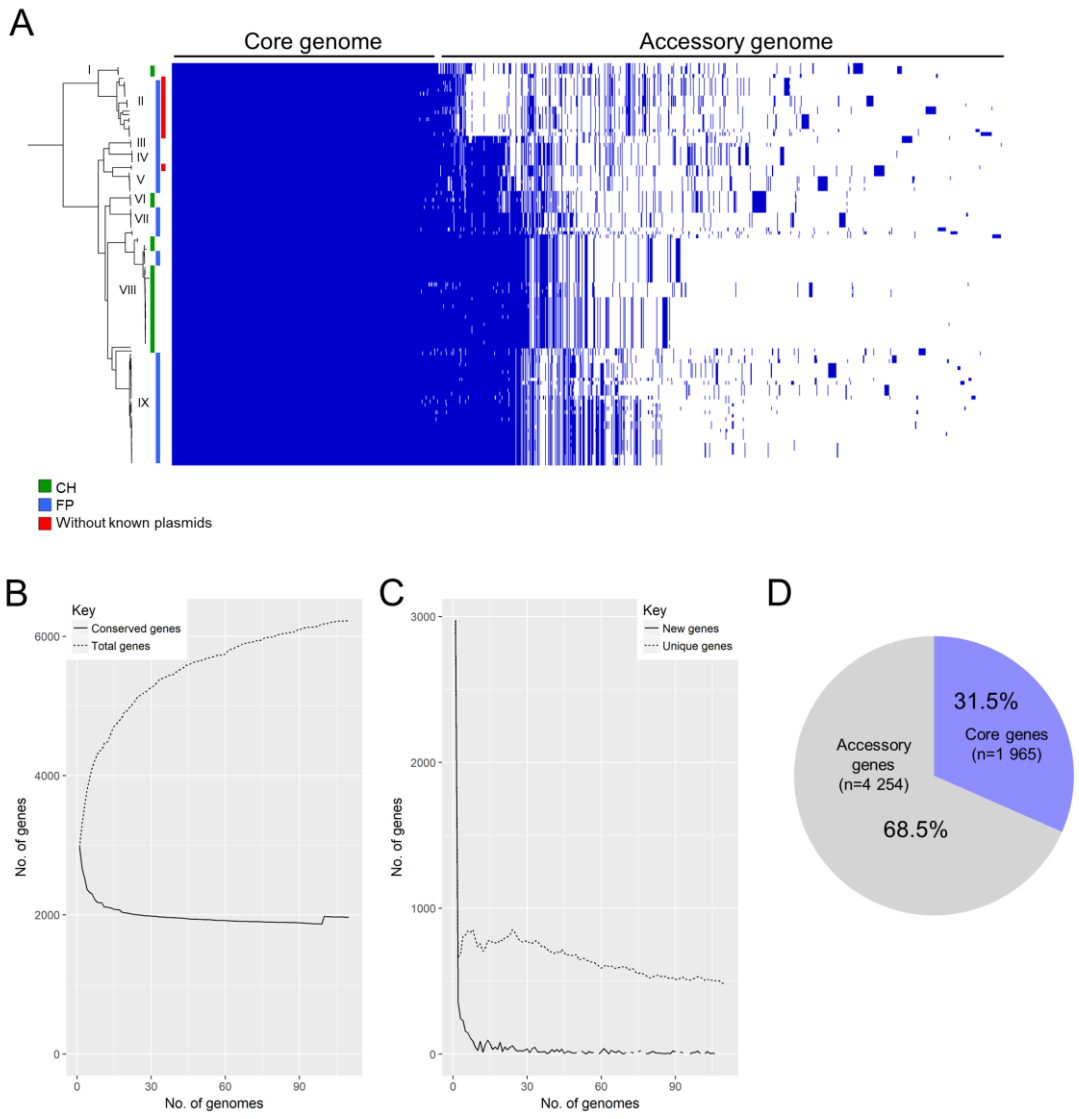


Figure 4.8: Pangenome analysis of 110 *C. perfringens* isolates in Chapter 4

(A) Linearised pangenome of 110 *C. perfringens* strains in this study including NCTC8239. Visualised by Phandango. Statistics showing addition of **(B)** conserved genes vs total genes and **(C)** New genes vs unique genes per genome during pangenome construction. **(D)** Proportion of core and accessory genes in the pangenome.

I first identified subset-specific genes using a bacterial pan-GWAS approach, with these genes further annotated based on NCBI RefSeq gene annotations and categorised under COG classes into 3 comparison groups: (1) CH vs FP; (2) FP outbreaks; (3) FP lineage I, FP lineage II-VII and CH-FP plasmid-CPE isolates (Figure 4.9A-D).

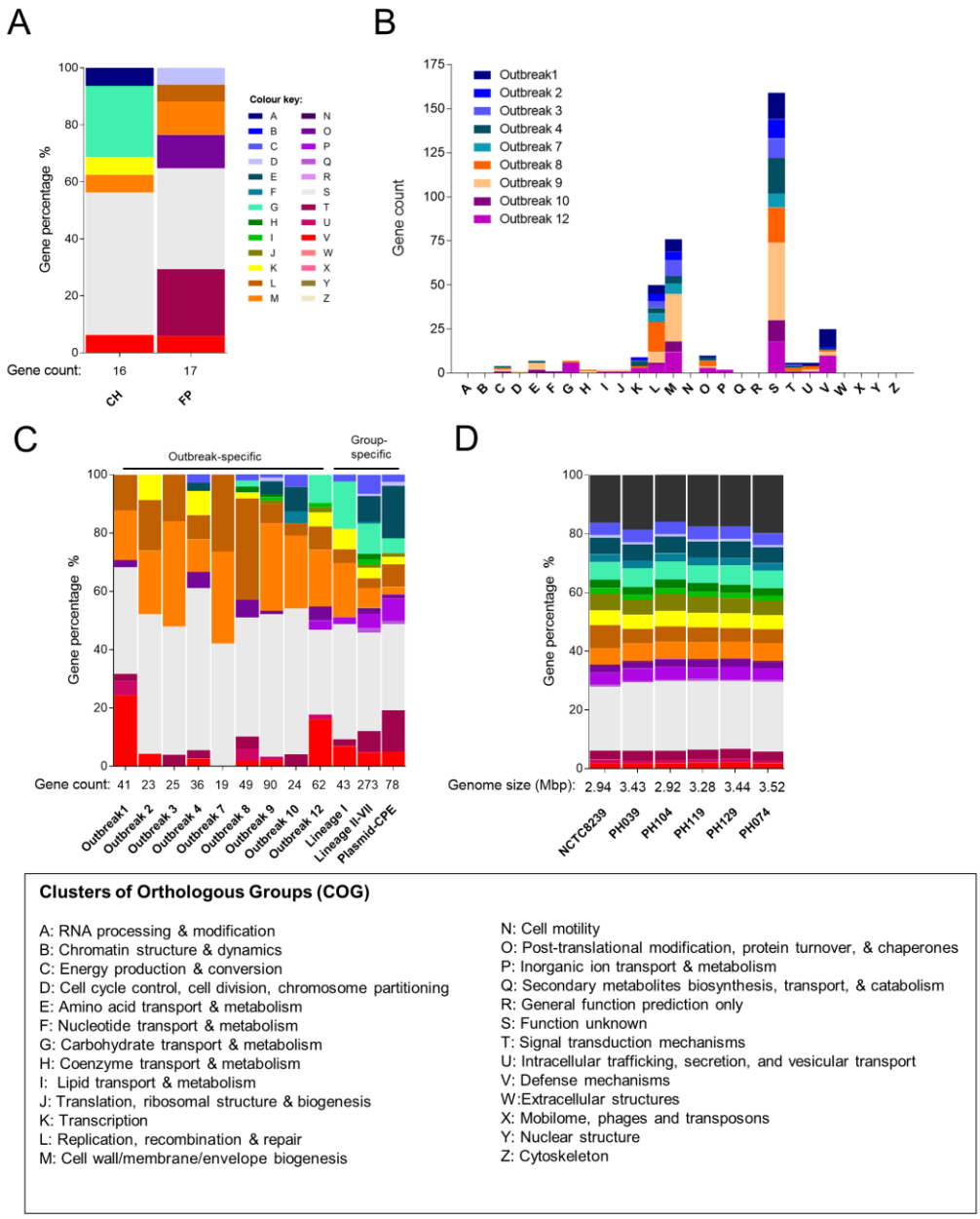


Figure 4.9: Functional analysis of *C. perfringens* genomes

(A) Subset-specific accessory genes comparison CH vs FP isolates. **(B)** Outbreak-specific isolate gene counts of COG-assigned food-poisoning accessory genomes. **(C)** Functional classification (COG) of representative outbreak- and group-specific accessory genes in FP isolates. **(D)** Whole-genome functional analysis of representative *C. perfringens* isolates including reference genome NCTC8239. Box (bottom): specific functional description of COG assignments used in this study.

Phosphotransferase system (PTS)-related genes (n=4) were encoded exclusively in CH isolates (Figure 4.9A and Appendix 2 Table S4.4). These genes might contribute to the isolates' fitness to utilise complex carbohydrates (COG category G) in competitive niches, like the GI tract [329]. PTS genes have also been related to virulence regulation in other pathogens including foodborne pathogen *Listeria monocytogenes* [330]. Heat-shock protein (Hsp70) DnaK co-chaperone was annotated in the FP-specific accessory genome, which may be involved in capsule formation [331].

Accessory genes specific to each FP outbreak were variable (Figure 4.9B-C), but three annotated functional classes were conserved; L (replication, recombination & repair), M (cell wall/membrane/envelope biogenesis) and V (defense mechanisms). Prominent genes detected in all isolates included phage-related proteins (n=49) (L, M and S), glycosyltransferases (n=37) (M), restriction modification systems (n=16) (V), transposases (n=9) (L) and integrases (n=8) (L). It was evident that most genes were associated with phages, seemingly a major source of MGEs, potentially showing isolates' adaptation to a particular niche.

Isolates within the FP lineage I (n=17) generally had significantly smaller genomes (genome size 2.95 ± 0.03 Mb including NCTC8239; Table 4.2), compared to remaining isolates in this study (3.39 ± 0.08 Mb; n=93; P<0.0001), and were most similar to reference genome NCTC8239 (ANI≥99.40%), and appeared to lack plasmids (Figure 4.6). Correspondingly, less group-specific accessory genes were present compared with other isolates in lineages II-VII (Figure 4.9C**).** Notably, multidrug transporter 'small multidrug resistance' genes were exclusively detected in FP lineage I isolates, whereas ABC transporters were more commonly encoded in plasmid-carrying isolates (virulence plasmids pCPF5603 and pCPF4969 carry various ABC transporter genes). The Mate efflux family protein gene was detected solely in lineage II-VII isolates.

4.6 Discussion

Using WGS data and genomic analysis this work investigated a representative sub-set of 109 gastrointestinal outbreak-associated *C. perfringens* isolates obtained between 2011-2017 in England and Wales in collaboration with PHE. Probing population variation, virulence-associated genes, plasmid contents and accessory genomes, this

study determined potential epidemic phylogenetic clusters and virulence determinants that were strongly associated with outbreak isolates.

Isolate	FP Lineage	Genome Size (bp)	GC (%)	ANI (%)
NCTC8239	I	3,008,497	28.19	100.00
PH004	I	2,923,443	27.87	99.42
PH005	I	2,934,006	27.93	99.43
PH006	I	2,928,275	27.89	99.42
PH007	I	2,946,980	27.89	99.42
PH135	I	2,912,738	27.90	99.40
PH136	I	2,905,546	27.91	99.40
PH115	I	2,977,232	27.87	99.51
PH127	I	2,924,057	27.93	99.47
PH080	I	2,983,246	27.90	99.47
PH078	I	2,987,740	27.92	99.47
PH081	I	2,984,289	27.89	99.47
PH028	I	2,994,736	27.96	99.64
PH029	I	2,954,967	27.92	99.65
PH104	I	2,929,035	27.90	99.64
PH107	I	2,914,363	27.91	99.66
PH105	I	2,944,409	27.93	99.65
Stats (mean ± SD)		**2.95 ± 0.03 Mb**	**27.92 ± 0.07**	**99.51 ± 0.10**

Table 4.2: Genome comparison of isolates within FP lineage I

Genome sizes were based on genome assemblies of isolates. ANI (%) of each genome was compared with reference genome NCTC8239.

Analysis of CH isolates (defined as non-foodborne outbreaks as no food samples were received by PHE that indicate foodborne illness) revealed a cluster of closely-related isogenic strains that carried plasmid pCPF5603, which encoded important toxin genes *cpe* and *cpb2* (Figure 4.3A-B). Moreover, these isolates also displayed identical virulence gene profiles, including the potential to produce up to 10 toxins, which are linked to symptoms such as diarrhoea and cramping. These findings were important within the context of disease control and prediction/ epidemiology tracking of transmission modes, as most recent outbreaks in North East England have been reported to occur within CHs, 83% (out of 46 outbreaks) reported in 2012-2014 [316]. Indeed, this cluster of clonal strains was geographically linked to CHs in North East

England, indicating a specific persistent clone was responsible for most gastrointestinal outbreaks in that region.

A recent investigation into food-chain fresh meat products (>200 samples) indicated significant contamination of *C. perfringens* especially in beef (~30%), followed by poultry and pork (both ~26%). Importantly, 90% of *C. perfringens* strains isolated from those meat products were *cpe*-positive, indicating food chain or farmed animals as potential reservoirs of entero-toxigenic *C. perfringens* strains [322,332]. Interestingly, 18% prevalence of *cpe*-positive *C. perfringens* strains had previously been reported in food handlers' faeces (confirmed via PCR), indicating a potential role of the human reservoir in FP outbreaks [333]. Moreover, presence of persistent identical *C. perfringens* genotypes within a CH setting was previously reported, which indicated that several individuals harboured identical strains (as determined via PFGE profiling) throughout a 9-month sampling period, which was linked to CH residence location, however none of these isolates were positive for the *cpe* gene [334]. Determination of these potential reservoirs in the spread of *cpe*-positive *C. perfringens* will necessitate a One-Health-approach, and large-scale WGS-based screening to ascertain the phylogenomics of strains isolated from diverse sources surrounding outbreak-linked vicinities, which is important in the context of public health, especially in immune-compromised populations including older adults residing within CHs.

Through computational analysis this study determined that plasmid-*cpe* (specifically pCPF5603 or pCPF4969) carrying strains predominated within the outbreak-related *C. perfringens* isolates (~82%). These two virulence plasmids, pCPF5603 and pCPF4969, encoded several important virulence genes including *cpe*, *cpb2* (pCPF5603 only), ABC transporter and adhesin (also known as collagen adhesion gene *cna*), that could contribute to enhanced fitness and colonisation potential of *C. perfringens* within the GI tract (Figure 4.7A) [50]. Interestingly, pCPF4969 also contains a putative bacteriocin gene that may allow *C. perfringens* to outcompete other bacteria in the wider gut microbiota, and thus overgrow, initiating disease symptoms [45].

Typical *C. perfringens*-associated FP was previously thought to be primarily caused by chromosomal-*cpe* strains, which commonly assembled into distinct clusters that lack the *pfo* gene, which was reflected in the FP lineage I data from this study [335-337]. This is due to their phenotypic capacity to withstand high temperatures (via production

of a protective small acid soluble protein), and high salt concentrations during the cooking process, in addition to the shorter generation time, when compared to plasmid-*cpe* carrying strains [60,70]. These chromosomal-*cpe* strains were also demonstrated in this study to possess a significantly smaller genome size compared to the rest of the isolates, and importantly they were not known to carry any plasmids. Nevertheless, plasmid-borne *cpe*-carrying strains (pCPF5603 or pCPF4969) have also been associated with previous FP outbreaks, albeit this was a relatively uncommon occurrence. Thus, it was surprising that these findings indicated nearly all CH outbreak isolated strains (97%; 34/35) carried a *cpe*-plasmid [323,324]. Notably, pCPF5603-carrying strains (encoding IS*1151*) have also been associated with FP in nursing homes in Japan (7 out of 9 isolates), and in this study, 34 out of 35 CH-associated isolates appeared to carry plasmid pCPF5603 [327]. The fact that plasmid-*cpe* strains could cause diverse symptoms including short-lived FP, and long-lasting non-foodborne diarrhoea, indicates a potential role for the gut microbiome in modulating *C. perfringens* infection. Previous studies indicate that elderly people residing in CHs have a less diverse and less robust microbiota when compared to those residing in their own homes. This extends to individuals colonised with *C. perfringens*, which may link to the ability for food-poisoning associated *C. perfringens* strains to overcome gut microbiota-associated colonisation resistance mechanisms and 'take-hold' and thus initiate disease pathogenesis [183,334,338].

Of note, 'chromosomal-*cpe*' was reported two decades ago to be encoded on a transposon-like element Tn*5565* (6.3 kb, including flanking copies of IS*1470*), which could form an independent and stable circular-form in culture extracts (losing both copies of IS*1470*) [292,326]. This transposition element TN*5565* was thought to be integrated into the chromosome at a specific site as a unit, however, I observed that Tn*5565* is potentially an 'episome' with both flanking IS*1470* excised, instead of chromosome-encoded, in 6 of the foodborne isolates, as indicated by computational analysis (Figure 4.7D). The fact that computational analysis failed to detect any *cpe,* IS*1469* (*cpe*-specific), and IS*1470* (Tn*5565*-specific) in the high-sequencing-coverage PH029 genome (317X sequencing depth/coverage) indicated that Tn*5565* can be lost or may be passed on to other *C. perfringens* cells. It should also be noted that flanking IS*1470* of Tn*5565* may not have been successfully assembled during the genome assembly process due to the repetitive nature of those sequences (short-read

sequencing). However, the loss of the whole element Tn*5565* was also observed in public genome NCTC8239 (Accession: SAMN02436239; Tn*5565* not detected in this 55-contig short-read assembly but isolate sequenced by NCTC3000 project with Tn*5565* detected in long-read PacBio assembly used in this study), supporting the hypothesis that this mobile genetic element with *cpe* can be potentially lost, possibly via sub-culturing as in the case of PH029. This strongly indicates the mobility of this transposable element, which might be critical for transmission and disease control. Hence, potential transfer of *cpe* on Tn*5565* to *cpe*-negative strains may be possible (although the underlying plasmid transfer mechanisms will need to be determined), thus enhancing spread of food-poisoning *C. perfringens* isolates. Further epidemiology studies and laboratory experiments would be required to fully probe these questions.

As WGS provides enhanced resolution to identify outbreak-specific clonal strains, this study highlighted the importance of implementing WGS for *C. perfringens* profiling in reference laboratories, in place of the conventional fAFLP [335,336,339]. Routine *C. perfringens* surveillance of the CH environment and staff could prove critical for vulnerable populations, as outbreaks could rapidly spread, and this approach could potentially pinpoint the sources of contamination, and eventually eliminate persistent *cpe*-strains in the environment [334]. In light of the potential rapid transmissibility of *C. perfringens cpe*-strains responsible for food-poisoning outbreaks, real-time portable sequencing approaches such as the MinION, could facilitate the rapid identification of outbreak strains, which had been recently been reported to identify outbreak *Salmonella* strains in <2h [314,340].

These data highlighted the genotypic and epidemiological relatedness of a collection of *C. perfringens* strains isolated from disease-associated cases from across England and Wales, and indicates potential circulation of these strains, and the potential impact of plasmid-associated-*cpe* dissemination, linked to outbreak cases. This study highlights that further WGS phylogenetic and surveillance studies of diversely-sourced *C. perfringens* isolates will be required to fully understand the potential reservoir of FP-associated strains, so that intervention or prevention measures can be devised to prevent the spread of epidemiologically important genotypes, particularly in vulnerable communities, including older adults residing in CHs.

4.7 Future research directions

Long-read sequencing of representative samples on PacBio or Nanopore platforms especially chromosomal-*cpe* isolates will confirm whether Tn*5565* that encodes *cpe* can be lost via extensive sub-culturing, as long-read techniques are more likely to capture the repetitive flanking regions of IS*1470* which was not detected in short-read (albeit very high coverage) genome assemblies.

Experimental confirmation of the presence of plasmids pCPF5603 and pCPF4969 can be determined in the laboratory via plasmid extraction and plasmid sequencing on a sub-set of CH and FP isolates. This will validate the findings using computational approaches in this study and potentially obtain novel plasmids not currently known in the databases. Additionally, transfer of these virulence plasmids can be tested using *in vitro* and *in vivo* models to understand the underlying mechanisms of FP and non-foodborne diarrhoea caused by *C. perfringens*.

There exists an opportunity to whole-genome sequence the complete collection of *C. perfringens* isolates at PHE to capture a better picture of the phylogeographical information of *C. perfringens* potential spillover and spread in the UK, which would significantly enhance the understanding of this prevalent gut pathogen in the context of public health. In addition, this approach will provide insights on the plasmid sub-types of *C. perfringens* associated with either general FP outbreaks, sporadic outbreaks or those linked with CHs as an extension of this study.

Care home outbreaks that involve mainly older adult communities has been an important issue over the years, with the rise of non-foodborne outbreaks detailed in this study, thus to monitor closely the spread of outbreak strains will require a closer collaboration with PHE, with additional samples obtained from various CHs [316]. The main limitation, with respect to in-depth interpretation of these data, was the lack of completeness of the clinical metadata obtained from PHE. To pinpoint the key association of clinical isolates and strain virulence will necessitate an appropriate public health surveillance system to be in place. A potential discussion with PHE to improve *C. perfringens* surveillance and clinical data curation will form a key part of the future work plan.

Importantly, a preliminary gastroenteritis model of *C. perfringens* has been set up during my PhD that is described in Chapter 8. Potentially, foodborne and non-

foodborne associated *C. perfringens* can be used to establish novel food-poisoning or non-foodborne diarrhoea infection models. This would allow the determination of disease mechanisms, and development of translational research including preventative measures and therapies, particularly setting up a long-lasting diarrhoea model, which will be key to tackle this disease that affects primarily elderly communities.

Currently there is an obvious lack of WGS information on outbreak-related *C. perfringens* strains. *C. perfringens* was ranked as one of the top foodborne pathogens in several countries including the UK, the USA and Canada (this could also include many developing countries were information is not available due to the short-duration and under-reporting of this illness) [170,317,341,342]. As *C. perfringens* foodborne illnesses is currently not linked to high mortality rates, this pathogen is understudied at the whole genome level, possibly due to limited research funding. However, foodborne illnesses is known to significantly contribute to morbidity in populations, which directly impacts on the country's economy, as reported previously that *C. perfringens* foodborne illnesses caused an estimated economic burden of USD382 million per annum in the USA [343,344]. To tackle this issue globally, the first step would be to start with setting up collaborations with national surveillance units in a number of countries, and to whole-genome sequence the collected isolates, thus providing a better understanding of the global epidemiology.

In conclusion, this is the first time *C. perfringens*-associated outbreaks have been investigated via WGS and bioinformatic techniques and demonstrated that isogenic strains caused multiple outbreaks particularly in CHs (non-foodborne outbreaks), possibly linked to geographical distribution. Moreover, these data indicated that plasmid pCPF5603 was associated with CH isolates, while plasmid pCPF4969 was exclusively linked to FP isolates. In addition, CH isolates encoded significantly more virulence genes (including colonisation-important sialidases, inferring a greater capacity in virulence) than FP isolates, which corresponded to the potentially longer duration of diarrhoea caused by these strains, compared to the more common short-lived foodborne diarrhoea. Further works will be required to address research questions relevant to the actual spread and impact of *C. perfringens* foodborne and non-foodborne infections.

Chapter 5 Poultry NE: Caecal microbiome and *C. perfringens*

5.1 Introduction

This chapter explores the microbial diversity in the caecal microbiome of healthy commercial boiler chickens, and those birds diagnosed with NE- or sub-clinical NE- (SNE). It also describes the phylogenetics and comparative genomics on *C. perfringens* isolates obtained from both case and control chickens (Figure 5.1). This study was conducted in collaboration with Dr Ron Dixon and Dr Joseph Brown at University of Lincoln in 2016. This study aims to understand if the caecal microbiomes of broiler chickens can be used to determine specific microbial markers associated with NE, and to identify NE-linked virulence genes in *C. perfringens* isolates using WGS and bioinformatics approaches. Field work and bacterial isolation were performed by Dr Joseph Brown. I performed all the bioinformatics analyses in this chapter unless otherwise indicated in the text or figure legends.

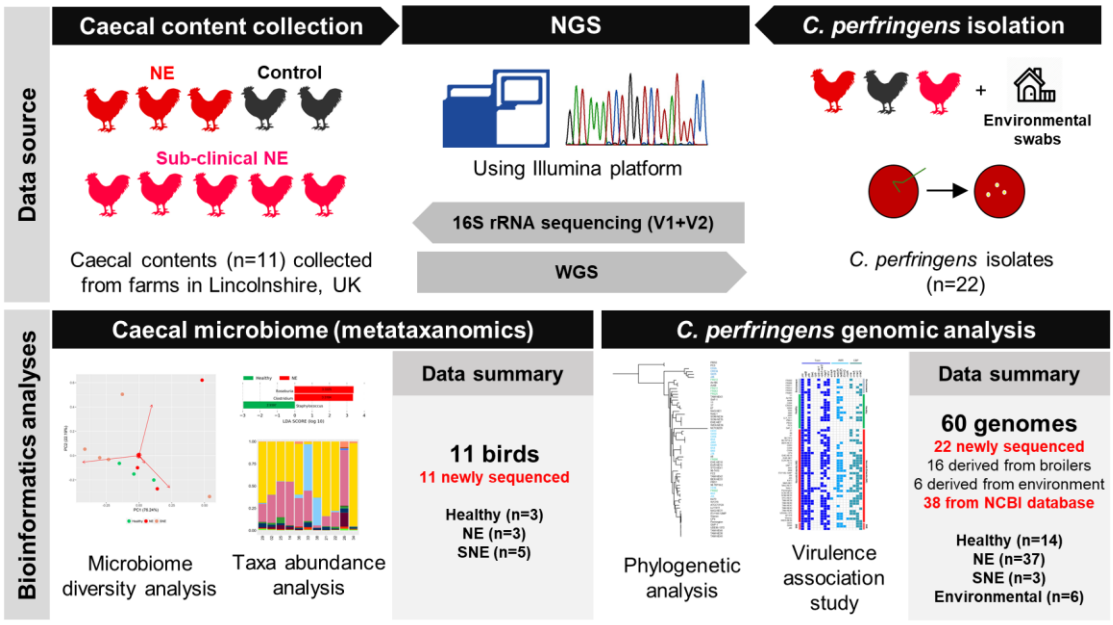


Figure 5.1: Schematic of data source and bioinformatics analyses described in Chapter 5

5.2 Background

Broiler chickens are solely bred for meat production, and represent a key global livestock member; estimated annual production of 50 billion birds worldwide [345]. As broilers reach slaughter weight at a young age (4-7 weeks) they are susceptible to several welfare and infection issues. Importantly, poultry NE, an inflammatory gut infection in chickens, is responsible for a loss of US$6 billion per annum in the poultry industry, with *C. perfringens* reported to be the main causative agent [1,122,124,138].

NE-associated pathologies are mainly characterised by gaseous lesions and mucosa necrosis in the gas-filled distended small intestine [145]. Proposed key *C. perfringens*-associated factors linked to NE include α-toxin, and more recently NetB and TpeL, (both pore-forming toxins) [146]. Other aetiological factors that have been shown to increase risk of NE include high-protein diets and environmental stressors, which may alters gut microbiota/host immunity [148]. In addition, SNE, which is a mild form of NE, is represented by poor growth and small intestinal ulcerative lesions and has also been associated with *C. perfringens* colonisation [146,346].

NetB is considered to be an essential *C. perfringens* virulence factor in poultry NE as determined in animal studies [122]. Expression of this pore-forming toxin was reported previously to be higher (92%) in NE chickens as compared to healthy chickens (29%), thus supporting its vital role in disease progression [123]. This toxin is known to be encoded exclusively on conjugative plasmids, indicating HGT may play a role in dissemination to NetB-negative strains [125]. Collagen adhesin (encoded by gene *cna*) is another candidate disease determinant which has been associated with chicken NE isolates in a recent bacterial genomic study [230].

The caecum represents the primary site for *C. perfringens* colonisation, which also contains the highest density of the chicken gut microbiota, therefore NE-induced alterations of this GI site are likely to reflect changes including those microbiota members [347]. Moreover, the chicken caecal microbiome is known to play a protective role in pathogen resistance to other enteric pathogens, including *Campylobacter jejuni*, and as such intestinal microbiota disruption may impact development of *C. perfringens*-associated NE in broiler chickens, although direct biological impact is yet to be confirmed [348-350].

To date, it is unclear if and how the chicken caecal microbiome changes during NE and SNE, therefore a microbiota profiling study was carried out to understand if there are any diseased-specific disturbances induced after *C. perfringens* infection. Moreover, at present there are only two WGS-based studies on chicken-associated *C. perfringens*; one study examined 30 isolates (from both broilers and turkeys), while the other profiled 59 isolates (in which 22 were newly sequenced genomes). Thus, to further our understanding on *C. perfringens* virulence in the context of NE, I also performed comparative genomics on known isolates (60 chicken/NE- and SNE-associated *C. perfringens* strains), alongside newly sequenced strains (22) specific to this project [230,290].

5.3 Hypothesis and aims

Broiler caecal contents exhibit microbiome signatures for poultry NE, and specific genotypes of chicken-associated *C. perfringens* are linked with poultry NE. To determine how the broiler gut microbiota changes during NE and SNE, and explore *C. perfringens* virulence potentials associated with disease via:

(1) 16S rRNA metataxonomic analysis of broiler caecal contents from 11 caecal samples, including healthy chickens (n=3), NE chickens (n=3), and SNE chickens (n=5).

(2) Perform comparative genomics on chicken-associated *C. perfringens* isolates using WGS technique from 60 bacterial genomes (22 newly sequenced and 38 public genomes).

5.4 Study population

5.4.1 Caecal microbiome analysis

Individual caecal samples (n=11; Table 5.1) were obtained in 2016-2017, processed (genomic DNA extraction and 16S rRNA library preparation performed by Ms Charlotte Leclaire) and sequenced via 16S rRNA sequencing (targeting V1/V2) at WTSI. All samples were obtained from poultry farms in Lincolnshire, UK.

Sample ID	Location	Host status
29	Lincolnshire, UK	Healthy control
2	Lincolnshire, UK	Healthy control
25	Lincolnshire, UK	Healthy control
14	Lincolnshire, UK	NE
33	Lincolnshire, UK	NE
36	Lincolnshire, UK	NE
21	Lincolnshire, UK	SNE
22	Lincolnshire, UK	SNE
26	Lincolnshire, UK	SNE
34	Lincolnshire, UK	SNE
38	Lincolnshire, UK	SNE

Table 5.1: Clinical data of broiler hosts

Health conditions of chickens of which caecal microbiomes were sequenced and investigated. NE: Necrotic Enteritis; SNE: Sub-clinical Necrotic Enteritis. Data: n=11.

5.4.2 Comparative genomics of *C. perfringens* isolates

C. perfringens isolates (n=16) were obtained from broiler intestinal samples (Table 5.2). Environmental strains (n=6) were isolated from chicken feed, surrounding straw and soil within the vicinities of poultry farms. 38 public genomes were retrieved from NCBI genome database (retrieved in July 2018) and subject to bioinformatics analysis, which made the total sample size 60 genomes. These isolates originated from countries including England, Belgium, Australia and USA and they can be further categorised into 4 groups based on the health conditions of the broiler hosts: healthy (n=14), NE (n=37), SNE (n=3) and environmental (n=6).

5.5 Results

5.5.1 NE and poultry caecal microbiome

5.5.1.1 Microbiome diversity analysis

Caecal contents of 11 individual chickens were sequenced on Illumina MiSeq platform to generate an average of 115,272 (range: 83,117-164,027) sequence reads per sample, with an average of 3,120 (range: 1,780-4,027) OTUs assigned (via QIIME) in each sample, clustering at 97% similarity (Appendix 3 Table S5.1).

Strain	Source	Origin	Source	Hosts
FR002	In this study	England	Soil	Environmental
FR006	In this study	England	Straw	Environmental
FR011	In this study	England	Soil	Environmental
FR013	In this study	England	Faeces	Environmental
FR025	In this study	England	Soil	Environmental
FR063	In this study	England	Feed	Environmental
11	NCBI	Belgium	Birds	Healthy
An185	NCBI	No information	Birds	Healthy
An68	NCBI	No information	Birds	Healthy
C056	In this study	England	Birds	Healthy
CR034	In this study	England	Birds	Healthy
G050	In this study	England	Birds	Healthy
I056	In this study	England	Birds	Healthy
I060	In this study	England	Birds	Healthy
J060	In this study	England	Birds	Healthy
LLY_N11	NCBI	No information	Birds	Healthy
PBD1	NCBI	Australia	Birds	Healthy
PBS5	NCBI	Australia	Birds	Healthy
PC5	NCBI	Australia	Birds	Healthy
SAF-1	NCBI	Australia	Birds	Healthy
37	NCBI	Belgium	Birds	NE
48	NCBI	Belgium	Birds	NE
67	NCBI	Belgium	Birds	NE
96-7415	NCBI	USA	Birds	NE
98.78718-2	NCBI	Denmark	Birds	NE
BER-NE33	NCBI	Australia	Birds	NE
C033	In this study	England	Birds	NE
C036	In this study	England	Birds	NE
C55A	In this study	England	Birds	NE
CP4	NCBI	Canada	Birds	NE
EHE-NE18	NCBI	Australia	Birds	NE
EHE-NE7	NCBI	Australia	Birds	NE
EUR-NE15	NCBI	Australia	Birds	NE
FC2	NCBI	USA	Birds	NE
GNP-1	NCBI	USA	Birds	NE
I055	In this study	England	Birds	NE
I058	In this study	England	Birds	NE
ITX1105-12MP	NCBI	USA	Birds	NE
J36	In this study	England	Birds	NE
J55	In this study	England	Birds	NE
K473	NCBI	Australia	Birds	NE
NAG-NE1	NCBI	Australia	Birds	NE
NAG-NE31	NCBI	Australia	Birds	NE
NobL1	NCBI	USA	Birds	NE
Pennington	NCBI	USA	Birds	NE
SOM-NE34	NCBI	Australia	Birds	NE
SOM-NE35	NCBI	Australia	Birds	NE
SYD-NE41	NCBI	Australia	Birds	NE
TAM-NE38	NCBI	Australia	Birds	NE
TAM-NE40	NCBI	Australia	Birds	NE
TAM-NE42	NCBI	Australia	Birds	NE
TAM-NE43	NCBI	Australia	Birds	NE
TAM-NE46	NCBI	Australia	Birds	NE
UDE95-1372	NCBI	USA	Birds	NE
W1319	NCBI	Australia	Birds	NE
Warren	NCBI	USA	Birds	NE
WER-NE36	NCBI	Australia	Birds	NE
G035	In this study	England	Birds	SNE
G049	In this study	England	Birds	SNE
J46	In this study	England	Birds	SNE

Table 5.2: Information of 60 *C. perfringens* isolates analysed in this study

Principal Component Analysis (PCA) indicated no distinct clustering of all 11 samples, indicating NE and SNE may not manifest in distinctive microbiome

'signatures' (Figure 5.2A). However, disease-specific profiles may be masked by the fact that a probiotic mix was given to these broilers as a preventative measure against NE development. Indeed, probiotic-associated *Bifidobacterium* (mean relative abundance: 54.17 ± 30.12%; mean ± SD) and *Lactobacillus* (mean relative abundance: 27.23 ± 19.27%) genera dominated the caecal microbiomes across all individual samples. Therefore, raw reads from these two genera were removed and a new PCA performed to understand the correlation of other secondary microbiota members (Figure 5.2B). This approach indicates a trend towards health-status clustering, however SNE microbiomes overlap with NE and control samples. Notably, healthy broiler caecal microbiomes seem to be positively associated with the genera *Blautia* and *Ruminococcus*, while an inverse relationship with *Enterococcus* is evident. NE-associated profiles appear to be dominated by the genera *Bacteroides*, *Clostridium*, *Faecalibacterium* and *Enterococcus*.

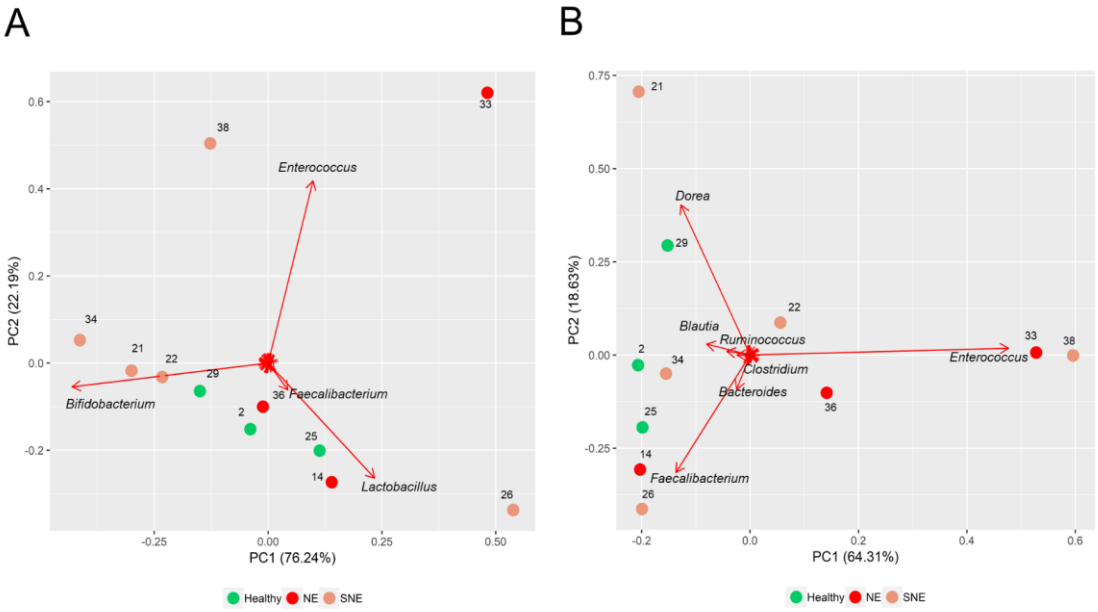


Figure 5.2: Clustering of caecal microbiomes via Principal Component Analysis

PCA demonstrated microbial genera relations with caecal microbiomes obtained from healthy and diseased broiler chickens. **(A)** PCA on 11 microbiome samples. Factor loadings correlates the bacterial species with caecal microbiomes. **(B)** PCA on probiotic-removed microbiomes, revealing secondary genus correlations and displayed better clustering. Figures generated in R, using packages *ggfortify* and *ggplot2*, and functions *autoplot* and *prcomp*.

Rarefaction analysis supports the availability of sufficient sequence reads to achieve asymptotic, i.e. optimal diversity of richness (at genus level; range: 15-25 genera) in each individual sample to represent each member of the microbiota (Figure 5.3). In

addition, Alpha-diversity analysis indicates there is no significant difference in genus diversity among 3 groups of samples including healthy, NE and SNE (mean range: 1.83-1.96; Figure 5.3).

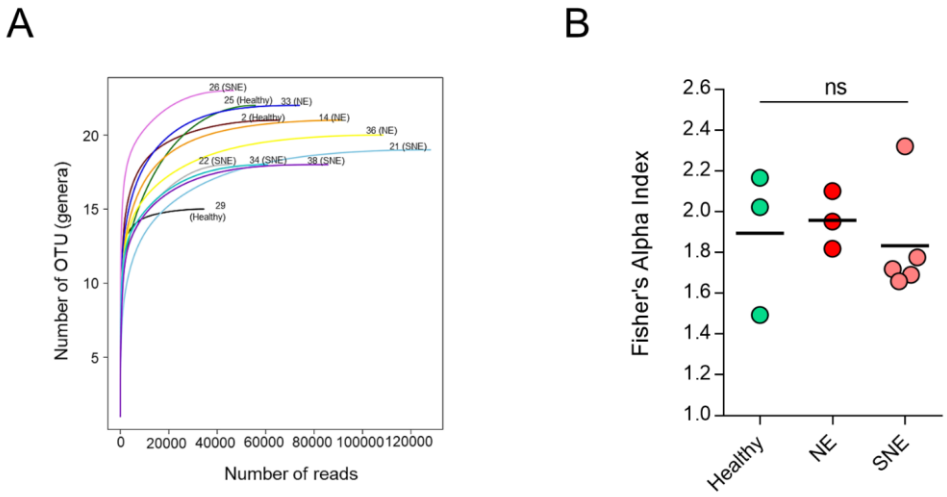


Figure 5.3: Species richness and diversity analysis of caecal microbiome samples

(A) Rarefaction curve (species richness) determined of each caecal sample. **(B)** Alpha-diversity of each group. Statistics generated by Mr Shabhonam Caim using R. Each curve is colour-labelled with sample ID. This rarefaction curve is generated based on the OTU (at genus level) assigned to the reads. Alpha-diversity is determined by Fisher's Alpha Index. Black bar represents mean of each group. Statistics test: Dunn's multiple comparisons test (two-tailed).

Relative bacterial genus abundance in each caecal sample was constructed to visualise microbiota profiles (Figure 5.4A). Thirty-eight genera were represented in the microbiome, with two genera most abundant, *Bifidobacterium* and *Lactobacillus*, which likely reflects the probiotic supplementation in the chicken feed. A number of minor genera, which are usual intestinal microbiota members, were detected in these samples (relative abundance <10% in each sample) including *Blautia, Coprococcus, Dorea* and *Oscillospira*. The microbiota member *Enterococcus*, which is widely used as veterinary probiotic (especially *Enterococcus faecium*), was found at high levels in broilers 33 (52.1%; NE) and 38 (31.3%; SNE).

Sequence reads of highly-dominant genera *Bifidobacterium* and *Lactobacillus* (Figure 5.4B), also *Enterococcus* (potentially probiotics in origin; Figure 5.4C) were removed to visualise minor gut microbiota members of broilers in each group. It appeared that *Clostridium* genus was more prominent in NE group (broilers 14 and 36) than healthy

chickens. Notably, *Blautia*, *Dorea, Oscillospira*, *Faecalibacterium* and *Ruminococcus* were present in all broiler microbiotas, although in low abundance.

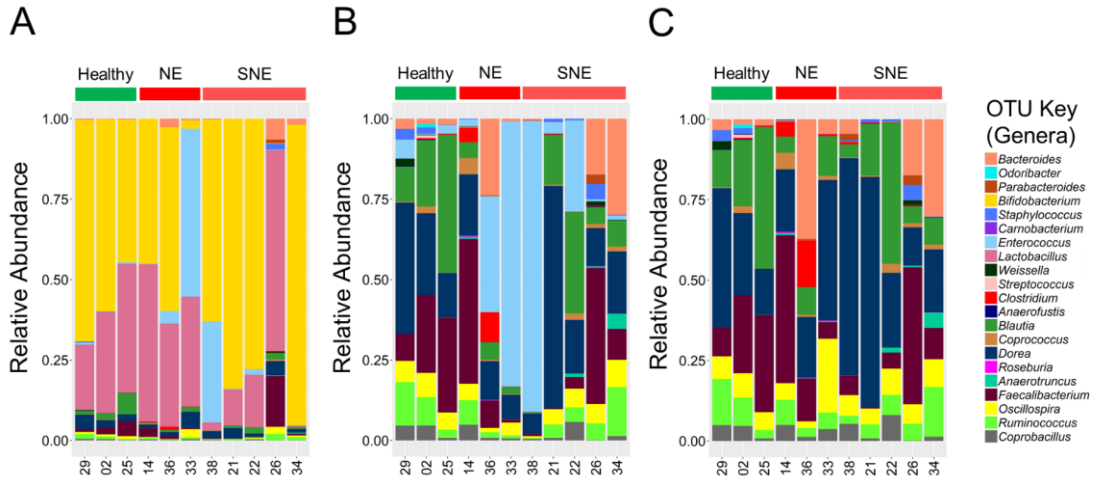


Figure 5.4: Caecal microbiomes of healthy, NE and SNE chickens

(A) Relative abundance of read counts on genus rank. **(B)** Major genera *Bifidobacterium* and *Lactobacillus* were removed (reads) in this figure. **(C)** Genus Enterococcus was further removed from (B) in this figure. Genera represented ≤0.10% of the total abundance reads are omitted in both figures. Broiler samples ID were indicated on the x-axes. Graphs were generated in R.

Using an additional paired-end BLASTN approach in taxa assignment to 16S rRNA sequences, a number of important genera were mapped to reference sequences at species level (Appendix 3 Figure S5.1) [299]. *Lactobacillus* genus were mainly composed of species *Lactobacillus reuteri* (common broiler gut member, also widely used as probiotic supplement), *Lactobacillus salivarius* (common swine gut microbiota member used as broiler probiotic that improves production and general health) and *Lactobacillus vaginalis* (frequently found in broiler gut and a persistent gut coloniser) [351-354]. *Enterococcus* genus primarily consisted of species *Enterococcus faecium*, a widely-used probiotic reported to promote broiler growth and suppress *C. jejuni* and *C. perfringens* infections, while stimulating the growth of *Lactobacillus* and *Bifidobacterium* [355,356]. Importantly, *Clostridium* genus was mainly assigned to *C. perfringens* sequences, denoting the potential NE-link of *C. perfringens* origins in NE-broilers 14 and 36.

5.5.1.2 Linear Discriminant Analysis (LDA) reveals NE-linked abundant taxa

To identify taxa (or, OTUs) that varied in abundance consistently within particular groups, LDA (based on Kruskal-Wallis test; $P<0.05$ is considered as significant) was

performed with pooled data based on normalised sequence reads of assigned OTUs at genus level (Figure 5.5). *Clostridium* was shown to be significantly enriched within NE caecal microbiomes (mean relative abundance: 0.45% vs 0.02% in healthy individuals; LDA score: 3.3), confirming the frequent link of *Clostridium,* especially *C. perfringens*, to chicken NE. Butyrate-producing genus *Roseburia* increased significantly (mean relative abundance: 0.02% vs 0.00% in healthy individuals; LDA score:3.3) compared to control group. Interestingly, *Staphylococcus* genus was significantly under-abundant in NE samples. None of the OTUs in the SNE group was identified to be significantly varied in abundance.

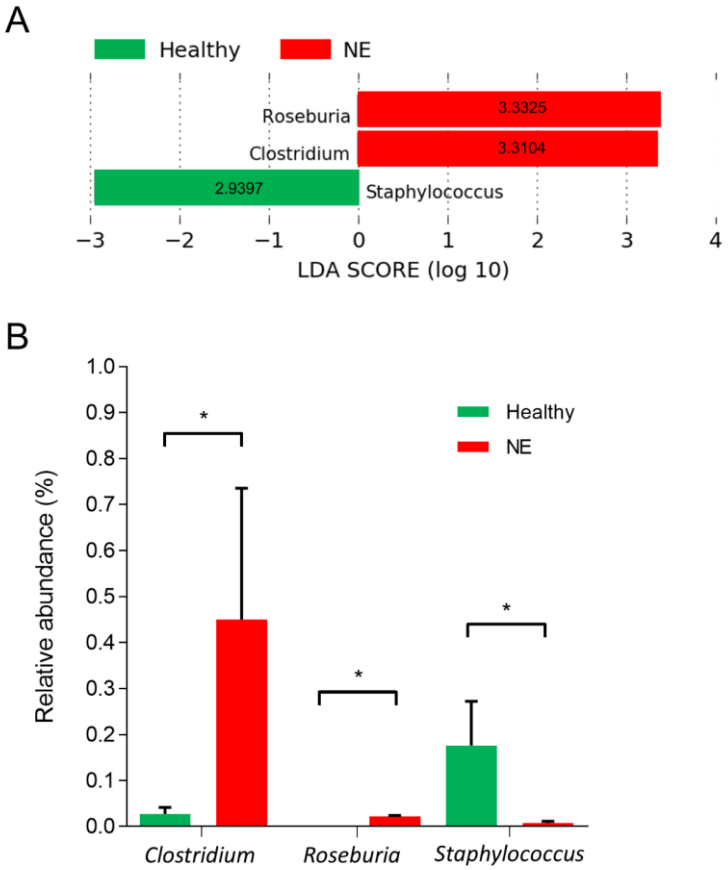


Figure 5.5: Differentially abundant OTUs identified via Linear Discriminant Analysis

LDA shows three differentially abundant OTUs in categorical groups (healthy vs NE) determined at genus level. **(A)** LDA plot generated via LEfSe. Significant LDA cut-off as 2.5. P<0.05 (Kruskal-Wallis test). **(B)** Group comparison (healthy vs NE) of differentially abundant OTUs via Mann Whitney statistical test. *P<0.05.

5.5.2 Chicken-associated *C. perfringens* and virulence potentials

5.5.2.1 Phylogenetic analysis reveals potential clonal expansion

To further explore *C. perfringens*-associated NE and SNE (including 6 environmental isolates from broiler farms), a phylogenetic tree of 60 chicken-associated isolates was constructed based on 7,750 SNPs extracted from core gene alignment (Figure 5.6). Clustering was assigned via hierBAPS analysis and these 60 isolates clustered into 5 major clades (hierBAPS level 1).

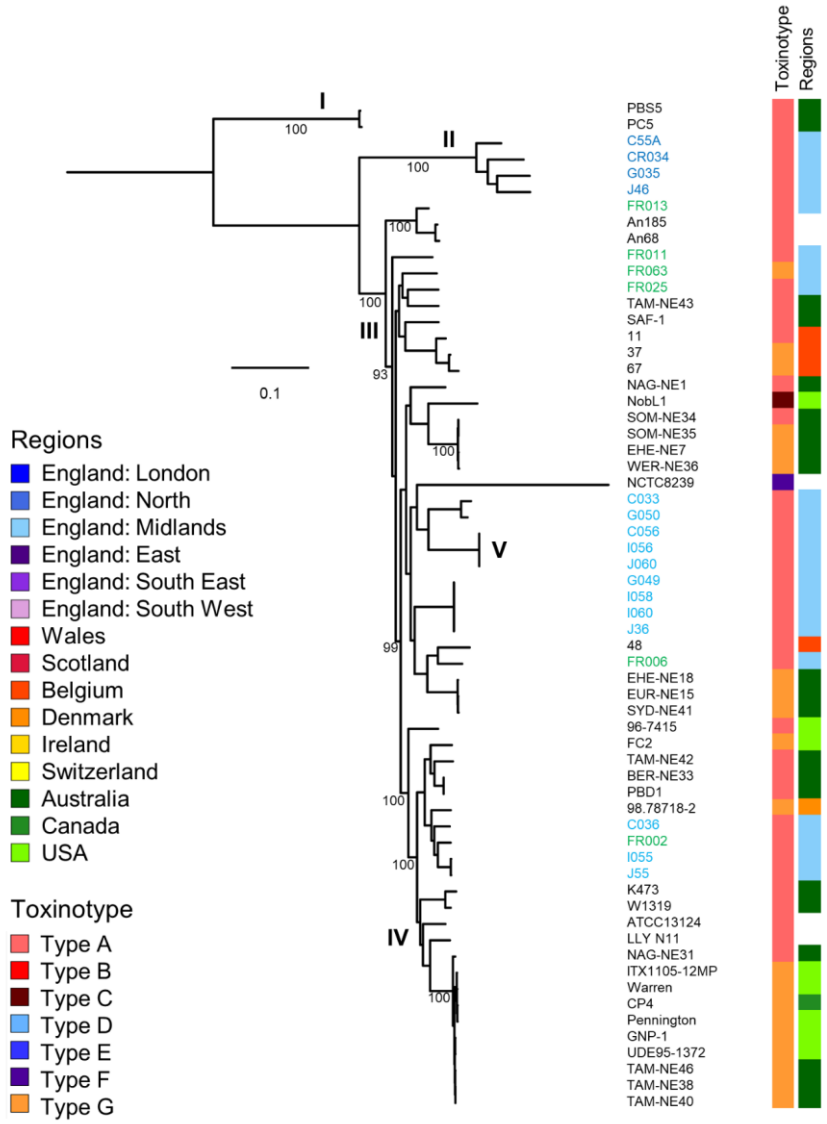


Figure 5.6: Genetic relationship of chicken-associated isolates

Mid-point rooted maximum-likelihood phylogeny constructed from 7,750 SNPs (in core gene alignment). Clusters assigned via hierBAPS analysis (level 1). ATCC13124 and NCTC8239 are reference genomes used in this tree. Bootstrap values are shown major nodes in the tree. Scale bar is indicative of nucleotide substitution per site. Newly sequenced genomes in this work are labelled in light blue. Green-labelled environmental isolates were obtained from chicken farms in Lincolnshire.

Two Australian isolates PBS5 and PC5 reside within a genetically distinct lineage I, suggesting a fundamental difference in genetic components of these isolates. Lincolnshire English isolates mainly reside within lineage V (12/16; 75%), and appear to converge within tight clusters, suggesting a close genetic relationship among these isolates. Toxinotype profiling indicated, out of 54 chicken isolates, 20 isolates were type G (37%; carry *netB*), 33 isolates type A (61%), and 1 isolate type C (carries *cpb*).

Via SNP analysis, identical/highly-similar genetic clusters were determined. Isolates C56, I56 (both from the same bird) and J60 (another bird; lineage V) obtained from farms in Lincolnshire were determined as identical strains with no SNP difference (0 SNPs). Another constricted cluster identified consisted of isolates G049, I058, I060 and J36 (from 4 individual birds in Lincolnshire; lineage V) were determined as identical strains (0 SNPs). Isolates J055 and I055 (lineage V), both from the same individual bird isolated from different intestinal sites were confirmed as highly-similar strains (2 SNPs in pair-wise comparison). Nevertheless, no environmental isolates were found to be genetically similar to any NE or SNE strains at SNP level.

Importantly, several publically available isolates also formed into highly-similar clusters. A sub-cluster in lineage IV, comprising 10 isolates obtained from three countries USA, Australia and Canada (Table 5.3) was determined to be an isogenic cluster (pair-wise SNP range: 1-86). Among these isolates, 4 strains (ITX1105-12MP, Warren, CP4 and Pennington) are near-identical (1-3 SNPs), suggesting a potential clonal expansion of a persistent genotype that warrants further investigation.

	CP4	GNP-1	ITX1105-12MP	NAG-NE31	Pennington	TAM-NE38	TAM-NE40	TAM-NE46	UDE95-1372	Warren
CP4	0	34	3	85	2	36	36	37	35	1
GNP-1	34	0	35	57	34	8	8	9	1	33
ITX1105-12MP	3	35	0	86	3	37	37	38	36	2
NAG-NE31	85	57	86	0	85	59	59	60	58	84
Pennington	2	34	3	85	0	36	36	37	35	1
TAM-NE38	36	8	37	59	36	0	0	5	9	35
TAM-NE40	36	8	37	59	36	0	0	5	9	35
TAM-NE46	37	9	38	60	37	5	5	0	10	36
UDE95-1372	35	1	36	58	35	9	9	10	0	34
Warren	1	33	2	84	1	35	35	36	34	0

Table 5.3: Pairwise-SNP comparison between isolates in clonal sub-lineage within lineage V

5.5.2.2 Virulence association study suggests combinatory role of toxin and adhesin genes in NE

Virulence profiles of these 60 chicken-associated isolates were also determined including toxins, AMR genes and collagen-binding protein/collagen adhesin (CBP; or, collagen adhesin), which was recently linked to *C. perfringens* NE-genotypes as represented in Figure 5.7 [230]. Using statistical test (Fisher's exact test), NE-trait associations were determined based on 14 control isolates and 37 NE isolates (Figure 5.8).

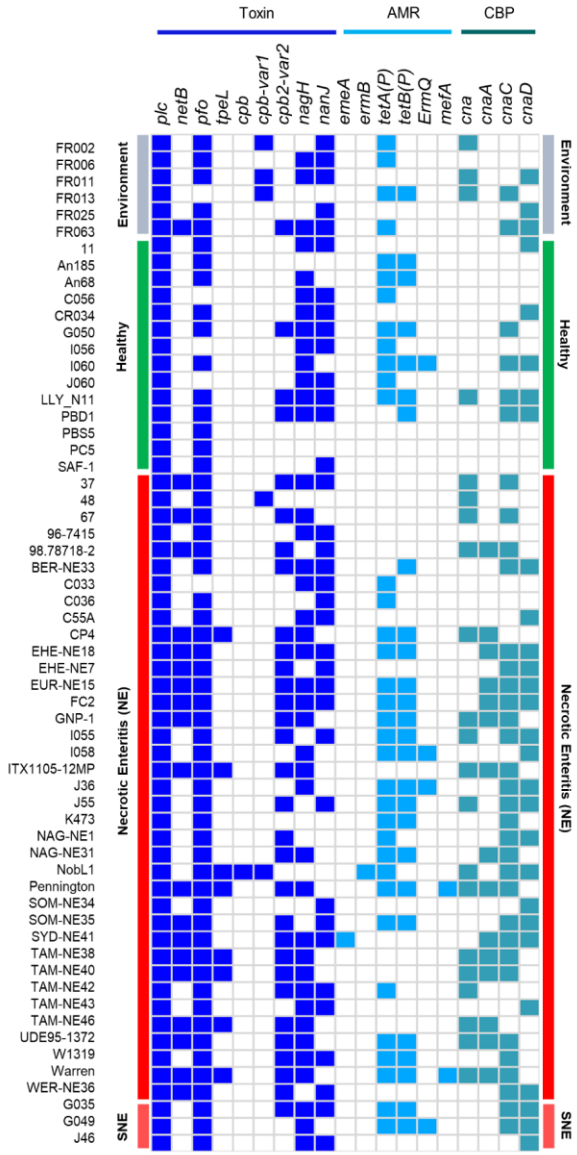


Figure 5.7: Toxin, AMR and collagen-binding protein profiling of chicken-associated isolates

Coloured cells indicate presence of genes, white cells represent absence of genes. CBP: Collagen-binding proteins (collagen adhesin). Data: n=60. Heatmap was generated in R.

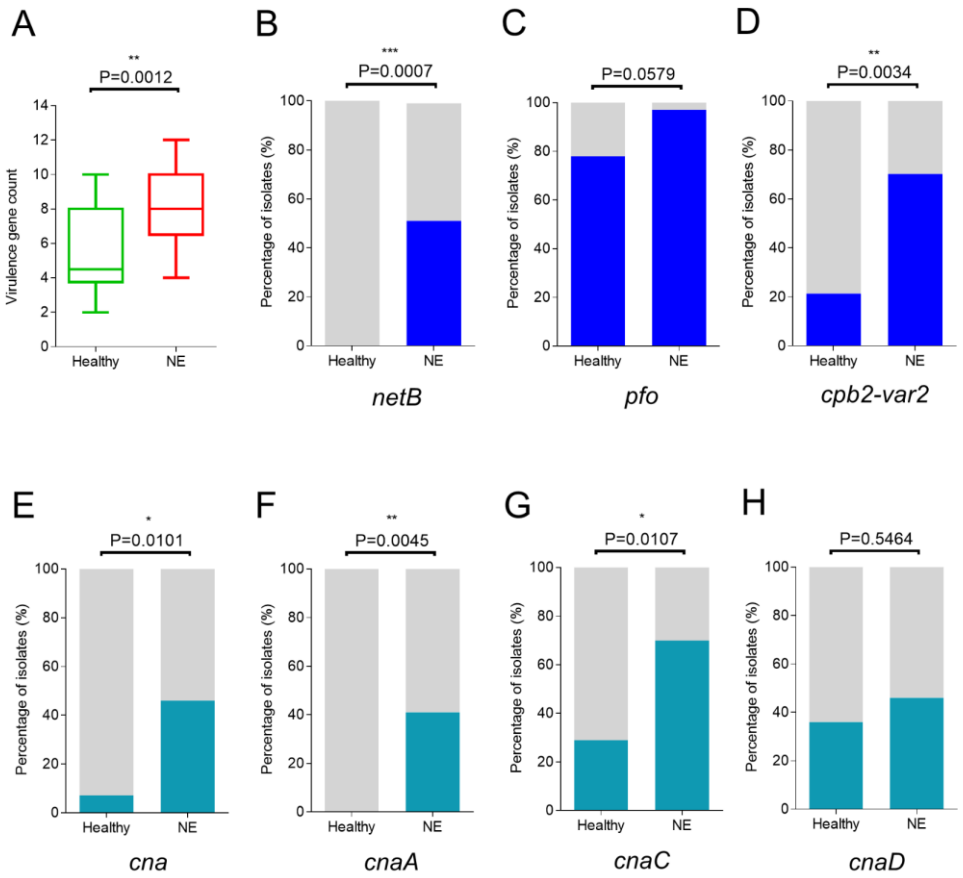


Figure 5.8: Enrichment of key virulence genes in control and NE case isolates

(A) Comparison of total virulence gene counts between control (n=14) and NE (n=37) isolates. Comparisons of gene enrichments in between groups as follows: **(B)** *netB* **(C)** *pfo* **(D)** *cpb2*-var2 **(E)** *cna* **(F)** *cnaA* **(G)** *cnaC* **(H)** *cnaD*. Data: (A) Non-parametric Mann-Whitney test *U* test, two-tailed. () to (H) Fisher's exact test, two-tailed. *P<0.05, **P<0.01, ***P<0.001.

NE determinant NetB toxin gene *netB* was significantly enriched in NE isolates (51%; 19/37; P<0.001), compared to control isolates (0%; 0/14) confirming previous findings on this NE-causing virulence gene (noting that environmental isolate FR063 also encoded *netB*) [122]. Haemolytic toxin PFO was encoded in all but one NE isolates (97.2%), suggesting its important role, although most control isolates (11/14; 78.5%) also carried this toxin. Accessory toxin gene CPB2 (variant 2, which was known to be >10 times more necrotic that CPB2-variant 1 in human cells) was encoded more abundantly in NE isolates (P<0.05). Importantly, collagen adhesin variant genes *cna* (P<0.05), *cnaA* (P<0.005) and *cnaC* (P<0.05) were all determined to be significantly enriched in NE isolates, with *cnaA* entirely absent from control isolates (0/14; 0%), but present in 40% (15/37) of the NE isolates. A similar trend was observed for *cna,* only 1/14 (7%) control isolates, but 46% (17/37) of NE isolates carry this gene, confirming the strong association as observed in a previous study [230].

AMR genes *emeA* (encodes multidrug-efflux pump) and *ermB* (Macrolide-Lincosamide-Streptogramin resistance; MLS) were each detected in one isolate (AMR genes frequently carried by *E. faecalis*) [229]. Other AMR genes including *tetA(P)*, *tetB(P)* and *ermQ* were frequently (31/51 isolates carried at least one of these genes) encoded in the examined isolates. Notably, 39.2% of the health and NE isolates (n=51) did not carry any acquired AMR genes.

Additionally, these data suggest that numerous virulence potentials in NE-associated isolates may play a combinatory role in disease pathogenesis, as NE isolates carry significantly more virulence genes (n=8) than control isolates (n=4.5; P<0.005), implying the vital role of accessory virulence genes, rather than a particular major causative toxin gene in disease pathology.

Potential combinatory effects of virulence genes were further investigated by Fisher's exact test (Figure 5.9). Co-occurrence of 4 virulence determinants including toxin genes *netB*, *pfo*, *cpb2* and CBP genes (≥1 *cna* variants) were significantly enriched in 51.3% (19/37) of NE isolates, but not in control isolates (P<0.01; Figure 5.9A); combinations of three (*pfo* + *cpb2* + CBP; Figure 5.9B) and two (CBP + either *pfo* or *cpb2*; Figure 5.9C-D) were also more abundantly detected in NE than control isolates (P<0.05). Hence, this data supports the crucial role of co-occurrent toxin and adhesin genes in *C. perfringens*-associated broiler NE.

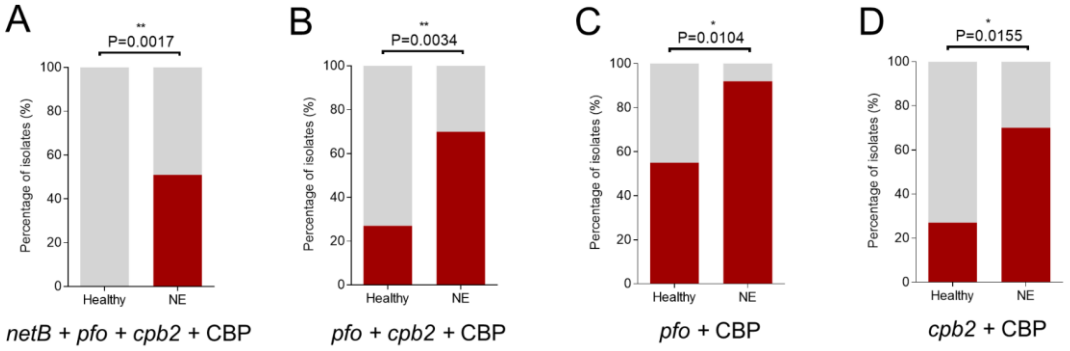


Figure 5.9: Comparison of key combinatory virulence genes in control and NE isolates

Comparison of total combinatory virulence gene counts between control (n=14) and NE (n=37) isolates as follows: **(A)** *netB* + *pfo* + *cpb2* + CBP **(B)** *pfo* + *cpb2* + CBP **(C)** *pfo* + CBP **(D)** *cpb2* + CBP. Data: Fisher's exact test, two-tailed. *P<0.05, **P<0.01. CBP: collagen-binding protein (n≥1). CBP: Collagen-binding proteins.

5.6 Discussion

Although this was a small-scale study, healthy broiler caecal microbiomes were found to have enhanced abundance of the genera *Blautia*. Members of the *Blautia* genus are known to be butyrate producers, and reductions in this genus have previously been associated with a *C. jejuni* infection model [357]. Moreover, *Blautia obeum* (later reclassified as *Ruminococcus*) has also been demonstrated to restrict the colonisation of gut pathogen *Vibrio cholerae* [358,359]. As butyrate is an important energy source for intestinal cells, these *Blautia* spp. may act as key beneficial microbiota members, serving to enhance intestinal health of chickens by strengthening the epithelial barrier, thus preventing pathogenic microbes successfully colonising and initiating disease.

Microbes associated with NE included *Clostridium*, *Faecalibacterium* and *Enterococcus*. Previous studies have indicated that *Enterococcus* (*E. cecorum*) is a poultry pathogen implicated in multiple disease representations such as sepsis and arthritis [360]. In this study, species-level taxa assignment of the caecal samples indicated that *E. faecium*, which is also commonly used as a probiotic (no corresponding dietary information for broilers sampled in this study mean it is unclear if this was used as a probiotic), was likely the dominant species (although this approach is not ideal given the short length of the 16S rRNA V1+V2 sequences). This suggests that *E. faecium* (which is also associated with human intestinal disease) may be associated with NE as evidenced by the PCA analysis [356]. Furthermore, *Clostridium* and *Faecalibacterium* were shown to increase in a *C. jejuni* chicken caecum infection model, implying their potential role in disease development [357].

Clostridium genus, as identified in caecal microbiome data via LDA analysis, was significantly enriched, albeit at low levels, in NE individuals. Further species-level assignment indicated that sample 33 *Clostridium* OTUs mapped to *C. perfringens* (89.6% reads mapped to *C. perfringens* at species level; 629/702 sequence reads), indicating that even a small proportion (relative abundance: 0.45 ± 0.28%; mean ± SEM) of *C. perfringens* could be lethal to broiler hosts. Therefore, microbiota profiling of *Clostridium* may be useful as a potential biomarker for NE-onset, however larger studies would be required to verify these findings.

Genus *Roseburia*, a normal intestinal microbiota member in both human and broilers, was frequently linked with healthy gut biomarkers as this genus of bacteria produces

the colonic fuel butyrate, which aids in epithelial barrier maintenance and helps restore beneficial gut microbiota [361,362]. The significant enrichment of beneficial genus *Roseburia*, unlikely to be involved in NE pathology, was detected in NE samples (relative abundance: 0.02 ± 0.002%; mean ± SEM) could be a biased result due to complete genus absence (relative abundance: 0.00%) in all control isolates which is likely a consequence of limited number of samples (n=3).

Probiotics, including *Bifidobacterium* and *Lactobacillus*, potentially also *Enterococcus*, have been frequently used in broiler farming primarily for the purpose of growth-promotion and prevention of bacterial infections [353,363,364]. These taxa of beneficial bacteria have been reflected in caecal microbiome analysis, with predominant OTU proportions been assigned to *Bifidobacterium* and *Lactobacillus* across all samples. No significant changes in the OTU proportions of these two probiotics across three groups (P>0.20). Nevertheless, species-level analysis showed that predominant taxa *L. reuteri* and *L. vaginalis* are both frequently found in broilers' gut and are persistent strains, therefore, I could not determine whether these strains were supplemented or inherent, due to the lack of reliable metadata [351,354]. In addition, these data also could not definitively verify the colonisation of these probiotic-associated genera in broilers' intestines, or whether the high levels are more transient in nature, as it is common practice in poultry farms to administrate these strains in large amounts within the feed. Interestingly, supplementation of *Lactobacillus* was reported to increase the intestinal abundance of beneficial gut member *Roseburia*, which was not observed in this study (0.00% abudance in all control birds), further suggesting that the presence of *Lactobacillus* might be transient rather than persistent, although this effect can also be strain-specific [365].

It has been consistently reported that *Clostridium*, particularly *C. perfringens*, is the primary infectious agent to cause chicken NE, further supported by the caecal microbiome profiles obtained in this study. As *C. perfringens* thrives at ambient bird body temperature (i.e. 40-42ºC), with a doubling time <8 mins *in vitro (*the shortest generation time known for a microorganism), this may link to its ability to rapidly overgrow and cause disease pathology [70]. Moreover, as *C. perfringens* can secrete >20 identified toxins, including NetB toxin which has been identified as an essential toxin in NE development, this also emphasises strong links between this pathogen and broiler-associated NE [121,125]. I also determined most NE isolates encoded NetB and

that isolate FR063 (obtained from chicken feed) carries this toxin, which suggests NE-associated strains may be sourced from the environment including feed and bedding, potentially acting as reservoirs, linked to NE outbreaks [366].

Several studies indicate that collagen adhesin (encoded by *cna*) is also an important NE virulence factor [306,367,230,290]. Collagen adhesin is known to facilitate bacterial colonisation within the gut, and in this study, this gene (including its variants *cnaA* and *cnaC*) was overabundant in NE-associated isolates ($P<0.05$), suggesting a strong association with NE outcome [367]. Therefore, successful colonisation mediated by collagen adhesins could play an essential role in chicken NE development and NetB-mediated pathology.

PFO has been linked with bovine haemorrhagic enteritis [87], and a trend for presence of this toxin was also apparent in NE isolates. Furthermore, CPB2, or beta2-toxin, particularly variant 2 (known to be 10 times more toxic than variant 1 in human Caco2 cells) was also found to be significantly enriched in NE isolates ($P<0.005$) [89], indicating that accessory toxins may also play an underrated role in chicken NE.

In silico analysis indicates that the virulence capacity of *C. perfringens* linked to NE is mediated by a range of genes encompassing colonisation factors and toxins. Indeed, in contrast to previous studies, half of NE isolates did not encode NetB, suggesting this toxin alone may not be 'essential' for NE-associated pathology, and other accessory toxins such as CPB2 may also be involved. As these other factors appear to be important for NE, it could be proposed that e.g. *cna* variants should form part of the routine typing scheme to trace *C. perfringens* strains. These findings were supported by combinatory effects analysis, which demonstrated that co-occurrent virulence genes including *netB*, *pfo*, *cpb2* and *cna* variant genes were associated with NE isolates ($P<0.05$) and may play an important role in NE disease progression [230].

An important finding from this small-scale WGS-based analysis indicates cross-continental clonal expansion in lineage (IV), as isogenic isolates were obtained from three different countries (USA, Australia and Canada). These data indicate global dissemination of NE-associated virulent genotypes, which is in agreement with a previous study that indicated clonal expansion of *C. perfringens* via multiple-locus variable-number tandem repeat analysis ($n=328$)[368]. To further determine the population structure of *C. perfringens* strains using WGS approaches, significantly

larger sample sizes will be required, which will allow a deeper understanding on the spread of *C. perfringens* in chicken farms worldwide, which is vital in the context of disease control.

An obvious limitation of the microbiome study is the limited number of caecal samples available to determine the microbiome diversity of broiler chickens (n=11). In addition, the overabundance of genera *Bifidobacterium* and *Lactobacillus* could have masked the microbiome signals from other bacterial members, consequently, relative abundance alone could not be used to understand the associations of NE traits with caecal microbiomes, which necessitates other approaches including LDA as performed to determine differentially enriched OTUs [239,369]. Importantly, poor metadata involving clinical statistics of the broiler hosts is also an apparent issue in further interpreting the isolate and caecal sequencing data presented in this chapter.

5.7 Future research directions

A more advanced, time-saving and in-depth shotgun metagenome technique can be applied to investigate disease-related caecal metagenome and related bacteria at species level. This will provide profound insights into the species- and strain-level information and functionalities of broiler caecal microbiome dynamics in diseased states. More importantly, with current single bacterial genome reconstruction *in silico* technique possible, *de novo* bacterial genomes can now confidently be extracted from shotgun metagenome data and subjected to further genomic analysis [313].

The need of a larger sample size on caecal contents is apparent for future studies, including longitudinal caecal samples overtime would be required to understand the microbiota changes leading to disease progression.

In conclusion, this small-scale caecal microbiome study identified *Clostridium* genus, particularly *C. perfringens*, as potential NE-associated caecal biomarker. Importantly, genomic analysis on 60 broiler-associated *C. perfringens* isolates indicates the potential farm-soil reservoir of NE-causing *netB*-positive strains, and the important role of co-occurrence of virulence genes (*netB, pfo, cpb2* and *cna* variants) in NE disease progression. Furthermore, global clonal expansion of NE-associated *C.*

perfringens highlights the need for further investigation, which requires a large worldwide dataset on NE-related *C. perfringens* isolates.

Chapter 6 Population genomics of *C. perfringens* suggests potential zoonotic relevance, disease-associated lineages and inter-regional disseminations

6.1 Introduction

C. perfringens is a multi-host pathogen involved in numerous intestinal diseases in both humans and animals [1]. Despite its medical importance, there is a significant lack of WGS-based studies probing the genomic relationships of this enteric pathogen. My most recent published work using phylogenetic-based investigation analysed up to 56 isolates (article reproduced in Appendix 7) identified a significant number of accessory genes (87.4%) within the pangenome. These data clearly highlighted that a much larger genomic study was required to fully explore *C. perfringens* genetic diversity. In this chapter, I have used WGS to examine an unprecedented 552 *C. perfringens* genomes obtained from a wide range of hosts (12 host species; from human adults to zoo animals) and multiple geographical regions (10 countries) spanning the past 90 years (1930s-2017), to explore the genomic population structure of this pathogen, determine diseased-linked phylogenetic clusters, and gain insights into its epidemiology and associated virulence features (Figure 6.1).

This genomic-based work encompasses 552 genomes: 71 genome assemblies were retrieved from NCBI assembly databases; 22 genomes of NCTC isolates were re-assembled for this analysis; 443 isolates were newly sequenced during my PhD including 101 (88 genomes +13 genomes analysed for within-host SNP variation) genomes described in Chapter 3, 109 genomes investigated in Chapter 4, 22 genomes examined in Chapter 5, and 249 are novel isolates analysed in this work, representing the largest phylogenetic and comparative genomics study on *C. perfringens* to date.

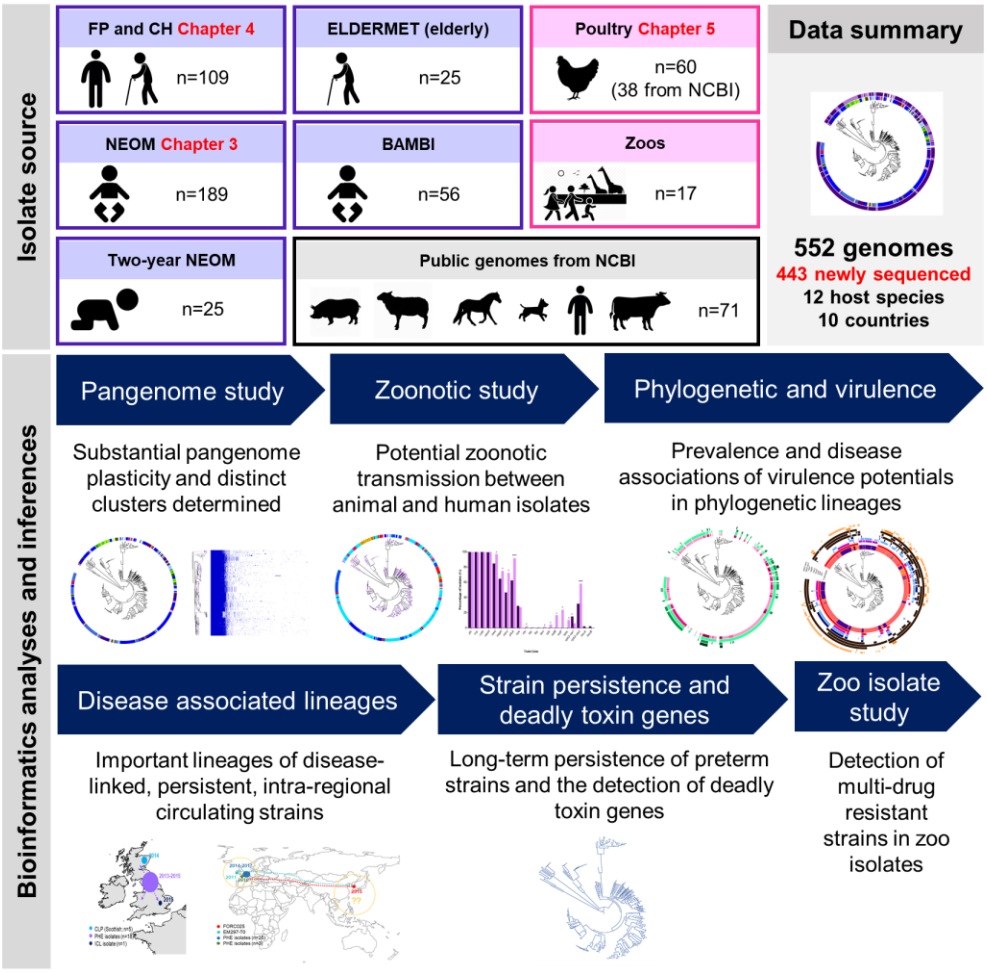


Figure 6.1: Schematic overview of data sources, analyses and inferences in Chapter 6

6.2 Background

C. perfringens-associated intestinal diseases have been described in humans and animals since 1890s when this pathogen was first isolated and named as *C. welchii* [370]. *C. perfringens* is known to mainly affect early- and later-life individuals, including NEC in preterm infants, haemorrhagic enterocolitis in calves, enteritis in foals, enteritis in dogs, NE in chickens, and FP in humans, particularly older adults in CHs [1,316]. Descriptions of these diseases have been covered in Chapter 1. Conventionally, clinically-obtained *C. perfringens* strains are currently typed and classified into 7 toxinotypes (A to G) according to the combination of typing toxins they encode (typing toxins are primarily carried on plasmids) to confirm disease symptoms and identify outbreak-associated strains [19,316]. This typing system has contributed significantly to our understanding of *C. perfringens* epidemiology in the past century, particularly in animal agriculture [125].

While WGS is increasingly being applied to bacterial pathogens in recent years, providing greater insights into the epidemiology and pathogenesis, very few studies (n<5) have applied this comprehensive approach to *C. perfringens* [335,336,371]. Moreover, WGS also offers better understanding of pathogens' genomic contents, including capturing information on virulence and AMR genes, and genetic relationships to infer spread and transmission of microbes, which surpasses the resolutions that can be obtained using routine laboratory typing methods, including fAFLP and multilocus sequence typing (MLST), as demonstrated in Chapter 3 to Chapter 5 [150,372,373].

Previous WGS-based studies on *C. perfringens* have investigated host-specific isolates (chickens), and limited numbers of isolates (n=56), which have contributed to our phylogenetic understanding [17,230,290]. In this work, I have applied WGS approaches to 552 *C. perfringens* strain genomes, which encompass publicly available sequences from the NCBI genome database (comprising historical isolates dated from the 1940s and contemporary strains, n=112), and a newly sequenced collection generated during my PhD (from across the UK and Ireland [n=443] via collaborations with numerous research groups). This study reports the novel findings of these large-scale comparative genomics and phylogenetic investigations.

6.3 Hypothesis and aims

C. perfringens is a genetically diverse, zoonotic and widely-disseminated pathogen. This chapter aims to understand the genomic population structure and specific disease-related traits of *C. perfringens* using a large international collection of isolates, specifically determining:

(1) The pangenome of 552 *C. perfringens* isolates, both newly sequenced and publicly available genomes.

(2) The population structure of *C. perfringens* isolates, targeted at identifying specific disease-related lineages.

(3) Potential zoonotic relationships between human- and animal-derived isolates, aiming to identify disease-linked virulence genes.

(4) In-depth genetic relationship between strains of specific medically important lineages on SNP level, to identify potential disseminations of *C. perfringens* strains.

6.4 Study population

The dataset encompasses a wide range of isolates obtained from humans (n=417), including adults, older adults, children and infants (both healthy and diseased), animals (n=113) including chickens (NE and healthy), dogs, horses, cows, pigs, sheep (primarily diseased individuals) and healthy zoo animals, 6 environmental soil isolates, and 14 isolates from public databases with source undetermined (Figure 6.2; Appendix 4 Table S6.1). Geographically, isolates originated from England (highest proportion of 68%; 380/552), Scotland (5), Wales (4), Ireland (17), Belgium (4), Denmark (1), Switzerland (2), Canada (12), USA (27) and Australia (21; Figure 6.4). Isolates were obtained from as early as 1930s (historical strains from NCTC collection) to contemporary isolates collected in 2017, and notably for this study include a comparable proportion of healthy and disease-isolated strains (healthy vs diseased: 43.6% vs 55.2%). These data provide the largest and most diverse genomic picture of *C. perfringens* to date, enabling construction of phylogenetic relationships between historical and present-day isolates from different geographical locations, and healthy versus disease-associated strains, thus allowing determination of certain disease-related phylogenetic clusters, which is vital for understanding *C. perfringens* dissemination and evolution.

All genome assemblies were determined (pair-wise alignment via mummer) to have a genome-wide ANI >95%, which confirms the *C. perfringens* species delineation (Appendix 4 Table S6.1) [217]. The majority of genome assemblies are of high quality within 1-100 contigs (89.3%; 493/552), with 75.7% (418/552) classified as very high-quality assemblies at 1-50 contigs (high integrity).

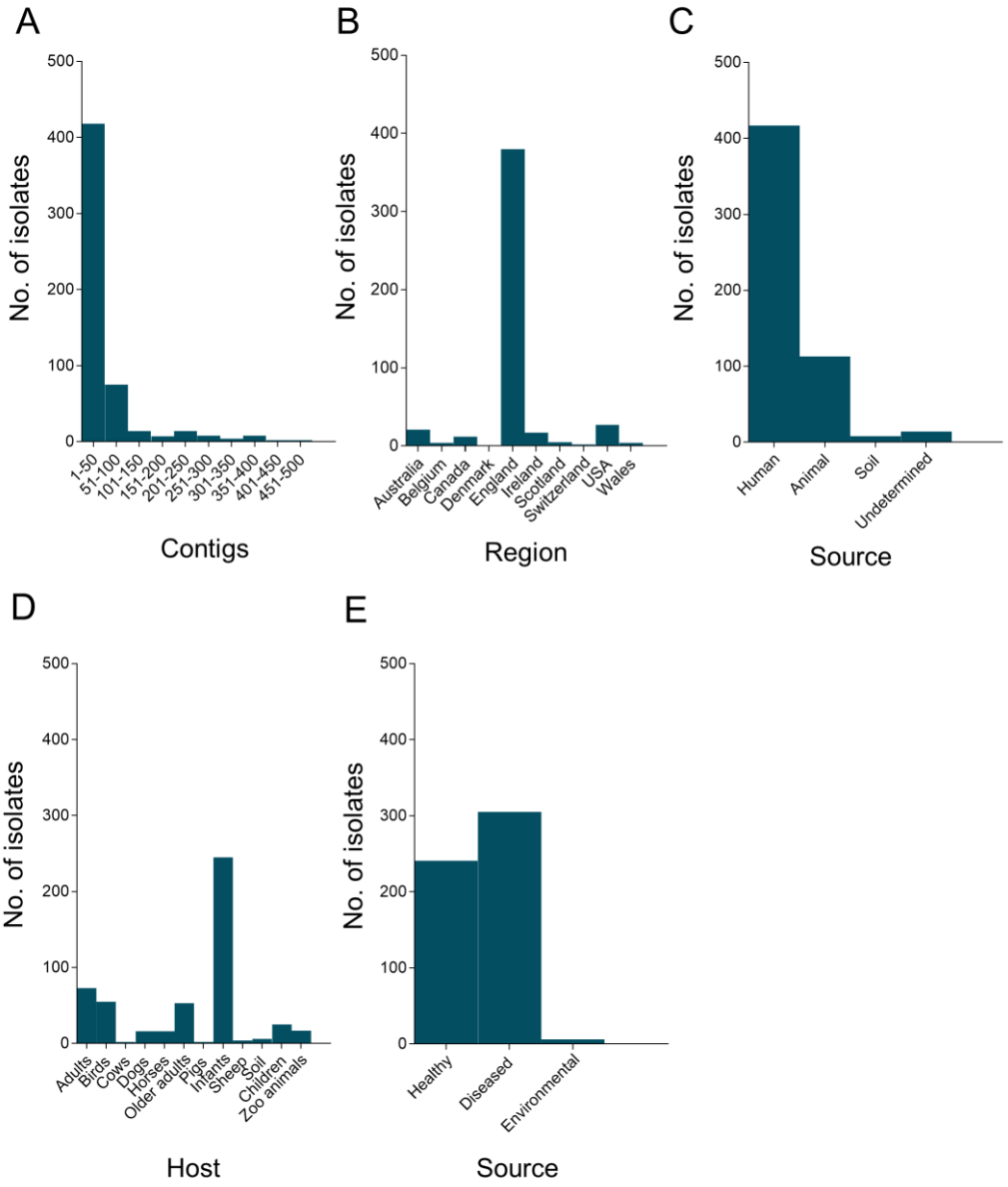


Figure 6.2: Statistics of *C. perfringens* genomes described in Chapter 6

(A) Number of contigs in the assemblies. **(B)** Isolate categories according to isolation region. **(C)** Isolate categories according to human/animal/environmental sources. **(D)** Isolate categories according to hosts. **(E)** Isolate categories according to health conditions.

6.5 Results

6.5.1 Construction of pangenome and phylogeny

6.5.1.1 Pangenome construction reveals genome plasticity

A total of 552 genomes were used to construct a pangenome of 1,251 core genes and 15,500 accessory genes (Figure 6.3). A phylogenetic tree was constructed based on SNPs in the core gene alignment (Figure 6.5). Both bootstrap values (well-supported

in all major nodes, values range: 93-100), and hierBAPS analysis confirmed accurate phylogenetic clustering of isolates.

Previously I reported a snapshot of *C. perfringens* pangenome plasticity (core genome: 12.6% in the pangenome) based upon 56 *C. perfringens* isolates [17]. Notably, the pangenome of 552 *C. perfringens* described in this section demonstrated a significantly higher diversity; 7.4% core genome (pangenome-wide), equivalent to ~40% core genes (1,251/3,000 genes) in a single *C. perfringens* genome (approximately 3,000 genes per genome).

6.5.1.2 Phylogeography demonstrate distinct sub-clusters

Using hierBAPS, all isolates (n=552) clustered into 11 major distinct clades (Figure 6.4A; I-XI). The data did not demonstrate strong phylogeographical clustering which was indicative of frequent transmission between regions. Clonal sub-clades appear to exist in several clusters including lineages II, III, X and XI. The complete phylogenetic tree with annotations is available on iTOL[13].

6.5.1.3 Distinct lineages implicate potential subspecies delineation

ANI was also examined for 552 *C. perfringens* genomes (pair-wise comparison). In relation to type strain ATCC13124 genome, ANI analysis indicated that all genomes belong to *C. perfringens* at species level of >95% (range: 95.95-99.20%; mean ± SD: 98.22 ± 0.69%). The majority of genomes (445, 80.6%) shared ANI > 98.24% with type strain ATCC13124, while the remaining 107 genomes (19.4%) had 95.95% ≤ ANI ≤ 97.89% (Figure 6.4B). This may indicate potential subspecies delineation of isolates within genetically distinct lineages I and II (ANI: 95-96%) and lineages III and V (ANI: 97%), while the difference of 0.35% ANI between the highest ANI in lineage V (isolate NCTC8449) and the rest of the lineages (isolate Q100.CN12; excluding lineages I, II and III), further supports the substantial genetic difference, which is also exemplified by the deep branching phylogeny.

[13] https://itol.embl.de/tree/14915521518111061154177278 6

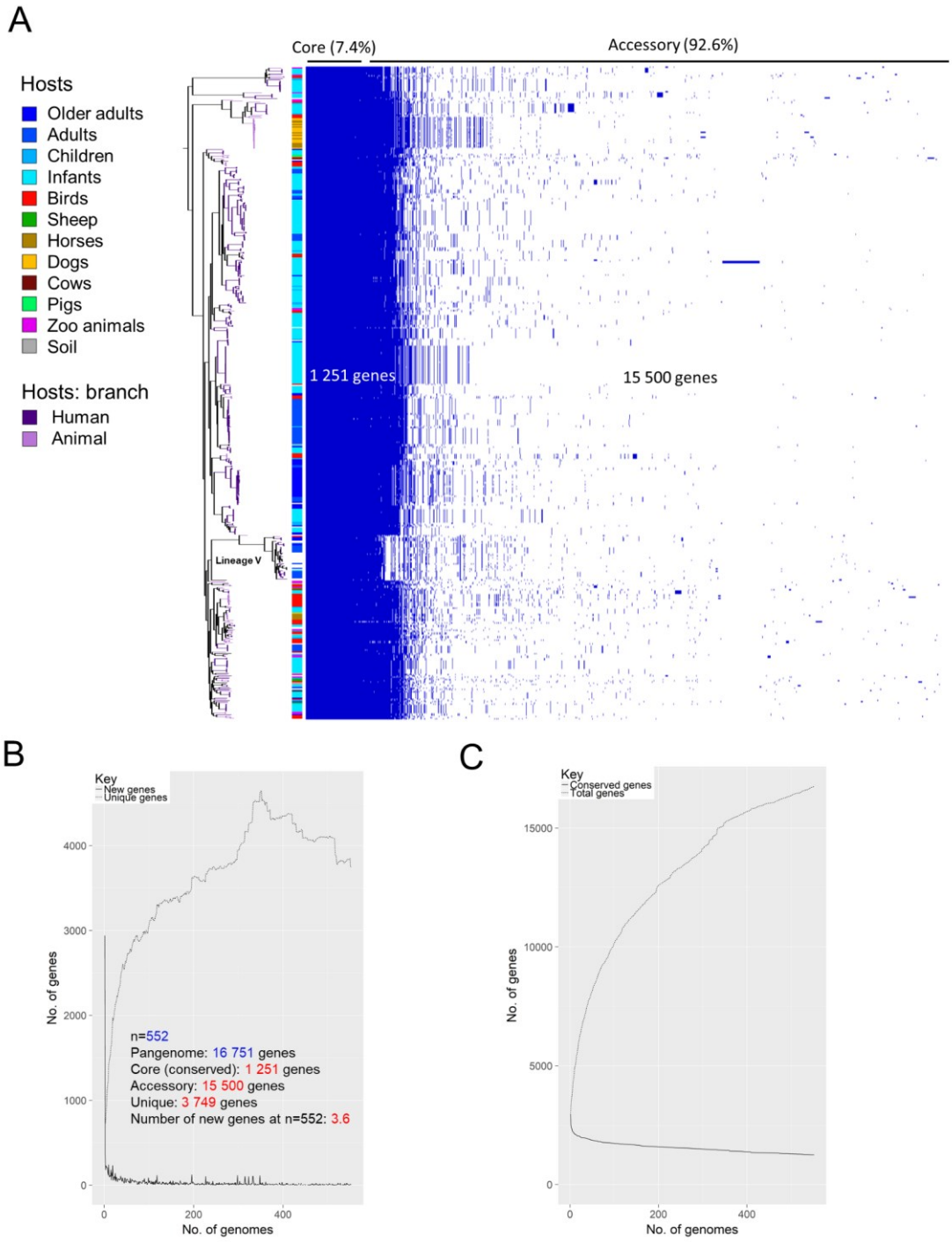


Figure 6.3: Pangenome constructed from 552 isolates of *C. perfringens*

(A) Linearised pangenome visualised by Phandango aligned with mid-point rooted core-genome phylogeny with host labels. **(B)** New genes vs unique gene statistics, with pangenome statistics. **(C)** Core genes (conserved genes) vs total genes statistics. Statistical plots generated via R.

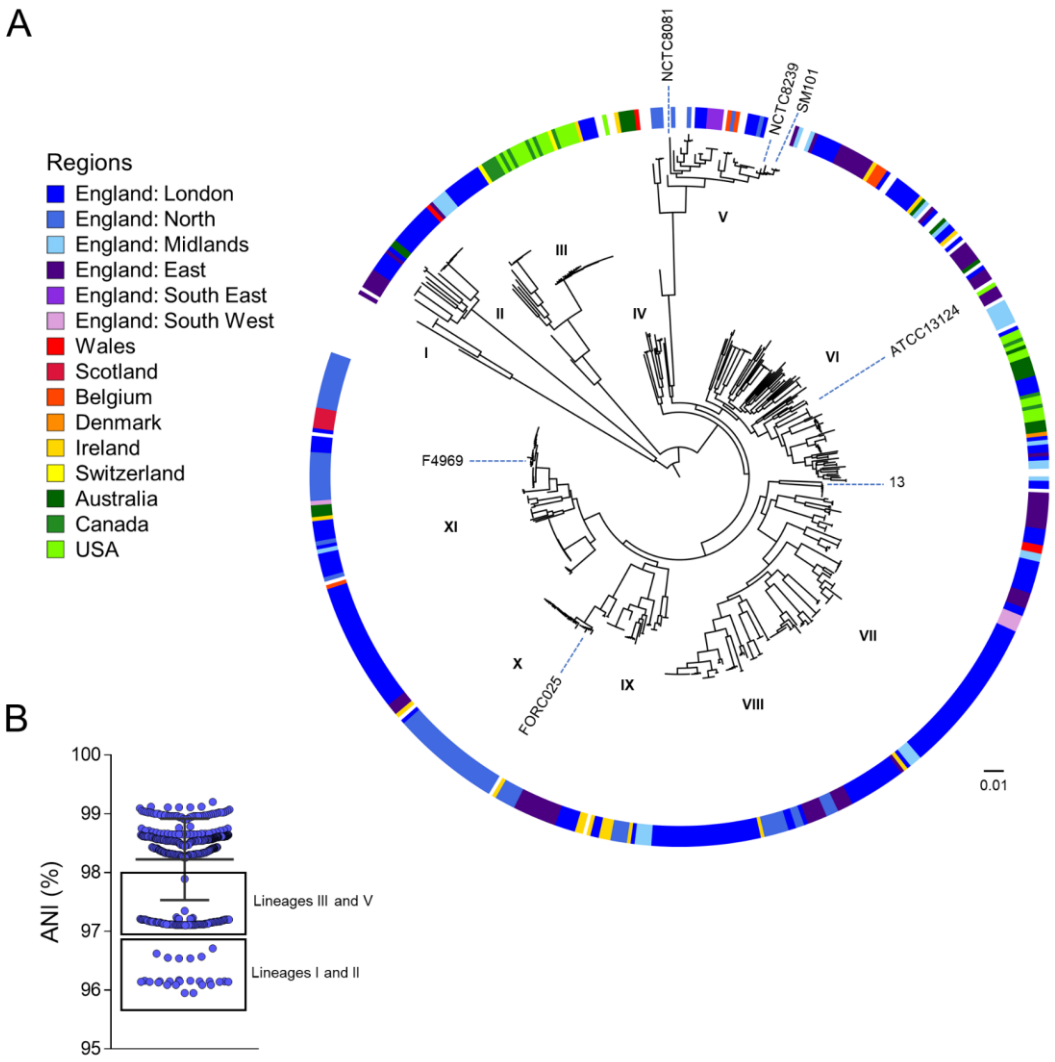


Figure 6.4: Phylogeography of 552 *C. perfringens* strains

(A) Mid-point rooted maximum-likelihood phylogeny inferred from 11,349 SNPs (from alignment of 1,251 core genes) identified in 552 *C. perfringens* genomes. The ring indicates the regions of isolation. Branch lengths are indicative of nucleotide substitution per site. Phylogenetic clustering identified via hierBAPS analysis. Some commonly known isolates are labelled in the tree including ATCC13124, str 13, F4969, FORC025, NCTC8081, NCTC8239 and SM101. **(B)** ANI (%) comparison of 552 isolates in this study in relation to reference genome ATCC13124. Isolates of specific distinct and deep-branching lineages (I, II, III and V) indicate potential subspecies delineation according to ANI.

6.5.2 Potential inference of zoonotic dissemination

C. perfringens has consistently been implicated in both human and animal diseases [1]. Nevertheless, no study to date has ever investigated the potential zoonotic transfer of this pathogen, inferred via WGS techniques, due to scarcity of genomic sequence data (only 58 *C. perfringens* genomes available prior to June 2018, which has recently risen to 112 in November 2018) [374]. To understand potential zoonotic reservoirs, an in-depth

pangenome and robust phylogenetic analysis is required and will be addressed in this chapter.

6.5.2.1 Phylogenetic analysis

The genetic diversity of *C. perfringens* was assessed via phylogenetic analysis. One striking initial observation was co-clustering and extensive mixing of animal and human isolates (Table 6.1; Figure 6.5), indicating potential zoonotic spread between humans and animals, especially within lineages IV and VI (Figure 6.5A). Strains in lineage VII-XI were mainly associated with humans (95.6%; 297/311), while lineage III was dominated by animal isolates (76.2%). Human isolates were also predominant in lineages I (83.3%), II (81.0%) and V (94.8%).

Lineage	Human isolates (%)	Animal isolates (%)	P-value
I	83.3 (5/6)	16.7 (1/6)	>0.99
II	81.0 (17/21)	19.0 (4/21)	>0.99
III	23.8 (10/42)	76.2 (32/42)	<0.01**
IV	40.0 (6/15)	60.0 (9/15)	<0.01**
VI	54.7 (58/106)	45.3 (48/106)	<0.01**
VII-XI	95.4 (297/311)	4.6 (14/311)	<0.01**

Table 6.1: Proportion comparison of human and animal isolates within various lineages

Lineage V was omitted from statistics as sources of 44.7% (17/38) was unable to be determined. 5 isolates from lineage VI, 2 from lineage IV and 1 from lineage XI were omitted due to undetermined sources. Data: Fisher's exact test, two-tailed; comparison with isolate proportion of assumed equal distribution in the major lineages.

The phylogenetic association of human and animal isolates was supported via a focused analysis on two highly similar sub-clusters in lineages VI and VII, comprising both human and animal isolates (ANI>99.0%; Table 6.2). Notably, the mixing of isolates from USA (FC2) and Denmark (98.78718-2) in lineage VI indicated the spread of *C. perfringens* was not confined to regional populations. The absence of accessory genes unique to either group (human and animal) determined via pan-GWAS analysis, indicates that *C. perfringens* may have a stable host-independent accessory genome, or may indicate the potential spread of this pathogen between different host populations.

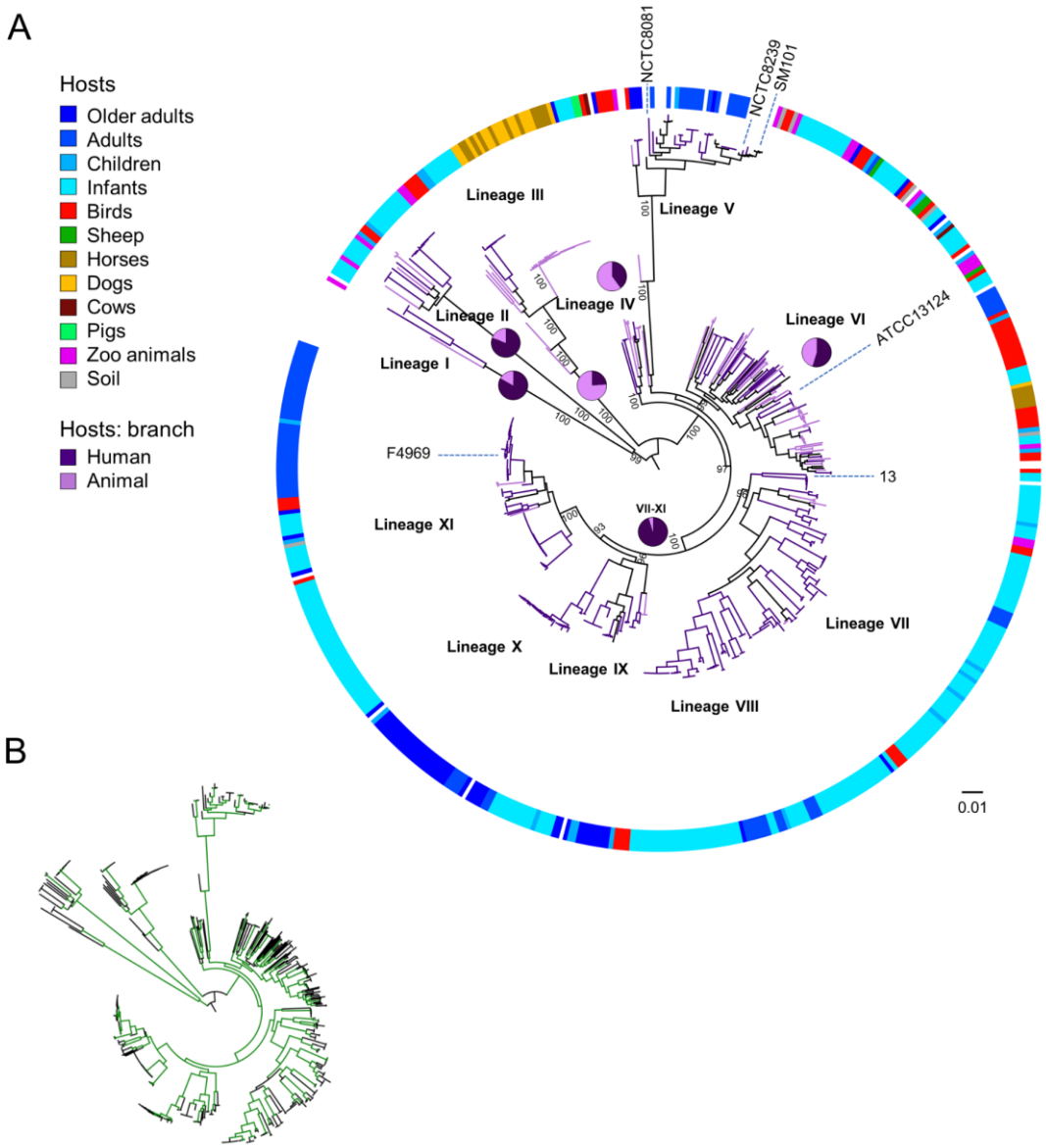


Figure 6.5: Isolate host distribution and proportion in the phylogenetic tree

Mid-point rooted maximum-likelihood phylogeny inferred from 11,349 SNPs (alignment of 1,251 core genes) identified in 552 *C. perfringens* genomes. **(A)** Colour ring indicates hosts of isolates. Coloured branches represent either human or animals' isolates. Pie charts represent the proportion of human- and animal-derived origins **(B)** Green branches and nodes represent bootstrap replicate values > 70 (well-supported branches). Phylogenetic clustering identified via hierBAPS analysis.

Lineage	Strain	Country	Host	ANI (%)
VI	Q137.2.2C	UK	Human	100.00
	FC2	USA	Animal	99.06
	98.78718-2	Denmark	Animal	99.15
	Q049.2.3C	UK	Human	99.48
	Q106.10C	UK	Human	99.29
	I055	UK	Animal	99.39
	J55	UK	Animal	99.39
	C036	UK	Animal	99.34
	Q164.1C	UK	Human	99.15
VII	M092.1C	UK	Human	100.00
	M128.2.3C	UK	Human	99.07
	Q106.8C	UK	Human	99.04
	Q107.4C	UK	Human	99.03
	LHZ155	UK	Animal	99.14
	LHZ158	UK	Animal	99.12
	C033	UK	Animal	99.72
	G050	UK	Animal	99.89
	M092.2C	UK	Human	99.99

Table 6.2: Isolates in lineages VI and VII sub-clusters identified as highly similar

6.5.2.2 Virulence analysis

Given the high number of virulence elements, the distribution of toxin and AMR genes in the pangenome of human and animal isolates was investigated (Figure 6.6).

Toxin genes including *plc* (encodes PLC), *ccp* (encodes clostripain) and *colA* (encodes collagenase) were present in all isolates. Importantly, several toxin genes were statistically enriched in each group of isolates (Fisher's exact test; as $P<0.05$ unless otherwise stated). Sialidase encoding genes *nanI* (76.3%), *nagH* (76.6%) and *nanJ* (70.4%) were enriched in animal isolates (n=113), with the important pore-forming toxin gene *pfoA* also enriched in animal isolates compared to human-derived strains (92.3% vs 62.6%; $P<0.0001$). Accessory toxin gene *cpb2* was statistically enriched, with variant 1 associated with humans ($P<0.05$), while variant 2 was linked with animals ($P<0.0001$). Further analysis into CPB2 variants, via comparing human and animal origins, indicated only 8.5% (6/71) of CPB2-positive animal-derived isolates harbour CPB2 variant 1, while 91.55% (65/71) of these carry CPB2 variant 2, demonstrating a strong veterinary association of CPB2 variant 2 ($P=0.0001$). In

addition, there was a comparable proportion of human-derived CPB2 variant 1 (31.75%; 60/189) and variant 2 (68.25%; 129/189) among CPB2-positive isolates.

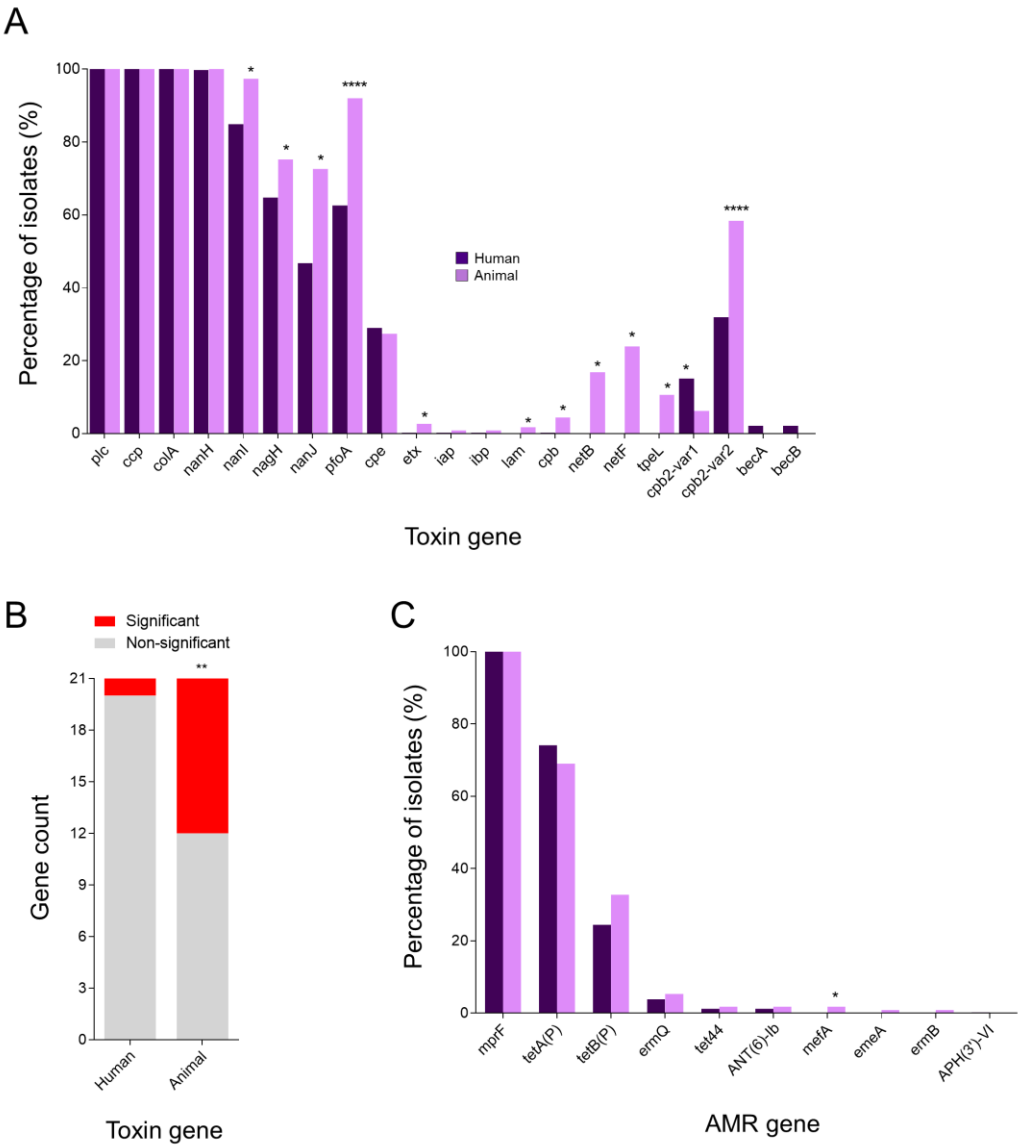


Figure 6.6: Predicted key virulence genes in human and animal *C. perfringens* isolates

(A) Predicted toxin genes. **(B)** Comparison of significantly-enriched toxin genes in both animal and human groups **(C)** Predicted AMR genes. * P<0.05, ** P<0.01, **** P<0.0001; Fisher's exact test (two-tailed).

Other statistically enriched toxin genes in animals included *etx*, *lam*, *cpb*, *netB*, *netF* and *tpeL* (P<0.05), corresponding with clinical findings that these toxin genes are primarily associated with animals, and rarely reported in human isolates [1,125,375]. Overall, toxin genes were more significantly enriched (P<0.001; Fisher's exact test) within animal isolates (42.8%; 9/21 genes), when compared to human isolates (4.7%;

1/21 genes), suggesting host-specificity and evolutionary host-adaptations of virulence toxin genes, which correlates with the fact that most of these enriched genes are plasmid-borne (Figure 6.6).

A total of 9 acquired AMR genes were identified in all genomes (n=530). The most common AMR gene was *tetA(P)* [70.1%; 387/552], a well-known tetracycline efflux pump in *C. perfringens* [376]. Other relatively common AMR genes include tetracycline-resistant *tetB(P)* [39.7%; 139/552] and erythromycin-resistant *ermQ* [3.9%; 22/552]. Importantly, only one AMR gene *mefA* (encodes macrolide efflux pump) was statistically increased in animal isolates (Fisher's exact test, P=0.045). These results support the possible common spread of *C. perfringens* between human and animal populations.

6.5.3 Prevalence and disease associations of virulence potentials

One prominent observation during analysis was that distribution of virulence genes in distinct lineages seemed to differ. This section will focus on discussing the overall distribution of virulence potentials across distinct phylogenetic lineages including toxinotypes, accessory toxins, AMR potentials and virulence plasmids, and associations with disease-phenotypes (Figure 6.7).

6.5.3.1 Toxinotypes

Toxinotype A was the most common type in this pangenome (Figure 6.7), encoding only one typing toxin *plc* (65.5%; 362/552), followed by type F (carries *cpe*; 28.8%; 159/552), type G (carries *netB*; 3.6%; 20/552), type C (0.9%; 5/552), type D (0.54%; 3/552), type E (0.36%; 2/552), and type B (0.18%; 1/552). This pangenome included all 7 toxinotypes, with 6 toxinotypes out of 7 (except type E) detected in lineage VI, suggesting substantial genomic diversity in this lineage. Lineage V consisted of primarily type F isolates (84.2%; 32/38; encode *cpe*). The presence of type F isolates in 8/11 (72.7%) major lineages, indicates the extensive mobility of the *cpe* gene, usually carried on virulence plasmids.

6.5.3.2 Accessory toxins

PFO, a haemolytic toxin, was also encoded in all but one isolate in lineage VI (99.13%; 115/116), while it was entirely absent from lineage V, a cluster linked to FP (0%;

0/38), and partially present in lineage VII and VIII (24%; 35/146), a lineage comprising mainly neonatal isolates (82.8%; 121/146). This important toxin was statistically enriched in diseased isolates (77.7%; 237/305) when compared to isolates obtained from healthy hosts (58.85%; 137/241; P<0.0001, Fisher's exact test), suggesting its role in disease pathogenesis.

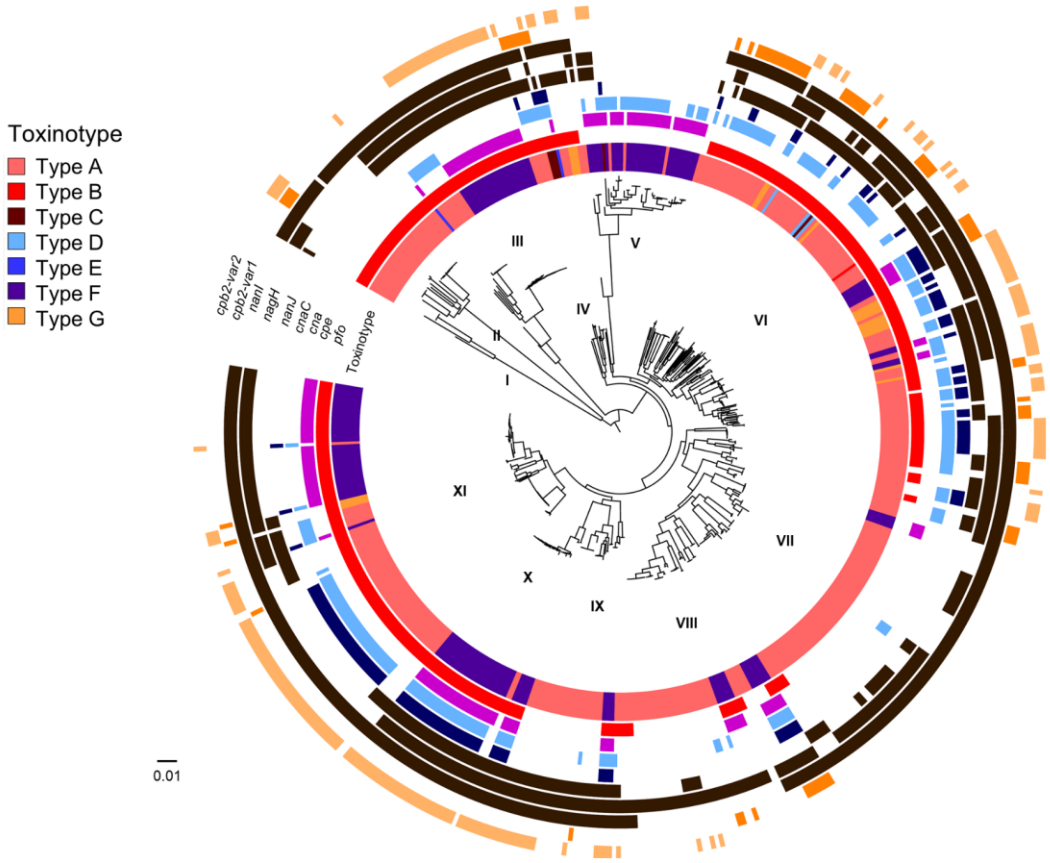


Figure 6.7: Distribution of key accessory virulence genes

Mid-point rooted maximum-likelihood phylogeny aligned with predicted key accessory virulence genes.

Another observation was the low prevalence of collagen binding protein encoding genes *cna* and *cnaC* in lineages I (0%), II (0%) and VII+VIII (23.9%; 35/146). There was a reduced proportion of *cna* in healthy-host-derived isolates (38.1%; 92/241), while *cna* was increased in disease-linked isolates (46.8%; 143/305; P=0.06). Importantly, *cnaC* was statistically enriched (P<0.001) in disease-associated isolates

(59.3%; 181/305) compared to healthy-host isolates (29.9%; 72/241), again implying the important role of *cnaC* in *C. perfringens* virulence.

Sialidase *nanI* was exclusively absent from lineage V (0%; 0/38), with low prevalence of sialidases *nagH* and *nanJ* (both 10.5%; 4/38), representing a distinctive feature of lineage V. This sialidase is known to contribute to human colonic cell adherence, and was not detected in any of the isolates in this FP-associated clade, which may link to the often short-duration of FP symptoms [182].

It was evident that only one variant of CPB2 could be encoded in each CPB2-positive genome (98.2%; 226/271), with the exception of 5 isolates (within lineage II and IV) that carry both variants. CPB2 variant 1 in lineage VI (33.6%; 39/116) was comparable to CPB2 variant 2 (40.5%; 47/116), a lineage known to comprise an extensive mixture of animal and human isolates, denoting the potential veterinary origin of CPB2 variant 1 (shown in previous section) [151]. Disease-association wise, CPB2 variant 2 was enriched (P=0.004) in diseased-host isolates (42.6%; 130/305) compared with healthy-host-derived isolates (29.8%; 72/241), suggesting a role of CPB2 variant 2 in disease development [377].

6.5.3.3 AMR potentials

The prevalence of most commonly acquired AMR genes, namely *tetA(P), tetB(P)* and *ermQ*, was examined in this 552 strain pangenome, with the aim to link specific clusters to AMR potentials (Figure 6.8).

It was determined that 71.1% (393/552) of the isolates in the pangenome encode *tetA(P)*, 25.7% encoded *tetB(P)*(142/552), and 3.9% of the isolates encoded *ermQ* (22/552).

Of note, isolates in lineage I, V and X encoded statistically less AMR genes (P=0.02; Mann-Whitney test) compared to other lineages. In lineage V, there was only 10.5% (4/38) isolates that carried 1 AMR gene *tetA(P)*, this could be due to the reason that 17 (out of 38) isolates in this lineage were obtained from NCTC collection of pre-antibiotic era (before 1950), and all these NCTC isolates (100%; 17/17) did not carry any identified acquired AMR genes. Similarly, most isolates in lineage X (78.9%; 30/38), mainly contemporary FP-linked CH isolates, were found not to carry any

acquired AMR genes either. Lineage I also displayed low prevalence of AMR carriage (16.6%; 1/6).

By contrast, the majority of isolates in remaining lineages encoded more acquired AMR genes (P=0.02) than lineages I, V and X, most commonly *tetA(P)* as represented in lineage II (71.4%; 15/21), III (88.1%; 37/42), IV (70.6%; 12/17), VI (50.0%; 58/116), VII + VIII (99.3%; 145/146), IX (55.9%; 29/34) and XI (81.4%; 48/59).

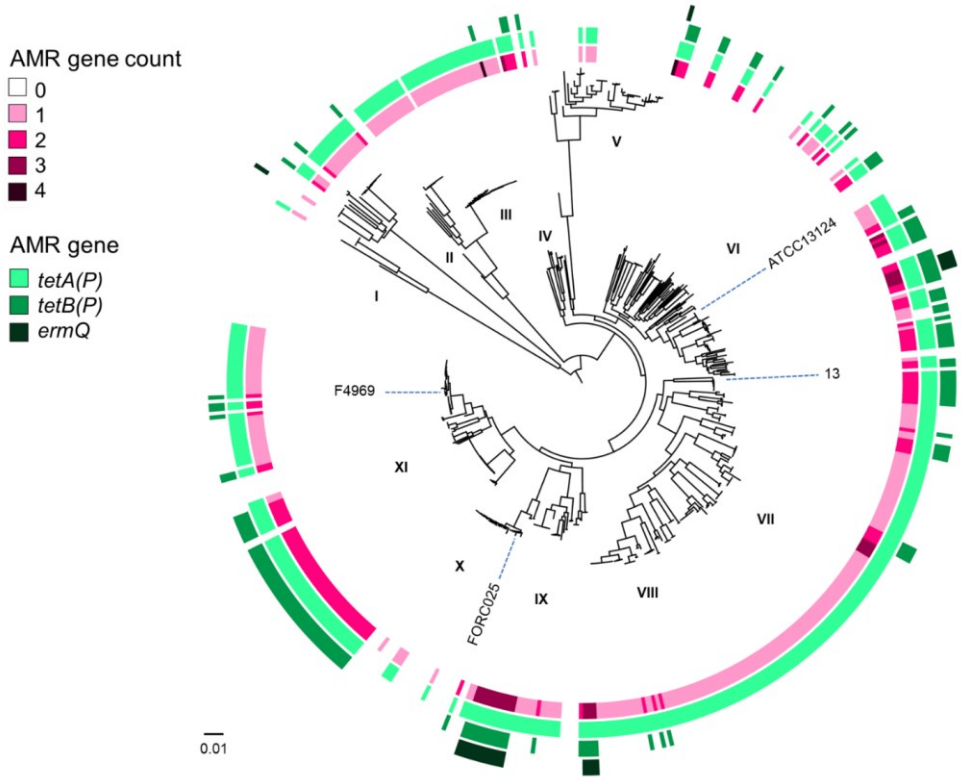


Figure 6.8: Distribution of key AMR genes in the phylogeny

Mid-point rooted maximum-likelihood phylogeny aligned with predicted key AMR genes and AMR gene count statistics.

Importantly, lineage VII and VIII, a lineage comprised primarily of neonatal isolates (82.8%; 121/146), displayed the highest prevalence of tetracycline efflux pump *tetA(P)* at 99.3%, suggesting a selective pressure on the acquisition of AMR genes within the hospital environment, especially NICU [378].

Multiple AMR genes (n=3), namely *tetA(P)*, *tetB(P)* and *ermQ* (a macrolide-lincosamide-streptogramin B resistance gene) were detected in several isolates in

lineage VI (n=4; neonatal), lineage VIII (n=4; chicken), and lineage IX (n=12; neonatal). Acquisition of erythromycin-resistant determinant *ermQ* may be linked to the common prescription of this antibiotic during pregnancy to prevent *Streptococcus* infections, similarly, erythromycin is also known to be frequently administered in the broiler chicken industry [379-381].

6.5.3.4 Prevalence of virulence plasmids

Virulence plasmids are an important aspect in the spread of disease-relevant virulence factors in *C. perfringens* including important toxins CPE (FP-linked) and NetB (NE-linked), and AMR genes [51,130,292]. Most virulence plasmids of *C. perfringens* are conjugative, carrying conjugative systems consisting of conjugative loci (*tcp* genes) [41,291]. To examine the prevalence of conjugative virulence plasmids in *C. perfringens*, conjugative-plasmid-specific Tcp systems (a total of 10 known Tcp loci, different plasmids carry different numbers of Tcp loci) were screened as these genes would indicate the presence of conjugative plasmids (Figure 6.9).

Around half (46.5%) of the strains analysed encoded conjugative plasmids, with low prevalence (P=0.017; Mann-Whitney test) detected in lineages II (28.5%; 6/21), V (10.5%;4/38) and VI + VII (15.0%; 22/146), when comparing with other lineages, e.g. lineages I (83.3%; 5/6), III (71.4%; 30/42), IV (64.7%; 11/17), VI (57.7%; 67/116), IX (94.1%; 32/34), X (86.8%; 33/38) and XI (83.1%;49/59).

Importantly, it was also observed that *cpe* was encoded on plasmids (100% *cpe*-Tcp pairing) in all lineages (III, IV, VI, VII, VIII, IX, X and XI) with the exception of lineage V (Figure 6.9). Lineage V contained 89.4% *cpe*-carrying isolates (34/38), however, only 8.8% (3/34) of these isolates appeared to co-encode conjugative plasmids, suggesting these *cpe* toxin genes may be encoded within the chromosome, supported by the presence of historical FP isolates SM101 and NCTC8239 in this cluster [382]. The presence of Tcp in all lineages (100%) implicated the widespread nature of *C. perfringens* conjugative plasmids.

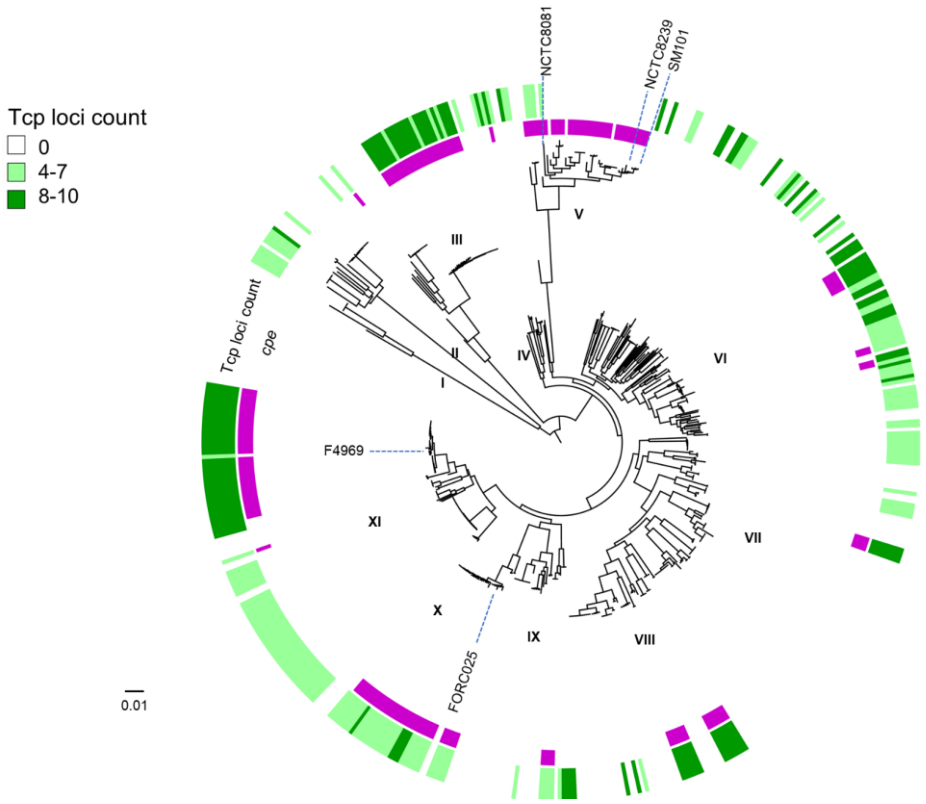


Figure 6.9: Distribution of predicted Tcp loci in the phylogeny

Mid-point rooted maximum-likelihood phylogeny aligned with predicted Tcp loci (conjugative loci: conjugative plasmid-related).

6.5.4 A persistent circulating genotype in the UK that causes both non-foodborne diarrhoea and FP

Sporadic diarrhoea isolate F4969, isolated in 1990s in Europe (exact location could not be determined), has long been used to investigate diarrhoea-associated virulence features in *C. perfringens* [116]. This strain carries a well-known plasmid pCPF4969 that encodes *cpe*, an enterotoxin linked to both foodborne and non-foodborne diarrhoea as described in Chapter 4 [22]. Notably, this isolate apparently clustered within lineage XIb (using hierBAPS), which consisted primarily of isolates from FP patients (81.5%; 31/38) in North East England (Figure 6.10).

Data further suggested a close genetic relationship between this sporadic diarrhoea strain F4969 and FP strains isolated between 2011-2017, as SNP analysis revealed genetic distance between isolates within lineage XIb to be 64.4 ± 22.0 SNPs (mean ± SD; range of 34-83 SNPs; pair-wise comparison) versus isolates outside of lineage XIb (XIa, closest sub-lineage) at 377.8 ± 55.9 SNPs (Figure 6.10C). Genome ANI

analysis supported the findings, all pair-wise ANI values were >99.77%, indicating highly-similar core-genome nucleotide identity (Table 6.3).

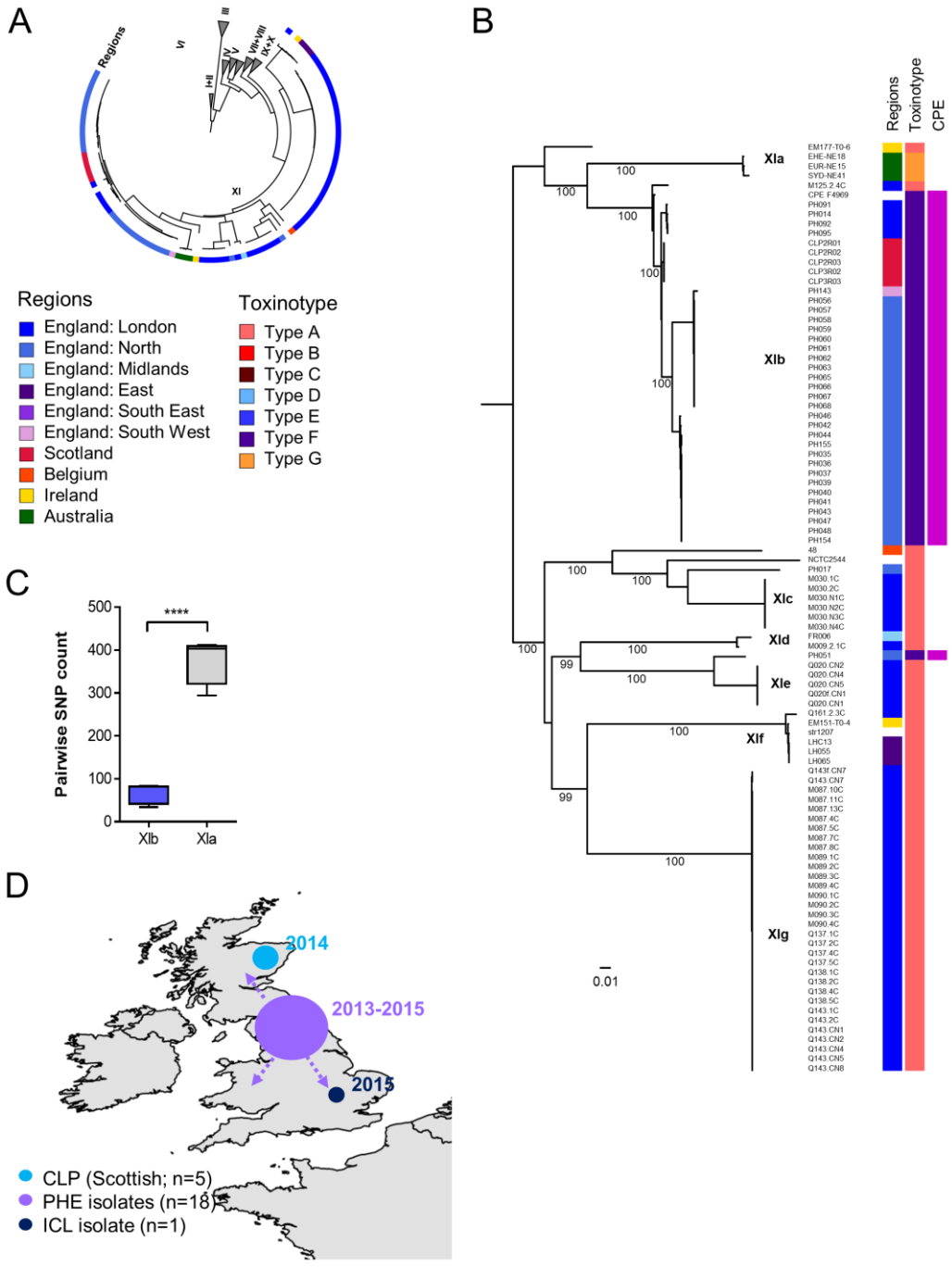


Figure 6.10: Genetic relationship of isolates nested within lineage XI

Figure above reveals potential regional circulating CPE-strains in North East England and Scotland. **(A)** Position of lineage XI in the pangenome tree. **(B)** zoomed-in mid-point rooted maximum-likelihood phylogeny constructed from 2,809 core-gene SNPs. Lineages assigned via hierBAPS analysis (level 2). Scale bar is indicative of nucleotide substitution per site. Bootstrap values are shown on the tree. **(C)** The pair-wise SNP comparison of lineage XIb and lineage Xa with respect to isolate F4969. **** P<0.0001 (Mann-Whitney test). **(D)** The depiction of potential spread of closely-related strains (lineage XIb) across UK between 2013-2015. Map generated via R using package *rworldmap*.

Importantly, 5 isolates (CLP isolates) obtained from Scotland (healthy individuals) nested within this North-East England clade, indicating this strain was regional-specific, potentially a diarrhoea-causing strain with CPE-encoding virulence plasmid. Nevertheless, as source information of strain F4969 was only available as 'Europe', it suggests this cluster may represent a persistent *C. perfringens* genotype, potentially widespread in Europe and possibly linked to North East England and Scotland regions (Figure 6.10D).

Strain	Year	Case	Region	ANI (%)	IS*1470*-like	SNP
CPE_F4969	1990s	Sporadic diarrhoea	Europe	100	Detected	0
PH014	2013	Food poisoning	England	99.84	Detected	43
PH091	2013	Food poisoning	England	99.83	Detected	43
PH092	2013	Food poisoning	England	99.83	Detected	42
PH095	2013	Food poisoning	England	99.82	Detected	42
M125.2.4C	2015	Healthy child	England	99.80	Not detected	54
CLP2R01	2014	Healthy adult	Scotland	99.80	Detected	34
CLP2R02	2014	Healthy adult	Scotland	99.80	Detected	34
CLP2R03	2014	Healthy adult	Scotland	99.80	Detected	34
CLP3R02	2014	Healthy adult	Scotland	99.80	Detected	34
CLP3R03	2014	Healthy adult	Scotland	99.80	Detected	34
PH044	2015	Food poisoning	England	99.78	Detected	83
PH155	2015	Food poisoning	England	99.77	Detected	83
PH046	2015	Food poisoning	England	99.77	Detected	83
PH042	2015	Food poisoning	England	99.77	Detected	83
PH035	2015	Food poisoning	England	99.78	Detected	82
PH036	2015	Food poisoning	England	99.78	Detected	82
PH037	2015	Food poisoning	England	99.77	Detected	82
PH039	2015	Food poisoning	England	99.78	Detected	82
PH040	2015	Food poisoning	England	99.78	Detected	82
PH041	2015	Food poisoning	England	99.78	Detected	82
PH043	2015	Food poisoning	England	99.78	Detected	82
PH047	2015	Food poisoning	England	99.78	Detected	82
PH048	2015	Food poisoning	England	99.78	Detected	82
PH154	2015	Food poisoning	England	99.77	Detected	82

Table 6.3: Comparative analysis of isogenic strains within F4969-related lineage XIb

SNP distance and ANI were performed with respect to strain F4969. IS1470-like is a specific insertion sequence for plasmid pCPF4969 that was first identified in strain F4969.

Of note, an English isolate M125.2.4C obtained from a child in London was only 54 SNPs different from the F4969 type strain, and 12 SNPs different from FP strains PH092 and PH095, indicating a wider and longer-term reservoir of this disease-causing *C. perfringens* genotype. However, this isolate does not carry the CPE plasmid that is considered as the causative agent of *C. perfringens*-associated diarrhoea (Figure 6.10D; refer to discussion in Chapter 4 for potential plasmid dissemination on FP-related isolates).

6.5.5 Lineage IX and X: revealing potential intercontinental transfer and care-home-derived persistent genotypes

C. perfringens can be associated with persistent diarrhoea and 'less healthy conditions' (i.e. long-term hospitalisation) in vulnerable communities, including older adults [289,334]. As FP in CHs were reported frequently in North East England between 2011-2014 (described in Chapter 4), it is therefore important to understand the spread of *C. perfringens* strains among this group of individuals, which is critical to limiting any associated public health impact [316].

Isolates (n=8; total n=17 sequenced) sourced from the Irish ELDERMET study (elderly microbiome), were detected to lie within lineages IX and X (Figure 6.11A; Table 6.6) [289]. Isolates shared a close genetic relationship within sub-lineages Xc and Xd (76.7 ± 37.1 SNPs; mean ± SD; range of 10-98 SNPs) versus isolates outside these lineages (lineages Xa and Xb) at 372.8 ± 128.5 SNPs (with respect to isolate FORC025; $P<0.0001$; Mann-Whitney test), as confirmed by pair-wise comparison of ANI of >99.78% (Table 6.4).

Notably, isolate FORC025, a FP isolate obtained in South Korea in 2015 (NCBI source data), demonstrated a strong genetic similarity to Irish ELDERMET isolate EM297-T0 (SNP difference: 13 SNPs; lineage Xc), and other English FP isolates (n=28; FP isolates obtained from CHs in North East England) in lineage Xc and Xd. However, these isolates (FORC025 and EM297-T0) did not carry *cpe* (causative toxin for FP encoded in virulence plasmids) in their genomes. This could indicate the potential loss of the CPE plasmid (predicted to be plasmid pCPF5603, confirmed absence of the plasmid by plasmid-specific IS search of IS*1151*, a unique sequence encoded in pCPF5603) in both FORC025 and EM297-T0, as implicated by the remaining 28

isolates (93%; 28/30) carrying CPE virulence plasmids, and the fact that FORC025 was an FP isolate can be attributed to its CPE toxin production.

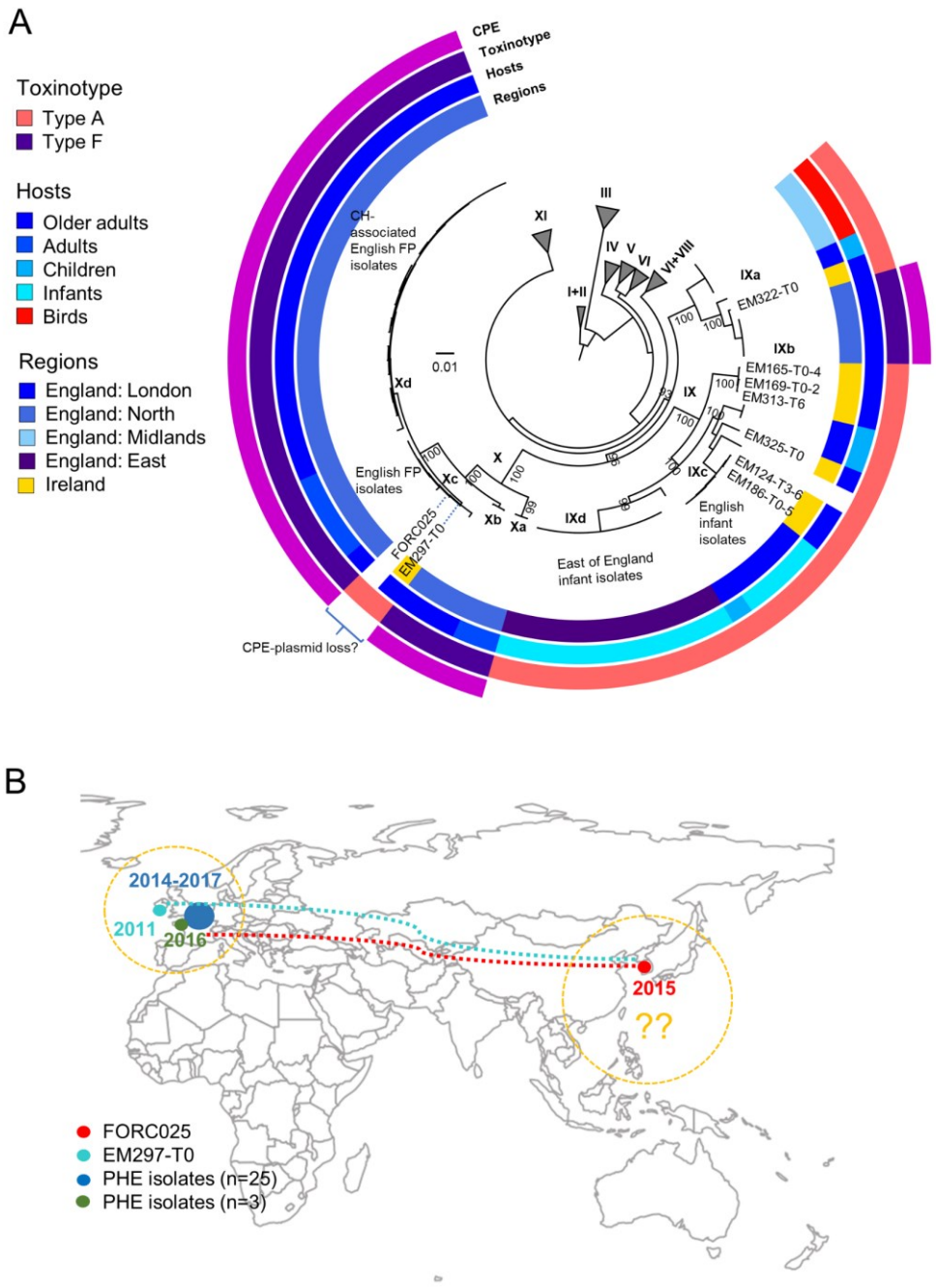


Figure 6.11: Potential intercontinental transfer and care-home-derived persistent genotypes

(A) Genetic relationship of isolates nested within lineages IX and X and associated virulence profiles and metadata. **(B)** Care home-related clade and potential interregional spread as far as far-east between 2011-2017. FORC025 is an FP isolate from South Korea (SAMN04209542). Clustering determined using hierBAPS (level 1& 2). Bootstrap values are represented in the tree. Map generated in R using package *rworldmap*.

Importantly, these findings imply the potential link between CH isolates in Ireland and North East England, as demonstrated by this close genetic relationship (Figure 6.11B). Furthermore, as FORC025 is a Korean FP isolate, this would suggest intercontinental transfer of a persistent genotype of FP-causing *C. perfringens* strain(s), sharing surprisingly high genomic similarity with those in Europe (10-191 SNPs; 74.2 ± 38.9 SNPs) in the same clade (Figure 6.11B).

ID	Year	Case	Region	ANI (%)	IS*1151*	Lineage	SNP
FORC025	2015	FP	Korea	100.00	Not detected	Xc	0
EM297-T0	2011	CH	Ireland	99.95	Not detected	Xc	13
PH123	2015	FP	England	99.91	Detected	Xc	10
PH124	2015	FP	England	99.91	Detected	Xc	10
PH072	2016	CH-FP	England	99.78	Detected	Xc	84
PH054	2015	FP	England	99.91	Detected	Xc	10
PH055	2015	FP	England	99.92	Detected	Xc	10
PH100	2014	CH-FP	England	99.92	Detected	Xc	55
PH016	2014	CH-FP	England	99.79	Detected	Xd	85
PH018	2014	CH-FP	England	99.79	Detected	Xd	85
PH019	2014	CH-FP	England	99.79	Detected	Xd	85
PH030	2014	CH-FP	England	99.81	Detected	Xd	85
PH031	2014	CH-FP	England	99.80	Detected	Xd	85
PH052	2015	CH-FP	England	99.78	Detected	Xd	87
PH053	2015	CH-FP	England	99.78	Detected	Xd	87
PH073	2016	CH-FP	England	99.78	Detected	Xd	86
PH074	2016	CH-FP	England	99.78	Detected	Xd	85
PH075	2016	CH-FP	England	99.78	Detected	Xd	86
PH076	2016	CH-FP	England	99.79	Detected	Xd	85
PH077	2016	CH-FP	England	99.79	Detected	Xd	86
PH079	2017	CH-FP	England	99.79	Detected	Xd	89
PH096	2013	CH-FP	England	99.78	Detected	Xd	87
PH097	2014	CH-FP	England	99.79	Detected	Xd	87
PH098	2014	CH-FP	England	99.79	Detected	Xd	86
PH101	2014	CH-FP	England	99.79	Detected	Xd	87
PH102	2014	CH-FP	England	99.66	Detected	Xd	191
PH103	2014	CH-FP	England	99.78	Detected	Xd	88
PH138	2016	CH-FP	Wales	99.83	Detected	Xd	98
PH139	2016	CH-FP	Wales	99.82	Detected	Xd	97
PH140	2016	CH-FP	Wales	99.82	Detected	Xd	97

Table 6.4: Comparative analysis of highly-similar strains within lineages Xc and Xd

Interestingly, three isolates in lineage IXb (EM165-T0-4, EM169-T0-2 and EM313-T6) and two isolates in IXc (EM124-T3-6 and EM186-T0-5) were highly similar at the SNP level (all isolates obtained from long-stay CH residents in the ELDERMET study; Table 6.5). In lineage IXb, pair-wise SNPs were shown to be 3.6 ± 3.5 SNPs

(mean ± SD; range of 0-7 SNPs; ANI>99.99%), while IXc two isolates differed by only 1 SNP, confirming highly-similar persistent strains (<7 SNPs; ANI>99.99%) circulating in long-stay Irish CHs and their potential relationship with English isolates (Figure 6.11). Notably, with respect to their virulence capacity, these elderly isolates were all type A according to current toxinotyping scheme, 53% (9/17), and did not harbour any other major typing/ accessory toxins, with an exception of EM200-T6 that carried CPB2-variant2 (Table 6.6).

EM165-T0 and EM169-T0 were previously demonstrated to be identical via PFGE technique [334]. In this current work, WGS provided a higher resolution showing 4 SNPs difference between these two CH isolates (Table 6.5), highlighting the robustness of WGS when applied in routine CH surveillance of *C. perfringens* strains.

Lineage	Strain	Host	ANI (%)	SNP
IXb	EM165-T0-4	Long-stay OA	100.00	0
	EM169-T0-2	Long-stay OA	99.99	4
	EM313-T6	Long-stay OA	99.99	7
IXc	EM124-T3-6	Long-stay OA	100.00	0
	EM186-T0-5	Long-stay OA	99.99	1

Table 6.5: Strain similarity comparison of Irish ELDERMET *C. perfringens* strains

Strain	*cpb*	*cpe*	*etx*	*iap*	*ibp*	*netB*	*pfo*	*plc*	Toxinotype
EM082-T6-5	-	-	-	-	-	-	+	+	A
EM095-T0-2	-	-	-	-	-	-	-	+	A
EM124-T3-6	-	-	-	-	-	-	-	+	A
EM151-T0-4	-	-	-	-	-	-	+	+	A
EM159-T0-8	-	-	-	-	-	-	+	+	A
EM159-T6-2	-	-	-	-	-	-	-	+	A
EM165-T0-4	-	-	-	-	-	-	-	+	A
EM169-T0-2	-	-	-	-	-	-	-	+	A
EM177-T0-6	-	-	-	-	-	-	+	+	A
EM186-T0-5	-	-	-	-	-	-	-	+	A
EM200-T6	-	-	-	-	-	-	+	+	A
EM297-T0	-	-	-	-	-	-	+	+	A
EM313-T6	-	-	-	-	-	-	-	+	A
EM322-T0	-	-	-	-	-	-	+	+	A
EM325-T0	-	-	-	-	-	-	-	+	A
EM331-T0	-	-	-	-	-	-	+	+	A
EM342-T0	-	-	-	-	-	-	+	+	A

Table 6.6: *In silico* toxinotyping of ELDERMET *C. perfringens* isolates

6.5.6 Genome features and potential intra-regional spread of food-poisoning-derived isolates within distinctive lineage V

Human FP can be caused by three sub-types of type-F *C. perfringens* strains as identified by *cpe* gene encoded exclusively on: (1) chromosome (or, more accurately carried on transposable element Tn*5565*, frequently found to be integrated into chromosome); (2) plasmid pCPF5603; or (3) plasmid pCPF4969, as described in Chapter 4 [60,322]. Historically, type-F FP isolates were mainly described as chromosomal-*cpe* strains that were determined to be genetically distinct from plasmid-carrying strains, characterised by the lack of *pfo* gene in these strains including strains NCTC8239 and SM101 [322,382]. This section describes the unique genome features and potential intra-regional spread of chromosomal-*cpe* strains using primarily isolates from NCTC collection that were not discussed in Chapter 4.

Historical isolates from NCTC collection (n=15) clustered in lineages Vc and Vd (n=33) as a genetically distinct lineage (lineage V, exemplified by the long branching), supported by bootstrap values of 100% on major divergent nodes (Figure 6.12A). All the isolates in FP-associated lineage Vc and Vd shared a close genetic relationship (pair-wise ANI>99.3%; within-lineage pair-wise SNP: 314.1 ± 110.1 SNPs, Table 6.7; Figure 6.12B), agreeing with previous observations [60,322]. These isolates did not encode accessory toxin PFO (0%; 0/37) and sialidase NanI (0%; 0/37). Data showed that 92% (34/37) of these isolates carried CPE as FP-causative toxin, with the exception of NCTC8678 (FP human isolate in 1951), Q146.2.5C (child isolate obtained in 2014) and PH029 (English FP isolate), possibly due to the loss of Tn*5565*-*cpe,* potentially during the sub-culturing processes as discussed previously in Chapter 4.

Notably, lineage Vb was genetically distinct (compared with lineage Vc and Vd, pair-wise SNP: 824.4 ± 169.4 SNPs; Figure 6.12B) from lineages Vc and Vd (within-lineage Vc and Vd pair-wise SNP: 314.1 ± 110.1 SNPs), and these three FP isolates in lineages Vb (PH015, PH032 and PH033) were obtained from English CHs, the other was a chicken isolate NCTC10240. These strains carried sialidase NanJ (100%;4/4) and CPB2 (75%;3/4), while isolates in lineages Vc and Vd did not (100%; 33/33).

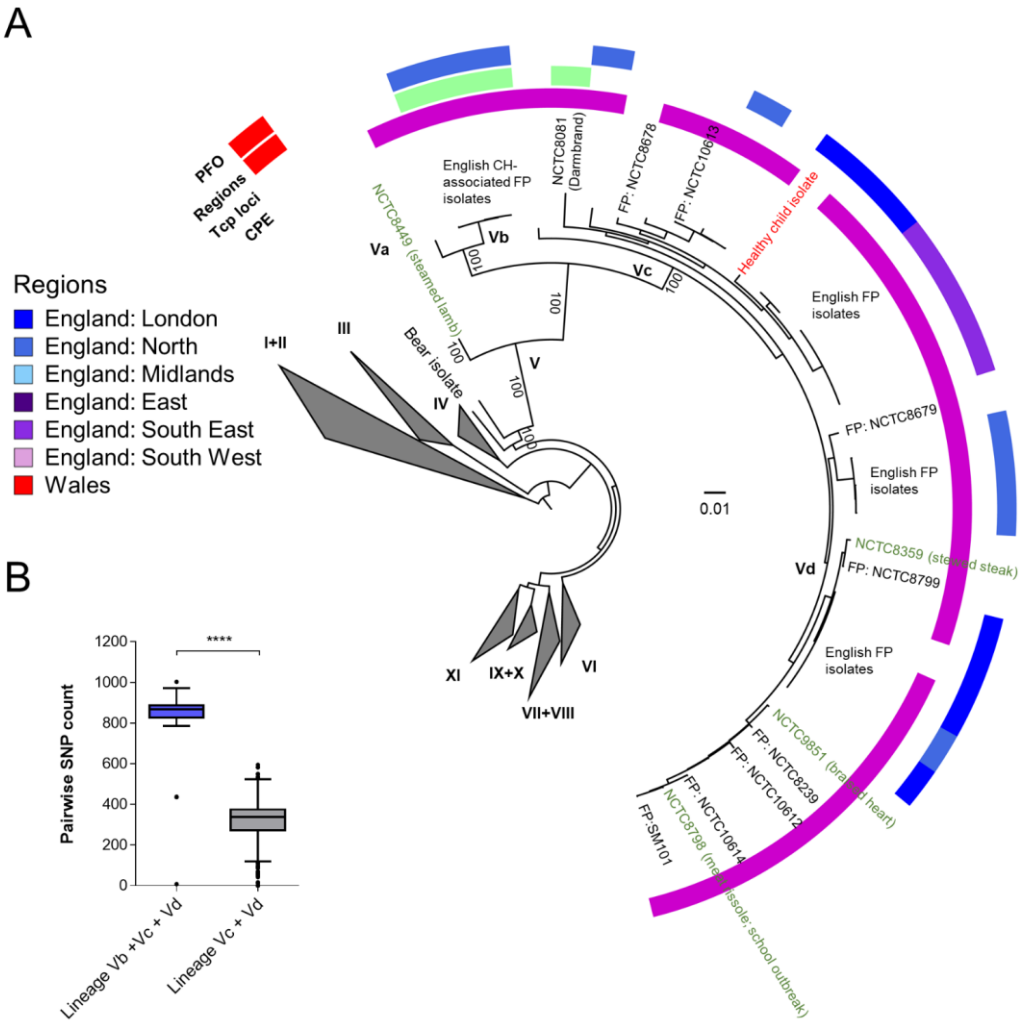


Figure 6.12: Genetic relationship of isolates within lineage V

(A) Genetic relationship of isolates nested within lineage V consists of mainly food-poisoning associated isolates. **(B)** Pair-wise SNP distances between isolates within and across sub-clades of lineage V shown in Tukey box plots. Origins of NCTC isolates are not determined in the labels as majority of the metadata are incomplete. Sub-lineages are assigned via hierBAPS analysis (level 2). Stats: Mann-Whitney test, two-tailed; **** P<0.0001.

With regard to AMR, only recent isolates PH135, PH136, Q146.2.5C and PH127 (n=4) were computationally predicted to encode acquired AMR gene *tetA(P)*, an efflux pump effective against tetracycline, while the rest of the isolates did not carry any acquired AMR gene (89%; 33/37).

The genomes of these isolates in lineage V had a smaller genome size (range 3.0-3.2 Mb vs average genome size of 3.6 Mb; except Darmbrand isolate NCTC8081 3.6 Mb),

implicating the lack of plasmid acquisition in these isolates or a distinct genome arrangement [17,383].

Human Darmbrand (also known as enteritis necroticans) isolate NCTC8081 is a heat-resistant type C isolate that carries conjugative-plasmid-derived CPB (or, beta-toxin; conjugative system detected; Table 6.7) responsible for malnutrition-associated necrotizing gut infections primarily in children post-war [158]. The inclusion of this isolate in lineage V might indicate the genetic relatedness to the rest of the FP-derived isolates, which agrees with previous findings [383]. NCTC8081 bore the general characteristic of lineage V isolates: lacking PFO, CPB2, NanI, NanJ and acquired AMR genes (it is a pre-antibiotic-era isolate). It had, however, a relatively larger genome of 3.6 Mb (vs 3.0 Mb), denoting a high acquisition of mobile genetic elements, potentially virulence plasmids that transformed this isolate from type F to type C which carried CPB [383]. Interestingly, NCTC8081 was the only conjugative-plasmid-carrying isolate in lineage Vc and Vd.

6.5.7 Two-year NEOM isolates: long-term persistence and detection of deadly toxin genes

The two-year NEOM project is a continuation of NEOM study (preterm infant study described in Chapter 3), which includes additional two year post NICU discharge samples from former NEOM participants that were screened for *C. perfringens* in faecal samples, aiming to determine potential persistence [12]. A total of 25 isolates were obtained from 23 of the original NEOM NICU preterm infants.

The two-year NEOM isolate Q135.2.1C shared ANI of 99.95% with preterm isolates from NEOM preterm infants Q124 (n=4) and Q125 (n=4), with only 44 SNPs difference, denoting potential within-ward transfer of *C. perfringens* (Figure 6.13). This finding gives a snapshot that certain hospital-acquired strains may persist long-term (i.e. two years post discharge from NICU) in the infant gut.

Strain	Case	*pfo*	*cpe*	Tcp loci	ANI (%)
NCTC8239	FP	Not detected	Detected	Not detected	100.00
NCTC10239	Undetermined	Not detected	Detected	Not detected	99.50
NCTC10612	FP	Not detected	Detected	Not detected	99.82
NCTC10613	FP	Not detected	Detected	Not detected	99.48
NCTC10614	FP	Not detected	Detected	Not detected	99.84
NCTC8081	Darmbrand	Not detected	Detected	Detected	99.20
NCTC8238	Undetermined	Not detected	Detected	Not detected	99.88
NCTC8247	Undetermined	Not detected	Detected	Not detected	99.57
NCTC8359	FP (food)	Not detected	Detected	Not detected	99.66
NCTC8678	FP	Not detected	Not detected	Not detected	99.49
NCTC8679	FP	Not detected	Detected	Not detected	99.50
NCTC8797	Undetermined	Not detected	Detected	Not detected	99.50
NCTC8798	FP (food)	Not detected	Detected	Not detected	99.86
NCTC8799	FP	Not detected	Detected	Not detected	99.66
NCTC9851	FP (food)	Not detected	Detected	Not detected	99.84
PH004	FP	Not detected	Detected	Not detected	99.43
PH005	FP	Not detected	Detected	Not detected	99.42
PH006	FP	Not detected	Detected	Not detected	99.42
PH007	FP	Not detected	Detected	Not detected	99.42
PH028	FP	Not detected	Detected	Not detected	99.64
PH029	FP	Not detected	Not detected	Not detected	99.65
PH078	FP	Not detected	Detected	Not detected	99.47
PH080	FP	Not detected	Detected	Not detected	99.47
PH081	FP	Not detected	Detected	Not detected	99.47
PH104	FP	Not detected	Detected	Not detected	99.64
PH105	FP	Not detected	Detected	Not detected	99.65
PH107	FP	Not detected	Detected	Not detected	99.66
PH115	FP	Not detected	Detected	Not detected	99.51
PH127	FP	Not detected	Detected	Not detected	99.47
PH135	FP	Not detected	Detected	Not detected	99.40
PH136	FP	Not detected	Detected	Not detected	99.40
Q146.2.5C	Healthy child	Not detected	Not detected	Not detected	99.39
SM101	FP	Not detected	Detected	Not detected	99.85

Table 6.7: Highly similar strains (ANI >99.3%) identified in lineage V

Another isolate Q154.2.6C (two-year NEOM study) shared ANI of >99.91% with NEOM isolate M087.1C (only 16 SNPs between these two isolates; Figure 6.13; Table 6.8). This genomic similarity indicates a larger or shared *C. perfringens* reservoir between two hospitals QCCH and SMH, as infant Q154 (isolate Q154.2.6C) was resident in QCCH during NEOM, with infant M087 a patient in SMH; yet notably they both harboured a genetically-similar *C. perfringens* isolate M087.1C, with infant

Q154 colonised with this strain (Q154.2.6C) two years after the NEOM study. However, as QCCH and SMH had frequent patient transfer, it is not surprising that certain strains of bacteria including *C. perfringens* may be circulating in the hospital settings or among healthcare staff.

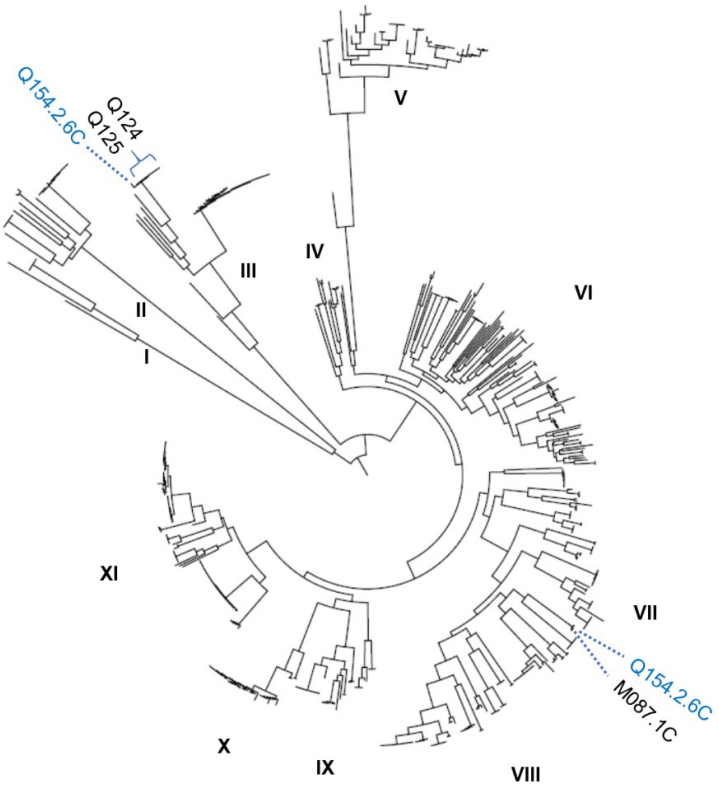


Figure 6.13: Phylogenetic positions of hospital-associated *C. perfringens* strains

Blue-labelled strains indicate isolates from infants 2 years after previous hospital admission (SMH and QCCH; 2-year NEOM) that are genetically closely related.

Importantly, isolate Q041.2.4C (type D) carried the highly virulent pore-forming toxin coding gene *etx*, while Q061.2.1C (type E) encoded ITX gene (Figure 6.8). As 76% (32/42) of isolates in lineage III consisted of animal isolates, this suggests these potentially lethal strains may be animal-derived. This was supported by previous findings which showed that toxinotypes D and E were associated with enterotoxaemia in animals including lambs (type D), calves and rabbits (type E), nevertheless these toxinotypes have never been linked with human diseases [109]. Furthermore, there has never been any report on healthy children carrying type D and E isolates, indicating a potential zoonotic/environmental transfer of veterinary-origin *C. perfringens* strains that necessitates a better understanding. WGS-based studies of *C. perfringens* genetic

diversity have been very limited to date, a larger study will be required to probe the origins of these potentially dangerous strains.

Strain	AMR gene			Typing toxin gene						Toxinotype
	tetA(P)	*tetB(P)*	*ermQ*	*plc*	*etx*	*iap*	*ibp*	*cpe*	*pfo*	
M009.2.1C	+	-	-	+	-	-	-	-	+	A
M040.2.1C	+	-	-	+	-	-	-	-	-	A
M125.2.4C	+	+	-	+	-	-	-	-	+	A
M128.2.3C	+	-	-	+	-	-	-	-	+	A
M135.2.3C	+	+	-	+	-	-	-	-	+	A
M137.2.1C	+	+	-	+	-	-	-	-	-	A
Q012.2.1C	+	+	-	+	-	-	-	-	-	A
Q034.2.1C	+	+	-	+	-	-	-	-	+	A
Q036.2.1C	-	-	-	+	-	-	-	-	+	A
Q041.2.4C	-	-	-	+	+	-	-	-	+	D
Q048.2.1C	+	-	-	+	-	-	-	-	-	A
Q049.2.3C	+	+	-	+	-	-	-	-	+	A
Q061.2.1C	+	-	-	+	-	+	+	+	+	E
Q063.2.1C	+	+	-	+	-	-	-	-	+	A
Q135.2.1C	+	-	-	+	-	-	-	-	+	A
Q137.2.2C	+	+	-	+	-	-	-	-	+	A
Q137.2.3C	+	+	-	+	-	-	-	-	+	A
Q138.2.5C	+	-	-	+	-	-	-	-	-	A
Q138.2.7C	+	-	-	+	-	-	-	-	-	A
Q146.2.5C	+	-	-	+	-	-	-	-	-	A
Q154.2.6C	+	-	-	+	-	-	-	-	-	A
Q161.2.3C	+	+	-	+	-	-	-	-	+	A
Q182.2.2C	-	-	-	+	-	-	-	-	+	A
Q210.2.1C	+	+	+	+	-	-	-	-	-	A
Q213.2.1C	+	-	-	+	-	-	-	-	+	A

Table 6.8: AMR and toxinotyping profiles of 25 two-year NeoM isolates with PFO screening

6.5.8 Zoo-associated *C. perfringens* strains: an exploratory study

As *C. perfringens* has been linked with several veterinary intestinal infections in animals including NE in avian species, the presence of virulent *C. perfringens* strains in animals' gut could pose a potential health risk [384]. Knowledge on *C. perfringens* diversity and their infectious role in wild animals is still poorly understood although significant progress has been made in recent years [385]. *C. perfringens* in zoo animals is still not well understood although a pathogenic role has been reported, including a typical *C. perfringens*-associated gastroenteritis, represented by intestinal necrosis and abdominal distention, was reported in young black-footed ferrets back in 1990 [386]. A total of 17 isolates were obtained from zoo animals to initially explore the genomics of these strains (Table 6.9). This section describes the virulence profiles of 17 isolates obtained from zoo animals kept in Africa Alive, Banham Zoo in Suffolk (UK) and Welsh Mountain Zoo in Conwy County (Wales, UK).

All 17 isolates were type A, and mostly encoded accessory toxin PFO (82%; 14/17). Importantly, 76% of the isolates (13/17) encoded AMR efflux pump *tetA(P)*, indicating the potential wide-spread tetracycline resistance in zoo animals. Importantly, a lion isolate LHZCP161 carried 4 acquired AMR genes, including gentamicin-resistant *ANT(6)-Ib,* tetracycline resistant *tet44* and *tetA(P),* and macrolide-lincosamide-streptogramin resistant gene *ermQ*, revealing a snapshot of potential multidrug-resistance-gene reservoirs in animals.

Notably, multiple strains (n=2) of *C. perfringens* were detected in 6/10 (60%) individuals, namely seriema, aardvark, European brown bear 1, Europe brown bear 2, lion and snow leopard, as determined by phylogenetic lineages. As only a maximum of 2 isolates were taken from these 7 samples, this indicates a larger diversity of strains could potentially be present within an individual animal.

Strain	Host	AMR gene					Toxin gene		Toxinotype	Lineage
		tet44	*ANT(6)-Ib*	*tetA(P)*	*tetB(P)*	*ermQ*	*plc*	*pfoA*		
LHZ459	Seriema	-	-	-	-	+	+	+	A	II
LHZ460	Seriema	-	-	-	-	-	+	+	A	VI
LHZ472	Aardvark	-	-	+	+	-	+	-	A	VI
LHZ474	Aardvark	-	-	+	-	-	+	+	A	III
LHZ476	Yellow mongoose	-	-	+	+	-	+	+	A	VI
LHZ513	Goose	-	-	+	+	-	+	+	A	VI
LHZ574	Colobus monkey	-	-	-	-	-	+	+	A	I
LHZCP154	European brown bear	-	-	+	-	-	+	+	A	IV
LHZCP155	European brown bear	-	-	+	-	-	+	+	A	VII
LHZCP157	European brown bear	-	-	+	-	-	+	+	A	III
LHZCP158	European brown bear	-	-	+	-	-	+	-	A	VII
LHZCP161	Lion	+	+	+	-	+	+	-	A	VI
LHZCP162	Lion	-	-	-	-	-	+	+	A	VI
LHZCP163	Snow leopard	-	-	+	-	-	+	+	A	VI
LHZCP164	Snow leopard	-	-	+	-	-	+	+	A	II
LHZCP170	Colobus monkey	-	-	+	+	-	+	+	A	VI
LHZCP171	Colobus monkey	-	-	+	+	-	+	+	A	VI

Table 6.9: AMR and toxin screening of zoo-associated *C. perfringens* strains

6.6 Discussion

Significant pangenome diversity was detected within 552 *C. perfringens*, representing the lowest bacterial core genome reported to date at 7.4%. This extreme trait is an uncommon observation in other bacteria, with previous studies demonstrating larger core genomes including *C. difficile* (30.3%; 40 genomes), *S. enterica* (16%; 206 genomes) and *E. coli* (13%; 307 genomes) [29,387,388]. In agreement with my previous smaller scale study, these data indicate that *C. perfringens* has a highly variable pangenome at 92.6% accessory genome[17], with an average of 3.6 new genes expected to be added to this open pangenome of 552 genomes for each new *C. perfringens* genome included. More isolates from different geographical regions around the globe should be included, which will further enhance our understanding of this potentially

unprecedented variability of the *C. perfringens* pangenome. This observation also implicates the potential adaptability of *C. perfringens* in difference ecological niches, i.e. acquiring new genes, which warrants further in-depth investigation.

Given the high variability of *C. perfringens* pangenome, there is a potential to assign certain lineages, especially I, II, III and V, into *C. perfringens* subspecies, based on both genomic clustering and ANI comparison. JFP isolates were obtained from foals and dogs, known to carry NetF toxin, linking to specific foal enteritis pathology, and were demonstrated to be phylogenetically distinct from common *C. perfringens* isolates using MLST, and confirmed by this WGS study [40]. Notably, food-poisoning related historical isolates in lineage V were evidently a deep-branching cluster, phenotypically they are known as ultra-heat-resistant types compared to common *C. perfringens* strains, possibly due to small acid proteins [60]. Potentially, these lineages of isolates might differ in pathogenicity, as seen in both JFP and food-poisoning isolates, with experimental validation required for the remaining isolates. Phylogenomically, this substantial distinction of these lineages is demonstrated for the first time in this study.

This large-scale, genomic-based study outlines some vital observations in the phylogenetical distributions of *C. perfringens* in a global and one health context. Strains of *C. perfringens* have long been considered as a zoonotic enteric pathogen, affecting primarily neonates of humans and animals, resulting in potentially fatal enterocolitis including, NE in chickens, haemorrhagic enterocolitis in calves and NEC in human preterm infants [1,389]. Specific toxinotype F strains, a main causative pathotype (encoding CPE) for self-limiting human FP diarrhoea, have also been implicated in diarrhoeal disease in dogs and neonatal pigs, presented with resembling symptoms of enteric mucosal necrosis to profuse diarrhoea [156,390,391]. Type F *C. perfringens* was suggested to be prevalent within the food chain, as CPE was detected in 1.4% of 887 retailed foods in America [392].

In this study, the substantial mixing and co-clustering of human and animal isolates (Figure 6.5) in several major lineages supports the potential zoonotic origin and spread of *C. perfringens* strains, confirmed by highly similar ANI values (>99.0%) in several sub-lineages, indicating the high similarities between animal- and human-derived strains. However, some clusters were dominated by either human (lineages VII-XI) or

animal-predominant (lineages I-II) strains, suggesting a fundamental difference in genomic components that allow colonisation of these strains in different host species, which may require further experimental analysis [382]. This also points to the potential niche or host adaption, impacting genome evolution and host range. This adaption may be due to various mechanisms including acquisition and loss of genes, genome rearrangement and inactivation of genes, and HGT (plasmids and phages), which has been observed in *Salmonella*, resulting in narrow-host-range serovars, i.e. certain serovars can only colonise limited range of hosts [393]. This genetic aspect of *C. perfringens* is currently not well understood and will require in-depth genomic analysis in the future, nevertheless, these current data demonstrated that isolates in lineage V have significant genome reduction compared to the remaining isolates, suggesting potential genetic and evolutionary adaptation. In brief, this robust phylogenetic analysis indicated no definite distinction between animal- and human-derived *C. perfringens* strains as has been observed in other bacterial species including *C. difficile.* This is further supported by the absence of overlapping accessory genes between these two sub-sets of isolates although potential niche adaptation appears to be occurring in limited lineages [394].

Data analysis indicates that more toxin genes were encoded by animal-associated isolates, with these toxin genes plasmid-borne rather than encoded chromosomally including *etx*, *lam*, *cpb*, *netB* and *netF.* As these toxins are mainly associated with clinical diseases in animals, e.g. ETX in sheep and goat enterotoxaemia (affects brain primarily), this suggests these plasmid carrying strains may be readily picked up by grazing animals from environmental sources i.e. soils [80,157]. However, these data are somewhat biased, as animal isolates in the public domain were mainly obtained from diseased individuals rather than healthy, although as reported in this study, a healthy child's isolate (toxinotype D) also encoded for the deadly toxin gene *etx*. To determine whether these plasmid-borne toxin genes are human, environment or animal-associated will require a large-scale genomic study.

Surprisingly, a limited number of AMR genes (maximum 4 acquired AMR genes in a single genome) were observed consistently, with the majority of isolates only encoding tetracycline resistance genes. Compared to *C. difficile*, a current serious issue in healthcare-associated diarrhoea, this species encodes significantly more AMR genes (up to 11 AMR genes in a single genome) including tetracycline, clindamycin

and aminoglycosides-resistant genes [29]. *C. perfringens* has also been implicated in AAD of hospitalised patients, which can potentially be attributed to its endospore forming nature that may circumvent the effect of antibiotics, however, no such isolates been sequenced to date [24,325]. This also indicates current antibiotics used in clinical settings are still effective against *C. perfringens* infections excluding tetracycline.

In the analysis, CPB2-var1 was associated with humans, while CPB2-var2 was linked to animal isolates. Although these two variants of toxin CPB2 (likely to be carried on plasmids) have similar (74%) nucleotide identity, their correlation with pathogenicity remains unclear [395]. In fact, the host-specific association of this toxin gene might not be as clear-cut as initially thought. CPB2-var1 (or, consensus CPB2), was first identified in *C. perfringens* str13 (human gas gangrene strain) [37]; this toxin variant can also be detected in both porcine (as high as 93%) and equine-related strains including D21/98 (accession: AJ537530) [395]. CPB2-var2 (or, atypical CPB2) was detected in both environmental strain FORC003 (source: aquarium water; accession: CP009558.1) and human FP isolate F5603 (accession: AB236337.1) [377]. However, this variant was also detected in chicken isolates EHE-NE18 (accession: CP025503.1), LLY_N11 (accession: CP023413.1) and Del1 (accession: CP019578.1). Using this (limited) available data, it appears that the host-specific association is not definitive, as both variants were detected in both human and animal hosts. Hence, the presence of the toxin variants should rely upon the spread of its encoding plasmids [41,125]. Moreover, there is more work to be done to understand the pathogenicity of this toxin in relation to disease development, even though current thinking in the field suggests this toxin may not be clinically important [151].

The widespread nature of *C. perfringens* conjugative plasmids was observed; with plasmids present in all genetic clusters, and in both human and animal isolates. *C. perfringens* toxin genes are known to be primarily plasmid-encoded, including the six major typing toxins (further in-depth details as described in Chapter 1) [41,42]. Particularly, the food-poisoning-related toxin CPE is plasmid-encoded, with *in vitro* transfer between *C. perfringens* strains, converting type A strain to type F strain via the CPE-encoding conjugative plasmid being previously demonstrated [51,292]. Importantly, it is postulated that conjugative transfer of plasmids is also likely to occur *in vivo* and the actual dissemination of conjugative plasmids might have been wider than previously thought [43,292]. This phenomenon may explain the high level of genetic

diversity observed in the pangenome analysis, and could play a crucial role in host adaptation, as these plasmids commonly encode colonisation-important collagen binding protein or related sialidase as discussed in Chapter 3 and Chapter 4 [115,305,396].

Notably, previously considered accessory (non-essential) toxins including PFO, collagen adhesins (*cna* and *cnaC*) and CPB2-var2 were all statistically enriched ($P<0.05$) in diseased-associated isolates, confirming the important roles these toxins may play as virulence determinants. CPB2-var2, also known as CPB2h1, was demonstrated experimentally to be >10 times more potent than CPB2-var1 (CPB2h2) in Caco-2 cells, supporting its statistical association with enteric diseases [89]. Colonisation factors including collagen adhesin genes *cna* and *cnaC* have previously been implicated in chicken and turkey NE, however their widespread presence in numerous *C. perfringens* isolates as determined in this study highlights the need for future in-depth experimental research [230].

Importantly, the 'type' human sporadic diarrhoeal isolate F4969 (encodes CPE; isolated in 1990s), isolated from a patient residing in Europe (specific region unknown), was shown to be genetically closely related (mean pair-wise SNP comparison: 64.4 ± 22.0 SNPs) to Scottish and London isolates (mainly FP isolates, 18/24; isolated between 2011-2017). These data indicate a significant public health concern, as F4969-like strains appear to be persistent FP-related *C. perfringens* strains, that have been in circulation for ~ 20 years, which is likely disseminated across North East England, or could be a regional persistent genotype (Europe) attributed to regional clonal expansion [337]. These data also demonstrated the wider reservoir, including colonisation within healthy individuals.

This analysis also indicated a more global spread of FP strains, as a South Korean isolate FORC025 (accession: SAMN04209542), isolated from a FP patient in 2015 (stool sample), was demonstrated to be genetically closely related (pair-wise SNP: 74.2 ± 38.98 SNPs; range: 10-191) to English and Welsh isolates (n=29; lineages Xc and Xd). This South Korean FP isolate, was highly similar to 4 English FP isolates at 10 SNPs difference, and one Irish CH isolate (EM297-T0; 13 SNPs) in the same sub-lineage Xd, indicating intercontinental transfer, the first time this has been implicated. Nevertheless, as sequencing data of *C. perfringens* is currently limited on the Asian

continent, it would be interesting to obtain further isolates and data to compare with European-derived FP isolates, to determine the extent of this dissemination.

Following these strain dynamics is particularly important in more vulnerable or disease-prone communities, such as elderly people residing in CHs, where FP-related genotypes of *C. perfringens* appear to be prevalent [334]. As *C. perfringens* is a serious issue in this setting, further precautions should be taken including handling the catering in these premises [316]. Furthermore, routine surveillance on human- and environment-derived *C. perfringens* could be implemented as part of the disease-control policy to reduce the incidence of CH-related FP cases.

Importantly, the robustness of SNP analysis in epidemiology was demonstrated throughout this study, including the capacity to link healthy-individual isolates with disease-associated isolates in specific regions. Currently WGS is not employed in *C. perfringens* surveillance on a national level (including UK and Canada), therefore the bigger picture of *C. perfringens* extra-regional/global circulation is unknown presently. Hence, it is proposed that public health agencies could implement WGS on outbreak-associated *C. perfringens* isolates to further understand this pathogen, which may contribute to the disease monitoring and control (a more thorough discussion in Chapter 4 on FP case isolates and WGS).

For many years, traditional typing methods including MLST have been used to define 'chromosomal-CPE', which indicated that these *C. perfringens* strains were genetically distinct from strains of other sources, including plasmid-CPE strains [382]. These strains did not carry PFO and were highly heat-resistant, which was attributed to presence of small acid proteins [60]. This WGS analysis confirmed these previous findings as chromosomal-CPE strains formed an exclusively distinct cluster, lineage Vc, comprising historical FP isolates and contemporary FP isolates from PHE. These isolates did not encode PFO, sialidases NanI, NanJ and NagH. Interestingly, these strains do not appear to carry any conjugative plasmids (except Darmbrand isolate NCTC8081) and had a smaller genome (approximately 3.0 Mb, compared to 3.2-3.8 Mb), indicating less evolutionary adaption via MGEs. Furthermore, lack of NanI in these strains may correlate with the characteristic self-limiting symptoms of FP diarrhoea, implying rapid pathogen clearance due to absence of key colonisation factors [116].

The long-term persistence and inter-hospital spread of *C. perfringens* demonstrated in two-year NEOM isolates raises significant concerns for at-risk preterm infants residing in healthcare units. As *C. perfringens* is strongly associated as a causative agent of preterm NEC, extra strategies should be implemented to eradicate this pathogen from neonatal wards [8]. Another healthcare-associated pathogen *C. difficile*, also a proficient spore former, causes AAD and in some patients pseudomembranous colitis, which has risen dramatically over the last decade, mainly due to the spread and persistence of hyper-virulent strains [30,397].

The detection of deadly toxin producers in a healthy child suggests the important role of host defence and gut microbiota in colonisation resistance [398,399]. This indicates that whether *C. perfringens* is pathogenic or a mere commensal in the gut hinges on the impact of the host immune system, and the role of other microbiota members as discussed in Chapter 3, including the findings that supplementation of certain probiotics significantly reduces incidence of preterm NEC. This reflects the disease susceptibilities in particular groups including early-life and later-life individuals. Furthermore, this may suggest the future infection research on *C. perfringens*-associated gut diseases should be directed towards understanding the impact of gut microbiota members on *C. perfringens* rather than merely the disease-causing toxins.

Due to the large numbers of preterm hospital-acquired preterm-associated isolates (33.5%; n=185) and FP-related isolates (19.7%; n=109), relative to other diseases, host, and geographical location of strains, this may have influenced interpretation of the global *C. perfringens* phylogeny. Thus, for an even further in-depth study of *C. perfringens* genomic structure and relationships, community-acquired human and animal isolates from a broad regional spread, would be required.

6.7 Future research directions

This study analysed an unprecedented WGS dataset of *C. perfringens* isolates (n=552), obtained from a wide range of hosts from a number of geographical regions. Nevertheless, to obtain an even wider picture of *C. perfringens* population structure will requires the inclusion of strains from under represented diseased populations including nonfood-borne diarrhoea, enterocolitis-associated isolates from various

animal populations, and isolates of different toxinotypes, to better understand the evolutionary relationships and circulation of virulence plasmids.

As *C. perfringens*-associated FP is a worldwide issue, the addition of FP-associated strains from other continents including Asia and Africa, especially equatorial countries, where temperatures may favour spread and growth of *C. perfringens*, would be interesting to include. Analysis of these strains would advance the understanding of the strains causing FP and help develop/design effective preventative measure for these issues [327,400,401].

This analysis has revealed our apparent lack of understanding with respect to animal-associated *C. perfringens* genotypes, although previous studies have frequently toxinotyped these strains. To better understand the potential zoonotic transfer of *C. perfringens* will require routine surveillance of animal-associated *C. perfringens* strains from farms and zoos, including pets, particularly from households with infected humans. This is critical as previous studies have indicated that for other pathogens, e.g. *C. difficile*, animals can potentially serve as reservoirs and carriers of hyper-virulent genotypes, which is important in the One Health context [394,402]. Of note, only a handful of sequences available to date are of animal-associated strains, albeit chicken isolates are well represented.

Although toxinotyping is used to rapidly identify toxinotypes, which do in some cases link well with disease representations (particularly veterinary field) [1,20]; WGS approaches in this study have proved a far more discriminatory method to comprehensively understand clinical strains including toxin carriage, specific virulence genes including AMR, assignment of sub-types via phylogenetic analysis, and tracking potential global/regional disseminations at SNP level. With the technological advancements of real-time WGS, it is possible to whole-genome sequence an isolate in under 2 h and obtain relevant information [314]. Hence, it is proposed that WGS should progressively replace the conventional labour-intensive and time-consuming PCR approach (toxinotyping only) and be routinely implemented for national surveillance of *C. perfringens* outbreaks to monitor these isolates at the strain level. This important advancement will unveil more detailed clinically relevant information including virulence plasmids, accessory toxin genes and AMR, all of which are deemed critical to clinicians and veterinary medicine.

In conclusion, this work has analysed a total of 552 genomes (443 newly sequenced) of *C. perfringens*, representing the largest population genomic study of *C. perfringens* to date. Pertinent findings include, significant pangenomic variations (7.4% core genome in the pangenome), zoonotic potentials, virulence determinants associated with sub-clusters of isolates, substantial conjugative plasmid distributions, intra-regional disseminations and long-term persistence, which have provided novel genomic and biological insights into the largest collection of *C. perfringens* genomes. Hypotheses generated from this analysis will allow in-depth genomic and experimental analyses to be carried out, thus advancing the understanding of this important enteric pathogen.

Chapter 7 Phenotypic characterisation of preterm-associated *C. perfringens* strains reveals NEC-associated hypervirulent phenotypes

7.1 Introduction

The key aims of the study presented in this chapter were to characterise the virulence traits of clinically important *C. perfringens* strains including gas production, hydrogen sulphide generation, cell toxicity and antibiotic resistance, to identify strain-specific virulence traits/genotypes associated with preterm-NEC. An additional aim was to identify hyper-virulent strains of *C. perfringens* to take forward for the development of a *C. perfringens* infection model detailed in Chapter 8 (Figure 7.1).

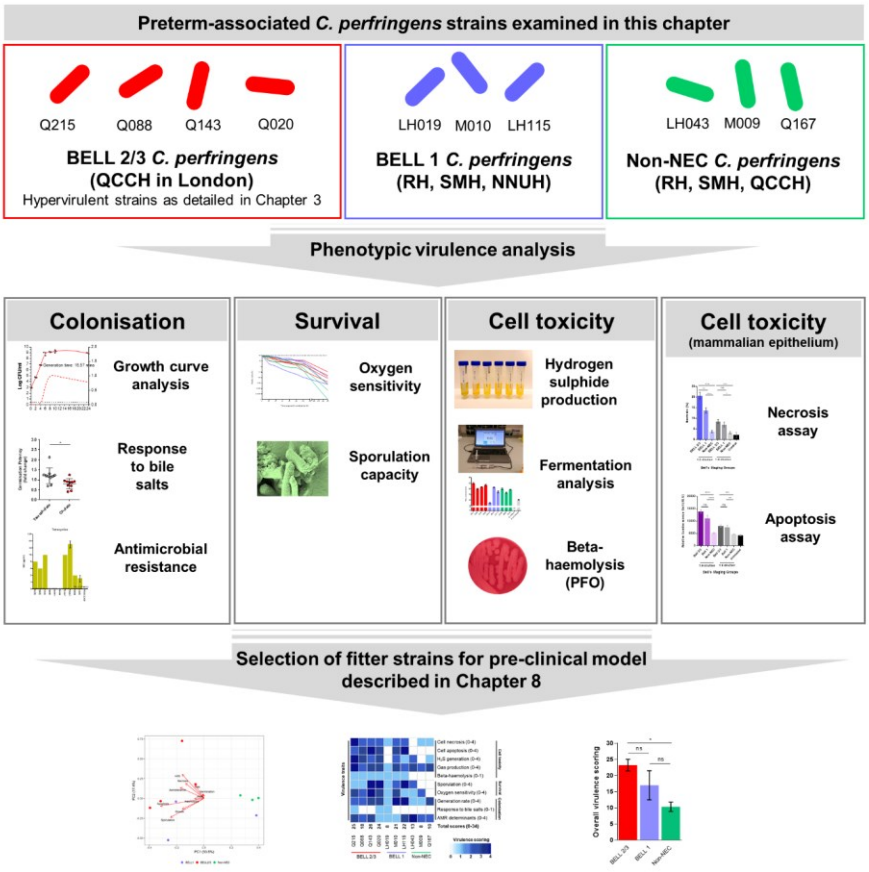


Figure 7.1: Schematic of examined isolates, experimental analysis and aims of Chapter 7

7.2 Background

C. perfringens has been linked with numerous intestinal diseases in both humans and animals, including as an emerging candidate linked to preterm-NEC [1,8,9]. This enteric pathogen is known for its resilience to stressful environmental conditions, and prolific production of an arsenal of >20 toxins [403].

Although *C. perfringens* is classified as an anaerobe, this pathogen possesses a complex system to handle oxidative stress including genes that encode superoxide dismutase and superoxide reductase, however the underlying mechanisms are still not well understood [404]. These aerotolerant properties, and the ability to form endospores links to *C. perfringens* ability to survive in aerobic environments for considerable periods of time, which is vital for survival and transmission, particularly in hospital environments [15,405,406].

C. perfringens is also capable of reducing sulphite to generate gas such as hydrogen sulphide, which is often used to confirm presence of *C. perfringens* using TSC differential agar, on which *C. perfringens* forms pitch black colonies <20 h incubation due to precipitations of ferrous sulphide in this agar [11,71]. Importantly, excessive hydrogen sulphide is proposed as a potential mediator of intestinal inflammation and is associated with the inflammatory bowel disease Ulcerative Colitis [72,407]. Importantly, excessive gas production by *C. perfringens* may be linked to specific enteric disease symptoms, and in the case of NEC may be associated with the extensive intestinal inflammation and formation of gas cysts that are characteristic of this disease [12,73].

C. perfringens has the potential to secrete a combination of >20 toxins, although no single strain is currently known to produce all toxins, and this potent virulence feature has long been considered as the primary pathogenesis trait of this bacterium [1,121]. Several studies have examined molecular toxicity of *C. perfringens* secreted molecules against epithelial cells including bovine epithelial cells and human cell line Caco-2, with these studies indicating that CPE can directly damage these cell types [85,114]. However, to date there are no studies specifically probing how different clinically relevant strains interact with gut epithelial cells, including preterm-NEC-associated isolates, thus our current understanding on this important *C. perfringens* pathogenesis trait is currently limited.

Comparative genomics in Chapter 3 indicates that NEC-associated isolates may be defined as 'hypervirulent' strains, as they encode significantly more toxin genes (particularly PFO encoding gene *pfo*), when compared with control isolates (potentially commensal *C. perfringens*). Nevertheless, specific phenotypes of strain-specific characteristics associated with NEC-linked isolates has yet to be determined. This chapter will explore specific phenotypic virulence traits associated with NEC-linked isolates including; oxygen sensitivity, gas generation, hydrogen sulphide production, molecular toxicity against mammalian intestinal cell lines, antibiotic resistance, growth kinetics and sporulation-related features, with the aim of linking genotype to virulence phenotype and additionally allowing for selection of 'hyper-virulent' strains for use in a mouse infection model (Chapter 8).

7.3 Hypothesis and aims

It is hypothesised that disease-associated *C. perfringens* isolates exhibit strain-specific virulence properties, and that certain strains are more virulent than other non-pathogenic strains. The aims of this chapter are:

(1) To characterise various *C. perfringens*-associated virulence factors in selected NEC or non-NEC *C. perfringens* isolates via multiple *in vitro* approaches and identify key phenotypic traits;

(2) To identify suitable hyper-virulent strains for use in a pre-clinical model that is described in Chapter 8.

7.4 Isolate selection and information

A total of 10 *C. perfringens* isolates were selected for their associations with preterm NEC: from patients in Bell Staging I (BELL 1), II and III (BELL 2/3), and from control samples (non-NEC). Isolates LH019, LH115 and LH043 were isolated from BAMBI samples and were newly sequenced. The relevant information of the isolates is detailed in Table 7.1.

	Isolate ID	Hospital	NEC staging	Gestation weeks	Sample collected	NEC day
1	Q215*	QCCH	BELL 2/3	28+1	40 days	40
2	Q088*	QCCH	BELL 2/3	27+1	24 days	24
3	Q143*	QCCH	BELL 2/3	31+1	19 days	19
4	Q020*	QCCH	BELL 2/3	27+1	19 days	19
5	LH019	RH	BELL 1	24+5	56 days	32
6	M010	SMH	BELL 1	27+4	16 days	16
7	LH115	NNUH	BELL 1	24+3	49 days	28
8	LH043	RH	Non-NEC	25+5	42 days	/
9	M009	SMH	Non-NEC	28+1	15 days	/
10	Q167	QCCH	Non-NEC	26+0	13 days	/

Table 7.1: Selected *C. perfringens* isolates with clinical metadata

* These isolates were characterised as *C. perfringens*-associated-NEC, hyper-virulent strains as defined in Chapter 3. RH: Rosie Hospital (Cambridge, UK); SMH: St Mary's Hospital (London, UK); QCCH: Queen Charlotte's and Chelsea Hospital (London, UK); NNUH: Norfolk and Norwich University Hospital (Norwich, UK).

7.5 Results

7.5.1 Growth kinetics assay

C. perfringens is reported to double in 12-17 mins at 37°C (log phase), and 8-12 mins at 43°C in optimal medium, making this pathogen the fastest-growing microorganism ever known [70]. Here the growth kinetics of these 10 isolates were investigated using rich media BHI at 37°C in anaerobic conditions, which is the normal mammalian body temperature, to compare growth dynamics and generation times.

Doubling time of all tested isolates ranged between 12-19 mins at 37°C, which is in agreement with previous studies analysing *cpe*-carrying strains [70]. Moreover, between NEC and non-NEC strains there was no statistical variations in mean generation time ($P>0.05$; Table 7.2). Notably, isolate LH115 (Bell 1) had the fastest generation time at 12.47 mins during log phase, while M009 (non-NEC) was the slowest at 19.09 mins (Figure 7.2). All isolates grew to confluency (~10^9 CFU/ml) from approximately 10^3 CFU/ml within 5-7 hours.

	Bell 2/3	Bell 1	Non-NEC
Minimum	14.05	12.47	13.43
Median	15.79	13.45	17.21
Maximum	16.68	18.84	19.09
Mean ± SD	15.58 ± 1.27	14.92 ± 3.43	16.58 ± 2.88

Table 7.2: Statistics of *C. perfringens* generation time based on Bell staging groups

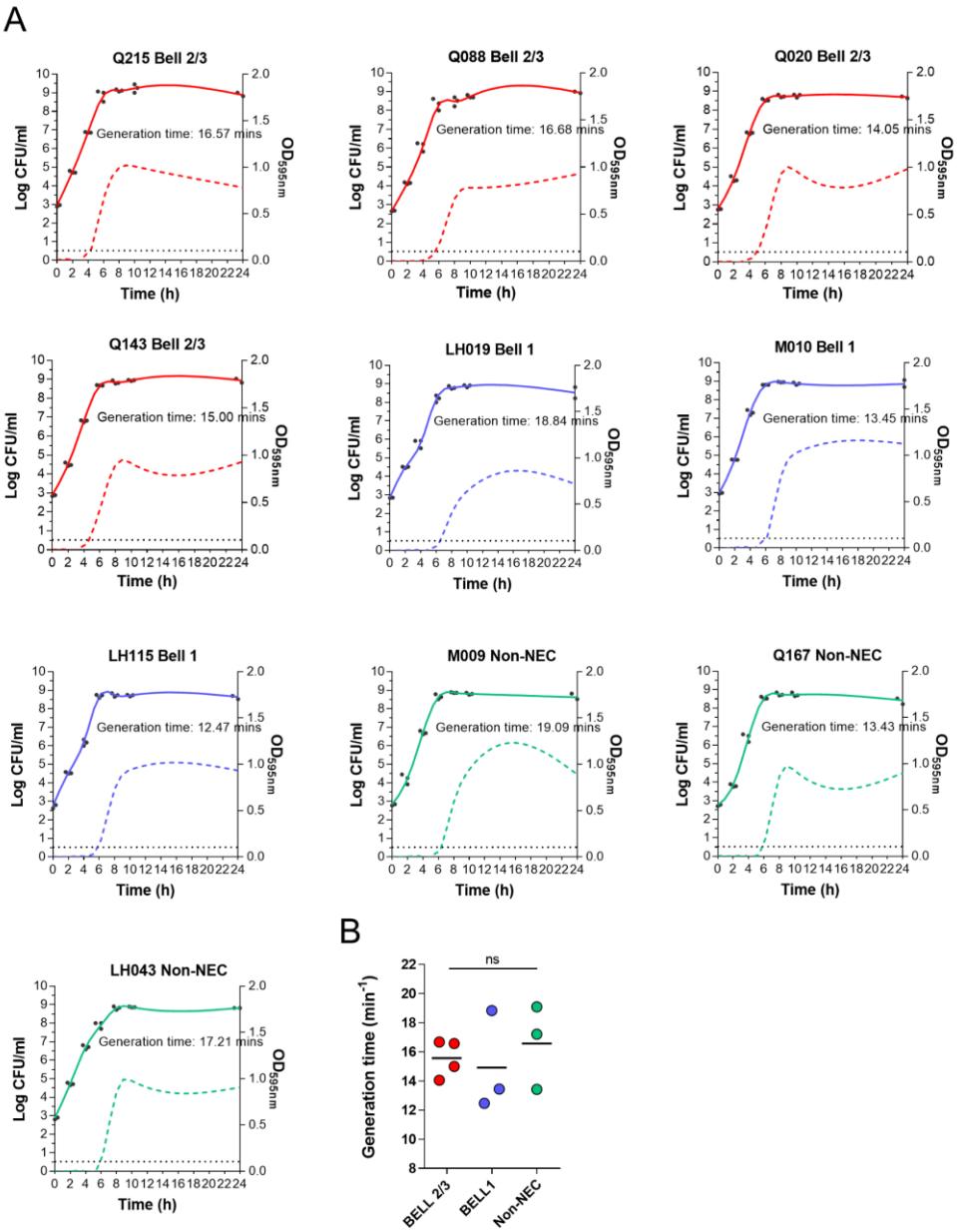


Figure 7.2: Growth kinetics of *C. perfringens* isolates and generation time comparison

(A) Growth kinetics of *C. perfringens* isolates in CFU and optical density over 24h. **(B)** Generation time comparisons between isolates in various groups. Mean generation time is calculated during the log growth phase. Data represent 3 biologically-independent cultures (n=3), one-way ANOVA, Dunn's multiple comparison test. OD: Optical Density.

7.5.2 Gas production

C. perfringens was first associated with human disease during World War I, gas gangrene, which was so-called due to the characteristic gas production and associated pathology [408,409]. Importantly, preterm-NEC has also been described as 'intestinal gas gangrene', reflecting the fulminant formation of gas cysts in the intestine, coupled with tissue necrosis, hallmark NEC representations [190]. Thus, it is hypothesised that certain extra-virulent strains may generate more gases than others, which may be associated with mechanical injuries within the intestinal lamina propria.

Gas production of all 10 selected strains was measured in the first 10 hours, as described in the method section. Strain-specific variations were observed: NEC-associated strains generated a larger gas volume than the non-NEC strains at seeding concentrations of 10^3 CFU/ml (Figure 7.3). Statistically, Bell 2/3 isolates generated significantly more gases than Bell 1 (mean: 22.2 ml vs 16.4 ml; P<0.01) and non-NEC isolates (mean: 22.2 ml vs 18.9 ml; P<0.01; Figure 7.3B). Positive control *E. coli* was observed to produce gas in anaerobic conditions at high CFU/ml (starting CFU at 10^6 CFU/ml). By contrast, *B. bifidum* was not observed to generate gas, even at high CFU seeding concentration (10^3-fold > *C. perfringens* strains at 10^6 CFU/ml).

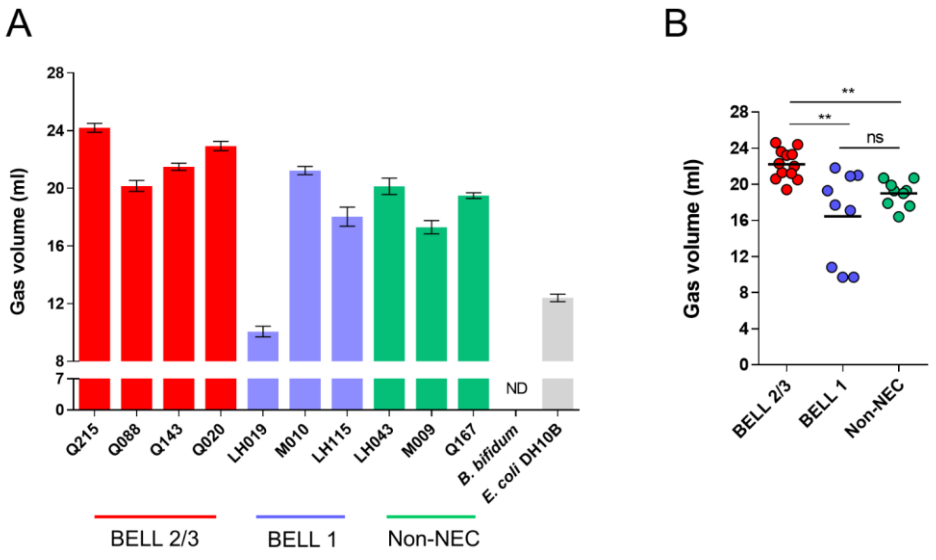


Figure 7.3: Gas production assay on selected *C. perfringens* isolates

(A) Accumulated gas production of *C. perfringens*. Confluent cultures were inoculated at 10^3 CFU/ml (exception: *B. bifidum* and *E. coli* at 10^6 CFU/ml) and sealed in 30ml-sterile air-tight gas vials (50ml) in anaerobic chamber prior to experiment. All cultures and equipment were pre-reduced overnight before experiment. Measurements were taken every hour and the experiments stopped at 10h. ND: Not Detectable. Data: means ± SEM (n=3). **(B)** Comparisons of isolates' gas production in groups. Each dot represents one technical replicate. Data: means ± SEM, one-way ANOVA (Kruskal-Wallis test, α=0.05; Dunn's multiple comparison test). ** P<0.01.

7.5.3 Hydrogen sulphide assay

Microorganisms are capable of producing hydrogen sulphide via utilisation of compounds containing organic sulphur, e.g. thiosulphate, sulphite and sulphate [410]. *C. perfringens*, is known to produce sulphite in TSC media to form pitch-black colonies, resulting from the precipitation of ferrous sulphide[71]. Importantly, hydrogen sulphide is also known to be cytotoxic to mammalian cells [411]. Here, I hypothesised that *C. perfringens* may initiate pathogenesis via production of hydrogen sulphide, which correlates with epithelial cell cytotoxicity (Figure 7.4) [72,407].

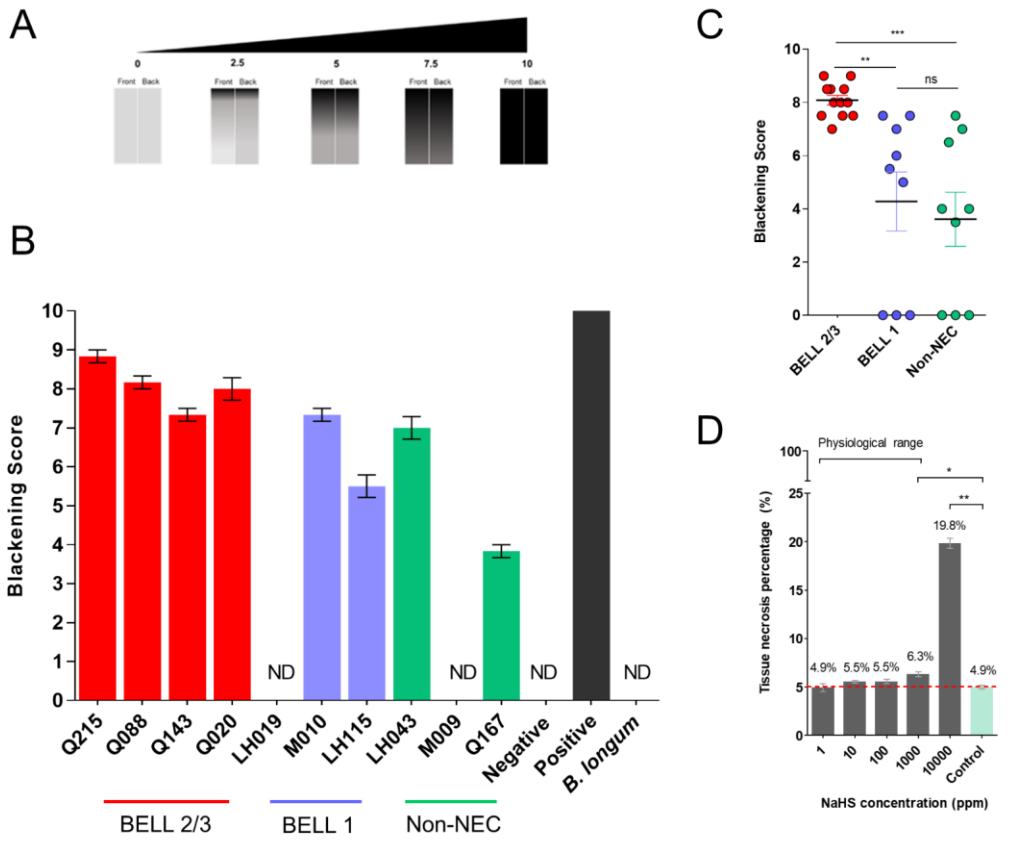


Figure 7.4: Hydrogen sulphide assay on selected *C. perfringens* isolates

(A) Blackening scoring system for H_2S detection. Data representation: mean ± SEM (n=3). *B. longum* was cultured for a continuous 20h (2x), no sign of black stain observed. ND: Not Detectable. This experiment was performed in 3 independent cultures for each isolate. **(B)** H_2S visual measurement test. Blackening scoring - 1: slight black stain on the strip; 5: 50%black strip; 10:100% black strip. The degree of blackening was used as a H_2S concentration/production gauge. **(C)** Comparison for blackening scores on H_2S production. Data representation: means ± SEM. One-way ANOVA Dunn's multiple comparisons test, *n*=3. ** P<0.01, *** P<0.001, ns: non-significant. **(D)** Necrotic effect of NaHS (H_2S donor) on IEC-6 monolayer system at 2h-incubation. Data: each bar represents necrosis percentage of tissues with percentages indicated on top of each bar. Physiological concentration of hydrogen sulphide in mammalian gut ranges from 1-1000ppm. * P<0.05, ** P<0.01 (two-tailed t-test, unpaired Welch's correction, n=3).

In this assay, visual-quantification was employed to determine relative concentration of hydrogen sulphide production (H_2S strips), and NaHS, a known hydrogen sulphide donor, was incubated with IEC-6 monolayers, a rat small intestine cell line [412]. Significant variation was observed over the 10 selected strains (Figure 7.4). Using a visual scoring method (Figure 7.4A) strains LH019 and M009 did not produce H_2S in the first 8 h, and *B. longum* (non-sulphite reducer) was not observed to generate H_2S as expected (Figure 7.4B). Group comparison on blackening scoring showed that Bell 2/3 isolates produced significantly higher concentration (8.0 ± 0.6; mean ± SD) of H_2S than Bell 1 (4.3 ± 3.3; $P<0.01$) and non-NEC isolates (3.6 ± 3.0; $P<0.001$; Figure 7.4C).

Toxicity of H_2S was tested in IEC-6 monolayers (Figure 7.4D), with H_2S concentration at 10^3 ppm causing slight necrosis (6.3% vs mean baseline 4.9%; $P<0.05$). At 10^4 ppm significant necrosis was observed (19.8%, $P<0.01$), which experimentally indicated that H_2S is directly cytotoxic to mammalian epithelial layers at high concentration.

7.5.4 Oxygen tolerance assay

Oxygen-sensitivity was tested over a period of 14 days of ambient-air exposure. This oxygen tolerance trait may be expected to facilitate host-to-host and/or nosocomial transmissions of particularly virulent strains [15]. Results indicated that Q143 (Bell 2/3) had $>10^4$-fold more viability, than the least aerotolerant strain Q167 (non-NEC) (Figure 7.5A), which had no detectable re-growth at day 14. Bell-staging group comparison indicated that isolates associated with Bell 2/3 were significantly more aerotolerant ($P<0.0001$), than non-NEC isolates at day 14 (Figure 7.5B). Strains with only a 100-fold CFU reduction (at 10^6 CFU/ml) after two weeks were Bell 2/3 isolates Q143, Q020, Q215, Q088 (100%) and Bell 1 isolate M010, and were thus identified as oxygen-resilient isolates.

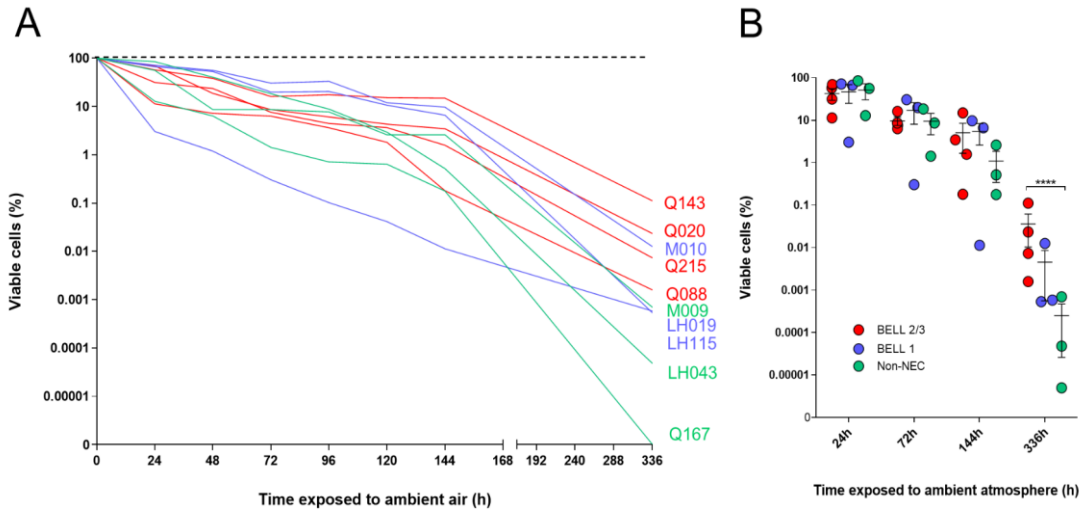


Figure 7.5: Oxygen tolerance assay on selected *C. perfringens* isolates

(A) Viability of *C. perfringens* strains exposed to ambient air over 14 days. Anaerobic NEC-associated *C. perfringens* isolates from various Bell stages and non-NEC individuals are shown in the brackets (n=3). The experiment stopped at 336h (14 days). Data: mean; n=3 (independent cultures). **(B)** Comparisons of *C. perfringens* viability (%) after exposure to ambient air at 4 time-points across experimental period according to clinical groups. Each dot represents SEM and mean percentage of viable cells compared to Day 0 control within each Bell staging group. For all time-points, no statistically significant differences were found in between diseased/non-diseased groups except time-point 336 h. Data: mean (n=3). Central bars represent group mean. One-way ANOVA Dunn's multiple comparisons test. **** P<0.0001.

7.5.5 Sporulation capacity

Sporulation capacity was determined as trait may directly link to *C. perfringens* transmission capacity and ability to survive through harsh environments and passed on to the next host [15]. *C. perfringens* is known to be a spore-forming species, however it has been reported to be relatively difficult to sporulate *in vitro,* unless cultured in Duncan-Strong (DS) medium (developed in the late 1960s) [61].

Initially to determine if *C. perfringens* strains could be sporulated in rich BHI media, two strains were randomly selected: Q215 and LH115. These two strains were incubated anaerobically for 6 days, then treated with 70% ethanol for 4h and plated out on BHI agar (Figure 7.6A). Results show that neither strain sporulated, which is in agreement with previous literature [61].

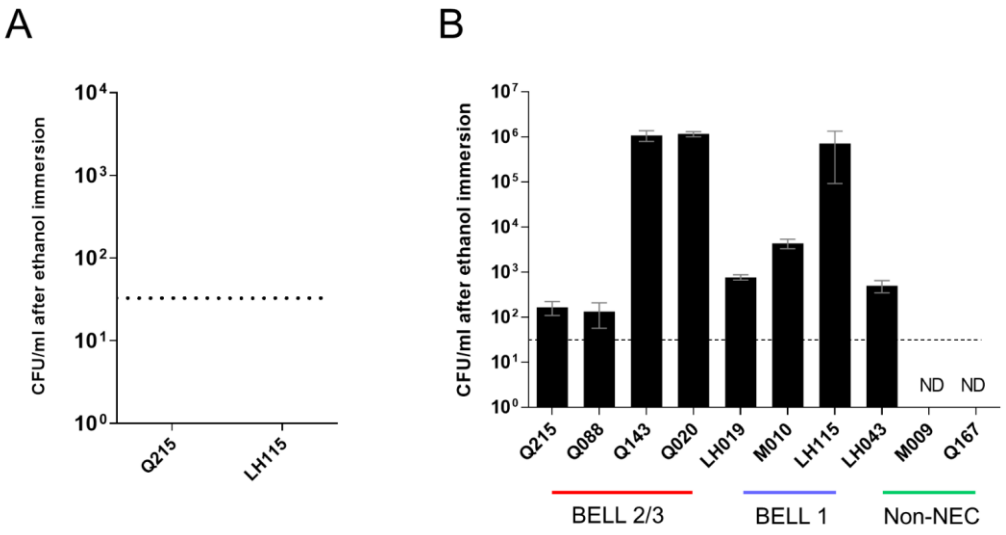


Figure 7.6: Sporulation assay on selected *C. perfringens* isolates

(A) Sporulation capacities of over-culturing of two strains in BHI media for 6 days. Two strains were incubated in rich media BHI for a consecutive 6 days, cultures were immersed in 70% ethanol for 4 h and plated on BHI agar. No sporulated colonies observed. **(B)** Sporulation capacity of *C. perfringens* strains. Pure bacterial cultures (24h Duncan-Strong medium culturing) were immersed in 70% ethanol for 4h to eliminate vegetative cells before being washed with PBS and spotted onto BHI growth medium supplemented with 0.1% sodium taurocholate as a germinant. Strains M009 and Q167 were below LOD (50 CFU/ml) as no single colony was detected (ND: non-detected) on BHI plates even after 48h DS medium culturing. Varied sporulation capacity of each strain was displayed in this assay. Data: mean ± SEM (n=3). Dotted line represents LOD.

DS modified medium was then employed to examine sporulation capacity of 10 selected strains (Figure 7.6B). After 24h-incubation, and 4h-70% ethanol treatment to eliminate vegetative cells, most strains displayed sporulation on 0.1% taurocholate supplemented BHI agar, with the exception of strains M009 and Q167 (both non-NEC) that did not sporulate, even after prolonged 48h-incubation in DS modified medium. These results demonstrated variation in sporulation capacities between strains as colonies $>10^5$ CFU/ml (or, spores/ml) were recovered in strains Q143, Q020 and LH115 (strains Q215 and LH115 did not sporulate in BHI for 6 days), while $<10^4$ CFU/ml, were recovered from the remaining strains, and both M009 and Q167 did not sporulate after 48h.

7.5.6 Bile salt assays

Primary bile salts including taurocholate are known to induce spore germination, which has previously been linked to enhanced pathogenicity in *C. difficile* [68,413]. Conversely, secondary bile salts (e.g. deoxycholate and chenodeoxycholate),

produced after de-conjugation of primary bile salts by certain gut microbiota members, have been reported to inhibit either the germination/growth of *C. perfringens* [413]. Thus, this assay was designed to understand the variation of *C. perfringens* spores' responses to bile salts, including primary bile salts taurocholate and cholate, and secondary inhibitory bile salts deoxycholate and chenodeoxycholate.

Most strains (7/8) exhibited higher mean germination potency (>1.0 fold) on taurocholate-supplemented BHI, when compared to control (BHI not supplemented with taurocholate; Figure 7.7A). Notably, spores of three strains (Q215, Q020 and LH019) had significantly enhanced germination on taurocholate when compared to cholate ($P<0.05$). Overall, taurocholate (1.2 ± 0.4-fold change) significantly improved germination in comparison to germination in the presence of cholate (0.8 ± 0.2-fold change; $P=0.017$; Figure 7.7B). Importantly, secondary bile salts including chenodeoxycholate and deoxycholate inhibited germination (or, growth) of all strains (8/8). A mixed-bile assay was carried out to determine whether more physiological-like conditions and concentrations of bile salts impacted *C. perfringens* spore germination. Strains Q215 and Q088 cultured overnight in BHI before being plated on BHI agar supplemented with 0.1% taurocholate, 0.1% cholate and 0.1% deoxycholate (secondary bile salt) had no growth after 20h incubation. Furthermore, different concentrations of deoxycholate were used to determine MIC, and this assay showed that 0.05% had limited inhibition activity on strain Q215 at 72h, indicating 0.1% is the MIC for *C. perfringens* inhibition (>7 days), which is a physiological concentration [414].

Finally, primary bile salts taurocholate and cholate were investigated for their ability to impact vegetative cell growth of strains Q215 and Q088, however no statistical difference was observed ($P=0.500$; Figure 7.7C).

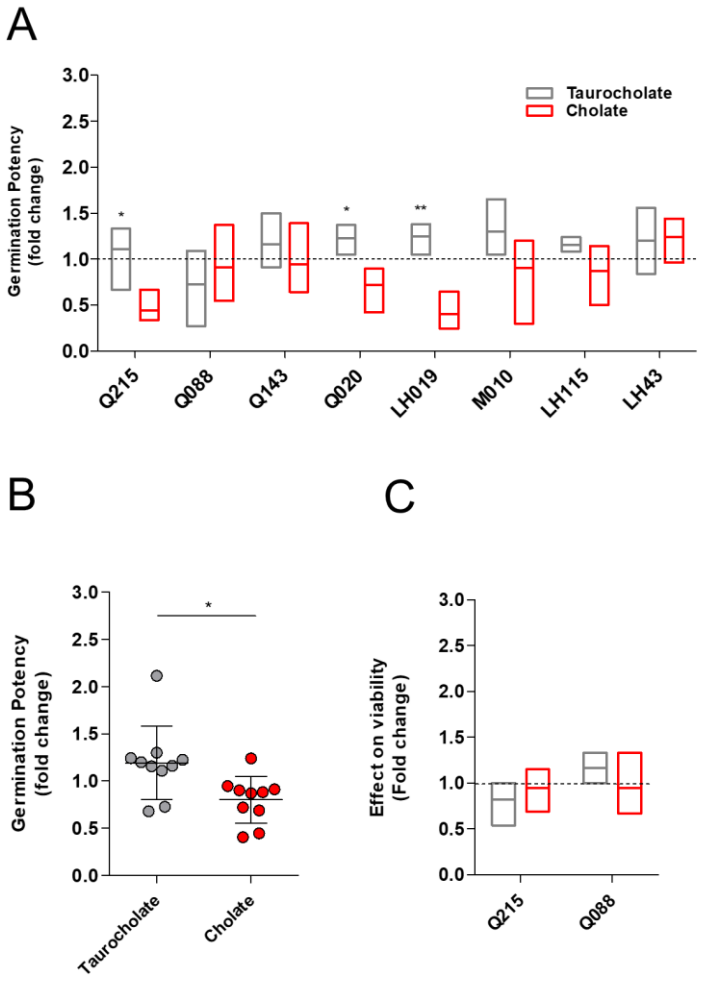


Figure 7.7: Bile salt assays on selected *C. perfringens* isolates

(A) Responses of *C. perfringens* spores to bile-salt germinants supplemented at 0.1%. The number of colony-forming units (CFU; representing germinated spores) present on plates in the presence of a germinant is expressed as a fold-change with respect to the number of CFU recovered on plates in the absence of a germinant. Tested strains were subjected to ethanol treatment (70% ethanol, 4h immersion) prior to being plated (*n*=3 biological replicates for each strain). A fold-change of one (represented by dotted black line) indicates that a germinant had no effect on the CFU recovered. Data: mean and range, Welch's unpaired non-parametric t-test (* P < 0.05, ** P<0.01). **(B)** Comparison of the effect of different bile salt supplementations on germination potency. Data: Mean ± SD, Welch's two-tailed unpaired t- test, * P<0.05. **(C)** Bile salt responses of vegetative cells. Data: Mean with range, no statistical difference observed.

7.5.7 Minimum inhibitory concentration assay

C. perfringens has a range of AMR profiles, with tetracycline resistance particularly widespread [126,415]. Thus, a MIC assay was performed to determine the resistance profiles of *C. perfringens* strains to tetracycline, gentamicin and penicillin, as these are commonly prescribed in neonatal clinical settings (i.e. NICU) [416].

Four NEC-associated strains (out of 7) were demonstrated to display tetracycline MIC ≥ 4µg/ml, while 3 out 7 strains had MIC ≤ 0.125 µg/ml (LOD; no growth was observed; Figure 7.8A). Non-NEC strain LH043 displayed the highest MIC (32 µg/ml).

Resistance towards gentamicin was high (MIC ≥ 64µg/ml) in all *C. perfringens* strains with the exception of strain Q088 (MIC ≥ 32µg/ml; Figure 7.8B). Seven out of 8 strains had MIC ≥ 128µg/ml towards gentamicin, which indicates that this antibiotic is generally ineffective towards inhibiting *C. perfringens*.

Importantly, all tested strains showed no resistance towards penicillin (MIC ≤ 0.125 µg/ml; Figure 7.8C). The lowest observed MIC was in strain LH019 (no detectable growth at LOD 0.015µg/ml).

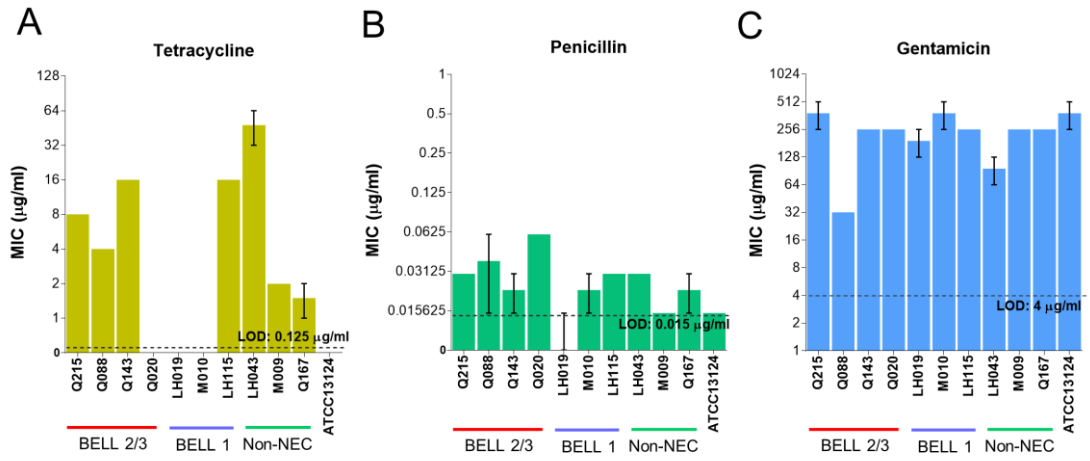


Figure 7.8: Minimum Inhibitory Concentrations of specific antibiotics against *C. perfringens*

(A) Tetracycline; **(B)** Penicillin; and **(C)** Gentamicin. ATCC13124 was used as a control strain. Data: Mean ± SEM (2 separate experiments). LOD: Limit of detection.

7.5.8 Toxicity against mammalian intestinal epithelial cells

7.5.8.1 Necrosis assay

C. perfringens strains are known to secret pore-forming toxins that could necrotise gut epithelial tissues [1,403]. Strain variation in pore-forming capacities has previously been hypothesised, as *C. perfringens* can be pathogenic or a commensal member of the gut microbiota [417]. A necrosis assay was therefore performed using intestinal epithelial monolayers (IEC-6 and Caco-2 cell lines), with co-incubation of sterile-filtered soluble media (10h-incubation with 1% starting inoculum across all 10 strains) to compare and determine cell toxicities.

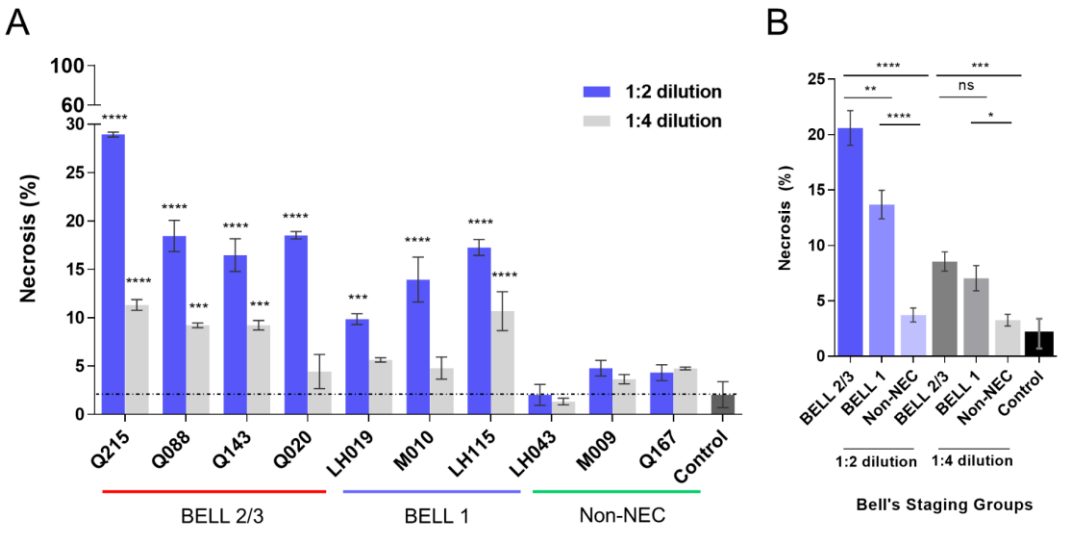


Figure 7.9: Necrosis assay performed on Caco-2 monolayer

(A) Necrosis percentage of Caco-2 monolayer co-cultured with *C. perfringens* sterile-filtered supernatants after 2h. **(B)** Statistics compare each group. Caco-2 P31 was used in this assay. Supernatants of *C. perfringens* (10 h-incubation with 1% starting inoculum) were sterile-filtered before used. Each well was co-incubated with 15% supernatants of either 1:2 or 1:4 dilution in the working media. Promega membrane integrity kit was employed to examine the necrosis percentage. (Data: mean ± SEM; n=3; **** P<0.0001, *** P<0.001, ** P<0.01, * P<0.05. One-way ANOVA with Bonferroni multiple comparison test). Normal cell death percentage is indicated by black dashed line (control).

In the Caco-2 model, a human colonic cell line, 7 strains including all Bell 2/3 isolates (Q215, Q088, Q143 and Q020) and all Bell 1 isolates (LH115, M010s and LH019), were shown to induce significant necrosis (Bell 2/3: 20.59 ± 5.40%; Bell 1: 13.68 ± 3.88%; negative control: 2.00 ± 2.32%; mean ± SD), compared to control at 1:2 dilution (P<0.0001; Figure 7.9A-B). Notably, Q215 (Bell 2/3 NEC case isolate) induced 28.93 ± 0.42% necrosis (mean ± SD, n=3 technical replicates) in under 2 h, which was the most potent necrotising isolate. By contrast, none of the non-NEC strains (LH043, M009 and Q167) induced significant necrosis on Caco-2 monolayers (3.72 ± 1.87% in 1:2 dilution; P>0.05).

In IEC-6 (rat small intestinal cell line) monolayers, comparable results were observed (1:4 dilution). Bell 2/3 isolates Q215, Q088, Q143 and Q020 (group mean ± SD: 39.71 ± 20.9%), Bell 1 group strains M010 and LH115 (64.81 ± 25.75%; except LH019) demonstrated significantly higher necrosis (P<0.01; control necrosis %: 2.18 ± 1.03%) (Figure 7.10A-B). Again, none of the non-NEC strains (mean necrosis ± SD: 13.61 ± 11.73%) induced significant necrosis (P>0.05). Overall, significantly higher necrosis

was observed in Bell 2/3 (P<0.01) and Bell 1 isolates (P<0.001), when compared to non-NEC isolates (Figure 7.10B).

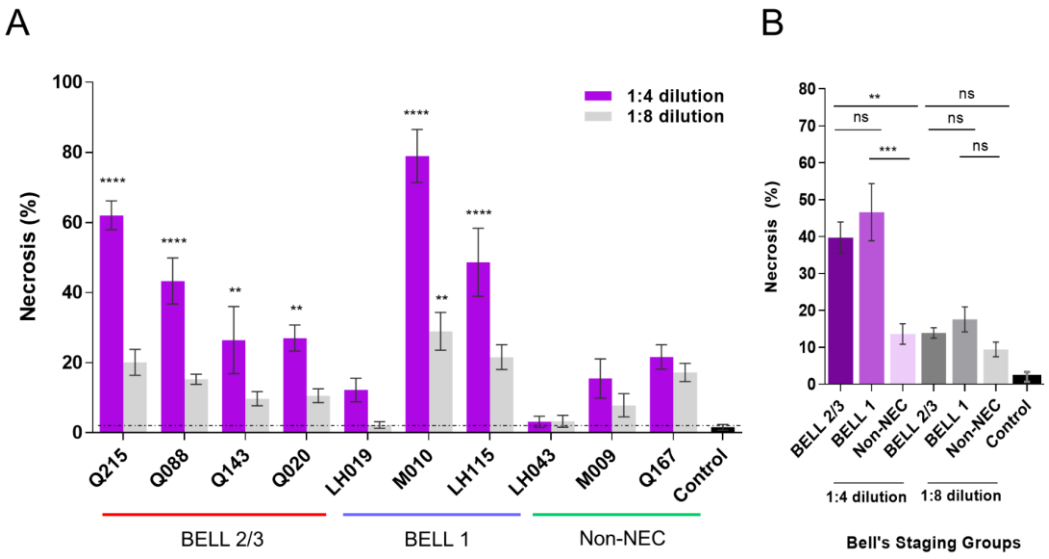


Figure 7.10: Necrosis assay performed on IEC-6 monolayer

(A) Necrosis percentage of IEC-6 monolayer co-cultured with *C. perfringens* sterile-filtered supernatants (2h incubation). **(B)** Statistics compare each group. Each well was co-incubated with either 1:4 or 1:8 dilution in the working media. Promega membrane integrity kit was employed to examine the necrosis percentage. (Data: mean ± SEM; n=6; **** P<0.0001, *** P<0.001, ** P<0.01. One-way ANOVA with Bonferroni multiple comparison test). Normal cell death percentage is indicated by black dashed line (control). Data were based on two independent experiments.

7.5.8.2 Apoptosis assay

A further cell death assay was performed to understand *C. perfringens* capacity to induce programmed necrosis i.e. apoptosis. Similar to what was described in the necrosis assay, supernatants of Bell 2/3 strains Q215, Q088, Q143 and Q020 induced significant levels of apoptosis-associated caspase enzymes (caspases 3 and 7) after 3h co-incubation with Caco-2 cells (P<0.01; 24376 ± 8661 RLU vs control background: 5230 ± 128 RLU) (Figure 7.11A). Bell 1 isolates M010 and LH115 (not LH019) also induced significant apoptosis in this model (P<0.0001; mean RLU: 31327 ± 8562 RLU). Overall, supernatants from NEC-associated isolates inducex significantly greater caspase activities (P<0.01) in Caco-2 monolayers as compared to non-NEC isolates (5593 ± 661 RLU), where no significant levels of apoptosis were detected (P>0.05; Figure 7.11B).

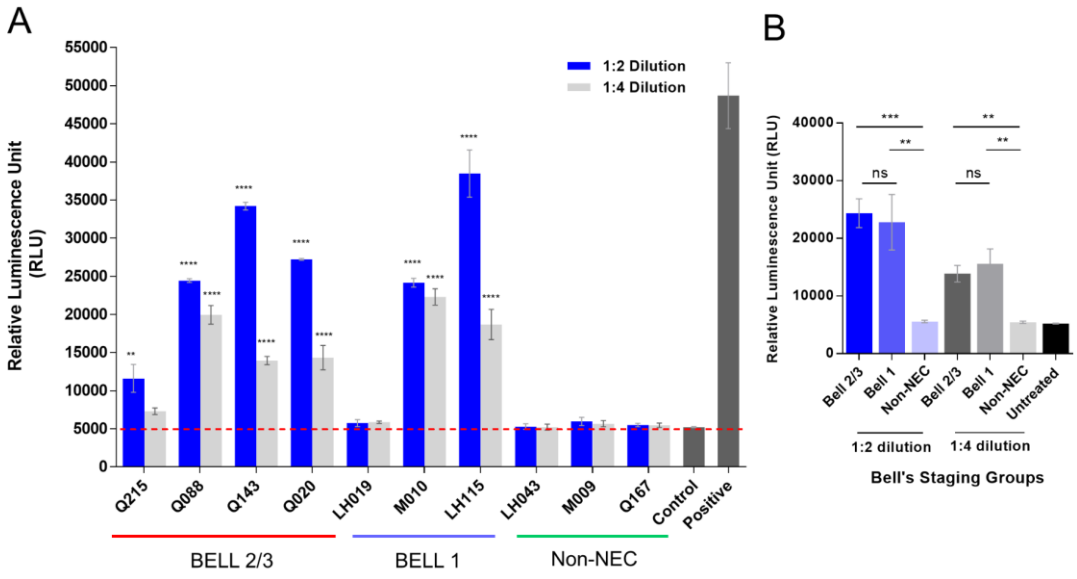


Figure 7.11: Apoptosis assay performed on Caco-2 monolayer

(A) Caspase activity measured via RLU (marker of apoptosis) of Caco-2 monolayer co-cultured with *C. perfringens* sterile-filtered supernatants (3h incubation). **(B)** Statistics compare each group. Bacterial supernatants diluted in BHI have negligible luminescence background. All data represent 3 biological triplicates (n=3) in a single experiment (data are shown as mean ± SEM). Staurosporine was used as the agent for positive control. Data: One-way ANOVA test was performed with Bonferroni's multiple comparison test to compare each group with control group. Significance: * P<0.05, ** P<0.01, *** P<0.001, **** P<0.0001.

Supernatants of *C. perfringens* strains showed similar patterns in caspase activity in IEC-6 monolayers, with all Bell 2/3 isolates (13900 ± 3582 RLU) and two Bell 1 isolates (M010 and LH115; 13951 ± 4411 RLU) inducing significant levels of apoptosis (P<0.0001; Figure 7.12). Again, spent media of all three non-NEC isolates did not induce any apoptotic capacity in this rat cell line model (P>0.05; 4736 ± 529 RLU).

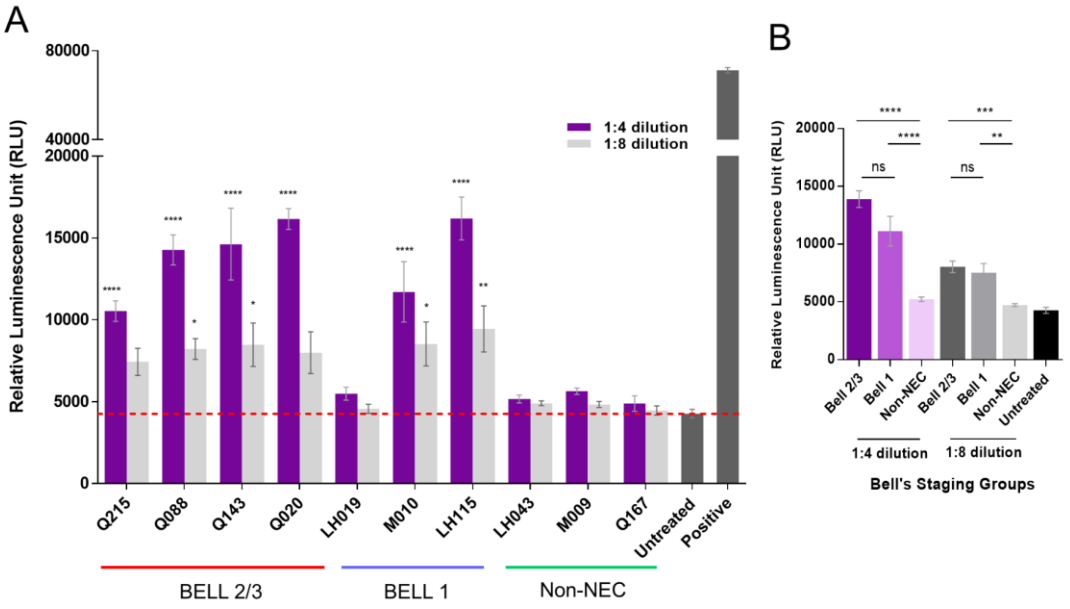


Figure 7.12: Apoptosis assay performed on IEC-6 monolayer

(A) Caspase activity measured via RLU (marker of apoptosis) of IEC-6 monolayer co-cultured with *C. perfringens* sterile-filtered supernatants (3h incubation). **(B)** Statistics compare each group. Bacterial supernatants diluted in BHI have negligible luminescence background. All data represent 6 biological triplicates (n=6) in two independent experiments (data are shown as mean ± SEM). Caspase-Glo 3/7 Assay (Promega) was used for measurement of caspase activity of the cells and performed according to the manufacturers' instructions. Staurosporine was used as the agent for positive control. Data: One-way ANOVA test was performed with Bonferroni's multiple comparison test to compare each group with control group. * P<0.05, ** P<0.01, *** P<0.001, **** P<0.0001.

7.5.9 PFO expression assay

As PFO appears to be a NEC-associated toxin (Chapter 3), which may correlate with the enhanced cell death observed in the cell line studies described above, presence of PFO was phenotypically confirmed to link to genotyped *pfo*-carrying strains. In brief, 5% sheep-blood-BHI-agar was used to demonstrate PFO production, as PFO is known to be haemolytic, and displays beta-haemolysis (complete haemolysis) on blood agar [87].

All *pfo*-carrying strains (8/8) exhibited beta-haemolysis on blood agar plates, confirming expression of *pfo* (Figure 7.13). In contrast, all non-NEC strains, not *pfo*-encoding, did not exhibit haemolytic effects (3/3; 100%).

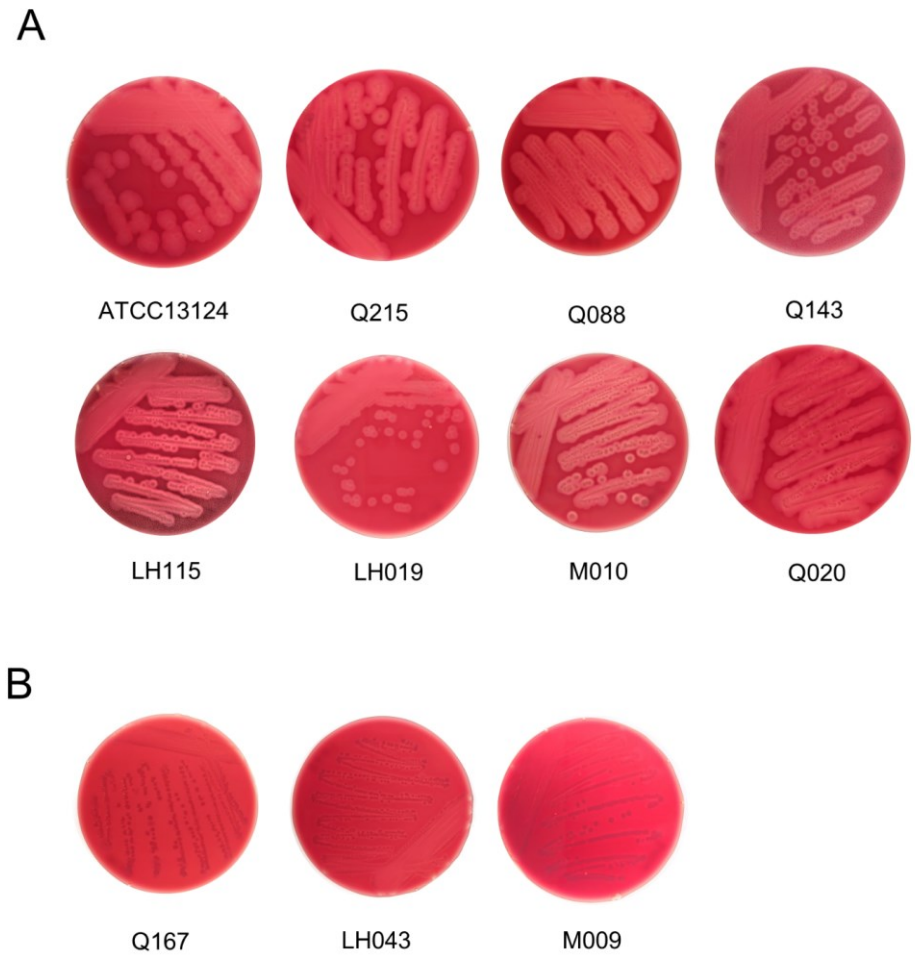


Figure 7.13: PFO expression assay on selected *C. perfringens* isolates

Expression of PFO secretion exhibiting beta-haemolysis (complete haemolysis) on sheep blood agar by **(A)** strains harbour PFO gene, and **(B)** strains do not harbour PFO gene. Strains were streaked on BHI-based sheep blood agar (5%) and incubated anaerobically for 20 h before visual examinations. Lysis of red blood cells can be observed in all (100%) PFO-carrying strains.

7.5.10 Virulence and AMR correlations: genotype vs phenotype

Necrotic tissue is the hallmark symptom of *C. perfringens* infections, including poultry NE and NEC, which is linked to secreted toxins [73]. Notably, necrotising capacities of these 10 strains in Caco-2 monolayers (Figure 7.14A), correlated with presence or absence of *pfo*. This data indicates that PFO may be the main necrosis-inducer, linking to cytotoxicity ability in both Caco-2 and IEC-6 models, and confirmed by secretion of PFO on blood agar (Figure 7.13).

Interestingly, *pfo*-negative, Beta-2-toxin-carrying strains M009 and LH43 did not show potent necrotic capacity. Notably, although LH043 possesses 8 toxin genes (excluding *pfo*), this strain showed the least cytotoxicity, highlighting that PFO is a

potent cytotoxin to colonic cells, which may link to NEC phenotypes (as described in Chapter 3).

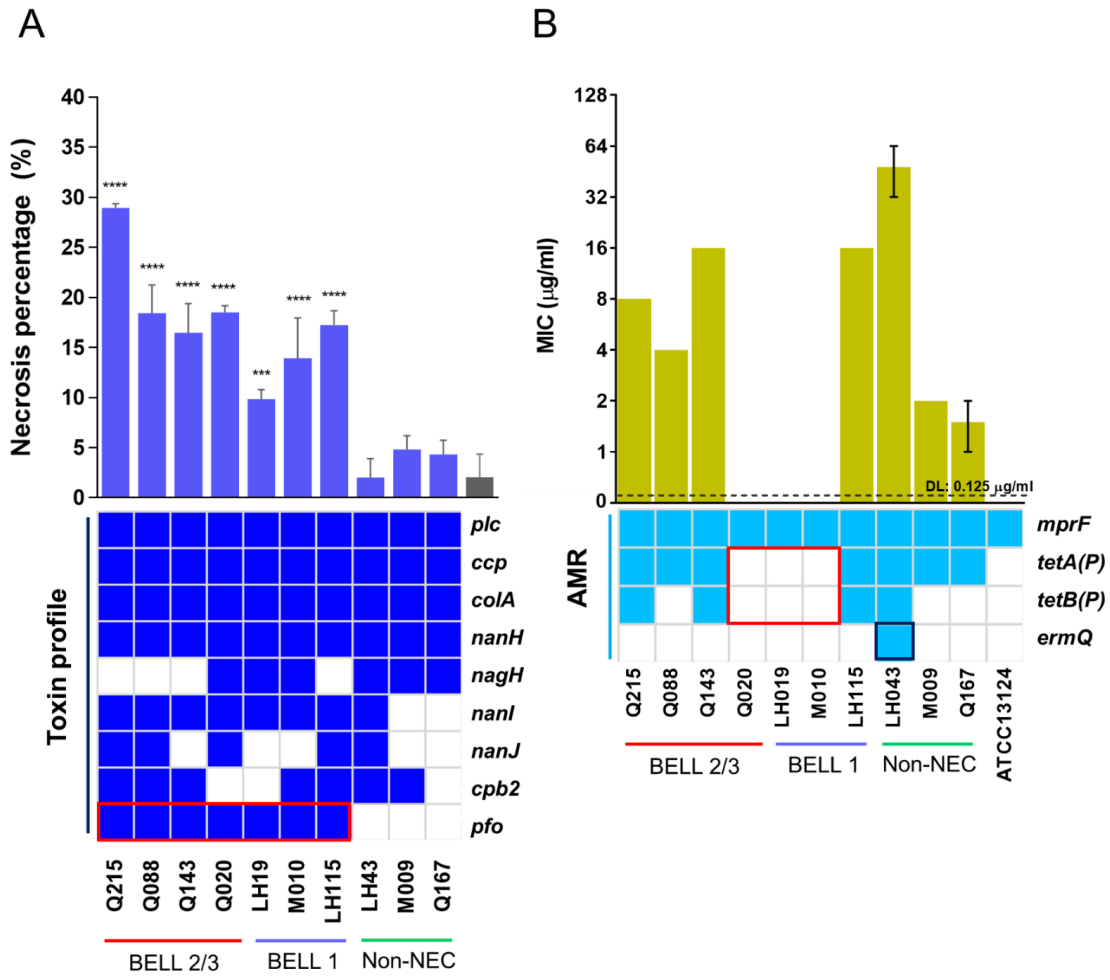


Figure 7.14: Phenotypes and genotypes virulence comparison

(A) Necrotising phenotypes expressed in Caco-2 cells, and **(B)** AMR phenotypes on tetracyclines, vs genotyping profiles of examined strains. Heatmaps were generated using R. AMR profiles were based on CARD database.

To associate exotoxin genes with necrosis traits, Fisher's exact test was performed to correlate enrichment of genes between significant-necrosis group (Q215, Q088, Q143, Q020, M010, LH115) and non-significant-necrosis group (LH019 considered as non-significant taking IEC-6 model into account, LH043, M009, Q167). PFO was shown to be enriched in significant-necrosis group (P=0.033; two-tailed, OR: 30), while other toxin genes (*cpb2, nanJ, nanI* and *nagH*) were not significantly associated with either group (P>0.05; 0.05<OR<14).

Importantly, tetracycline resistance was directly correlated with presence of acquired AMR genes *tetA(P)* and *tetB(P)* as isolates (Q020, LH019 and M010), that do not encode these two genes were highly susceptible to this antibiotic (MIC < 0.125 µg/ml; Figure 7.14). Multidrug-resistant clinical strain LH043 (also encoded erythromycin-resistant gene *ermQ*) displayed the highest resistance against tetracycline at MIC 32 µg/ml. Importantly, tetracycline ribosomal-protection gene *tetB(P)* seemingly could enhance isolates resistance against tetracycline, as all 4 isolates that harboured this gene had MIC ≥ 8 µg/ml (range 8-32 µg/ml), while three isolates that only encoded *tetA(P)* demonstrated a relatively lower MIC ≤ 4 µg/ml (range: 1-4 µg/ml).

7.5.11 Selection of strains for pre-clinical models

This section describes the selection of suitable hyper-virulent strains for use in establishing a pre-clinical model that is described in Chapter 8.

PCA analysis, based on phenotypic traits as demonstrated in this chapter, (n=10) revealed that non-NEC strains exhibited a distinctly less virulent phenotypic profile, when compared to NEC-associated strains (Figure 7.15A). Bell 2/3 isolates Q215, Q088, Q143 and Q020 formed a cluster, (Figure 7.15A) with LH019 clustering more closely with non-NEC strains.

A virulence scoring system was used to gauge the virulence of these strains based on subjective ranking of pathogenicity based on the data from the comparative analyses (primarily 0-4 system, 4 showed strongest characteristic; otherwise 0-1 in binary characteristic; Figure 7.15B). Strains Q215 and Q143 (both from Bell 2/3) were ranked as the top-two strains, specifically strain Q215 displayed the strongest necrotising capacity (4/4). Hence, these two strains (Q215 and Q143) were selected as the hyper-virulent strains to be used in pre-clinical models. In addition, Bell 2/3 isolates exhibited higher virulence scoring ($P<0.05$) compared to non-NEC isolates (Figure 7.15C-D). This suggested Bell 2/3 isolates as hyper-virulent strains in contrast to non-NEC strains, supported by hyper-virulent genotypes as described in Chapter 3.

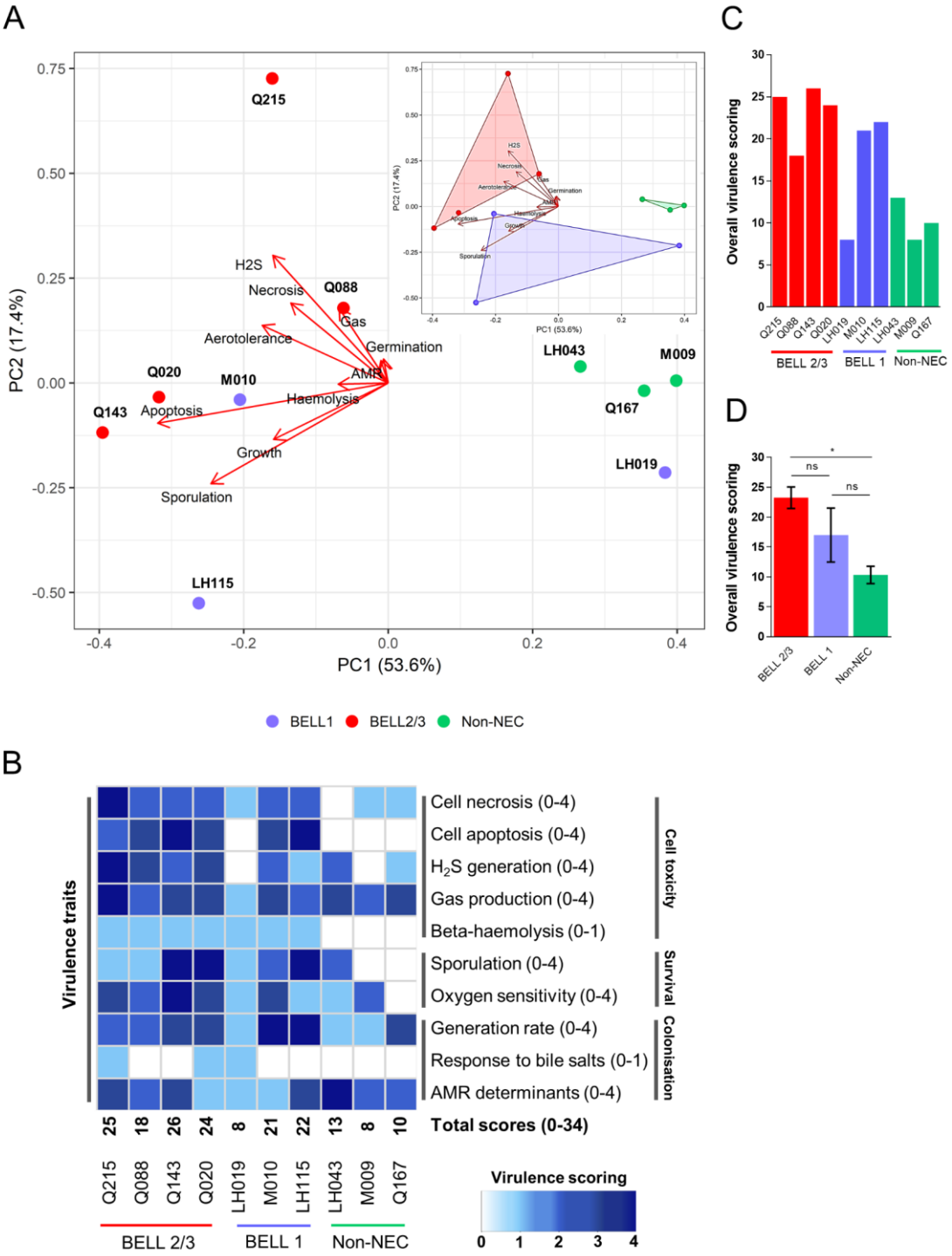


Figure 7.15: Overall phenotypic virulence traits of all selected *C. perfringens* isolates

(A) PCA based on phenotypic traits. Embedded figure showing clustering of isolates (top right corner of the figure). **(B)** Virulence scoring of each strain. Data: total virulence scoring (0-34). **(C)** Comparisons of groups. Data: mean ± SEM. One-way ANOVA, Bonferroni's multiple comparisons test. * P<0.05, ns: non-significant. **(D)** Individual virulence scoring breakdowns. Overall virulence traits (n=10) examined using scoring system 0-4/0-1; where 0 is least and 4 (or, 1) shows strongest, or binary characteristics. Heatmap generated in R.

7.6 Discussion

Most *C. perfringens*-related studies have concentrated on characterising the vast array of toxins this pathogen can secrete [21,121,418,419]. However, in the wider context of pathogenesis, *C. perfringens* is expected to encode and display a range of features that may enhance its virulence capabilities, particularly within hospital settings. Thus, in this chapter I sought to perform in-depth phenotyping on 10 selected *C. perfringens* strains based on 10 key virulence traits (classified by cell toxicity, survival and colonisation), with data indicating significant virulence variations between strains, including presence of non-pathogenic strains and hypervirulent strains that may link to clinical symptoms [420].

C. perfringens is the fastest growing microbes characterised to date [70], with growth kinetic studies indicating that the strains profiled in this study had an average 15.68 ± 2.33 mins generation time. Although individual isolate differed in doubling rates, there were no statistical differences between NEC vs non-NEC strains. However, this rapid-growing trait could potentially contribute to its pathogenic role in clinical contexts, and correlates with the rapid symptom initiation observed in NEC (often <8h) and food-poisoning diarrhoea (within 8-12 h) allowing *C. perfringens* to proliferate and release toxic factors [73,319,323,421].

The gas-producing capacity of *C. perfringens* (using batch fermentation) has not been reported in the literature to date. Notably, overall NEC strains had enhanced gas production relative to non-NEC strains which may correlate with distinctive NEC symptoms, namely abdominal distension and pneumatosis intestinalis- the formation of gas cysts on the gut wall, both of which have been associated with *C. perfringens* overgrowth. This mechanical damage, brought about by gas build-up, may directly damage the intestinal barrier, leading to tissue necrosis and the severe pathology observed in preterm infants diagnosed with Bell 2/3 NEC.

The type of gas produced by *C. perfringens* will also have a direct impact on tissue necrosis, via direct cytotoxicity to epithelial cell lining the GI tract. These studies indicated that NEC-associated strains produced more hydrogen sulphide than non-NEC isolates, which is clinically important as previous studies have linked H_2S with Ulcerative Colitis, and direct necrosis of the gut epithelium, and also enhancement of the detrimental inflammatory processes observed in these patients [72,422-424]. These

pathogenesis factors are also characteristic of preterm-NEC, highlighting that hydrogen sulphide production by *C. perfringens* may be linked to key NEC pathology, which is the first time this has been reported.

As an anaerobe, *C. perfringens* does not grow in aerobic conditions; however, previous studies and analysis of NEC and non-NEC strains as described in this chapter indicates this pathogen can survive in aerobic environments >2 weeks, suggesting a strong detox-catalyst system, which appears to be strain-specific [56]. Importantly, this oxygen insensitivity trait poses a significant threat in clinical contexts and links directly to bacterial transmission. As *C. perfringens* can survive on ambient surfaces, spread of this pathogen could potentially be in the form of vegetative cells, thus increasing infection risk in vulnerable patients i.e. preterm infants and elderly patients [8,334]. More classically appreciated is *C. perfringens* transmission being linked to its ability to form environmentally resistant spores, and interestingly NEC-associated Bell 2/3 strains appeared to more readily sporulate, when compared to non-NEC strains highlighting another effective transmission mechanism for potentially more virulent strains and initiation of disease once resident in the gut.

From an intervention perspective, this would indicate that specific hygiene practices are required to prevent ready transmission of *C. perfringens* between wards and patients, as in the case of *C. difficile* endospores that can be detected on inanimate surface, washed bedsheets and even room air in the hospital wards [425-428]. Moreover, as *C. perfringens* spores survived extended time-periods in 70% ethanol, which is currently extensively used as disinfectant, other measures e.g. hydrogen peroxide decontamination on ward surfaces, or a more advanced disinfection technology cold-air plasma sterilisation (also, nonthermal plasma) should be implemented to introduce more effective decontamination approach in order to prevent dissemination of this pathogen in healthcare settings [425,426,429,430].

Bile salts have been extensively studied in relation to *C. difficile* infection, nevertheless relatively fewer studies have addressed the impact of bile salts in *C. perfringens* germination [68,431]. Primary bile salt taurocholate is a known germinant in spore formers including *C. difficile*, and this work revealed that *C. perfringens* can germinate in the presence of this bile acid. Importantly, accumulation of bile salts (including taurocholate) in preterm-infant intestines (primarily ileum, the common site

of NEC), possibly due to limited enterohepatic circulation capacity, has been implicated in the development of NEC, as well as in a rodent model [432-434]. This hypothesis was further supported by a recent study showing that bile salt concentrations (i.e. taurocholate, detected as early as the first three days) are higher in preterm infants who later developed NEC [435,436]. This could further explain why preterm infants are susceptible to *C. perfringens*-associated NEC, as *C. perfringens* spores can be potentially inoculated from the environment, germinate in the presence of high concentration of bile salt taurocholate and thrive in the gut. To probe this aspect further will require a mechanistic study in *C. perfringens* mouse model and bile salt administration.

Importantly, secondary bile salts (deoxycholate and chenodeoxycholate) have been linked to the inhibition of *C. difficile,* and *C. perfringens* in this study [431,437]. Secondary bile salts are produced by certain gut microbiota members through metabolism of primary bile salts (taurocholate), which is linked to a key colonisation resistance mechanism of the gut microbiota [438,439]. Importantly, deoxycholate was shown to exert inhibitory effects on *C. perfringens* at physiological concentrations i.e. 0.1% (>7 days), and at 0.05% for up to 72h *in vitro*, even in mixed bile conditions, i.e. taurocholate and deoxycholate, the growth of *C. perfringens* was inhibited due to the bacteriostatic effect of deoxycholate. Interestingly, these data may link to the effectiveness of current probiotic strategies in NICUs, where preterm infants are given probiotic bacteria including *Bifidobacterium*, a bacterial species know to produce secondary bile salt via bile salt hydrolase. Crucially, administration of *B. bifidium* has been shown to significantly reduce NEC incidence (ongoing Hall lab BAMBI study), which is potentially associated with the inhibitory effect of secondary bile salts against *C. perfringens* and other pathogens [301,439,440].

A global emergency is the rise of AMR and emergence of 'super bugs' that are refractory to currently available antibiotics. *C. perfringens* strains are widely reported to possess tetracycline resistance elements *tetA(P)* and *tetB(P)*, and analysis of the isolates tested in this chapter indicated widespread tetracycline-resistance, which correlated with presence of one or two *tet* elements [17,126,415,441]; while susceptible strains (M010, LH019, Q020) did not encode any *tet* genes. Clinical isolate LH043 had the highest tetracycline MIC at 32μg/ml; with the lowest MIC of *tet*-carrying strains at 2 μg/ml (Q167), further highlighting tetracycline should not be used to treat

C. perfringens infection [17]. Moreover, *C. perfringens* strains appear intrinsically resistant to the broad-spectrum antibiotic gentamicin, which is still widely used in NICUs as a prophylactic antibiotic. Conversely, penicillin appears to be an effective option to treat *C. perfringens* infection [442], with no beta-lactamase genes encoded in any of the *C. perfringens* genomes in this study, which may explain the effectiveness of penicillin observed.

Toxicity of *C. perfringens* sterile-filtered secretory supernatants were tested against both rodent small intestinal (IEC-6) and human colonic models (Caco-2) and demonstrated similar trends: Bell 2/3 isolates (n=4) were significantly more necrotising than Bell 1 ($P<0.01$) and non-NEC isolates ($P<0.001$). This virulence trait is expected to link directly to *C. perfringens* association with NE and NEC, both diseases characterised by necrosis of tissues [1]. Importantly, PLC (a cell membrane damaging enzyme), known to be secreted by all *C. perfringens* strains [76], does not link to necrotising capacity, as non-NEC strain supernatants do not induce epithelial cell death. This suggested necrosis of cells is related to certain toxins, and previous genome analysis (Chapter 3) indicated that most non-NEC strains do not encode any *pfo* genes (gene enriched in NEC-associated isolates) that transcribes pore-forming haemolytic toxin PFO. This toxin could be critical to disease initiation of both NEC and NE, as previous studies have indicated that PFO may act synergistically with PLC to damage (bovine) epithelial cells, and therefore presence of this gene may distinguish between pathogenic and commensal *C. perfringens* [25,87]. In this study, only the *pfo* gene was significantly enriched ($P<0.05$) in those isolates that could damage intestinal epithelial monolayers, further highlighting a PFO cytotoxicity link. However, to definitively prove this virulence gene, molecular Koch's postulates would have to be fulfilled by carrying out detailed *in vivo* and mutant experiments. CPB2 has previously been associated with NEC, however, non-NEC strains M009 and LH043 that encoded CPB2 toxin gene did not significantly induce necrosis in epithelial monolayers, and were not associated with necrosis-phenotypes ($P>0.05$) [8].

Overall, these data support the hypothesis that NEC-associated strains display a hyper-virulent phenotype, which correlates with the intestinal pathology characteristic of preterm-NEC patients, further supporting the data presented in Chapter 3.

7.7 Future research directions

Full profiling of the toxins secreted by *C. perfringens* strains could be conducted via a large-scale protein profiling study of case and control isolates in the preterm cohort. Initial work had been carried out, seeking to identify the full profile of secreted toxins in spent media using mass spectrometry approaches, however, data resolution was quite low as too many proteins were detected, making some toxins undetectable due to noise. Toxin probes might need to be developed to be able to quantify toxin levels and correlate back to genomic sequence data to gain in-depth understanding of strain-specific toxin secreting capacity.

To better mimic the biological context of *C. perfringens* infection, *in vitro* models could be optimised, e.g. using micro-aerobic *in vitro* human intestinal infection models (such as IVOCs), to enable *C. perfringens* to grow in anaerobic or microaerophilic conditions, while most intestinal cell lines cannot be cultured anaerobically. In this way, investigation of colonisation and infection could be observed, and a larger-scale phenotypic study could be conducted involving multiple *C. perfringens* strains, even looking into the infection mechanisms.

Furthermore, to further develop preventative measure against *C. perfringens* infection, *in vitro* continuous culture human colonic models could be utilised to understand the real impact of the wider microbiota on *C. perfringens* overgrowth, and test bile salt/probiotic therapies based on the secondary bile salt inhibition hypothesis.

In addition, it will be important to understand the oxygen systems in *C. perfringens* where NGS approaches can be applied i.e. RNAseq. As *C. perfringens* strains are cultured in aerobic condition, RNA can be extracted at different time points to investigate the expressed genes under aerobic conditions in order to pinpoint the genes involved in detoxifying oxygen.

Crucially, to fully explore pathogenesis of *C. perfringens*, animal models (such as a mouse model) are required to more closely replicate complex host responses. Development of a novel pre-clinical model for *C. perfringens* infection will be described in the following chapter.

Chapter 8 A novel oral-challenge mouse model of *C. perfringens* infection unveils underlying immune-mediated disease pathologies

8.1 Introduction

The primary aim of the study presented in this chapter was to establish a *C. perfringens* oral-challenge infection model to probe *C. perfringens* pathogenesis using fluorescently-tagged hyper-virulent NEC-linked strains identified in Chapter 7 (Figure 8.1).

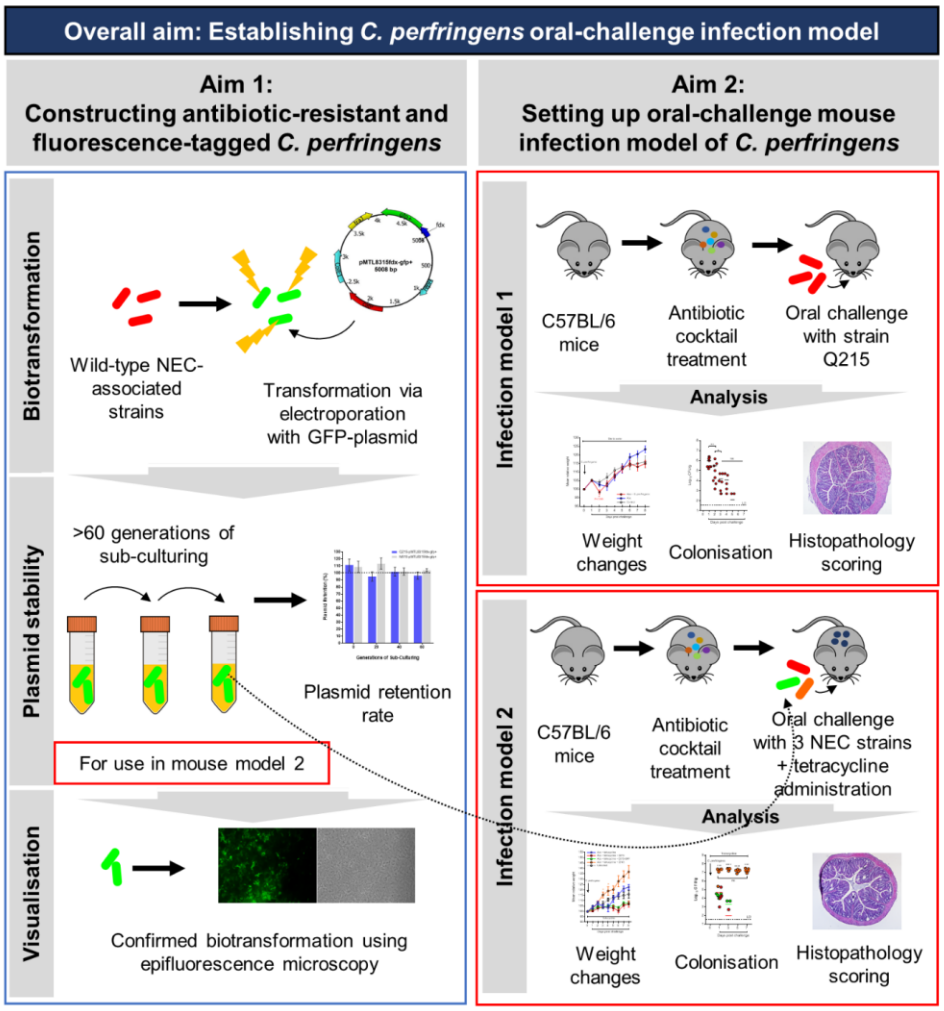


Figure 8.1: Schematic of study aims, experiments and analyses presented in Chapter 8

8.2 Background

C. perfringens has been recognised as the causative agent for numerous intestinal diseases in humans including preterm NEC and FP (gastroenteritis), as discussed in Chapter 3 and Chapter 4 respectively [1]. Furthermore, this notorious food-borne pathogen is also linked with AAD, which is a prolonged gastroenteritis (commonly >3 days, up to 60-70 days) frequently diagnosed in antibiotic-treated elderly patients residing within hospital wards, with symptoms attributed to the CPE toxin [24,325,443].

Primary clinical symptoms in *C. perfringens*-FP patients include rapid appearance (8-14h) of watery diarrhoea (usually self-limiting, low hospitalisation rate), intestinal cramp without fever or vomiting, which symptoms are distinguishable from other food pathogens such as *Salmonella* [1]. While in *C. perfringens*-NEC patients, abdominal distention (frequently <24h) and tissue necrosis in the intestine is commonly observed, which leads to high mortality rates (~40%) [73].

Though various animals, including rabbits (mainly intestinal loop model), rats and chickens have been studied as models for *C. perfringens* intestinal infections [122,347,444,445]; small animals (i.e. mice), which represent a more economical and tractable model system, have never been shown to be intestinally-infected by *C. perfringens*. Previous studies have used mice for different *C. perfringens* diseases, with a focus on toxins, namely CPE (intestinal loop model) [446] and CPB (IV injection model) [447]. However, to date no mouse model has been established using oral challenge, the natural mode of transmission in humans, which may be due to difficulties of *C. perfringens* colonisation in the murine gut, although a recent American study showed that up to 12% (48/416) of house mice (*Mus musculus*; non-laboratory) carried *C. perfringens* as commensals (a PCR-based study) [448-450].

The closest oral-challenge *C. perfringens* mouse model, to mimic human infection, is the type C-enterotoxaemia infection model, where BALB/c mice are intra-gastrically infected with *C. perfringens* type C strains. This model successfully induces disease <48h, however this study did not determine the long-term colonisation of *C. perfringens*, and was not optimised as a model to study intestinal infection, rather the focus was on systemic infection enterotoxaemia associated with CPB production from *C. perfringens* type C strains[80].

Another enterotoxaemia oral-route infection model, using BALB/c mice, was developed to determine rapid (<8 h) systematic disease symptoms after challenge with ETX-producing *C. perfringens* type D strains [451]. Notably, this study involved sealing the anus of infected mice indicating a failure to establish a reproducible disease model using 'normal' unsealed mice (which did not display symptoms), despite pre-treating with 1-7 days of broad-spectrum antibiotics, fasting and diet changes. Furthermore, this study did not explore murine gut colonisation of *C. perfringens* strains, highlighting the need for another model that more closely mimics clinically-relevant intestinal disease symptoms.

Development of a new *in vivo* mouse model could utilise the most common inbred mouse line, C57BL/6 mice (never reported before), which leverages a multitude of reagents and KO models, which could be used to understand colonisation and pathology of *C. perfringens* after natural oral infection. Establishment of such a clinically relevant model would allow a platform for e.g. vaccine development, testing of novel anti-infection treatments, and probing disease pathogenesis and mechanisms of protective immunity, that have relevance for serious intestinal *C. perfringens* infections such as gastroenteritis (in older adult communities) and NEC (in premature infants).

It is apparent that there is currently a lack of an appropriate mouse intestinal *C. perfringens* infection model, which is limiting the research community's ability to further understand underlying disease mechanisms. In this current chapter, a preliminary yet reproducible mouse model (using C57BL/6 mice) of *C. perfringens*-associated gastroenteritis was demonstrated with significant weight loss and marked signs of colitis determined by histology. Moreover, successful and persistent intestinal colonisation of *C. perfringens* was assessed and confirmed for the first time.

8.3 Hypothesis and aims

It is hypothesised that hyper-virulent *C. perfringens* strains could induce NEC-like or gastroenteritis disease symptoms in mice. The aims of this study are:

(1) Construct antibiotic-resistant and fluorescence-tagged *C. perfringens* strains for real-time colonisation tracking using *in vivo* imaging.

(2) Set up preliminary *C. perfringens* infection mouse model using hyper-virulent NEC-associated *C. perfringens* strains identified in Chapter 7.

8.4 Results

8.4.1 Construction of fluorescence tagged *C. perfringens*

This work aimed to generate GFP-tagged *C. perfringens* strains for use in real-time tracking and imaging *in vivo*. The GFP plasmid (5,008 bp-GFP-tagged plasmid; Figure 8.2A) was a gift from Dr Duncan Gaskin, and based on a reconstructed *Clostridium sporogenes* plasmid (original plasmid backbone pMTL8315 obtained from ClosTron), which was successfully introduced into *C. perfringens* strains Q215 and M010 by electroporation as described in the method section [248].

E. coli GFP-tagged stocks were confirmed to fluoresce brightly, and *C. perfringens* ATCC13124 reference strain was confirmed not to fluoresce under epifluorescence microscopy using GFP-filter (Figure 8.2B). Transformed strains Q215-GFP and M010-GFP were also confirmed to fluoresce brightly under epifluorescence microscopy (comparable to positive control), validated by the positive antibiotic selection (thiamphenicol resistance in anaerobic conditions; Figure 8.2C). In addition, both transformants were examined for plasmid stabilities by sub-culturing over 60 generations over 4 consecutive days, with >90% plasmid retention (n=3; Figure 8.2D). Strain Q215-GFP was then employed in infection model 2 (described later in this chapter).

8.4.2 Establishing a *C. perfringens* oral-challenge infection model

8.4.2.1 Isolate selection

Isolates Q215, Q143 (both hyper-virulent NEC strains) and Q088 (another NEC-associated strain; used in the preliminary study only) were selected based upon the phenotypic virulence scoring scheme as described in Chapter 7 and were used to establish an infection model in 3-5 weeks old C57BL/6 mice.

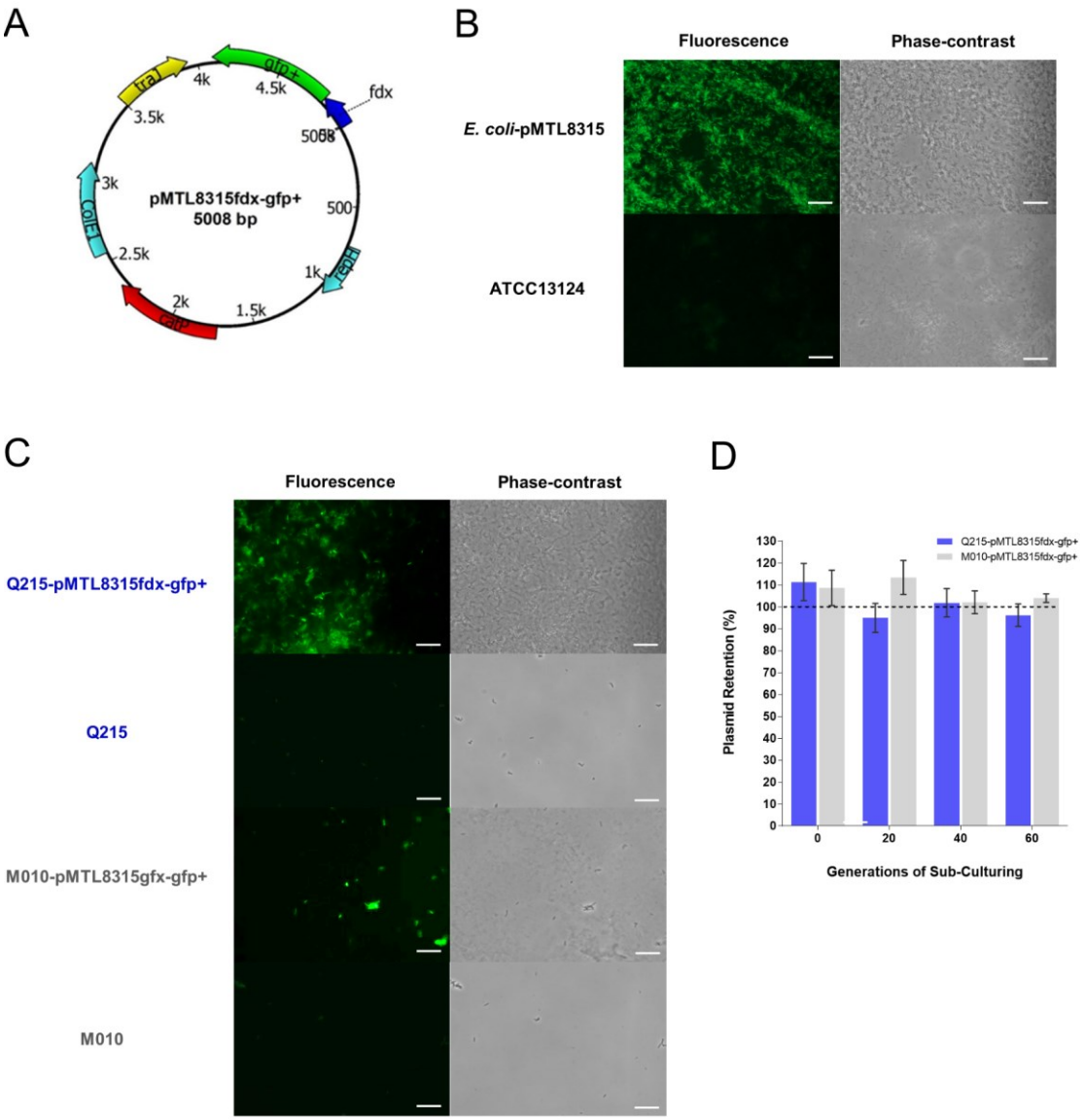


Figure 8.2: Construction of GFP-*C. perfringens* transformants Q215-GFP and M010-GFP

(A) Plasmid map for pMTL8315fdx-gfp+. **(B)** Fluorescence image of stock GFP-*E. coli* and reference strain *C. perfringens* ATCC13124. **(C)** Comparing images wild-type and GFP-*C. perfringens*. **(D)** Plasmid retention tests in transformants. Plasmids remained stable (>90%) in both transformants over 60 generations of *in vitro* sub-culturing. Experiment stopped at 60 generations. White bars represent 20µm. Data: Mean ± SEM (n=3).

8.4.2.2 Antibiotic pre-treatment is required for *C. perfringens* colonisation in mice

An initial study, used non-antibiotic treated C57BL/6 mice (aged 7-9 weeks), however these mice did not show substantial colonisation of *C. perfringens* strain Q215 (day 1 around 10^3 CFU/g; n=3) as CFUs were undetectable after day 1 (Figure 8.3A). Based on these data, a pretreatment with a combination of antibiotics routinely used in

preterm infants i.e. gentamicin and benzylpenicillin was administered to mice prior to *C. perfringens* oral gavage, however strain Q215 was again undetectable after day 1 (Figure 8.3B). These data indicate that human-associated *C. perfringens* do not efficiently colonise the intestine of adult C57BL/6 mice and indicated a stronger antibiotic scheme may be required to allow *C. perfringens* strain Q215 murine gut colonisation and subsequent infection/pathology.

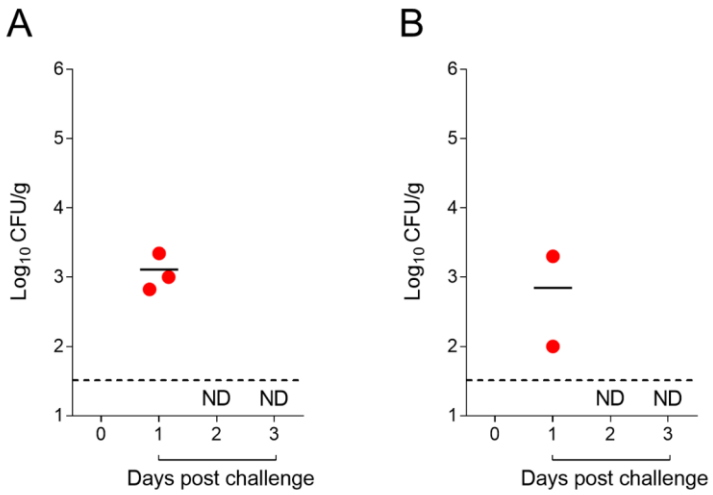


Figure 8.3: Intestinal colonisation of *C. perfringens* isolate Q215 in mice post oral challenge

(A) Mice without antibiotic pre-treatment. **(B)** Mice treated with gentamycin 5mg/kg (0.1mg per mouse), benzylpenicillin 25mg/kg (1mg per mouse) for 4 days, followed by 2 days sterile water, and then oral challenge with *C. perfringens*. ND: Not Detectable. Dotted lines indicate limit of CFU detection (50 CFU/g). 7- to 9-week-old C57BL/6 mice were used in this experiment. Black bars indicate means (n=3).

A murine model for the related pathogen *C. difficile* has been developed that uses a cocktail of antibiotics (kanamycin, gentamicin, colistin, metronidazole and vancomycin followed by single oral dose of clindamycin after two-day cessation of antibiotic cocktail) to establish and maintain colonisation [448,452]. Juvenile mice (3-4 weeks old) were used in place of adult mice (7-9 weeks old) in the following studies, aiming to mimic preterm intestinal infection in humans as these younger mice would have a relatively immature intestinal microbiota and immune system that should facilitate pathogenic infection. An initial study was carried out using this established antibiotic scheme for *C. difficile* infection [448] (Figure 8.4A), using *C. perfringens* strains Q215, Q143 and Q088. All challenged mice did not show significant differences in weight gain nor CFU from faecal pellets for the first three days post challenge ($P>0.05$; Figure 8.4B-D). Notably, high CFUs were recovered from all three groups (n=3 in each group; $>10^6$ CFU/g) indicating stable colonisation of *C*

perfringens after antibiotic administration, highlighting the feasibility of this antibiotic scheme for subsequent mouse work.

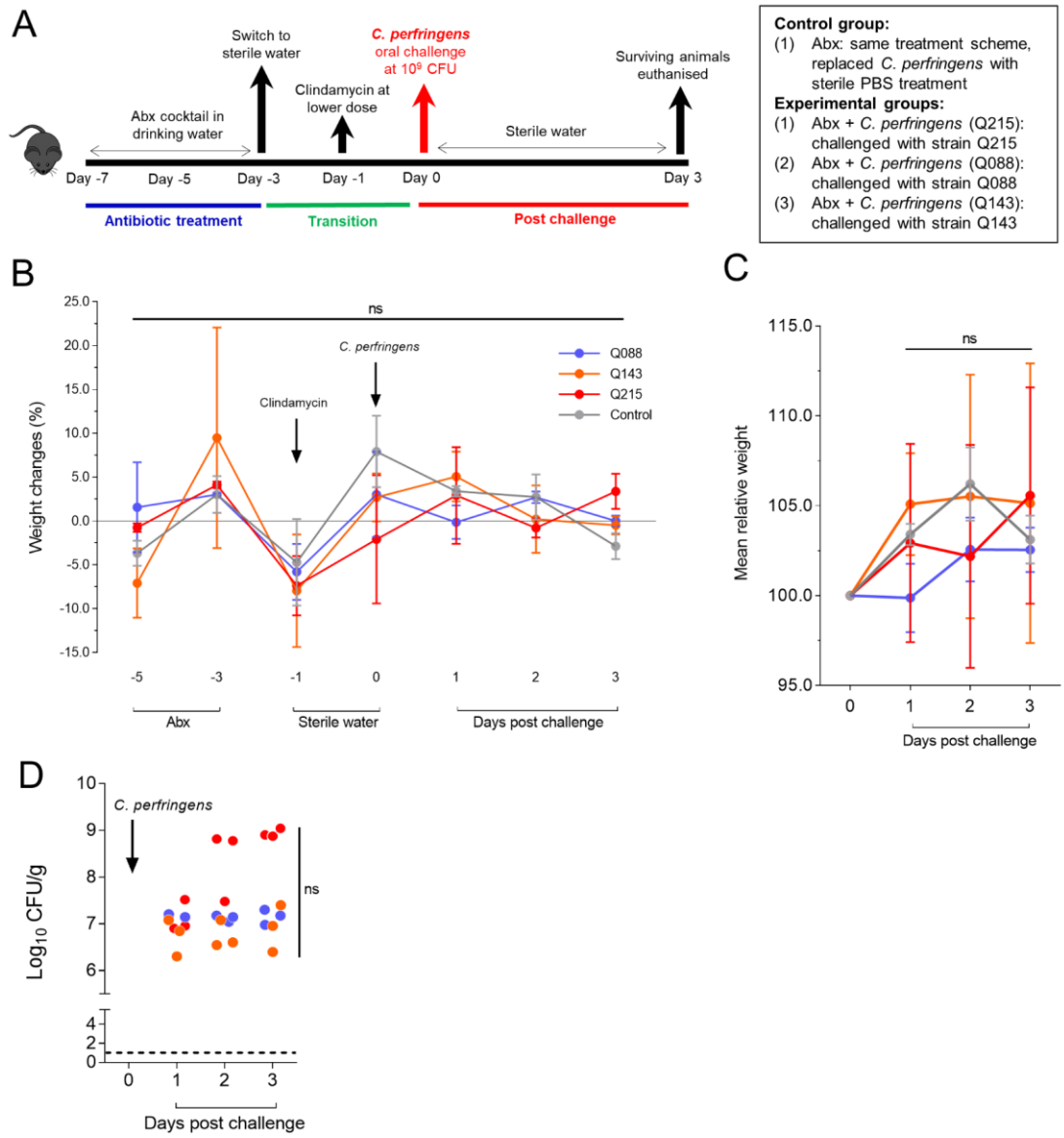


Figure 8.4: Preliminary colonisation study on *C. perfringens* in juvenile mice

(A) Scheme of experimental design. Three NEC-associated hypervirulent strains Q215, Q088 and Q143 were tested in mice, divided into 4 groups including control group (n=3 in each group). **(B)** Daily weight changes of mice over experimental period (pre- and post-challenge). **(C)** Relative weight change over time did not show significant changes. Weight change is based on weight on day 0. **(D)** Intestinal colonisation (CFU/g) of *C. perfringens* in 3- to 4-week mice over 3 days post challenge. NEC-associated isolate Q215(red) shows highest colonisation among all. Dotted line indicates LOD (50 CFU/g).

Data: (B) and (C) mean ± SEM (n=3). (D) Each dot represents mean (of 3 technical replicates) of each mouse. One-way ANOVA, Dunn's multiple comparison test. Abx: Antibiotic. Antibiotics used in pre-treatment: kanamycin (0.40 mg/ml), gentamicin (0.035 mg/ml), colistin (850U/ml), metronidazole (0.215 mg/ml), and vancomycin (0.045 mg/ml), followed by clindamycin (30 mg/kg).

8.4.2.3 Infection model (1)

C. perfringens strain Q215 was selected for this infection study due to the slightly higher CFUs obtained, when compared to the other strains after preliminary testing (as above). The experimental scheme of this study is illustrated in Figure 8.5A. Eight mice were challenged with 10^9 CFU of *C. perfringens* strain Q215 vegetative cells after antibiotic treatment, with the control group (n=6) also treated with the antibiotic cocktail for 4 days, followed by PBS gavage. Another untreated (i.e. no antibiotics) control group (n=5) was also used for comparison. Animals were monitored daily for *C. perfringens* colonisation (Figure 8.5B-C), weight changes (Figure 8.5E-F) and potential symptoms of disease. In the *C. perfringens* infection group, *C. perfringens* was cleared completely in all mice after 5 days post challenge (as determined by CFU counts), with most mice cleared of *C. perfringens* after day 4 (6/8 mice; Figure 8.5B-C). Significant weight loss ($P<0.05$; Figure 8.5E-F) was observed on day 2 in *C. perfringens* infected mice, when compared to mice in the antibiotic-only group and untreated group. All mice (n=8) in the *C. perfringens* infection group recovered from the initial weight loss (comparable to untreated group) and did not display an overt sign of disease throughout the experimental period.

To determine if loss of *C. perfringens* at later time-points was due to microbiota re-colonisation after antibiotic administration, overall microbiota load (as determined by 16S rRNA qPCR) was monitored. Notably there was no significant change in the *C. perfringens*-infection group from day 1 to 8 post challenge ($P>0.05$), indicating gut clearance of *C. perfringens* may be due to intrinsic resistance from other intestinal microbiota members (Figure 8.5D). There was a significant increase in 16S rRNA copies from day 7 prior to oral infection ($P<0.001$), suggesting a less diverse microbiota in young mice at 3 weeks old, which developed overtime to a more complex adult-like gut microbiome.

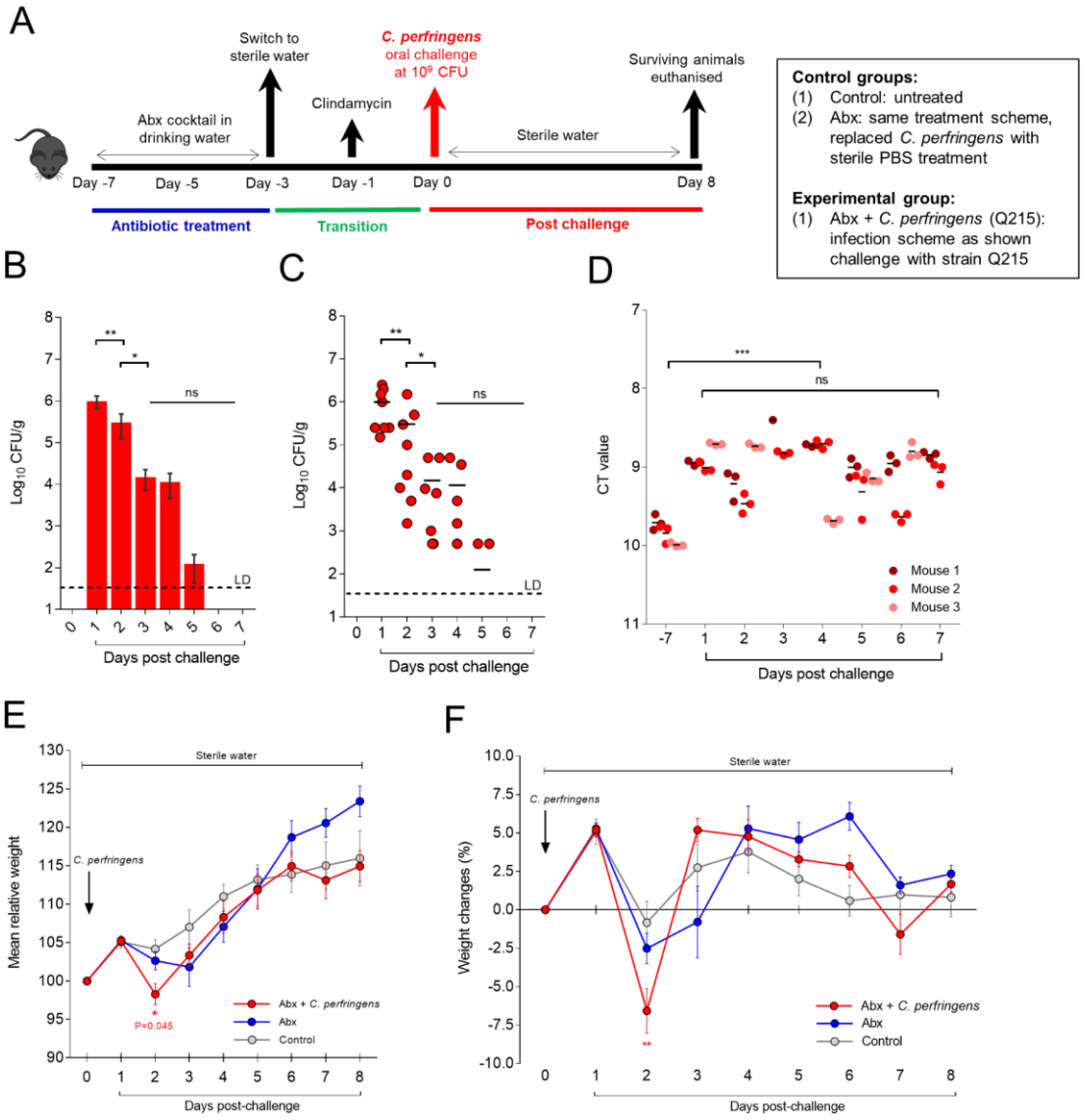


Figure 8.5: Induction of *C. perfringens*-associated infection in juvenile mice

(A) Scheme of experimental design. Strain Q215 was tested in mice (n=8; 3-4 weeks old), divided into 3 groups including control group (n=5) and Abx-only group (n=6). Abx-only and Abx+ *C. perfringens* groups were performed in two separate experiments. **(B)** and **(C)** Intestinal colonisation (CFU/g) of *C. perfringens* in 3 to 4-week adult mice (infection group, n=8) over 7 days post challenge. Dotted line indicates LOD (50 CFU/g). CFU could not be detected after 5 days post challenge. **(D)** Relative comparison of overall microbiota load between 3 mice post challenge using 16S rRNA qPCR (via 16S rRNA universal primers). Lower CT values indicate higher PCR-targeted products, i.e. microbial species. **(E)** Relative weight post challenge in all mice. Relative weight is based on the weight on day 0. **(F)** Daily weight changes of mice (%) over experimental period (post-challenge), based on weight on the previous day.

Data: (B) mean ± SEM (n=8). (C) Each bar represents mean, and each dot represents one individual mouse (mean of three technical replicates). Paired t test, two-tailed. * P<0.05, ** P<0.01, ns=non-significant. (D) Each dot represents one technical replicate of one individual mouse. One-way ANOVA, Dunn's multiple comparison test. *** P<0.001, ns=non-significant. (E) and (F) Data shown are means ± SEM. Abx: Antibiotic. Antibiotics used in pre-treatment: kanamycin (0.40 mg/ml), gentamicin (0.035 mg/ml), colistin (850U/ml), metronidazole (0.215 mg/ml), and vancomycin (0.045 mg/ml), followed by clindamycin (150 mg/kg).

Next, I sought to explore whether there were any changes in intestinal histopathology of mice challenged with *C. perfringens*, which may correspond to the evident initial weight loss. Histology (Figure 8.6A-C; raw data shown in Appendix 5 Table S8.1) staining indicated that colonic sections from mice infected with *C. perfringens* strain Q215 exhibited mild to moderate colitis (pathological score range: 7.0-11.0/14.0; mean ± SD: 9.3 ± 1.6; $P<0.001$) compared to untreated (score range: 0.5-4.0; mean ± SD: 1.7 ± 1.3) and antibiotic-only treated mice (score range: 1.5-4.5; mean ± SD: 3.2 ± 1.5; Figure 8.6D-E).

Colonic tissues from mice treated with antibiotic cocktail alone and untreated mice appeared normal, although some sections displayed minimal to mild pathological changes including epithelial hyperplasia in antibiotic-treated mice (Figure 8.6A-B); while tissue sections from mice challenged with *C. perfringens* showed moderate immune cell infiltrations (mean ± SD: 2.5 ± 0.5/4.0), substantial epithelial hyperplasia and marked goblet cell loss (mean ± SD: 3.2 ± 0.9/5.0), and mucosal ulcerations (mean ± SD: 3.5 ± 0.4/5.0; Figure 8.6C), indicating *C. perfringens*-associated pathogenesis in these mice.

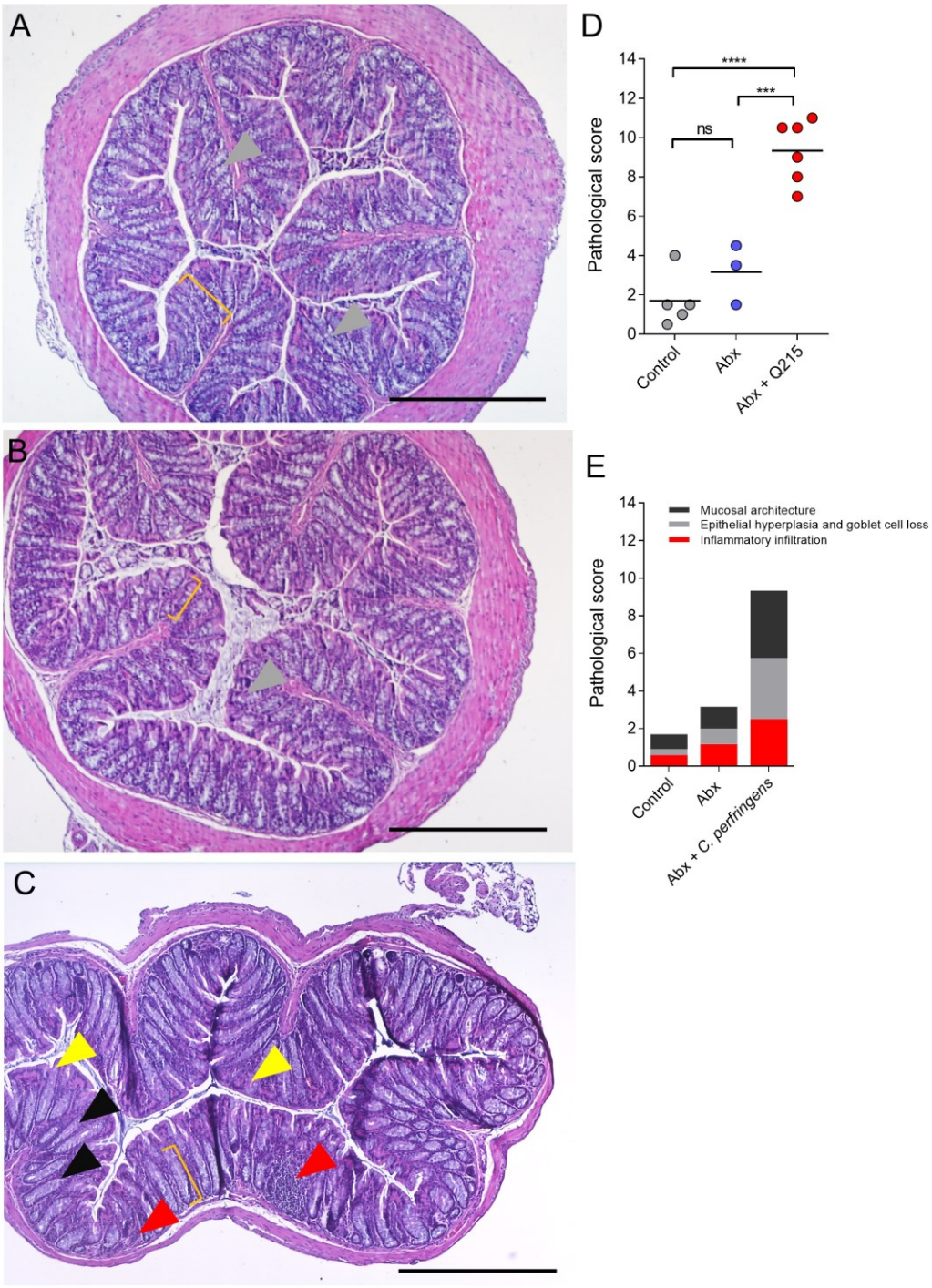


Figure 8.6: Histology on colon sections and pathological scoring.

Representative H&E-stained mice colon sections of **(A)** untreated group **(B)** Abx-only group, and **(C)** infection group (treated with *C. perfringens* strain Q215), showing epithelial changes (hyperplasia and goblet cell loss), intestinal inflammation (immune cell infiltration) and altered mucosal architecture (ulceration). Grey arrowhead- normal colonic goblet cells in untreated mice; yellow bracket- the length of normal colonic crypts as in (A); yellow arrowhead- erosion; black arrowhead- loss of goblet cells; red arrowhead- immune cell infiltration. Magnification: x40, scale bar 50 µm. **(D)** Pathological score of colons from mice of each group (control, Abx and infection groups). Each dot represents one mouse (bar: mean). One-way ANOVA, Dunn's multiple comparison test. *** P<0.001, **** P<0.0001. **(E)** Pathological score of colons from mice of each group, graded according to three general features (total 0-14): inflammatory infiltration (0-4), epithelial hyperplasia and goblet cell loss (0-5), and mucosal architecture (0-5). Mean score of each group is represented in the bars.

8.4.2.4 Infection model (2): tetracycline model

As *C. perfringens* strain Q215 was observed to be cleared from the mouse intestines after day 4 or day 5, I next sought to prolong the colonisation period, as it is hypothesised that overgrowth of *C. perfringens* in the gut initiates pathogenesis with associated disease symptoms. To provide this extended infection in the presence of a depleted gut microbiota, tetracycline was administered to mice in drinking water post *C. perfringens* gavage.

The experimental scheme is illustrated in Figure 8.7A, with three strains employed in this study including Q215, Q215-GFP (derived from Q215, described in section 8.4), and Q143, which all carry tetracycline resistance genes i.e. *tet* resistance determinants as described in Chapter 3, confirmed by phenotypic approaches (Chapter 7).

Similar to the previous study, mice infected with Q215 showed significant weight loss ($P<0.05$) on day 2, 6 (mean daily weight loss: 2.4%) and 7, and slower overall weight gain as compared to the group treated with antibiotic alone (Figure 8.7B-C). Infection with the recombinant strain Q215-GFP, also induced significant ($P<0.01$) weight loss on day 8, confirming the strain colonisation-effect (Figure 8.7B-C). The additional administration of tetracycline changed the weight dynamics, suggesting the potential role of other microbiota members in contributing to pathological changes.

Notably, mice challenged with Q143 consistently gained weight throughout the study period after oral infection, although apparently there had a slower growth rate from day 3 to day 5 post challenge.

Both Q215 and Q215-GFP groups appeared to be cleared of *C. perfringens* on day 5, which is in agreement with previous data, suggesting that tetracycline supplementation did not help prolong the colonisation of strain Q215 (Figure 8.7D). Surprisingly, although there was no weight loss or overt symptoms in mice infected with Q143, CFUs for strain Q143 remained significantly higher in mice ($>10^6$ CFU/g; $P<0.001$) throughout the post-challenge period (7 days).

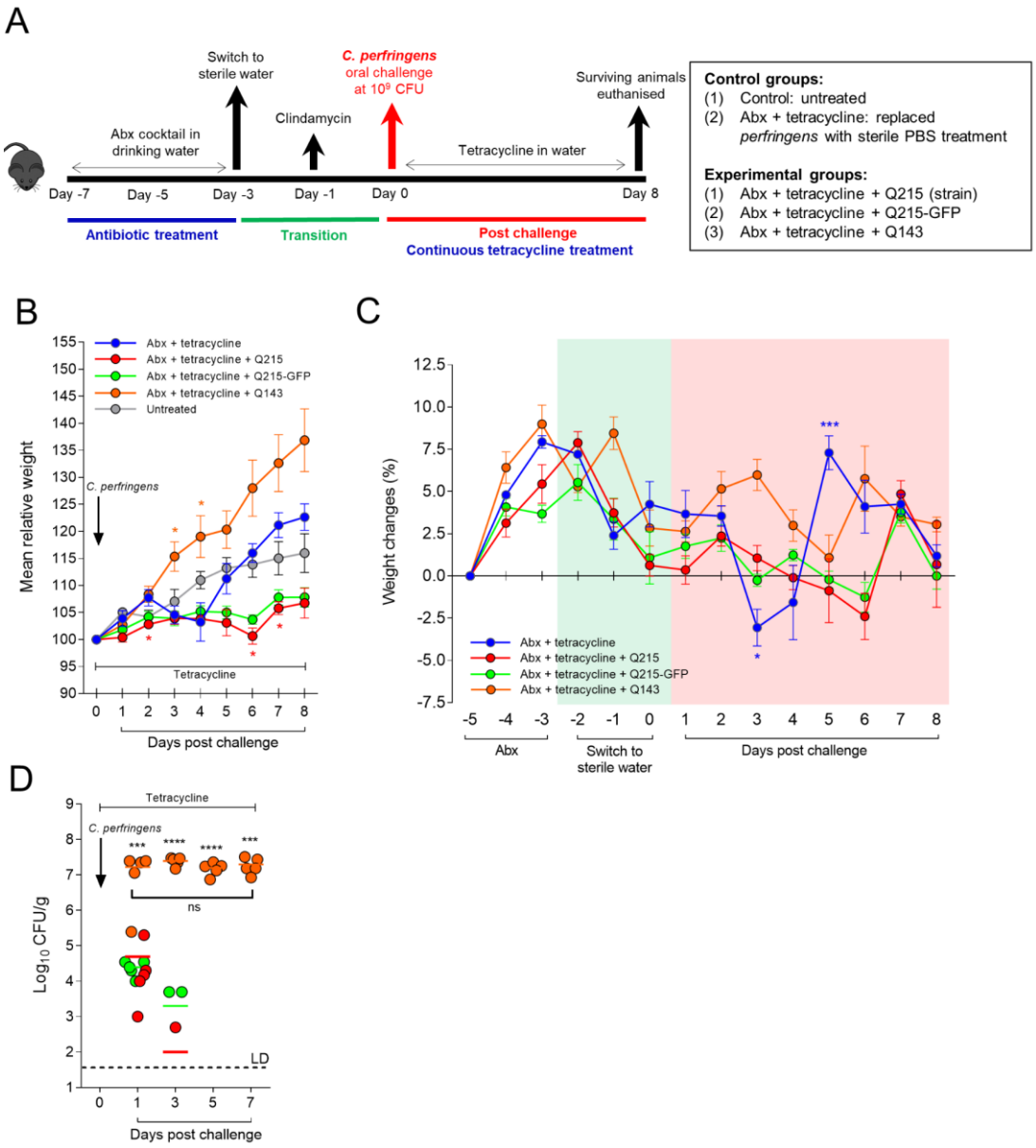


Figure 8.7: Induction of *C. perfringens* infection with tetracycline supplementation

(A) Scheme of experimental design (n=5 in each group). This study was performed in one single experiment. **(B)** Relative weight post challenge in all mice. Relative weight is based on the weight on day 0. **(C)** Daily weight changes of mice (%) over experimental period (pre- and post-challenge), based on weight on the previous day. **(D)** Intestinal colonisation (CFU/g) of *C. perfringens* in each group (n=5 in each group; 3-4 weeks old mice) over 7 days post challenge. Dotted line indicates LOD (50 CFU/g).

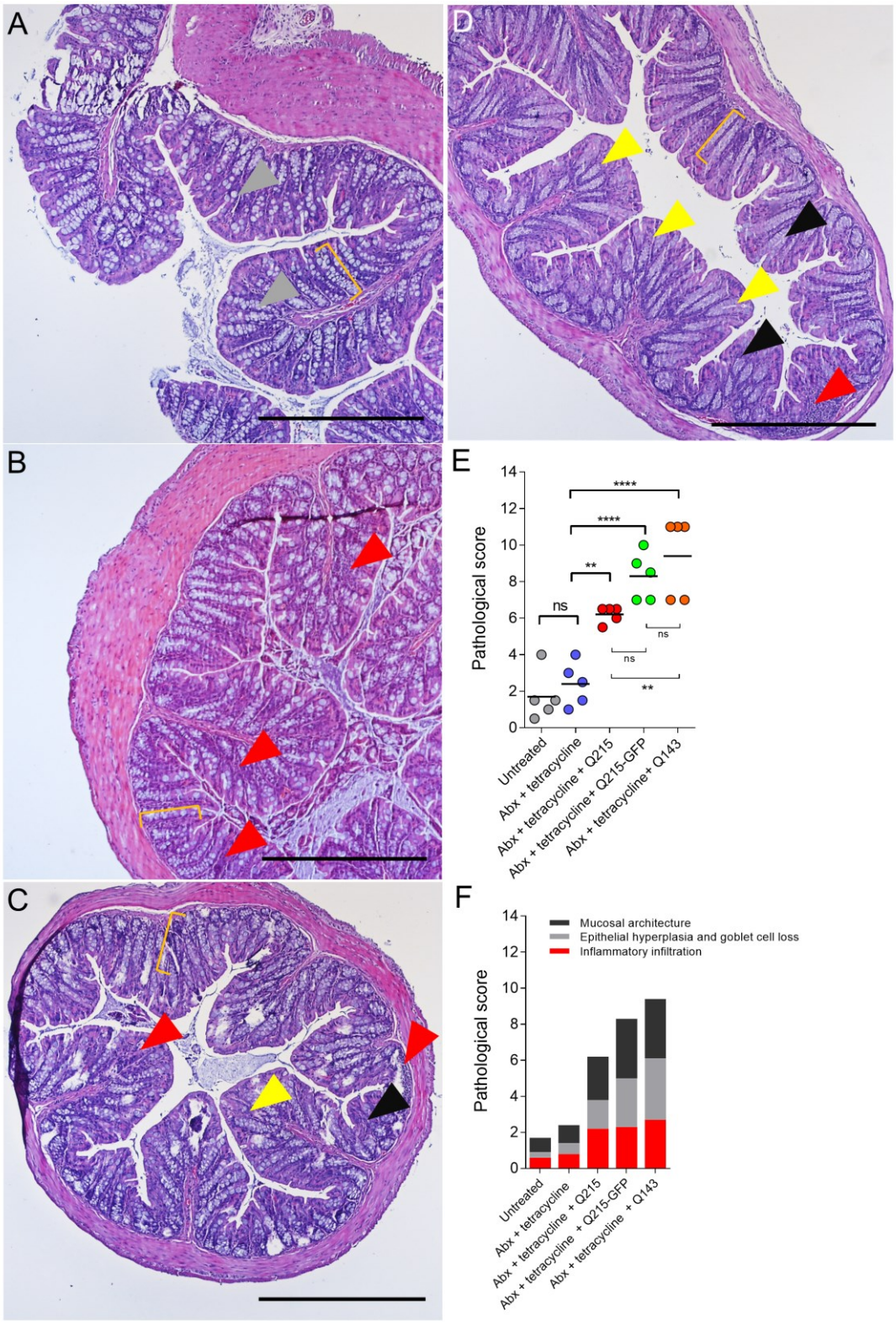
Data: (B) mean ± SEM (n=5 in each group). One-way ANOVA, Dunn's multiple comparison test (compared with Abx group as control). * P<0.05. (C) mean ± SEM (n=5 in each group). Dunn's multiple comparison test. * P<0.05, *** P<0.001. (D) Each dot represents CFU/g of each individual mouse from each experimental group (colour-coded). Dunn's multiple comparison test. *** P<0.001, **** P<0.0001, ns=non-significant. CFU/g recovered from Q143 group was significantly higher than the other two infection groups (Q215 and Q215-GFP) throughout the whole post-challenge period. Pre-antibiotic treatment: kanamycin (0.40 mg/ml), gentamicin (0.035 mg/ml), colistin (850U/ml), metronidazole (0.215 mg/ml), and vancomycin (0.045 mg/ml), followed by clindamycin (150 mg/kg). Tetracycline supplemented (0.001mg/ml) in drinking water post-challenge.

Colonic histology indicated that all mice infected with the three strains *C. perfringens* exhibited mild to moderate colitis (Figure 8.8A-D; raw data shown in Appendix 5 Table S8.2), having significantly higher pathological scores (pathological score range: 5.5-11.0/14.0; mean ± SD: 7.9 ± 1.9; P<0.0001) when compared to mice treated with antibiotics alone (score range: 1.0-4.0; mean ± SD: 2.4 ± 1.2; Figure 8.8E-F). Importantly, colonic tissues from group Q143 displayed moderate immune cell infiltration (mean score: 2.7/4.0), epithelial hyperplasia and goblet cell loss (3.4/5.0) and mucosal ulcerations (3.3/5.0), having the highest pathological mean score (9.4 ± 2.2/14.0) compared to group Q215 (6.2 ± 0.4/14.0) and group Q215-GFP (8.3 ± 1.3/14.0).

(Figure printed on next page)

Figure 8.8: Comparison of histology on colon sections and pathological scoring

Representative H&E-stained mice colon sections of **(A)** Abx-only group **(B)** Q215 group **(C)** Q215-GFP group, and **(D)** Q143 group (all treated with Abx cocktail prior to oral challenge and supplemented with tetracycline post challenge), showing epithelial changes (hyperplasia and goblet cell loss), intestinal inflammation and altered mucosal architecture (ulceration). Grey arrowhead- normal colonic goblet cells in untreated mice; yellow bracket- the length of 'normal' (uninfected) colonic crypts; yellow arrowhead- erosion; black arrowhead- loss of goblet cells; red arrowhead- immune cell infiltration. Magnification: x40, scale bar 50 µm. **(E)** Pathological score of colons from mice of each group (including data of untreated group from previous experiment for comparison). Each dot represents one mouse (bar: mean). One-way ANOVA, Dunn's multiple comparison test. ** P<0.01, **** P<0.0001, ns=non-significant. **(F)** Pathological score of colons from mice of each group, graded according to three general pathological features (total 0-14): inflammatory infiltration (0-4), epithelial hyperplasia and goblet cell loss (0-5), and mucosal architecture (0-5). Mean score of each group is represented in the bars.


A
D
B
E
C
F
Pathological score
Untreated
Abx + tetracycline
Abx + tetracycline + Q215
Abx + tetracycline+ Q215-GFP
Abx + tetracycline+ Q143
ns
**
****
Mucosal architecture
Epithelial hyperplasia and goblet cell loss
Inflammatory infiltration

8.5 Discussion

Oral infection of 3-week old C57BL/6 mice with NEC-associated type-A *C. perfringens* strains induced colitis-like histological changes in colonic sections, in tandem with significant weight loss on days 2 and 6. This is the first time *C. perfringens* has been shown to naturally infect and colonise C57BL/6 mice (via oral route) with associated intestinal pathology.

To date, no clinically-relevant small animal infection models have been developed for *C. perfringens*-associated enterocolitis. Initial studies highlighted a robust microbiota was present in mice housed within the UEA DMU, as indicated by efficient colonisation resistance against *C. perfringens* infection. Thus, an antibiotic pre-treatment (5-antibiotic cocktail + clindamycin) regimen was used to deplete the resident microbiota, as has previously been used for *C. difficile*-infection in C57BL/6 mice [453]. Faecal CFU data demonstrated that *C. perfringens* strain Q215 successfully colonised for 4-5 days in mice with an antibiotic-disturbed gut microbiota, which correlated with significant colonic histopathological changes. Importantly, a similar antibiotic regimen was also used in a highly-reproducible *C. jejuni* mouse infection model that mimics several human infection symptoms and several immunological aspects after high *C. jejuni* CFU colonisation [449]. The importance of antibiotic-induced microbiota disturbances was further highlighted in infection model 2, with the addition of tetracycline. Notably, weight loss was more pronounced in this model suggesting prolonged administration of tetracycline directly impacts disease severity, which correlated with enhanced histopathology, likely linked to depleted colonisation resistance abilities of the resident gut microbiota. Associated systemic biomarkers should be examined to determine the systemic impact of this pathogen, and how this links to poor growth. Importantly, this model (model 2) was only performed in a single experiment, thus further work is required to investigate the impact of different antibiotic regimens on the mouse microbiome, which will help with development of a robust and clinically relevant *C. perfringens* infection model.

Notably, serious *C. perfringens* infection is often associated with vulnerable individuals, that are expected to have a disturbed microbiota as a result of prolonged antibiotic usage, including preterm infants and elderly CH residents [289,334]. From these data it also appears that *C. perfringens* required a disturbed (murine) gut microbiota to initiate successful colonisation after the natural route of oral infection [8,289].

Significant weight loss ($P<0.05$; 6.6% daily weight loss) was observed on day 2, which is similar to findings previously described for the *C. difficile* relapse mouse model [452]. Nevertheless, all infected mice regained body weight from day 3 onwards, which may be correlated with the gut clearance of *C. perfringens* from day 3. The significant pathological changes observed in colonic sections was somewhat unexpected, as mice did not display any overt distress symptoms throughout study period. However frequent diarrhoea was observed, which does suggest intestinal changes and is similar to what is observed for the mouse pathogen *Citrobacter rodentium*, which is used as a model for human enteropathogenic *E. coli* [454-458].

General histopathological changes were observed in *C. perfringens*-challenged mice, including extensive goblet cell loss, hyperplasia in the crypts, immune cell infiltration, mucosal erosion and a certain degree of oedema. These observations are similar to those observed in the *C. rodentium* infection model [454]. In fact, pathological changes seen during *C. rodentium* infection are driven by the host immune system rather than direct damage. Goblet cell reduction, and associated mucus depletion is a pathogen-mediated immune subversion that facilitates enhanced colonisation of *C. rodentium* and is similar to what was observed in this *C. perfringens* model [459]. These pathological signs may primarily be the result of 'indirect immune damage' rather than 'direct damage' during *C. perfringens* infection, as only limited mucosal erosions (caused by pore-forming toxins) were observed [451,455]. Furthermore, significant immune cell infiltration observed exclusively in *C. perfringens*-challenged mice (8 days post-challenge) also supports the idea of immune-mediated pathology. This implicates the importance of the under-studied immune-driven aspect of *C. perfringens* intestinal infections, in contrast to the conventional thought of *C. perfringens* pathology as the sole result of toxin-driven direct damage [82,460].

However, although NEC strains were used in these studies, these symptoms do not clinically replicate the very serious intestinal pathology associated with preterm-NEC infection e.g. complete intestinal necrosis and abdominal distention [73]. There are several possible factors that could explain lack of NEC disease symptoms such as diet and age. As *C. perfringens* is known to require a high-protein diet to thrive in the gut, the normal mouse chow diet may not favour this pathogen, thus impacting overall growth and expression of certain virulence factors e.g. toxins. In addition, although young mice were used (3- to 4-week-old) these mice are already weaned and therefore

considered as 'adult' mice with an adult-like gut microbiota and host immune system, that would be expected to impact on optimal *C. perfringens* colonisation and disease pathogenesis. Correspondingly, serious *C. perfringens* intestinal infection is rarely seen in healthy human adults, with the exception of FP (large oral consumption) and AAD (weaker host immunity) cases [1,322]. Therefore, it will be necessary to further optimise this experimental protocol for a NEC-specific model.

Based on disease activities and pathologies it appears this infection regimen is more similar to gastroenteritis-associated *C. perfringens* infection, and therefore may represent an appropriate model for testing associated pathogenesis. The rapid intestinal clearance (4-5 days) of *C. perfringens* strain Q215, and recovery of weight gain (day 3) in infection model 1 closely resembles the self-limiting symptoms of human FP (24-48h) and may link to the spontaneous clearance observed in these patients, however to note strain Q215 is not a FP-related type F strain as it does not encode CPE. The intrinsic clearance of pathogen *C. perfringens* within the murine gut, which is likely host immune-mediated, warrants further investigation, also linking to potential bacterial colonisation factors, which are expected to vary from strain to strain [116,182,454]. Indeed, strain-level colonisation variations was demonstrated in the tetracycline model, as strain Q143 was maintained at high CFU throughout the study period. Notably, strain Q143 was shown to be highly-sporulant, exhibiting $>10^4$ log compared to Q215 (Chapter 7). Consequently, sporulation capacity could be an important factor for *C. perfringens* colonisation in mice. Importantly, *C. difficile* is known to form spores more readily than *C. perfringens*, which is directly linked to its ability to persist in both the human and murine gut, as shown in clinical and mouse studies [448,452].

8.6 Future research directions

To further optimise this infection model, it is necessary to modify the oral challenge protocol to include administration of *C. perfringens* spores instead of vegetative cells to potentially optimise the colonisation of strain Q215. Furthermore, more *C. perfringens* strains could be tested to understand whether commensal *C. perfringens* (not disease-associated) can also induce disease symptoms in antibiotic-treated mice.

The data obtained could be correlated with the genotyping data and link genotype to phenotypes, thus answering several questions postulated in Chapter 3 and Chapter 7.

Potentially, a different antibiotic scheme could also be applied. There are currently several antibiotic pre-treatment regimens employed in *C. difficile* studies, including the cefoperazone model which have successfully established infection models [461].

Following the successful colonisation of *C. perfringens* in mice, and associated enteric infection pathology, there is a possibility to develop a rigorous oral-challenge gastroenteritis model of FP or long-term AAD using FP-related type F strains (obtained from PHE). This model would allow development of preventative measures and testing of treatments such a probiotic supplementation or vaccines which could be clinically vital to older adult communities [289]. This potential FP or AAD infection model would represent the first of its kind in the field of *C. perfringens* research.

In addition, an optimised model could be used to probe specific pathogenesis traits of *C. perfringens* including exploring the role of toxin gene *pfo,* associated with clinical NEC representations, as presented in Chapter 3, via construction of *pfo*-knockout and *pfo*-complementing mutant strains (using technique e.g. CRISPR-Cas9 system as demonstrated in *C. difficile*) to confirm the essentiality of this virulence gene linking to diseases under Koch's molecular postulate [462]. This established model will be valuable in probing other accessory virulence genes including those related to host colonisation.

Faecal content DNA of infection model 1 has been sent for 16S rRNA sequencing, with the aim to identify taxa related to clearance of *C. perfringens*, and also explore the overall impact of the antibiotic cocktail on the murine gut microbiome, including recovery following antibiotic pre-treatment. These data will allow a greater understanding of the changes in bacterial taxa and may facilitate further optimisation of this infection protocol.

In conclusion, this is the first time that *C. perfringens* strains have been shown to naturally infect C57BL/6 mice and induced mild-to-moderate intestinal colitis. Further work is required to characterise the host immune-driven biomarkers of this infection model to probe pathological mechanisms, and to optimise this disease model to more closely mimic human infections including preterm-NEC, human FP (gastroenteritis)

and AAD, which is essential for downstream *C. perfringens* mechanistic studies and therapy development.

Chapter 9 Perspectives and final discussion

9.1 Background

Since the discovery of *C. perfringens* (previously named as *C. welchii*) >120 years ago, this pathogen has been associated with numerous histotoxic diseases, including the notorious gas gangrene [408,463]. Due to its importance in military research (as gas gangrene was one of the most common medical issues in traumatic soldiers during World War I and II), scientists focused on understanding the disease progressions of *C. perfringens*-mediated gas gangrene, resulting in development of a gas gangrene mouse model in early 1990s and sequencing of *C. perfringens* strain 13 (gas gangrene strain), which was the first anaerobic Gram-positive pathogen to be whole-genome sequenced back in 2002 [37,77,84].

Apart from gas gangrene, *C. perfringens* has been frequently linked with numerous intestinal diseases in both humans and animals including preterm-NEC, which has been termed as 'gas gangrene of the bowels' [74]. As such, *C. perfringens*-related typing toxins have been extensively studied in the past few decades, and successfully linked with human and veterinary disease representations. Nevertheless, in this genomic era only a small number of *C. perfringens* strains (n<120) have been whole-genome sequenced, perhaps the most genomically understudied enteric pathogen, highlighting the lack of understanding of *C. perfringens* genomes compared to other important pathogens including *C. difficile* and *Salmonella,* both of which have been extensively sequenced leading to subsequent development of disease models and therapeutics [1]. Importantly, no appropriate *C. perfringens*-associated small animal infection models have ever been described, thus hindering our understanding of natural *C. perfringens* intestinal infection mechanisms in mammals. Studies have implicated that *C. perfringens*-associated diseases are not purely caused by key typing toxins alone, there exists potentially a combination of several virulence elements including colonisation factors encoded in certain *C. perfringens* strains, or perhaps, some unidentified novel virulence genes that might be involved in disease progressions [116,118,230,290].

This underlying background drove the initial foundation of this PhD project, with the provision of *C. perfringens* isolates from key collaborators Prof J S Kroll (preterm cohort; n=189), Dr C Amar (FP and CH cohorts; n=109), and Dr R Dixon (poultry NE

cohort; n=22), this project developed to investigate and explore extensively the virulence-associated determinants and epidemiology of *C. perfringens* from various aforementioned cohorts, to probe global spread and cluster-specific genome features of *C. perfringens*, additionally phenotypic pathogenicity of selected strains was determined and more importantly, a novel mouse infection model was established, contributing to the new knowledge of *C. perfringens* in both human and veterinary medicine via multiple *in silico*, *in vitro* and *in vivo* approaches.

9.2 Impacts and future perspectives of key findings

WGS-based phylogenetic clustering of a huge collection (n=552) of *C. perfringens* isolates was visualised for the very first time in Chapter 6, suggesting the potential re-assignment of sub-species due to the unprecedented core-genome genetic variability observed. Indeed, this genetic variation was expected as a number of previously published studies indicated significant polymorphisms observed in *C. perfringens* housekeeping genes via various PCR-based typing studies including fAFLP and PFGE [27,150,325,382,464]. In agreement with previous studies, chromosomal-*cpe C. perfringens* strains are identified as a highly distinct cluster (96%<ANI<97%; lineage V) in the phylogeny, also reported to possess distinguished phenotypic expression including high heat-resistance, suggesting the potential sub-species assignment to this cluster of *C. perfringens* as they displayed similar-range of genome nucleotide identity [382].

Moreover, *C. perfringens* has been described frequently as a gut 'commensal' as this bacterial species is often isolated from healthy individuals, however, no specific investigations have been undertaken to understand what defines a commensal *C. perfringens*. Importantly, I have shown that several phylogenetic clusters can be considered 'commensal' *C. perfringens*, encoding significantly less virulence genes, and colonisation factors, with all isolates originating from 'healthy' individuals. For the first time, this WGS data has highlighted that specific clusters of strains could potentially be avirulent *C. perfringens,* and therefore a normal member of the gut microbiota [241]. Nevertheless, carriage rates of commensal *C. perfringens* in normal healthy populations is currently unknown, and this warrants further genomic investigations via a large-scale study.

With regard to the current toxinotyping scheme, updated recently to 7 toxinotypes (*C. perfringens* is the only known pathogen to be commonly typed according to the combination of key toxins), the various studies presented in this thesis strongly indicate that this method should not be solely relied on for determining disease associations, especially in human-related intestinal diseases, due to the low-resolution of information garnered about a particular isolate. A full genomic profile is thus preferred to comprehensively define virulence capacity and disease links. This is evidenced by the fact that type A isolates (only encode typing toxin PLC) were commonly diagnosed and associated with diseases (preterm NEC, chicken NE, human FP etc), highlighting the important role 'secondary' virulence factors may play, as discussed throughout e.g. PFO, CPB2, colonisation elements collagen adhesins and NanI [115,182,395,465], with these important virulence genes largely ignored in current typing schemes. WGS techniques should gradually replace all marker-gene-based typing methods especially in surveillance laboratories, to facilitate epidemiological investigation, as demonstrated in Chapter 3 and Chapter 4 of this thesis. This is particularly important in the context of public health surveillance and disease control on a national level.

Clonal expansion of certain disease-related *C. perfringens* case isolates has been reported for the first time in this thesis. This expansion may link to optimal fitness of the clonal strains, or attributed to the source of contaminations, as in the cases of food-poisoning, CH-associated diarrhoea and NICU wards. Further molecular work will be necessary to shed light on the underlying virulence arsenal of hypervirulent disease-causing *C. perfringens* strains.

In addition, the widespread nature and reservoirs of *C. perfringens* plasmids is yet to be determined. This area of research is of greater interest to *C. perfringens* researchers, as shown in Chapter 4, as plasmids are directly linked to specific disease cohorts. Importantly, *C. perfringens* virulence plasmids are critical for the transfer of important virulence genes among *C. perfringens* strains including colonisation factors, toxins and AMR genes [41,291]. Large-scale plasmid investigations should be obtained from a wider population of *C. perfringens* isolates, and WGS applied to understand the distribution of virulence plasmids, as currently only 48 sequenced plasmids are on the NCBI genome database (November 2018).

The identification of hypervirulent strains associated with preterm-NEC, from a multicentre study, has shed new light on how *C. perfringens* genotypes may be linked to NEC disease progression. Since the 1970s this pathogen has been considered as one of the major culprits behind this fatal infection, which has been further confirmed by recent sequencing-based large-scale studies [8,9]. NEC disease initiation is suggested to be multifactorial, including an immature gut microbiota and immune system, in tandem with more virulent *C. perfringens,* which is similar to other neonatal animal cases that share similar disease representations [73,199,466]. Hence, *C. perfringens* routine surveillance is necessary, especially in NICUs, given that this pathogen is crucial to the development of NEC – the deadliest mortality cause in premature neonates. This suggestion is supported by the fact that hypervirulent genotypes appear to be circulating in NICU wards (as demonstrated in Chapter 3), which necessitates future in-depth sample collections to confirm this finding.

Notably, I demonstrated that toxinotype D and E strains were detected in two healthy children, indicating reservoirs of certain animal-associated plasmids (especially plasmids that carry ETX, ITX and CPB) may be more widely disseminated than previously thought. Interestingly, toxinotypes D and E strains (carrying plasmids encoding ETX and ITX) mainly infect cattle and sheep, which typically having a higher body temperature (39.0-40.0ºC) than humans (36.5-37.5ºC), indicating that temperature might play an important role in *C. perfringens* 'overgrowing', and dictating whether a virulent strain can proliferate and cause disease. Body temperature may also explain the fact that *C. perfringens* predominately affects animals, rather than humans, and typically chickens (body temperature 41.0-45.0ºC), cows (38.3-40.0ºC), foals (37.7-38.8ºC), piglets (38.0-40.0ºC) [70].

This factor may also clarify the difficulties encountered establishing a mouse infection model of *C. perfringens*. As mice has a similar body temperature to human (36.5-37.5ºC), this model is more closely expected to mimic *C. perfringens*-associated human infections. In this body of work, I demonstrated a novel oral-challenged *C. perfringens* mouse model that exhibited significant colitis induced by *C. perfringens* NEC strains, albeit the expected NEC-like lesions were not observed. This may also suggest that strains without CPE could potentially induce gastroenteritis in humans, which is supported by the finding in Chapter 4 that certain FP case isolates did not encode CPE. This novel study will be improved once the microbiome is considered

(samples undergoing sequencing), thus allowing future development of preventative measures against gastroenteritis, such as probiotic supplementation. This new disease model offers the opportunity to study disease mechanisms of *C. perfringens* after optimisation.

The food chain as a reservoir of *C. perfringens* is an intriguing aspect of research, as previously highly toxigenic strains (encoding *cpe, cpb2* and *netB*) have been isolated from raw meats (detected up to 12.9% of the 232 samples) [467,468]. This can potentially be attributed to accidental contamination from human handling, however, this reservoir and route of spread of *C. perfringens* remains to be deciphered [333]. Notably, the dissemination of this pathogen may be universal, as indicated by the zoonotic findings, therefore I postulated that *C. perfringens* is a zoonotic agent that has no host boundaries, that can thrive in a wide host range.

Overall, my PhD work has significantly enhanced our understanding of multi-host pathogen *C. perfringens.* Using computational methods, I delved into the genomic contents and SNP-based epidemiological tracking, and also determined the phenotypic strain-specific virulence via *in vitro* approaches and finally, developed a novel *in vivo* model, which may allow more translational aspects of *C. perfringens*-related intestinal diseases to be addressed. These findings lay the groundwork for further investigations on *C. perfringens* as a medically important pathogen, particularly the important contribution of WGS data on *C. perfringens* isolates to the wider research community.

9.3 Translational aspects of research

Pathogen genomes convey important genetic information on their potential phenotypic virulence, which is crucial in medical microbiology. Currently >20 toxin genes have been identified and characterised in *C. perfringens*, and with these additional WGS information other novel toxin genes or factors may be discovered, as represented by the recently characterised foal-NE-associated NetF toxin [102].

Phage therapy represents a promising strategy to treat infections, while minimising the occurrence of multi-drug resistance in the microbiome, that may lead to secondary complications [469,470]. Probing bacteriophage contents in *C. perfringens* genomes could lead to further understanding of the role they play in virulence, as well as the

identification of temperate/lytic phages within bacterial genomes [293,295,471]. Potentially, novel *C. perfringens*-specific phages could be screened, isolated and characterised for therapeutic applications, owing to the huge culture collections of *C. perfringens* strains currently in place and its associated genomic information to replicate the successful trials in a recent poultry study [472].

From a diagnostic perspective, it is vital to rapidly detect *C. perfringens* with full virulence profiles at early stages, including rapid onset NEC. With the current technology of real-time Nanopore sequencing, samples can be metagenome-sequenced within <7h from sample reception, identification of bacterial population at species/strain level, and full virulence profiles including AMR potentials [315]. This promising pipeline could be used for *C. perfringens* rapid diagnosis in clinical settings, thus allowing quick clinical decisions with respect to treatment strategies. Furthermore, with the current binning-based genome-reconstruction techniques, single genomes of *C. perfringens* can be confidently extracted from shotgun metagenome data to infer full pathogenicity profiles in under 1-2 h [313].

Lastly, the novel *in vivo* model described in this thesis is an important step forward in *C. perfringens* pre-clinical research, particularly for future mechanistic understanding of disease development. Confirmation of disease-essential virulence elements including the hypothesised PFO will require the fulfilment of Koch's molecular postulates via testing of *pfo* knock-out and complemented strains in this disease model. Further therapeutic approaches including phage therapy can potentially be explored using the optimised models, which is crucial in pre-clinical trials.

Overall, this PhD research has contributed significantly to the genomic, phenotypic, pre-clinical and clinical knowledge of *C. perfringens.* These data should pave the way for future in-depth research to combat this notorious yet understudied pathogen, particularly in vulnerable communities including neonates and the elderly, via modern therapies or preventative measures.

References

1 Kiu, R. & Hall, L. J. An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerging Microbes & Infections* **7**, 141, doi:10.1038/s41426-018-0144-8 (2018).

2 Hassan, K. A. *et al.* Genomic analyses of *Clostridium perfringens* isolates from five toxinotypes. *Res. Microbiol.* **166**, 255-263, doi:10.1016/j.resmic.2014.10.003 (2015).

3 Welch, W. H. & Nuttall, G. H. F. A gas-producing bacillus (Bacillus aerogenes capsulatus, Nov, Spec.) capable of rapid development in the body after death. *Bull John Hopkins Hosp Baltimore* **3**, 81-91 (1891).

4 Veillon, M. M. & Zuber, W. Recherches sur quelques microbes strictement anaerobies et leur vole en pathologic. *Arch Med Exp Anat Pathol* **10**, 517-545 (1898).

5 Migula, W. *System der bakterien. Handbuch der morphologie, entwicklungsgeschichte und systematik der bakterien*. (Fischer, 1897).

6 Breed, R. S., Murray, E. G. D. & Hitchens, A. P. *Bergey's Manual of Determinative Bacteriology*. 6th Ed edn, (The Williams & Wilkins Company, 1948).

7 Lugli, G. A. *et al.* Ancient bacteria of the Otzi's microbiome: a genomic tale from the Copper Age. *Microbiome* **5**, 5, doi:10.1186/s40168-016-0221-y (2017).

8 Sim, K. *et al.* Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clin. Infect. Dis.* **60**, 389-397, doi:10.1093/cid/ciu822 (2015).

9 Heida, F. H. *et al.* A necrotizing enterocolitis-associated gut microbiota is present in the meconium: results of a prospective study. *Clin. Infect. Dis.*, doi:10.1093/cid/ciw016 (2016).

10 Revitt-Mills, S. A., Rood, J. I. & Adams, V. *Clostridium perfringens* extracellular toxins and enzymes: 20 and counting. *Microbiology Australia*, 114-117 (2015).

11 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).

12 Sim, K. *Defining the gastrointestinal microbiota in premature neonates: Its development and relation to necrotizing enterocolitis* Doctor of Philosophy (PhD) thesis, Imperial College London, (2015).

13 Nagler, F. P. O. Observations on a reaction between the lethal toxin of *Cl. welchii* (Type A) and human serum. *Br J Exp Pathol* **20**, 473-485 (1939).

14 Hansen, M. & Elliott, L. New presumptive identification test for *Clostridium perfringens*: reverese CAMP test. *J. Clin. Microbiol.* **12**, 617-619 (1980).

15 Browne, H. P. *et al.* Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543-546, doi:10.1038/nature17645 (2016).

16 van Asten, A. J., van der Wiel, C. W., Nikolaou, G., Houwers, D. J. & Grone, A. A multiplex PCR for toxin typing of *Clostridium perfringens* isolates. *Vet. Microbiol.* **136**, 411-412, doi:10.1016/j.vetmic.2008.11.024 (2009).

17 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, doi:10.3389/fmicb.2017.02485 (2017).

18 Amar, C. Fluorescent amplified fragment length polymorphism (fAFLP) analysis of *Listeria monocytogenes*. *Methods Mol Biol* **1157**, 95-101, doi:10.1007/978-1-4939-0703-8_8 (2014).

19 Rood, J. I. *et al.* Expansion of the *Clostridium perfringens* toxin-based typing scheme. *Anaerobe*, doi:https://doi.org/10.1016/j.anaerobe.2018.04.011 (2018).

20 Petit, L., Gibert, M. & Popoff, M. *Clostridium perfringens*: toxinotype and genotype. *Trends Microbiol.* **7**, 104-110 (1999).

21 Nagahama, M. *et al.* Recent insights into *Clostridium perfringens* Beta-toxin. *Toxins* **7**, 396-406 (2015).

22 Freedman, J. C., Shrestha, A. & McClane, B. A. *Clostridium perfringens* Enterotoxin: Action, Genetics, and Translational Applications. *Toxins (Basel)* **8**, doi:10.3390/toxins8030073 (2016).

23 Yonogi, S. *et al.* Development and application of a multiplex PCR assay for detection of the *Clostridium perfringens* enterotoxin-encoding genes cpe and becAB. *J. Microbiol. Methods* **127**, 172-175, doi:10.1016/j.mimet.2016.06.007 (2016).

24 Kim, Y. J. *et al.* Prevalence of *Clostridium perfringens* toxin in patients suspected of having antibiotic-associated diarrhea. *Anaerobe* **48**, 34-36, doi:10.1016/j.anaerobe.2017.06.015 (2017).

25 Verherstraeten, S. *et al.* Perfringolysin O: The underrated *Clostridium perfringens* toxin? *Toxins* **7**, 1702-1721 (2015).

26 Deprez, P. *Clostridium perfringens* infections - a diagnostic challenge. *Vet. Rec.* **177**, 388-389, doi:10.1136/vr.h5428 (2015).

27 Milton, A. A. P. *et al.* Prevalence and molecular typing of *Clostridium perfringens* in captive wildlife in India. *Anaerobe* **44**, 55-57, doi:10.1016/j.anaerobe.2017.01.011 (2017).

28 Myers, G. S. *et al.* Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. *Genome Res.* **16**, 1031-1040, doi:10.1101/gr.5238106 (2006).

29 Knight, D. R., Squire, M. M., Collins, D. A. & Riley, T. V. Genome Analysis of *Clostridium difficile* PCR Ribotype 014 Lineage in Australian Pigs and Humans Reveals a Diverse Genetic Repertoire and Signatures of Long-Range Interspecies Transmission. *Front Microbiol* **7**, 2138, doi:10.3389/fmicb.2016.02138 (2016).

30 Kumar, N. *et al.* Genome-Based Infection Tracking Reveals Dynamics of *Clostridium difficile* Transmission and Disease Recurrence. *Clin. Infect. Dis.*, civ1031, doi:10.1093/cid/civ1031 (2015).

31 Dingle, K. E. *et al.* Effects of control interventions on *Clostridium difficile* infection in England: an observational study. *Lancet Infect. Dis.* **17**, 411-421, doi:10.1016/S1473-3099(16)30514-X (2017).

32 He, M. *et al.* Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat. Genet.* **45**, 109-113, doi:10.1038/ng.2478 (2013).

33 Burnham, C. A. & Carroll, K. C. Diagnosis of *Clostridium difficile* infection: an ongoing conundrum for clinicians and for clinical laboratories. *Clin. Microbiol. Rev.* **26**, 604-630, doi:10.1128/CMR.00016-13 (2013).

34 Collins, J. *et al.* Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature*, doi:10.1038/nature25178 (2018).

35 Chewapreecha, C. *et al.* Global and regional dissemination and evolution of *Burkholderia pseudomallei*. *Nat Microbiol* **2**, 16263, doi:10.1038/nmicrobiol.2016.263 (2017).

36 Raven, K. E. *et al.* A decade of genomic history for healthcare-associated *Enterococcus faecium* in the United Kingdom and Ireland. *Genome Res.* **26**, 1388-1396, doi:10.1101/gr.204024.116 (2016).

37 Shimizu, T. *et al.* Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc Natl Acad Sci U S A* **99**, 996-1001, doi:10.1073/pnas.022493799 (2002).

38 Mollby, R., Holme, T., Nord, C. E., Smyth, C. J. & Wadstrom, T. Production of phospholipase C (alpha-toxin), haemolysins and lethal toxins by *Clostridium perfringens* types A to D. *J Gen Microbiol* **96**, 137-144, doi:10.1099/00221287-96-1-137 (1976).

39 Hobbs, B. C., Smith, M. E., Oakley, C. L., Warrack, G. H. & Cruickshank, J. C. *Clostridium welchii* food poisoning. *J Hyg (Lond)* **51**, 75-101 (1953).

40 Gohari, I. M. *et al.* NetF-producing *Clostridium perfringens*: Clonality and plasmid pathogenicity loci analysis. *Infect Genet Evol* **49**, 32-38, doi:10.1016/j.meegid.2016.12.028 (2017).

41 Freedman, J. C. *et al. Clostridium perfringens* type A-E toxin plasmids. *Res. Microbiol.* **166**, 264-279, doi:10.1016/j.resmic.2014.09.004 (2015).

42 Li, J. *et al.* Toxin plasmids of *Clostridium perfringens*. *Microbiol. Mol. Biol. Rev.* **77**, 208-233, doi:10.1128/MMBR.00062-12 (2013).

43 Wisniewski, J. A. & Rood, J. I. The Tcp conjugation system of *Clostridium perfringens*. *Plasmid* **91**, 28-36, doi:10.1016/j.plasmid.2017.03.001 (2017).

44 Grant, K. A. *et al.* The identification and characterization of *Clostridium perfringens* by real-time PCR, location of enterotoxin gene, and heat resistance. *Foodborne Pathog Dis* **5**, 629-639, doi:10.1089/fpd.2007.0066 (2008).

45 Miyamoto, K. *et al.* Complete sequencing and diversity analysis of the enterotoxin-encoding plasmids in *Clostridium perfringens* type A non-food-borne human gastrointestinal disease isolates. *J. Bacteriol.* **188**, 1585-1598, doi:10.1128/JB.188.4.1585-1598.2006 (2006).

46 Sayeed, S., Li, J. & McClane, B. A. Characterization of virulence plasmid diversity among *Clostridium perfringens* type B isolates. *Infect. Immun.* **78**, 495-504, doi:10.1128/IAI.00838-09 (2010).

47 Sayeed, S., Li, J. & McClane, B. A. Virulence plasmid diversity in *Clostridium perfringens* type D isolates. *Infect. Immun.* **75**, 2391-2398, doi:10.1128/IAI.02014-06 (2007).

48 Gurjar, A., Li, J. & McClane, B. A. Characterization of toxin plasmids in *Clostridium perfringens* type C isolates. *Infect. Immun.* **78**, 4860-4869, doi:10.1128/IAI.00715-10 (2010).

49 Li, J., Miyamoto, K. & McClane, B. A. Comparison of virulence plasmids among *Clostridium perfringens* type E isolates. *Infect. Immun.* **75**, 1811-1819, doi:10.1128/IAI.01981-06 (2007).

50 Miyamoto, K. *et al.* Identification of novel *Clostridium perfringens* type E strains that carry an iota toxin plasmid with a functional enterotoxin gene. *PLoS One* **6**, e20376, doi:10.1371/journal.pone.0020376 (2011).

51 Bannam, T. L. *et al.* Necrotic enteritis-derived *Clostridium perfringens* strain with three closely related independently conjugative toxin and antibiotic resistance plasmids. *MBio* **2**, doi:10.1128/mBio.00190-11 (2011).

52 Ebersbach, G. & Gerdes, K. Plasmid segregation mechanisms. *Annu. Rev. Genet.* **39**, 453-479, doi:10.1146/annurev.genet.38.072902.091252 (2005).

53 Watts, T. D., Johanesen, P. A., Lyras, D., Rood, J. I. & Adams, V. Evidence that compatibility of closely related replicons in *Clostridium perfringens* depends on linkage to parMRC-like partitioning systems of different subfamilies. *Plasmid* **91**, 68-75, doi:10.1016/j.plasmid.2017.03.008 (2017).

54 McClane, B. A., Uzal, F. A., Miyakawa, M. E. F., Lyerly, D. & Wilkins, T. in *The Prokaryotes: A handbook on the biology of bacteria* Vol. 4 (eds M. Dworkin *et al.*) 698-752 (Springer, 2006).

55 Briolat, V. & Reysset, G. Identification of the *Clostridium perfringens* genes involved in the adaptive response to oxidative stress. *J. Bacteriol.* **184**, 2333-2343 (2002).

56 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).

57 Machida, Y. *et al.* [An outbreak of enterocolitis due to *Clostridium perfringens* in a hospital for the severely disabled]. *Kansenshogaku Zasshi* **63**, 410-416 (1989).

58 Li, J., Paredes-Sabja, D., Sarker, M. R. & McClane, B. A. *Clostridium perfringens* Sporulation and Sporulation-Associated Toxin Production. *Microbiol Spectr* **4**, doi:10.1128/microbiolspec.TBS-0022-2015 (2016).

59 Orsburn, B., Melville, S. B. & Popham, D. L. Factors contributing to heat resistance of *Clostridium perfringens* endospores. *Appl. Environ. Microbiol.* **74**, 3328-3335, doi:10.1128/AEM.02629-07 (2008).

60 Li, J. & McClane, B. A. A novel small acid soluble protein variant is important for spore resistance of most *Clostridium perfringens* food poisoning isolates. *PLoS Pathog* **4**, e1000056, doi:10.1371/journal.ppat.1000056 (2008).

61 Duncan, C. L. & Strong, D. H. Improved medium for sporulation of *Clostridium perfringens*. *Appl Microbiol* **16**, 82-89 (1968).

62 Varga, J., Stirewalt, V. L. & Melville, S. B. The CcpA protein is necessary for efficient sporulation and enterotoxin gene (cpe) regulation in *Clostridium perfringens*. *J. Bacteriol.* **186**, 5221-5229, doi:10.1128/JB.186.16.5221-5229.2004 (2004).

63 Li, J. & McClane, B. A. Evaluating the involvement of alternative sigma factors SigF and SigG in *Clostridium perfringens* sporulation and enterotoxin synthesis. *Infect Immun* **78**, 4286-4293, doi:10.1128/IAI.00528-10 (2010).

64 Harry, K. H., Zhou, R., Kroos, L. & Melville, S. B. Sporulation and enterotoxin (CPE) synthesis are controlled by the sporulation-specific sigma factors SigE and SigK in *Clostridium perfringens*. *J. Bacteriol.* **191**, 2728-2742, doi:10.1128/JB.01839-08 (2009).

65 Li, J., Chen, J., Vidal, J. E. & McClane, B. A. The Agr-like quorum-sensing system regulates sporulation and production of enterotoxin and beta2 toxin by *Clostridium perfringens* type A non-food-borne human gastrointestinal disease strain F5603. *Infect. Immun.* **79**, 2451-2459, doi:10.1128/IAI.00169-11 (2011).

66 Li, J., Freedman, J. C., Evans, D. R. & McClane, B. A. CodY Promotes Sporulation and Enterotoxin Production by *Clostridium perfringens* Type A Strain SM101. *Infect. Immun.* **85**, doi:10.1128/IAI.00855-16 (2017).

67 Setlow, P. Spore germination. *Curr. Opin. Microbiol.* **6**, 550-556 (2003).

68 Sorg, J. A. & Sonenshein, A. L. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J. Bacteriol.* **190**, 2505-2512, doi:10.1128/JB.01765-07 (2008).

69 Modi, N. & Wilcox, M. H. Evidence for antibiotic induced *Clostridium perfringens* diarrhoea. *J. Clin. Pathol.* **54**, 748-751 (2001).

70 Li, J. & McClane, B. A. Further comparison of temperature effects on growth and survival of *Clostridium perfringens* type A isolates carrying a chromosomal or plasmid-borne enterotoxin gene. *Appl. Environ. Microbiol.* **72**, 4561-4568, doi:10.1128/AEM.00177-06 (2006).

71 Fuchs, A. & Bonde, G. J. The availability of sulphur for *Clostridium perfringens* and an examination of hydrogen sulphide production. *J gen Microbiol* **16**, 330-340 (1957).

72 Rowan, F. E., Docherty, N. G., Coffey, J. C. & O'Connell, P. R. Sulphate-reducing bacteria and hydrogen sulphide in the aetiology of ulcerative colitis. *Br J Surg* **96**, 151-158, doi:10.1002/bjs.6454 (2009).

73 Neu, J. & Walker, A. W. Necrotizing enterocolitis. *The New England Journal of Medicine* **364**, 255-264 (2011).

74 Pedersen, P. V., Hansen, F. H., Halveg, A. B. & Christiansen, E. D. Necrotising enterocolitis of the newborn--is it gas-gangrene of the bowel? *Lancet* **2**, 715-716 (1976).

75 Macfarlane, M. G. & Knight , B. C. The biochemistry of bacterial toxins: The lecithinase activity of Cl. *welchii* toxins. *Biochem J.* **35**, 884-902 (1941).

76 Titball, R. W., Naylor, C. E. & Basak, A. K. The *Clostridium perfringens* alpha-toxin. *Anaerobe* **5**, 51-64, doi:10.1006/anae.1999.0191 (1999).

77 Awad, M. M., Bryant, A. E., Stevens, D. L. & Rood, J. I. Virulence studies on chromosomal alpha-toxin and theta-toxin mutants constructed by allelic exchange provide genetic evidence for the essential role of alpha-toxin in *Clostridium perfringens*-mediated gas gangrene. *Mol. Microbiol.* **15**, 191-202 (1995).

78 Takehara, M. *et al. Clostridium perfringens* alpha-Toxin Impairs Innate Immunity via Inhibition of Neutrophil Differentiation. *Sci Rep* **6**, 28192, doi:10.1038/srep28192 (2016).

79 Timoney, J. F., Gillespie, J. H., Scott, F. W. & Barlough, J. E. *Hagan and Bruner's microbiology and infectious diseases of domestic animals*. (Comstock Publishing Associates, 1988).

80 Uzal, F. A. *et al.* Development and application of new mouse models to study the pathogenesis of *Clostridium perfringens* type C Enterotoxemias. *Infect. Immun.* **77**, 5291-5299, doi:10.1128/IAI.00825-09 (2009).

81 Murrel, T. G. C. Pigbel in Papua New Guinea: An ancient disease rediscovered. *Int. J. Epidemiol.* **12**, 211-214 (1983).

82 Ma, M. *et al.* Synergistic effects of *Clostridium perfringens* enterotoxin and beta toxin in rabbit small intestinal loops. *Infect. Immun.* **82**, 2958-2970, doi:10.1128/IAI.01848-14 (2014).

83 Hunter, S. E., Brown, J. E., Oyston, P. C., Sakurai, J. & Titball, R. W. Molecular genetic analysis of beta-toxin of *Clostridium perfringens* reveals sequence homology with alpha-toxin, gamma-toxin, and leukocidin of Staphylococcus aureus. *Infect. Immun.* **61**, 3958-3965 (1993).

84 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).

85 Goossens, E. *et al.* Rethinking the role of alpha toxin in *Clostridium perfringens*-associated enteric diseases: a review on bovine necro-haemorrhagic enteritis. *Vet. Res.* **48**, 9, doi:10.1186/s13567-017-0413-x (2017).

86 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, doi:10.1084/jem.20041215 (2004).

87 Verherstraeten, S. *et al.* The synergistics necrohemorrhagic action of *Clostridium perfringens* perfringolysin and alpha toxin in the bovine intestine and against bovine endothelial cells. *Vet. Res.* **44**, 45 (2013).

88 Gibert, M., Jolivet-Reynaud, C. & Popoff, M. R. Beta2 toxin, a novel toxin produced by *Clostridium perfringens*. *Gene* **203**, 65-73 (1997).

89 Fisher, D. J. *et al.* Association of beta2 toxin production with *Clostridium perfringens* type A human gastrointestinal disease isolates carrying a plasmid enterotoxin gene. *Mol. Microbiol.* **56**, 747-762, doi:10.1111/j.1365-2958.2005.04573.x (2005).

90 Vilei, E. M. *et al.* Antibiotic-induced expression of a cryptic cpb2 gene in equine beta2-toxigenic *Clostridium perfringens*. *Mol. Microbiol.* **57**, 1570-1581, doi:10.1111/j.1365-2958.2005.04789.x (2005).

91 Theoret, J. R., Uzal, F. A. & McClane, B. A. Identification and characterization of *Clostridium perfringens* beta toxin variants with differing trypsin sensitivity and in vitro cytotoxicity activity. *Infect. Immun.* **83**, 1477-1486, doi:10.1128/IAI.02864-14 (2015).

92 Nagahama, M. *et al.* Cellular vacuolation induced by *Clostridium perfringens* epsilon-toxin. *FEBS J.* **278**, 3395-3407, doi:10.1111/j.1742-4658.2011.08263.x (2011).

93 Fernandes da Costa, S. P. *et al.* Identification of a key residue for oligomerisation and pore-formation of *Clostridium perfringens* NetB. *Toxins (Basel)* **6**, 1049-1061, doi:10.3390/toxins6031049 (2014).

94 Jin, F. *et al.* Purification, characterization, and primary structure of *Clostridium perfringens* lambda-toxin, a thermolysin-like metalloprotease. *Infect. Immun.* **64**, 230-237 (1996).

95 Ohno-Iwashita, Y., Iwamoto, M., Mitsui, K., Kawasaki, H. & Ando, S. Cold-labile hemolysin produced by limited proteolysis of theta-toxin from *Clostridium perfringens*. *Biochemistry* **25**, 6048-6053 (1986).

96 Manich, M. *et al. Clostridium perfringens* delta toxin is sequence related to beta toxin, NetB, and Staphylococcus pore-forming toxins, but shows functional differences. *PLoS One* **3**, e3764, doi:10.1371/journal.pone.0003764 (2008).

97 Matsushita, O., Yoshihara, K., Katayama, S., Minami, J. & Okabe, A. Purification and characterization of *Clostridium perfringens* 120-kilodalton collagenase and nucleotide sequence of the corresponding gene. *J. Bacteriol.* **176**, 149-156 (1994).

98 Roggentin, P., Rothe, B., Lottspeich, F. & Schauer, R. Cloning and sequencing of a *Clostridium perfringens* sialidase gene. *FEBS Lett.* **238**, 31-34 (1988).

99 Jiang, Y. F., Kulkarni, R. R., Parreira, V. R. & Prescott, J. F. Immunization of Broiler Chickens Against *Clostridium perfringens*-Induced Necrotic Enteritis Using Purified Recombinant Immunogenic Proteins. *Avian Dis.* **53**, 409-415 (2009).

100 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, doi:10.1128/IAI.01759-14 (2014).

101 Mehdizadeh Gohari, I. *et al.* A novel pore-forming toxin in type A *Clostridium perfringens* is associated with both fatal canine hemorrhagic gastroenteritis and fatal foal necrotizing enterocolitis. *PLoS One* **10**, e0122684, doi:10.1371/journal.pone.0122684 (2015).

102 Gohari, I. M. *et al.* Plasmid Characterization and Chromosome Analysis of Two netF+ *Clostridium perfringens* Isolates Associated with Foal and Canine Necrotizing Enteritis. *PLoS One* **11**, e0148344, doi:10.1371/journal.pone.0148344 (2016).

103 Popoff, M. R. Epsilon toxin: a fascinating pore-forming toxin. *FEBS J.* **278**, 4602-4615, doi:10.1111/j.1742-4658.2011.08145.x (2011).

104 Tamai, E. *et al.* Accumulation of *Clostridium perfringens* epsilon-toxin in the mouse kidney and its possible biological significance. *Infect. Immun.* **71**, 5371-5375 (2003).

105 U.S. Bureau of Industry and Security. (ed U.S. Bureau of Industry and Security) (2017).

106 UK Home Office & Department for International Trade. (ed Department for International Trade) (Crown, 2017).

107 Hilger, H. *et al.* The long-lived nature of *Clostridium perfringens* iota toxin in mammalian cells induces delayed apoptosis. *Infect. Immun.* **77**, 5593-5601, doi:10.1128/IAI.00710-09 (2009).

108 Stiles, B. G., Wigelsworth, D. J., Popoff, M. R. & Barth, H. Clostridial binary toxins: iota and C2 family portraits. *Front Cell Infect Microbiol* **1**, 11, doi:10.3389/fcimb.2011.00011 (2011).

109 Stiles, B. G., Barth, G., Barth, H. & Popoff, M. R. *Clostridium perfringens* epsilon toxin: a malevolent molecule for animals and man? *Toxins (Basel)* **5**, 2138-2160, doi:10.3390/toxins5112138 (2013).

110 Obana, N., Nomura, N. & Nakamura, K. Structural requirement in *Clostridium perfringens* collagenase mRNA 5' leader sequence for translational induction through small RNA-mRNA base pairing. *J. Bacteriol.* **195**, 2937-2946, doi:10.1128/JB.00148-13 (2013).

111 Awad, M. M. *et al.* Construction and virulence testing of a collagenase mutant of *Clostridium perfringens*. *Microb. Pathog.* **28**, 107-117, doi:10.1006/mpat.1999.0328 (2000).

112 Shinoda, T. *et al.* Structural basis for disruption of claudin assembly in tight junctions by an enterotoxin. *Sci Rep* **6**, 33632, doi:10.1038/srep33632 (2016).

113 Eichner, M. *et al.* In Colon Epithelia, *Clostridium perfringens* Enterotoxin Causes Focal Leaks by Targeting Claudins Which are Apically Accessible Due to Tight Junction Derangement. *J. Infect. Dis.* **217**, 147-157, doi:10.1093/infdis/jix485 (2017).

114 Chakrabarti, G. & McClane, B. A. The importance of calcium influx, calpain and calmodulin for the activation of CaCo-2 cell death pathways by *Clostridium perfringens* enterotoxin. *Cell. Microbiol.* **7**, 129-146, doi:10.1111/j.1462-5822.2004.00442.x (2005).

115 Llanco, L. A., Nakano, V. & Avila-Campos, M. J. Sialidase production and genetic diversity in *Clostridium perfringens* type A isolated from chicken with necrotic enteritis in Brazil. *Curr. Microbiol.* **70**, 330-337, doi:10.1007/s00284-014-0722-5 (2015).

116 Li, J. & McClane, B. A. NanI Sialidase Can Support the Growth and Survival of *Clostridium perfringens* Strain F4969 Using Sialyated Host Macromolecules (Mucin) or Caco-2 Cells. *Infect. Immun.*, doi:10.1128/IAI.00547-17 (2017).

117 Flores-Diaz, M. *et al.* A cellular deficiency of gangliosides causes hypersensitivity to *Clostridium perfringens* phospholipase C. *J. Biol. Chem.* **280**, 26680-26689, doi:10.1074/jbc.M500278200 (2005).

118 Theoret, J. R. *et al.* Native or Proteolytically Activated NanI Sialidase Enhances the Binding and Cytotoxic Activity of *Clostridium perfringens* Enterotoxin and Beta Toxin. *Infect. Immun.* **86**, doi:10.1128/IAI.00730-17 (2018).

119 Chiarezza, M. *et al.* The NanI and NanJ sialidases of *Clostridium perfringens* are not essential for virulence. *Infect. Immun.* **77**, 4421-4428, doi:10.1128/IAI.00548-09 (2009).

120 Li, J., Uzal, F. A. & McClane, B. A. *Clostridium perfringens* Sialidases: Potential Contributors to Intestinal Pathogenesis and Therapeutic Targets. *Toxins (Basel)* **8**, doi:10.3390/toxins8110341 (2016).

121 Keyburn, A. L., Bannam, T. L., Moore, R. J. & Rood, J. I. NetB, a pore-forming toxin from necrotic enteritis strains of *Clostridium perfringens*. *Toxins (Basel)* **2**, 1913-1927, doi:10.3390/toxins2071913 (2010).

122 Keyburn, A. L. *et al.* NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. *PLoS Pathog* **4**, e26, doi:10.1371/journal.ppat.0040026 (2008).

123 Abildgaard, L., Sondergaard, T. E., Engberg, R. M., Schramm, A. & Hojberg, O. In vitro production of necrotic enteritis toxin B, NetB, by netB-positive and netB-negative *Clostridium perfringens* originating from healthy and diseased broiler chickens. *Vet. Microbiol.* **144**, 231-235, doi:10.1016/j.vetmic.2009.12.036 (2010).

124 Lepp, D. *et al.* Identification of novel pathogenicity loci in *Clostridium perfringens* strains that cause avian necrotic enteritis. *PLoS One* **5**, e10795, doi:10.1371/journal.pone.0010795 (2010).

125 Rood, J. I., Keyburn, A. L. & Moore, R. J. NetB and necrotic enteritis: the hole movable story. *Avian Pathol.* **45**, 295-301, doi:10.1080/03079457.2016.1158781 (2016).

126 Johnstone, F. R. & Cockcroft, W. H. *Clostridium welchii* resistance to tetracycline. *Lancet* **1**, 660-661 (1968).

127 Rood, J. I., Buddle, J. R., Wales, A. J. & Sidhu, R. The occurrence of antibiotic resistance in *Clostridium perfringens* from pigs. *Aust. Vet. J.* **62**, 276-279 (1985).

128 Lyras, D. & Rood, J. I. Genetic organization and distribution of tetracycline resistance determinants in *Clostridium perfringens*. *Antimicrob. Agents Chemother.* **40**, 2500-2504 (1996).

129 Rood, J. I., Maher, E. A., Somers, E. B., Campos, E. & Duncan, C. L. Isolation and characterization of multiply antibiotic-resistant *Clostridum perfringens* strains from porcine feces. *Antimicrob. Agents Chemother.* **13**, 871-880 (1978).

130 Soge, O. O., Tivoli, L. D., Meschke, J. S. & Roberts, M. C. A conjugative macrolide resistance gene, mef(A), in environmental *Clostridium perfringens* carrying multiple macrolide and/or tetracycline resistance genes. *J. Appl. Microbiol.* **106**, 34-40, doi:10.1111/j.1365-2672.2008.03960.x (2009).

131 Osman, K. M. & Elhariri, M. Antibiotic resistance of *Clostridium perfringens* isolates from broiler chickens in Egypt. *Rev Sci Tech* **32**, 841-850 (2013).

132 Kouassi, K. A., Dadie, A. T., N'Guessan, K. F., Dje, K. M. & Loukou, Y. G. *Clostridium perfringens* and *Clostridium difficile* in cooked beef sold in Cote d'Ivoire and their antimicrobial susceptibility. *Anaerobe* **28**, 90-94, doi:10.1016/j.anaerobe.2014.05.012 (2014).

133 Akhi, M. T. *et al.* Antibiotic Sensitivity of *Clostridium perfringens* Isolated From Faeces in Tabriz, Iran. *Jundishapur J Microbiol* **8**, e20863, doi:10.5812/jjm.20863v2 (2015).

134 Yadav, J. P. *et al.* Molecular characterization and antimicrobial resistance profile of *Clostridium perfringens* type A isolates from humans, animals, fish and their environment. *Anaerobe* **47**, 120-124, doi:10.1016/j.anaerobe.2017.05.009 (2017).

135 Park, J. Y. *et al.* Characterization of *Clostridium perfringens* isolates obtained from 2010 to 2012 from chickens with necrotic enteritis in Korea. *Poultry Science* **94**, 1158-1164, doi:10.3382/ps/pev037 (2015).

136 Tyrrell, K. L. *et al.* In vitro activities of daptomycin, vancomycin, and penicillin against *Clostridium difficile*, *C. perfringens*, *Finegoldia magna*, and *Propionibacterium acnes*. *Antimicrobial Agents and Chemotherapy* **50**, 2728-2731, doi:10.1128/Aac.00357-06 (2006).

137 Slavic, D. *et al.* Antimicrobial susceptibility of *Clostridium perfringens* isolates of bovine, chicken, porcine, and turkey origin from Ontario. *Can J Vet Res* **75**, 89-97 (2011).

138 Li, C., Yan, X. & Lillehoj, H. S. Complete Genome Sequence of *Clostridium perfringens* LLY_N11, a Necrotic Enteritis-Inducing Strain Isolated from a Healthy Chicken Intestine. *Genome Announc* **5**, doi:10.1128/genomeA.01225-17 (2017).

139 Ngamwongsatit, B. *et al.* Multidrug resistance in *Clostridium perfringens* isolated from diarrheal neonatal piglets in Thailand. *Anaerobe* **38**, 88-93, doi:10.1016/j.anaerobe.2015.12.012 (2016).

140 Bryan, L. E., Kowand, S. K. & Van Den Elzen, H. M. Mechanism of aminoglycoside antibiotic resistance in anaerobic bacteria: *Clostridium perfringens* and *Bacteroides fragilis*. *Antimicrob. Agents Chemother.* **15**, 7-13 (1979).

141 Udhayavel, S., Thippichettypalayam Ramasamy, G., Gowthaman, V., Malmarugan, S. & Senthilvel, K. Occurrence of *Clostridium perfringens* contamination in poultry feed ingredients: Isolation, identification and its antibiotic sensitivity pattern. *Anim Nutr* **3**, 309-312, doi:10.1016/j.aninu.2017.05.006 (2017).

142 Parish, W. E. Necrotic enteritis in the fowl (Gallus gallus domesticus). I. Histopathology of the disease and isolation of a strain of *Clostridium welchii*. *J Comp Pathol* **71**, 377-393 (1961).

143 Zahoor, I., Ghayas, A. & Basheer, A. Genetics and genomics of susceptibility and immune response to necrotic enteritis in chicken: a review. *Mol. Biol. Rep.*, doi:10.1007/s11033-017-4138-8 (2017).

144 Wade, B. & Keybun, A. The true cost of necrotic enteritis. *Poultry World* (2015). <https://www.poultryworld.net/Meat/Articles/2015/10/The-true-cost-of-necrotic-enteritis-2699819W/>.

145 Timbermont, L., Haesebrouck, F., Ducatelle, R. & Van Immerseel, F. Necrotic enteritis in broilers: an updated review on the pathogenesis. *Avian Pathol.* **40**, 341-347, doi:10.1080/03079457.2011.590967 (2011).

146 Olkowski, A. A., Wojnarowicz, C., Chirino-Trejo, M., Laarveld, B. & Sawicki, G. Sub-clinical necrotic enteritis in broiler chickens: novel etiological consideration based on ultra-structural and molecular changes in the intestinal tissue. *Res. Vet. Sci.* **85**, 543-553, doi:10.1016/j.rvsc.2008.02.007 (2008).

147 Coursodon, C. F., Glock, R. D., Moore, K. L., Cooper, K. K. & Songer, J. G. TpeL-producing strains of *Clostridium perfringens* type A are highly virulent for broiler chicks. *Anaerobe* **18**, 117-121, doi:10.1016/j.anaerobe.2011.10.001 (2012).

148 Drew, M. D., Syed, N. A., Goldade, B. G., Laarveld, B. & Van Kessel, A. G. Effects of dietary protein source and level on intestinal populations of *Clostridium perfringens* in broiler chickens. *Poult Sci* **83**, 414-420 (2004).

149 Nowell, V. J. *et al.* Genome sequencing and analysis of a type A *Clostridium perfringens* isolate from a case of bovine clostridial abomasitis. *PLoS One* **7**, e32271, doi:10.1371/journal.pone.0032271 (2012).

150 Nakano, V. *et al.* Multilocus sequence typing analyses of *Clostridium perfringens* type A strains harboring tpeL and netB genes. *Anaerobe* **44**, 99-105, doi:10.1016/j.anaerobe.2017.02.017 (2017).

151 Chan, G. *et al.* The epidemiology of *Clostridium perfringens* type A on Ontario swine farms, with special reference to cpb2-positive isolates. *BMC Vet Res* **8**, 156, doi:10.1186/1746-6148-8-156 (2012).

152 Johansson, A. *et al.* Genetic diversity of *Clostridium perfringens* type A isolates from animals, food poisoning outbreaks and sludge. *BMC Microbiol.* **6**, 47, doi:10.1186/1471-2180-6-47 (2006).

153 Larson, H. E. & Borriello, S. P. Infectious diarrhea due to *Clostridium perfringens*. *J. Infect. Dis.* **157**, 390-391 (1988).

154 Gohari, I. M. *et al.* Characterization of *Clostridium perfringens* in the feces of adult horses and foals with acute enterocolitis. *Can J Vet Res* **78**, 1-7 (2014).

155 Diab, S. S. *et al.* Pathology of *Clostridium perfringens* type C enterotoxemia in horses. *Vet. Pathol.* **49**, 255-263, doi:10.1177/0300985811404710 (2012).

156 Songer, J. G. & Uzal, F. A. Clostridial enteric infections in pigs. *J Vet Diagn Invest* **17**, 528-536 (2005).

157 Niilo, L. *Clostridium perfringens* Type C Enterotoxemia. *Can. Vet. J.* **29**, 658-664 (1988).

158 Hansen, K. *et al. Darmbrand - Enteritis Necroticans*. (George Thieme, 1949).

159 Sterne, M. & Warrack, G. H. The Types of *Clostridium perfringens*. *J Pathol Bacteriol* **88**, 279-283 (1964).

160 Cooke, R. The pathology of pig bel. *P N G Med J* **22**, 35-38 (1979).

161 Murrell, T. G., Roth, L., Egerton, J., Samels, J. & Walker, P. D. Pig-bel: enteritis necroticans. A study in diagnosis and management. *Lancet* **1**, 217-222 (1966).

162 Murrell, T. G., Egerton, J. R., Rampling, A., Samels, J. & Walker, P. D. The ecology and epidemiology of the pig-bel syndrome in man in New Guinea. *J Hyg (Lond)* **64**, 375-396 (1966).

163 Sakurai, J. & Duncan, C. L. Some properties of beta-toxin produced by *Clostridium perfringens* type C. *Infect. Immun.* **21**, 678-680 (1978).

164 Foster, W. D. The bacteriology of necrotizing jejunitis in Uganda. *East Afr. Med. J.* **43**, 550-553 (1966).

165 Gan, K. H., Sukono, D., Satari, S., Sujudi, R. W. & Njooo, S. N. First outbreak of necrotising enteritis caused by *Clostridium welchii* in Indonesia. *Nadj Kedkt* **1**, 802-806 (1967).

166 Headington, J. T., Sathornsumathi, S., Simark, S. & Sujatanond, W. Segmental infarcts of the small intestine and mesenteric adenitis in Thai children. *Lancet* **1**, 802-806 (1967).

167 Murrell, T. G. Enteritis necroticans in Nepal. *Lancet* **1**, 279 (1979).

168 Gui, L., Subramony, C., Fratkin, J. & Hughson, M. D. Fatal enteritis necroticans (pigbel) in a diabetic adult. *Mod Pathol* **15**, 66-70, doi:10.1038/modpathol.3880491 (2002).

169 Know, R. & MacDonald, E. Outbreaks of foodpoisoning in certain Leicester institutions. *Med Offr* **69**, 21-22 (1943).

170 Scallan, E. *et al.* Foodborne illness acquired in the United States--major pathogens. *Emerg Infect Dis* **17**, 7-15, doi:10.3201/eid1701.P11101 (2011).

171 Da Silva Felicio, M. T. *et al.* Risk ranking of pathogens in ready-to-eat unprocessed foods of non-animal origin (FoNAO) in the EU: initial evaluation using outbreak data (2007-2011). *Int. J. Food Microbiol.* **195**, 9-19, doi:10.1016/j.ijfoodmicro.2014.11.005 (2015).

172 DuPont, H. L. Clinical practice. Bacterial diarrhea. *N Engl J Med* **361**, 1560-1569, doi:10.1056/NEJMcp0904162 (2009).

173 Collie, R. E. & McClane, B. A. Evidence that the enterotoxin gene can be episomal in *Clostridium perfringens* isolates associated with non-food-borne human gastrointestinal diseases. *J. Clin. Microbiol.* **36**, 30-36 (1998).

174 Borriello, S. P. *et al.* Enterotoxigenic *Clostridium perfringens*: a possible cause of antibiotic-associated diarrhoea. *Lancet* **1**, 305-307 (1984).

175 Hogenauer, C., Hammer, H. F., Krejs, G. J. & Reisinger, E. C. Mechanisms and management of antibiotic-associated diarrhea. *Clin. Infect. Dis.* **27**, 702-710 (1998).

176 Brett, M. M., Rodhouse, J. C., Donovan, T. J., Tebbutt, G. M. & Hutchinson, D. N. Detection of *Clostridium perfringens* and its enterotoxin in cases of sporadic diarrhoea. *J. Clin. Pathol.* **45**, 609-611 (1992).

177 Mpamugo, O., Donovan, T. & Brett, M. M. Enterotoxigenic *Clostridium perfringens* as a cause of sporadic cases of diarrhoea. *J. Med. Microbiol.* **43**, 442-445, doi:10.1099/00222615-43-6-442 (1995).

178 Wong, S. *et al.* Use of antibiotic and prevalence of antibiotic-associated diarrhoea in-patients with spinal cord injuries: a UK national spinal injury centre experience. *Spinal Cord* **55**, 583-587, doi:10.1038/sc.2016.193 (2017).

179 Ayyagari, A., Agarwal, J. & Garg, A. Antibiotic associated diarrhoea: infectious causes. *Indian J Med Microbiol* **21**, 6-11 (2003).

180 Asha, N. J. & Wilcox, M. H. Laboratory diagnosis of *Clostridium perfringens* antibiotic-associated diarrhoea. *J. Med. Microbiol.* **51**, 891-894, doi:10.1099/0022-1317-51-10-891 (2002).

181 Asha, N. J., Tompkins, D. & Wilcox, M. H. Comparative analysis of prevalence, risk factors, and molecular epidemiology of antibiotic-associated diarrhea due to *Clostridium difficile*, *Clostridium perfringens*, and *Staphylococcus aureus*. *J. Clin. Microbiol.* **44**, 2785-2791, doi:10.1128/JCM.00165-06 (2006).

182 Li, J. & McClane, B. A. Contributions of NanI sialidase to Caco-2 cell adherence by *Clostridium perfringens* type A and C strains causing human intestinal disease. *Infect. Immun.* **82**, 4620-4630, doi:10.1128/IAI.02322-14 (2014).

183 Larcombe, S., Hutton, M. L. & Lyras, D. Involvement of Bacteria Other Than *Clostridium difficile* in Antibiotic-Associated Diarrhoea. *Trends Microbiol.* **24**, 463-476, doi:10.1016/j.tim.2016.02.001 (2016).

184 Pituch, H. *et al.* Laboratory diagnosis of antibiotic-associated diarrhea: a Polish pilot study into the clinical relevance of *Clostridium difficile* and *Clostridium perfringens* toxins. *Diagn. Microbiol. Infect. Dis.* **58**, 71-75, doi:10.1016/j.diagmicrobio.2006.12.007 (2007).

185 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, doi:10.1016/j.jpedsurg.2009.12.002 (2010).

186 Lim, J. C., Golden, J. M. & Ford, H. R. Pathogenesis of neonatal necrotizing enterocolitis. *Pediatr Surg Int* **31**, 509-518 (2015).

187 Bisquera, J. A., Cooper, T. R. & Berseth, C. L. Impact of necrotizing enterocolitis on length of stay and hospital charges in very low birth weight infants. *Pediatrics* **109**, 423-428 (2002).

188 Berdon, W. E. *et al.* Necrotizing Enterocolitis in the Premature Infant. *Radiology* **83**, 879-887, doi:10.1148/83.5.879 (1964).

189 Butel, M. J. *et al.* Clostridial pathogenicity in experimental necrotising enterocolitis in gnotobiotic quails and protective role of bifidobacteria. *Journal of Medical Microbiology* **47**, 391-399 (1998).

190 Hanke, C. A. *et al. Clostridium perfringens* intestinal gas gangrene in a preterm newborn. *Eur J Pediatr Surg* **19**, 257-259, doi:10.1055/s-2008-1038958 (2009).

191 Kosloske, A. M., Ulrich, J. A. & Hoffman, H. Fulminant necrotising enterocolitis associated with clostridia. *Lancet* **312**, 1014-1016 (1978).

192 Yu, V. Y., Joseph, R., Bajuk, B., Orgill, A. & Astbury, J. Necrotizing enterocolitis in very low birthweight infants: a four-year experience. *Aust Paediatr J* **20**, 29-33 (1984).

193 Warren, S., Schreiber, J. R. & Epstein, M. F. Necrotizing enterocolitis and hemolysis associated with *Clostridium perfringens*. *American journal of diseases of children (1960)* **138**, 686-688 (1984).

194 Blakey, J. L. *et al.* Development of gut colonisation in pre-term neonates. *J. Med. Microbiol.* **15**, 519-529, doi:10.1099/00222615-15-4-519 (1982).

195 Kosloske, A. M., Ball Jr, W. S., Umland, E. & Skipper, B. Clostridial necrotizing enterocolitis. *J. Pediatr. Surg.* **20**, 155-159 (1985).

196 Hoy, C. *et al.* Quantitative changes in faecal microflora preceding necrotising enterocolitis in premature neonates. *Arch. Dis. Child.* **65**, 1057-1059 (1990).

197 de la Cochetiere, M. F. *et al.* Early intestinal bacterial colonization and necrotizing enterocolitis in premature infants: the putative role of *Clostridium*. *Pediatr Res* **56**, 366-370, doi:10.1203/01.PDR.0000134251.45878.D5 (2004).

198 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).

199 Schlapbach, L. J., Ahrens, O., Klimek, P., Berger, S. & Kessler, U. *Clostridium perfringens* and necrotizing enterocolitis. *J Pediatr* **157**, 175, doi:10.1016/j.jpeds.2010.03.027 (2010).

200 Izumiya, H. *et al.* Whole-genome analysis of *Salmonella enterica* serovar Typhimurium T000240 reveals the acquisition of a genomic island involved in multidrug resistance via IS1 derivatives on the chromosome. *Antimicrob. Agents Chemother.* **55**, 623-630, doi:10.1128/AAC.01215-10 (2011).

201 Ramos, P. I. *et al.* Comparative analysis of the complete genome of KPC-2-producing *Klebsiella pneumoniae* Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. *BMC Genomics* **15**, 54, doi:10.1186/1471-2164-15-54 (2014).

202 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, doi:10.1093/gbe/evx210 (2017).

203 Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **173**, 697-703 (1991).

204 Sim, K. *et al.* Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing. *PLoS One* **7**, e32543, doi:10.1371/journal.pone.0032543 (2012).

205 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335-336, doi:10.1038/nmeth.f.303 (2010).

206 Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620, doi:10.1093/bioinformatics/btt593 (2014).

207 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461, doi:10.1093/bioinformatics/btq461 (2010).

208 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).

209 Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol* **12**, e1004957, doi:10.1371/journal.pcbi.1004957 (2016).

210 R Development Core Team. in *R Foundation for Statistical Computing* (Vienna, Austria, 2010).

211 Page, A. J. *et al.* Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb Genom* **2**, e000083, doi:10.1099/mgen.0.000083 (2016).

212 Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821-829, doi:10.1101/gr.074492.107 (2008).

213 Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579, doi:10.1093/bioinformatics/btq683 (2011).
214 Nadalin, F., Vezzi, F. & Policriti, A. GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics* **13 Suppl 14**, S8, doi:10.1186/1471-2105-13-S14-S8 (2012).
215 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).
216 Davis, M. P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. & Enright, A. J. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* **63**, 41-49, doi:10.1016/j.ymeth.2013.06.027 (2013).
217 Kim, M., Oh, H. S., Park, S. C. & Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **64**, 346-351, doi:10.1099/ijs.0.059774-0 (2014).
218 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).
219 Lawson, P. A., Citron, D. M., Tyrrell, K. L. & Finegold, S. M. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prevot 1938. *Anaerobe* **40**, 95-99, doi:10.1016/j.anaerobe.2016.06.008 (2016).
220 Sasi Jyothsna, T. S., Tushar, L., Sasikala, C. & Ramana, C. V. *Paraclostridium benzoelyticum* gen. nov. sp. nov., isolated from marine sediment and reclassification of *Clostridium bifermentans* as *Paraclostridium bifermentans* comb. nov. Proposal of a new genus *Paeniclostridium* gen. nov. to accommodate *Clostridium* sordellii and *Clostridium ghonii*. *Int. J. Syst. Evol. Microbiol.*, doi:10.1099/ijsem.0.000874 (2016).
221 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691-3693, doi:10.1093/bioinformatics/btv421 (2015).
222 Hadfield, J. & Harris, S. *Interactive visualization of genome phylogenies - Phandango*, <https://jameshadfield.github.io/phandango/> (2017).
223 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
224 Page, A. J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* **2**, e000056, doi:10.1099/mgen.0.000056 (2016).
225 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490, doi:10.1371/journal.pone.0009490 (2010).
226 Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242-245, doi:10.1093/nar/gkw290 (2016).
227 Tonkin-Hill, G., John A., L., Stephen D., B., Simon D. W., F. & Jukka, C. RhierBAPS: An R Implementation of the Population Clustering Algorithm hierBAPS. *Wellcome Open Research* **3**, 93 (2018).
228 Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* **3**, e000131, doi:10.1099/mgen.0.000131 (2017).
229 Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566-D573, doi:10.1093/nar/gkw1004 (2017).
230 Ronco, T. *et al.* Genome analysis of *Clostridium perfringens* isolates from healthy and necrotic enteritis infected chickens and turkeys. *BMC Res Notes* **10**, 270, doi:10.1186/s13104-017-2594-9 (2017).
231 Brynildsrud, O., Bohlin, J., Scheffer, L. & Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biology* **17**, doi:10.1186/s13059-016-1108-8 (2016).
232 Roosaare, M., Puustusmaa, M., Mols, M., Vaher, M. & Remm, M. PlasmidSeeker: identification of known plasmids from bacterial whole genome sequencing reads. *PeerJ* **6**, e4588, doi:10.7717/peerj.4588 (2018).
233 Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945 (2000).
234 Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009-1010, doi:10.1093/bioinformatics/btr039 (2011).

235 Arndt, D., Marcu, A., Liang, Y. & Wishart, D. S. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. *Brief Bioinform*, doi:10.1093/bib/bbx121 (2017).

236 Bland, C. *et al.* CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209, doi:10.1186/1471-2105-8-209 (2007).

237 Huerta-Cepas, J. *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286-293, doi:10.1093/nar/gkv1248 (2016).

238 Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.*, doi:10.1093/molbev/msx148 (2017).

239 Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60, doi:10.1186/gb-2011-12-6-r60 (2011).

240 Harmon, S. M., Kautter, D. A. & Peeler, J. T. Improved medium for enumeration of *Clostridium perfringens*. *Appl Microbiol* **22**, 688-692 (1971).

241 Brady, J. *et al.* Toxinotyping of necrotic enteritis-producing and commensal isolates of *Clostridium perfringens* from chickens fed organic diets. *Avian Pathol.* **39**, 475-481, doi:10.1080/03079457.2010.521141 (2010).

242 Miles, A. A., Misra, S. S. & Irwin, J. O. The estimation of the bactericidal power of the blood. *Journal of Hygiene* **38**, 732-749, doi:10.1017/S002217240001158X (1938).

243 Theodorou, M. K., Williams, B. A., Dhanoa, M. S., McAllan, A. B. & France, J. A simple gas production method using a pressure transducer to determine the fermentation kentics of ruminant feeds. *Anim. Feed Sci. Technol.* **48**, 185-197 (1994).

244 Rotilie, C. A., Fass, R. J., Prior, R. B. & Perkins, R. L. Microdilution technique for antimicrobial susceptibility testing of anaerobic bacteria. *Antimicrob. Agents Chemother.* **7**, 311-315 (1975).

245 Zhao, Y., Biggs, T. D. & Xian, M. Hydrogen sulfide (H2S) releasing agents: chemistry and biological applications. *Chem Commun (Camb)* **50**, 11788-11805, doi:10.1039/c4cc00968a (2014).

246 O'Brien, D. K. & Melville, S. B. Effects of *Clostridium perfringens* alpha-toxin (PLC) and perfringolysin O (PFO) on cytotoxicity to macrophages, on escape from the phagosomes of macrophages, and on persistence of *C. perfringens* in host tissues. *Infect. Immun.* **72**, 5204-5215, doi:10.1128/IAI.72.9.5204-5215.2004 (2004).

247 Brunt, J. *et al.* Functional characterisation of germinant receptors in *Clostridium botulinum* and *Clostridium sporogenes* presents novel insights into spore germination systems. *PLoS Pathog* **10**, e1004382, doi:10.1371/journal.ppat.1004382 (2014).

248 Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. & Minton, N. P. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* **70**, 452-464, doi:10.1016/j.mimet.2007.05.021 (2007).

249 Hughes, K. R. *et al.* Bifidobacterium breve reduces apoptotic epithelial cell shedding in an exopolysaccharide and MyD88-dependent manner. *Open Biol* **7**, doi:10.1098/rsob.160155 (2017).

250 Erben, U. *et al.* A guide to histomorphological evaluation of intestinal inflammation in mouse models. *Int J Clin Exp Pathol* **7**, 4557-4576 (2014).

251 RStudio: Integrated Development for R. RStudio, Inc. (Boston, MA, 2015).

252 Schmid, K. O. & Quaiser, K. *Über eine besonders schwer verlaufende Form von Enteritis beim Säugling*. 114 (1953).

253 Obladen, M. Necrotizing enterocolitis - 150 years of fruitless search for the cause. *Neonatology* **96**, 203-210, doi:10.1159/000215590 (2009).

254 Lin, P. W. & Stoll, B. J. Necrotising enterocolitis. *Lancet* **368**, 1271-1283, doi:10.1016/s0140-6736(06)69525-1 (2006).

255 Gordon, P. V. & Swanson, J. R. Necrotizing enterocolitis is one disease with many origins and potential means of prevention. *Pathophysiology* **21**, 13-19, doi:10.1016/j.pathophys.2013.11.015 (2014).

256 Bell, M. J. *et al.* Neonatal necrotizing enterocolitis: therapeutic decisions based upon clinical staging. *Annals of Surgery* **187**, 1-7 (1978).

257 Walsh, M. C. & Kliegman, R. M. Necrotizing enterocolitis: treatment based on staging criteria. *Pediatr Clin North Am* **33**, 179-201 (1986).

258 Musemeche, C. A., Kosloske, A. M., Bartow, S. A. & Umland, E. T. Comparative effects of ischemia, bacteria, and substrate on the pathogenesis of intestinal necrosis. *J. Pediatr. Surg.* **21**, 536-538 (1986).

259 Claud, E. C. Neonatal Necrotizing Enterocolitis -Inflammation and Intestinal Immaturity. *Antiinflamm Antiallergy Agents Med Chem* **8**, 248-259, doi:10.2174/187152309789152020 (2009).

260 Kindley, A. D., Rboerts, P. J. & Tulloch, W. H. Neonatal necrotising enterocolitis. *Lancet* **1**, 649 (1977).

261 Kliegman, R. M., Fanaroff, A. A., Izant, R. & Speck, W. T. Clostridia as pathogens in neonatal necrotizing enterocolitis. *Journal of Pediatrics* **95**, 287-289 (1979).

262 Zhou, Y. *et al.* Longitudinal analysis of the premature infant intestinal microbiome prior to necrotizing enterocolitis: a case-control study. *PLoS One* **10**, e0118632, doi:10.1371/journal.pone.0118632 (2015).

263 Brower-Sinning, R. *et al.* Mucosa-associated bacterial diversity in necrotizing enterocolitis. *PLoS One* **9**, e105046, doi:10.1371/journal.pone.0105046 (2014).

264 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).

265 Egan, C. E. *et al.* Toll-like receptor 4-mediated lymphocyte influx induces neonatal necrotizing enterocolitis. *J. Clin. Invest.* **126**, 495-508, doi:10.1172/JCI83356 (2016).

266 Clark, N. C., Hill, B. C., O'Hara, C. M., Steingrimsson, O. & Cooksey, R. C. Epidemiologic typing of *Enterobacter sakazakii* in two neonatal nosocomial outbreaks. *Diagn. Microbiol. Infect. Dis.* **13**, 467-472 (1990).

267 Walker, A. W. & Lawley, T. D. Therapeutic modulation of intestinal dysbiosis. *Pharmacol. Res.* **69**, 75-86, doi:10.1016/j.phrs.2012.09.008 (2013).

268 Warner, B. B. *et al.* Gut bacteria dysbiosis and necrotising enterocolitis in very low birthweight infants: a prospective case-control study. *The Lancet* **387**, 1928-1936, doi:10.1016/s0140-6736(16)00081-7 (2016).

269 Ferraris, L. *et al.* Clostridia in premature neonates' gut: incidence, antibiotic susceptibility, and perinatal determinants influencing colonization. *Plos One* **7**, doi:10.1371/journal.pone.0030594 (2012).

270 Coggins, S. A., Wynn, J. L. & Weitkamp, J. H. Infectious causes of necrotizing enterocolitis. *Clinics in Perinatology* **42**, 133-154 (2015).

271 Turroni, F. *et al.* Diversity of bifidobacteria within the infant gut microbiota. *PLoS One* **7**, e36957, doi:10.1371/journal.pone.0036957 (2012).

272 Lee, J. S. & Polin, R. A. Treatment and prevention of necrotizing enterocolitis. *Semin Neonatol.* **8**, 449-459 (2003).

273 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, doi:10.1542/peds.2009-3149 (2010).

274 Cassir, N. *et al.* Clostridium butyricum Strains and Dysbiosis Linked to Necrotizing Enterocolitis in Preterm Neonates. *Clin. Infect. Dis.* **61**, 1107-1115, doi:10.1093/cid/civ468 (2015).

275 Han, V. K., Sayed, H., Chance, G. W., Brabyn, D. G. & Shaheed, W. A. An outbreak of *Clostridium difficile* necrotizing enterocolitis: a case for oral vancomycin therapy? *Pediatrics* **71**, 935-941 (1983).

276 Alfa, M. J. *et al.* An outbreak of necrotizing enterocolitis associated with a novel *Clostridium* species in a neonatal intensive care unit. *Clin. Infect. Dis.* **35**, S101-105, doi:10.1086/341929 (2002).

277 Benamar, S., Cassir, N. & La Scola, B. Genome Sequence of a *Clostridium neonatale* Strain Isolated in a Canadian Neonatal Intensive Care Unit. *Genome Announc* **4**, doi:10.1128/genomeA.01431-15 (2016).

278 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).

279 Overturf, G. D., Sherman, M. P., Scheifele, D. W. & Wong, L. C. Neonatal necrotizing enterocolitis associated with delta toxin-producing methicillin-resistant Staphylococcus aureus. *Pediatr. Infect. Dis. J.* **9**, 88-91 (1990).

280 Romano-Keeler, J. *et al.* Early life establishment of site-specific microbial communities in the gut. *Gut Microbes* **5**, 192-201, doi:10.4161/gmic.28442 (2014).

281 Emami, C. N., NMittal, R., Wang, L., Ford, H. R. & Prasadarao, N. V. Role of neutrophils and macrophages in the pathogenesis of necrotizing entercolitis caused by Cronobacter sakazakii. *J. Surg. Res.* **172**, 18-28 (2012).

282 Bell, M. J., Feigin, R. D., Ternberg, J. L. & Brotherton, T. Evaluation of gastrointestinal microflora in necrotizing enterocolitis. *J Pediatr* **92**, 589-592 (1978).

283 Leigh, L., Stoll, B. J., Rahman, M. & McGowan, J., Jr. Pseudomonas aeruginosa infection in very low birth weight infants: a case-control study. *Pediatr. Infect. Dis. J.* **14**, 367-371 (1995).

284 Pumberger, W. & Novak, W. Fatal neonatal Salmonella enteritidis sepsis. *J Perinatol* **20**, 54-56 (2000).

285 Waligora-Dupriet, A. J., Dugay, A., Auzeil, N., Huerre, M. & Butel, M. J. Evidence of clostidial implication in necrotizing enterocolitis through bacterial fermentation in a gnotobiotic quail model. *Pediatr Res* **58**, 629-635 (2005).

286 Silverman, M. A., Konnikova, L. & Gerber, J. S. Impact of Antibiotics on Necrotizing Enterocolitis and Antibiotic-Associated Diarrhea. *Gastroenterol Clin North Am* **46**, 61-76, doi:10.1016/j.gtc.2016.09.010 (2017).

287 Alexander, V. N., Northrup, V. & Bizzarro, M. J. Antibiotic exposure in the newborn intensive care unit and the risk of necrotizing enterocolitis. *J Pediatr* **159**, 392-397, doi:10.1016/j.jpeds.2011.02.035 (2011).

288 Abt, M. C., McKenney, P. T. & Pamer, E. G. *Clostridium difficile* colitis: pathogenesis and host defence. *Nat. Rev. Microbiol.* **14**, 609-620, doi:10.1038/nrmicro.2016.108 (2016).

289 Lakshminarayanan, B. *The ELDERMET biobank: Isolation and characterization of the intestinal microbiota from elderly Irish subject* PhD thesis, University College Cork, (2014).

290 Lacey, J. A. *et al.* Whole genome analysis reveals the diversity and evolutionary relationships between necrotic enteritis-causing strains of *Clostridium perfringens*. *BMC Genomics* **19**, 379, doi:10.1186/s12864-018-4771-1 (2018).

291 Park, M. *et al.* Diversity of *Clostridium perfringens* isolates from various sources and prevalence of conjugative plasmids. *Anaerobe* **38**, 25-35, doi:10.1016/j.anaerobe.2015.11.003 (2015).

292 Brynestad, S., Sarker, M. R., McClane, B. A., Granum, P. E. & Rood, J. I. Enterotoxin plasmid from *Clostridium perfringens* is conjugative. *Infect. Immun.* **69**, 3483-3487, doi:10.1128/IAI.69.5.3483-3487.2001 (2001).

293 Sekulovic, O., Meessen-Pinard, M. & Fortier, L. C. Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. *J. Bacteriol.* **193**, 2726-2734, doi:10.1128/JB.00787-10 (2011).

294 Kim, K. P. *et al.* Inducible *Clostridium perfringens* bacteriophages PhiS9 and PhiS63: Different genome structures and a fully functional sigK intervening element. *Bacteriophage* **2**, 89-97, doi:10.4161/bact.21363 (2012).

295 Zimmer, M., Scherer, S. & Loessner, M. J. Genomic analysis of *Clostridium perfringens* bacteriophage phi3626, which integrates into guaA and possibly affects sporulation. *J. Bacteriol.* **184**, 4359-4368 (2002).

296 Zeaki, N., Susilo, Y. B., Pregiel, A., Radstrom, P. & Schelin, J. Prophage-Encoded Staphylococcal Enterotoxin A: Regulation of Production in *Staphylococcus aureus* Strains Representing Different Sea Regions. *Toxins (Basel)* **7**, 5359-5376, doi:10.3390/toxins7124889 (2015).

297 Stewart, A. W. & Johnson, M. G. Increased numbers of heat-resistant spores produced by two strains of *Clostridium perfringens* bearing temperate phage s9. *J Gen Microbiol* **103**, 45-50, doi:10.1099/00221287-103-1-45 (1977).

298 Adlerberth, I. & Wold, A. E. Establishment of the gut microbiota in Western infants. *Acta Paediatr* **98**, 229-238, doi:10.1111/j.1651-2227.2008.01060.x (2009).

299 Alcon-Giner, C. *et al.* Optimisation of 16S rRNA gut microbiota profiling of extremely low birth weight infants. *BMC Genomics* **18**, 841, doi:10.1186/s12864-017-4229-x (2017).

300 Kwak, M. J. *et al.* Evolutionary architecture of the infant-adapted group of *Bifidobacterium* species associated with the probiotic function. *Syst. Appl. Microbiol.* **39**, 429-439, doi:10.1016/j.syapm.2016.07.004 (2016).

301 Thomas, J. P., Raine, T., Reddy, S. & Belteki, G. Probiotics for the prevention of necrotising enterocolitis in very low-birth-weight infants: a meta-analysis and systematic review. *Acta Paediatrica* **106**, 1729-1741, doi:10.1111/apa.13902 (2017).

302 Gibson, G. R. & Wang, X. Regulatory effects of bifidobacteria on the growth of other colonic bacteria. *J Appl Bacteriol* **77**, 412-420 (1994).

303 Songer, J. G. Clostridial enteric diseases of domestic animals. *Clin. Microbiol. Rev.* **9**, 216-234 (1996).

304 Fernandez-Miyakawa, M. E., Jost, B. H., Billington, S. J. & Uzal, F. A. Lethal effects of *Clostridium perfringens* epsilon toxin are potentiated by alpha and perfringolysin-O toxins in a mouse model. *Vet. Microbiol.* **127**, 379-385, doi:10.1016/j.vetmic.2007.09.013 (2008).

305 Jost, B. H., Billington, S. J., Trinh, H. T. & Songer, J. G. Association of genes encoding beta2 toxin and a collagen binding protein in *Clostridium perfringens* isolates of porcine origin. *Vet. Microbiol.* **115**, 173-182, doi:10.1016/j.vetmic.2006.01.012 (2006).

306 Wade, B. *et al.* The adherent abilities of *Clostridium perfringens* strains are critical for the pathogenesis of avian necrotic enteritis. *Vet. Microbiol.* **197**, 53-61, doi:10.1016/j.vetmic.2016.10.028 (2016).

307 Drudy, D. *et al.* Human antibody response to surface layer proteins in *Clostridium difficile* infection. *FEMS Immunol. Med. Microbiol.* **41**, 237-242, doi:10.1016/j.femsim.2004.03.007 (2004).

308 Kirk, J. A. *et al.* New class of precision antimicrobials redefines role of *Clostridium difficile* S-layer in virulence and viability. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aah6813 (2017).

309 Alam, S. I., Bansod, S., Kumar, R. B., Sengupta, N. & Singh, L. Differential proteomic analysis of *Clostridium perfringens* ATCC13124; identification of dominant, surface and structure associated proteins. *BMC Microbiol.* **9**, 162, doi:10.1186/1471-2180-9-162 (2009).

310 Raven, K. E. *et al.* Complex Routes of Nosocomial Vancomycin-Resistant *Enterococcus faecium* Transmission Revealed by Genome Sequencing. *Clin. Infect. Dis.* **64**, 886-893, doi:10.1093/cid/ciw872 (2017).

311 Rivera, G. *et al.* Extensively drug-resistant Acinetobacter baumannii isolated in a university hospital: Role of inter-hospital transmission. *J Infect Dev Ctries* **10**, 96-99, doi:10.3855/jidc.6713 (2016).

312 Tajeddin, E. *et al.* The role of the intensive care unit environment and health-care workers in the transmission of bacteria associated with hospital acquired infections. *J Infect Public Health* **9**, 13-23, doi:10.1016/j.jiph.2015.05.010 (2016).

313 Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31**, 533-538, doi:10.1038/nbt.2579 (2013).

314 Quick, J. *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol* **16**, 114, doi:10.1186/s13059-015-0677-2 (2015).

315 Leggett, R. M. *et al.* Rapid profiling of the preterm infant gut microbiota using nanopore sequencing aids pathogen diagnostics. *bioRxiv*, doi:10.1101/180406 (2018).

316 Dolan, G. P. *et al.* An epidemiological review of gastrointestinal outbreaks associated with *Clostridium perfringens*, North East of England, 2012-2014. *Epidemiol. Infect.* **144**, 1386-1393, doi:10.1017/S0950268815002824 (2016).

317 O'Brien, S. J. *et al.* Modelling study to estimate the health burden of foodborne diseases: cases, general practice consultations and hospitalisations in the UK, 2009. *BMJ Open* **6**, e011119, doi:10.1136/bmjopen-2016-011119 (2016).

318 UK Food Standards Agency. Foodborne Disease Strategy 2010-2015. (2011).

319 Fernandez Miyakawa, M. E., Pistone Creydt, V., Uzal, F. A., McClane, B. A. & Ibarra, C. *Clostridium perfringens* enterotoxin damages the human intestine in vitro. *Infect. Immun.* **73**, 8407-8410, doi:10.1128/IAI.73.12.8407-8410.2005 (2005).

320 Gaucher, M. L. *et al.* Recurring Necrotic Enteritis Outbreaks in Commercial Broiler Chicken Flocks Strongly Influence Toxin Gene Carriage and Species Richness in the Resident *Clostridium perfringens* Population. *Front Microbiol* **8**, 881, doi:10.3389/fmicb.2017.00881 (2017).

321 Acheson, P., Bell, V., Gibson, J., Gorton, R. & Inns, T. Enforcement of science-using a *Clostridium perfringens* outbreak investigation to take legal action. *J Public Health (Oxf)* **38**, 511-515, doi:10.1093/pubmed/fdv060 (2016).

322 Lindstrom, M., Heikinheimo, A., Lahti, P. & Korkeala, H. Novel insights into the epidemiology of *Clostridium perfringens* type A food poisoning. *Food Microbiol.* **28**, 192-198, doi:10.1016/j.fm.2010.03.020 (2011).

323 Lahti, P., Heikinheimo, A., Johansson, T. & Korkeala, H. *Clostrdium perfringens* type A strains carrying a plasmid-borne enterotoxin gene (Genotype IS1151-cpe or IS1470-like-cpe) as a common cause of food poisoning. *J. Clin. Microbiol.* **46**, 371-373, doi:10.1128/Jcm.01650-07 (2008).

324 Tanaka, D. *et al.* An outbreak of food-borne gastroenteritis caused by *Clostridium perfringens* carrying the cpe gene on a plasmid. *Japanese Journal of Infectious Diseases* **56**, 137-139 (2003).

325 Sparks, S. G., Carman, R. J., Sarker, M. R. & McClane, B. A. Genotyping of enterotoxigenic *Clostridium perfringens* fecal isolates associated with antibiotic-associated diarrhea and food poisoning in North America. *J. Clin. Microbiol.* **39**, 883-888, doi:Doi 10.1128/Jcm.39.3.883-888.2001 (2001).

326 Cornillot, E. *et al.* The enterotoxin gene (cpe) of *Clostridium perfringens* can be chromosomal or plasmid-borne. *Mol. Microbiol.* **15**, 639-647 (1995).

327 Tanaka, D. *et al.* Genotyping of *Clostridium perfringens* isolates collected from food poisoning outbreaks and healthy individuals in Japan based on the cpe locus. *Japanese Journal of Infectious Diseases* **60**, 68-69 (2007).

328 Miyamoto, K., Li, J. H. & McClane, B. A. Enterotoxigenic *Clostridium perfringens*: Detection and Identification. *Microbes Environ.* **27**, 343-349, doi:10.1264/jsme2.ME12002 (2012).

329 Gera, K., Le, T., Jamin, R., Eichenbaum, Z. & McIver, K. S. The phosphoenolpyruvate phosphotransferase system in group A Streptococcus acts to reduce streptolysin S activity and lesion severity during soft tissue infection. *Infect. Immun.* **82**, 1192-1204, doi:10.1128/IAI.01271-13 (2014).

330 Mertins, S. *et al.* Interference of components of the phosphoenolpyruvate phosphotransferase system with the central virulence gene regulator PrfA of Listeria monocytogenes. *J. Bacteriol.* **189**, 473-490, doi:10.1128/JB.00972-06 (2007).

331 Genevaux, P., Wawrzynow, A., Zylicz, M., Georgopoulos, C. & Kelley, W. L. DjlA is a third DnaK co-chaperone of *Escherichia coli*, and DjlA-mediated induction of colanic acid capsule requires DjlA-DnaK interaction. *J. Biol. Chem.* **276**, 7906-7912, doi:10.1074/jbc.M003855200 (2001).

332 Hu, W. S., Kim, H. & Koo, O. K. Molecular genotyping, biofilm formation and antibiotic resistance of enterotoxigenic *Clostridium perfringens* isolated from meat supplied to school cafeterias in South Korea. *Anaerobe* **52**, 115-121, doi:10.1016/j.anaerobe.2018.06.011 (2018).

333 Heikinheimo, A., Lindstrom, M., Granum, P. E. & Korkeala, H. Humans as reservoir for enterotoxin gene--carrying *Clostridium perfringens* type A. *Emerg Infect Dis* **12**, 1724-1729, doi:10.3201/eid1211.060478 (2006).

334 Lakshminarayanan, B. *et al.* Prevalence and characterization of *Clostridium perfringens* from the faecal microbiota of elderly Irish subjects. *J. Med. Microbiol.* **62**, 457-466, doi:10.1099/jmm.0.052258-0 (2013).

335 Ashton, P. M. *et al.* Identification of *Salmonella* for public health surveillance using whole genome sequencing. *PeerJ* **4**, e1752, doi:10.7717/peerj.1752 (2016).

336 Schmid, D. *et al.* Whole genome sequencing as a tool to investigate a cluster of seven cases of listeriosis in Austria and Germany, 2011-2013. *Clin. Microbiol. Infect.* **20**, 431-436, doi:10.1111/1469-0691.12638 (2014).

337 Deguchi, A. *et al.* Genetic characterization of type A enterotoxigenic *Clostridium perfringens* strains. *PLoS One* **4**, e5598, doi:10.1371/journal.pone.0005598 (2009).

338 Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178-184, doi:10.1038/nature11319 (2012).

339 Keto-Timonen, R., Heikinheimo, A., Eerola, E. & Korkeala, H. Identification of *Clostridium* species and DNA fingerprinting of *Clostridium perfringens* by amplified fragment length polymorphism analysis. *J. Clin. Microbiol.* **44**, 4057-4065, doi:10.1128/JCM.01275-06 (2006).

340 Gardy, J., Loman, N. J. & Rambaut, A. Real-time digital pathogen surveillance - the time is now. *Genome Biol* **16**, 155, doi:10.1186/s13059-015-0726-x (2015).

341 Havelaar, A. H. *et al.* World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. *PLoS Med* **12**, e1001923, doi:10.1371/journal.pmed.1001923 (2015).

342 Thomas, M. K. *et al.* Estimates of foodborne illness-related hospitalizations and deaths in Canada for 30 specified pathogens and unspecified agents. *Foodborne Pathog Dis* **12**, 820-827, doi:10.1089/fpd.2015.1966 (2015).

343 Uzal, F. A. *et al.* Towards an understanding of the role of *Clostridium perfringens* toxins in human and animal disease. *Future Microbiol* **9**, 361-377 (2014).

344 Scharff, R. L. Economic burden from health losses due to foodborne illness in the United States. *J. Food Prot.* **75**, 123-131, doi:10.4315/0362-028X.JFP-11-058 (2012).

345 Ciwf.org.uk. *About chickens*, <https://www.ciwf.org.uk/farm-animals/chickens/> (2018).

346 Palliyeguru, M. W. C. D. & Rose, S. P. Sub-clinical necrotic enteritis: its aetiology and predisposing factors in commercial broiler production. *Worlds Poultry Science Journal* **70**, 803-815, doi:10.1017/S0043933914000865 (2014).

347 Stanley, D., Keyburn, A. L., Denman, S. E. & Moore, R. J. Changes in the caecal microflora of chickens following *Clostridium perfringens* challenge to induce necrotic enteritis. *Vet. Microbiol.* **159**, 155-162, doi:10.1016/j.vetmic.2012.03.032 (2012).

348 Awad, W. A. *et al.* Age-Related Differences in the Luminal and Mucosa-Associated Gut Microbiome of Broiler Chickens and Shifts Associated with *Campylobacter jejuni* Infection. *Front Cell Infect Microbiol* **6**, 154, doi:10.3389/fcimb.2016.00154 (2016).

349 Moore, R. J. Necrotic enteritis predisposing factors in broiler chickens. *Avian Pathol.* **45**, 275-281, doi:10.1080/03079457.2016.1150587 (2016).

350 Stanley, D., Wu, S. B., Rodgers, N., Swick, R. A. & Moore, R. J. Differential responses of cecal microbiota to fishmeal, Eimeria and *Clostridium perfringens* in a necrotic enteritis challenge model in chickens. *PLoS One* **9**, e104739, doi:10.1371/journal.pone.0104739 (2014).

351 Nakphaichit, M. *et al.* The effect of including *Lactobacillus reuteri* KUB-AC5 during post-hatch feeding on the growth and ileum microbiota of broiler chickens. *Poult Sci* **90**, 2753-2765, doi:10.3382/ps.2011-01637 (2011).

352 Tinrat, S., Saraya, S. & Traidej Chomnawang, M. Isolation and characterization of *Lactobacillus salivarius* MTC 1026 as a potential probiotic. *J. Gen. Appl. Microbiol.* **57**, 365-378 (2011).

353 Messaoudi, S. *et al.* In vitro evaluation of the probiotic potential of *Lactobacillus salivarius* SMXD51. *Anaerobe* **18**, 584-589, doi:10.1016/j.anaerobe.2012.10.004 (2012).

354 Wang, L., Lilburn, M. & Yu, Z. Intestinal Microbiota of Broiler Chickens As Affected by Litter Management Regimens. *Front Microbiol* **7**, 593, doi:10.3389/fmicb.2016.00593 (2016).

355 Cao, G. T. *et al.* Effects of a probiotic, *Enterococcus faecium*, on growth performance, intestinal morphology, immune response, and cecal microflora in broiler chickens challenged with *Escherichia coli* K88. *Poult Sci* **92**, 2949-2955, doi:10.3382/ps.2013-03366 (2013).

356 Svetoch, E. A. *et al.* Diverse antimicrobial killing by *Enterococcus faecium* E 50-52 bacteriocin. *J. Agric. Food Chem.* **56**, 1942-1948, doi:10.1021/jf073284g (2008).

357 Thibodeau, A. *et al.* Chicken Caecal Microbiome Modifications Induced by *Campylobacter jejuni* Colonization and by a Non-Antibiotic Feed Additive. *PLoS One* **10**, e0131978, doi:10.1371/journal.pone.0131978 (2015).

358 Hsiao, A. *et al.* Members of the human gut microbiota involved in recovery from Vibrio cholerae infection. *Nature* **515**, 423-426, doi:10.1038/nature13738 (2014).

359 Lawson, P. A. & Finegold, S. M. Reclassification of *Ruminococcus obeum* as *Blautia obeum* comb. nov. *Int. J. Syst. Evol. Microbiol.* **65**, 789-793, doi:10.1099/ijs.0.000015 (2015).

360 Stalker, M. J., Brash, M. L., Weisz, A., Ouckama, R. M. & Slavic, D. Arthritis and osteomyelitis associated with *Enterococcus cecorum* infection in broiler and broiler breeder chickens in Ontario, Canada. *J Vet Diagn Invest* **22**, 643-645, doi:10.1177/104063871002200426 (2010).

361 Hou, Q. *et al.* Differential fecal microbiota are retained in broiler chicken lines divergently selected for fatness traits. *Sci Rep* **6**, 37376, doi:10.1038/srep37376 (2016).

362 Tamanai-Shacoori, Z. *et al.* Roseburia spp.: a marker of health? *Future Microbiol* **12**, 157-170, doi:10.2217/fmb-2016-0130 (2017).

363 Mountzouris, K. C. *et al.* Evaluation of the efficacy of a probiotic containing *Lactobacillus, Bifidobacterium, Enterococcus*, and *Pediococcus* strains in promoting broiler performance and modulating cecal microflora composition and metabolic activities. *Poult Sci* **86**, 309-317, doi:10.1093/ps/86.2.309 (2007).

364 Gervasi, T., Curto, R. L., Minniti, E., Narbad, A. & Mayer, M. J. Application of *Lactobacillus johnsonii* expressing phage endolysin for control of *Clostridium perfringens*. *Lett. Appl. Microbiol.* **59**, 355-361 (2014).

365 De Cesare, A. *et al.* Effect of dietary supplementation with *Lactobacillus acidophilus* D2/CSL (CECT 4529) on caecum microbioma and productive performance in broiler chickens. *PLoS One* **12**, e0176309, doi:10.1371/journal.pone.0176309 (2017).

366 Torok, V. A., Hughes, R. J., Ophel-Keller, K., Ali, M. & Macalpine, R. Influence of different litter materials on cecal microbiota colonization in broiler chickens. *Poult Sci* **88**, 2474-2481, doi:10.3382/ps.2008-00381 (2009).

367 Wade, B., Keyburn, A. L., Seemann, T., Rood, J. I. & Moore, R. J. Binding of *Clostridium perfringens* to collagen correlates with the ability to cause necrotic enteritis in chickens. *Vet. Microbiol.* **180**, 299-303, doi:10.1016/j.vetmic.2015.09.019 (2015).

368 Sawires, Y. S. & Songer, J. G. *Clostridium perfringens*: insight into virulence evolution and population structure. *Anaerobe* **12**, 23-43, doi:10.1016/j.anaerobe.2005.10.002 (2006).

369 Tkacz, A., Hortala, M. & Poole, P. S. Absolute quantitation of microbiota abundance in environmental samples. *Microbiome* **6**, 110, doi:10.1186/s40168-018-0491-7 (2018).

370 Bennetts, H. W. *Bacillus welchii* and bowel lesions. With special reference to a case of coccidiosis. *Aust. Vet. J.* **6**, 153-154 (1930).

371 Mather, A. E. *et al.* Genomic Analysis of *Salmonella enterica* Serovar Typhimurium from Wild Passerines in England and Wales. *Appl. Environ. Microbiol.* **82**, 6728-6735, doi:10.1128/AEM.01660-16 (2016).

372 Wong, V. K. *et al.* Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nat. Genet.* **47**, 632-639, doi:10.1038/ng.3281 (2015).

373 Mather, A. E. *et al.* New Variant of Multidrug-Resistant *Salmonella enterica* Serovar Typhimurium Associated with Invasive Disease in Immunocompromised Patients in Vietnam. *MBio* **9**, doi:10.1128/mBio.01056-18 (2018).

374 Geer, L. Y. *et al.* The NCBI BioSystems database. *Nucleic Acids Res.* **38**, D492-496, doi:10.1093/nar/gkp858 (2010).

375 Lebrun, M., Mainil, J. G. & Linden, A. Cattle enterotoxaemia and *Clostridium perfringens*: description, diagnosis and prophylaxis. *Vet. Rec.* **167**, 13-22, doi:10.1136/vr.b4859 (2010).

376 Bannam, T. L. *et al.* The *Clostridium perfringens* TetA(P) efflux protein contains a functional variant of the Motif A region found in major facilitator superfamily transport proteins. *Microbiology* **150**, 127-134, doi:10.1099/mic.0.26614-0 (2004).

377 Jost, B. H., Billington, S. J., Trinh, H. T., Bueschel, D. M. & Songer, J. G. Atypical cpb2 genes, encoding beta2-toxin in *Clostridium perfringens* isolates of nonporcine origin. *Infect. Immun.* **73**, 652-656, doi:10.1128/IAI.73.1.652-656.2005 (2005).

378 Clark, R. H., Bloom, B. T., Spitzer, A. R. & Gerstmann, D. R. Reported medication use in the neonatal intensive care unit: data from a large national data set. *Pediatrics* **117**, 1979-1987, doi:10.1542/peds.2005-1707 (2006).

379 Berryman, D. I., Lyristis, M. & Rood, J. I. Cloning and sequence analysis of ermQ, the predominant macrolide-lincosamide-streptogramin B resistance gene in *Clostridium perfringens*. *Antimicrob. Agents Chemother.* **38**, 1041-1046 (1994).

380 Kuang, L. H., Mu, L. Y., Su, M., Zhou, W. & Jiang, Y. M. [Epidemiology and Resistance Mechanisms of Group B Streptococci in Late-pregnant Maternal Birth Canal]. *Sichuan Da Xue Xue Bao Yi Xue Ban* **46**, 692-696 (2015).

381 Mehdi, Y. *et al.* Use of antibiotics in broiler production: Global impacts and alternatives. *Animal Nutrition* **4**, 170-178, doi:10.1016/j.aninu.2018.03.002 (2018).

382 Xiao, Y., Wagendorp, A., Moezelaar, R., Abee, T. & Wells-Bennik, M. H. A wide variety of *Clostridium perfringens* type A food-borne isolates that carry a chromosomal cpe gene belong to one multilocus sequence typing cluster. *Appl. Environ. Microbiol.* **78**, 7060-7068, doi:10.1128/AEM.01486-12 (2012).

383 Ma, M., Li, J. & McClane, B. A. Genotypic and phenotypic characterization of *Clostridium perfringens* isolates from Darmbrand cases in post-World War II Germany. *Infect. Immun.* **80**, 4354-4363, doi:10.1128/IAI.00818-12 (2012).

384 Shivaprasad, H. L. *et al.* Ulcerative enteritis-like disease associated with *Clostridium perfringens* type A in bobwhite quail (Colinus virginianus). *Avian Dis.* **52**, 635-640, doi:10.1637/8341-050108-Reg.1 (2008).

385 Silva, R. O. & Lobato, F. C. *Clostridium perfringens*: A review of enteric diseases in dogs, cats and wild animals. *Anaerobe* **33**, 14-17, doi:10.1016/j.anaerobe.2015.01.006 (2015).

386 Schulman, F. Y., Montali, R. J. & Hauer, P. J. Gastroenteritis associated with *Clostridium perfringens* type A in black-footed ferrets (Mustela nigripes). *Vet. Pathol.* **30**, 308-310, doi:10.1177/030098589303000316 (1993).

387 McInerney, J. O., McNally, A. & O'Connell, M. J. Why prokaryotes have pangenomes. *Nat Microbiol* **2**, 17040, doi:10.1038/nmicrobiol.2017.40 (2017).

388 Kiu, R. & Hall, L. J. Response: Commentary: Probing Genomic Aspects of the Multi-Host Pathogen *Clostridium perfringens* Reveals Significant Pangenome Diversity, and a Diverse Array of Virulence Factors. *Frontiers in Microbiology* **9**, doi:10.3389/fmicb.2018.01857 (2018).

389 Songer, J. G. Clostridia as agents of zoonotic disease. *Vet. Microbiol.* **140**, 399-404, doi:10.1016/j.vetmic.2009.07.003 (2010).

390 Kruth, S. A., Prescott, J. F., Welch, M. K. & Brodsky, M. H. Nosocomial diarrhea associated with enterotoxigenic *Clostridium perfringens* infection in dogs. *J. Am. Vet. Med. Assoc.* **195**, 331-334 (1989).

391 Collins, J. E. *et al.* Diarrhea associated with *Clostridium perfringens* type A enterotoxin in neonatal pigs. *J Vet Diagn Invest* **1**, 351-353, doi:10.1177/104063878900100414 (1989).

392 Wen, Q. & McClane, B. A. Detection of enterotoxigenic *Clostridium perfringens* type A isolates in American retail foods. *Appl. Environ. Microbiol.* **70**, 2685-2691 (2004).

393 Foley, S. L., Johnson, T. J., Ricke, S. C., Nayak, R. & Danzeisen, J. Salmonella pathogenicity and host adaptation in chicken-associated serovars. *Microbiol. Mol. Biol. Rev.* **77**, 582-607, doi:10.1128/MMBR.00015-13 (2013).

394 Knetsch, C. W. *et al.* Zoonotic Transfer of *Clostridium difficile* Harboring Antimicrobial Resistance between Farm Animals and Humans. *J. Clin. Microbiol.* **56**, doi:10.1128/JCM.01384-17 (2018).

395 Farzan, A. *et al.* An investigation into the association between cpb2-encoding *Clostridium perfringens* type A and diarrhea in neonatal piglets. *Can J Vet Res* **77**, 45-53 (2013).

396 Li, J., Sayeed, S., Robertson, S., Chen, J. & McClane, B. A. Sialidases affect the host cell adherence and epsilon toxin-induced cytotoxicity of *Clostridium perfringens* type D strain CN3718. *PLoS Pathog* **7**, e1002429, doi:10.1371/journal.ppat.1002429 (2011).

397 Sim, J. H. *et al.* Determining the cause of recurrent *Clostridium difficile* infection using whole genome sequencing. *Diagn. Microbiol. Infect. Dis.* **87**, 11-16, doi:10.1016/j.diagmicrobio.2016.09.023 (2017).

398 Lawley, T. D. & Walker, A. W. Intestinal colonization resistance. *Immunology* **138**, 1-11, doi:10.1111/j.1365-2567.2012.03616.x (2013).

399 Buffie, C. G. & Pamer, E. G. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat. Rev. Immunol.* **13**, 790-801, doi:10.1038/nri3535 (2013).

400 Sutton, R. G. & Hobbs, B. C. Food poisoning caused by heat-sensitive *Clostridium welchii*. A report of five recent outbreaks. *J Hyg (Lond)* **66**, 135-146 (1968).

401 Abdul-Mutalib, N. A., Syafinaz, A. N., Sakai, K. & Shirai, Y. An overview of foodborne illness and food safety in Malaysia. *International Food Research Journal* **22**, 896-901 (2015).

402 French, E., Rodriguez-Palacios, A. & LeJeune, J. T. Enteric bacterial pathogens with zoonotic potential isolated from farm-raised deer. *Foodborne Pathog Dis* **7**, 1031-1037, doi:10.1089/fpd.2009.0486 (2010).

403 Revitt-Mills, S. A., Rood, J. I. & Adams, V. *Clostridium perfringens* extracellular toxins and enzymes: 20 and counting. *Microbiology Australia* **36**, 114-117 (2015).

404 Jean, D., Briolat, V. & Reysset, G. Oxidative stress response in *Clostridium perfringens*. *Microbiology* **150**, 1649-1659, doi:10.1099/mic.0.27017-0 (2004).

405 Paredes-Sabja, D. & Sarker, M. R. *Clostridium perfringens* sporulation and its relevance to pathogenesis. *Future microbiology* **4**, 519-525 (2009).

406 O'Brien, D. K. & Melville, S. B. The anaerobic pathogen *Clostridium perfringens* can escape the phagosome of macrophages under aerobic conditions. *Cell. Microbiol.* **2**, 505-519 (2000).

407 Li, L., Bhatia, M. & Moore, P. K. Hydrogen sulphide--a novel mediator of inflammation? *Curr Opin Pharmacol* **6**, 125-129, doi:10.1016/j.coph.2005.10.007 (2006).

408 Stevens, D. L. & Bryant, A. E. The role of clostridial toxins in the pathogenesis of gas gangrene. *Clin. Infect. Dis.* **35**, S93-S100, doi:10.1086/341928 (2002).

409 Lucey, B. P. & Hutchins, G. M. William H. Welch, MD, and the discovery of Bacillus *welchii*. *Arch Pathol Lab Med* **128**, 1193-1195, doi:10.1043/1543-2165(2004)128<1193:WHWMAT>2.0.CO;2 (2004).

410 Flannigan, K. L., McCoy, K. D. & Wallace, J. L. Eukaryotic and prokaryotic contributions to colonic hydrogen sulfide synthesis. *Am J Physiol Gastrointest Liver Physiol* **301**, G188-193, doi:10.1152/ajpgi.00105.2011 (2011).

411 Reiffenstein, R. J., Hulbert, W. C. & Roth, S. H. Toxicology of hydrogen sulfide. *Annu. Rev. Pharmacol. Toxicol.*, 109-134 (1992).

412 Li, B. *et al.* Intestinal epithelial cell injury is rescued by hydrogen sulfide. *J. Pediatr. Surg.* **51**, 775-778, doi:10.1016/j.jpedsurg.2016.02.019 (2016).

413 Heredia, N. L., Labbe, R. G., Rodriguez, M. A. & Garcia-Alvarado, J. S. Growth, sporulation and enterotoxin production by *Clostridium perfringens* type A in the presence of human bile salts. *FEMS Microbiol. Lett.* **68**, 15-21 (1991).

414 Northfield, T. C. & McColl, I. Postprandial concentrations of free and conjugated bile acids down the length of the normal human small intestine. *Gut* **14**, 513-518 (1973).

415 Park, M. *et al.* Phenotypic and genotypic characterization of tetracycline and minocycline resistance in *Clostridium perfringens*. *Arch. Microbiol.* **192**, 803-810, doi:10.1007/s00203-010-0605-5 (2010).

416 Soll, R. F. & Edwards, W. H. Antibiotic use in neonatal intensive care. *Pediatrics* **135**, 928-929, doi:10.1542/peds.2015-0707 (2015).

417 Pruteanu, M. & Shanahan, F. Digestion of epithelial tight junction proteins by the commensal *Clostridium perfringens*. *Am J Physiol Gastrointest Liver Physiol* **305**, G740-748, doi:10.1152/ajpgi.00316.2012 (2013).

418 Garcia, J. P. *et al.* Epsilon toxin is essential for the virulence of *Clostridium perfringens* type D infection in sheep, goats, and mice. *Infect. Immun.* **81**, 2405-2414, doi:10.1128/IAI.00238-13 (2013).

419 Yasugi, M. *et al.* In vitro cytotoxicity induced by *Clostridium perfringens* isolate carrying a chromosomal cpe gene is exclusively dependent on sporulation and enterotoxin production. *Microb. Pathog.* **85**, 1-10, doi:10.1016/j.micpath.2015.04.003 (2015).

420 Schonherr-Hellec, S. *et al.* Clostridial Strain-Specific Characteristics Associated with Necrotizing Enterocolitis. *Appl. Environ. Microbiol.* **84**, doi:10.1128/AEM.02428-17 (2018).

421 Saitoh, Y. *et al.* Strucutral insight into tight junction disassembly by *Clostridium perfringens* enterotoxin. *Science* **347**, 775-778 (2015).

422 Pitcher, M. C., Beatty, E. R. & Cummings, J. H. The contribution of sulphate reducing bacteria and 5-aminosalicylic acid to faecal sulphide in patients with ulcerative colitis. *Gut* **46**, 64-72 (2000).

423 Bhatia, M. Role of hydrogen sulfide in the pathology of inflammation. *Scientifica (Cairo)* **2012**, 159680, doi:10.6064/2012/159680 (2012).

424 Levine, J., Ellis, C. J., Furne, J. K., Springfield, J. & Levitt, M. D. Fecal hydrogen sulfide production in ulcerative colitis. *Am J Gastroenterol* **93**, 83-87, doi:10.1111/j.1572-0241.1998.083_c.x (1998).

425 Andersen, B. M. *et al.* Decontamination of rooms, medical equipment and ambulances using an aerosol of hydrogen peroxide disinfectant. *J. Hosp. Infect.* **62**, 149-155, doi:10.1016/j.jhin.2005.07.020 (2006).

426 Claro, T., Cahill, O. J., O'Connor, N., Daniels, S. & Humphreys, H. Cold-air atmospheric pressure plasma against *Clostridium difficile* spores: a potential alternative for the decontamination of hospital inanimate surfaces. *Infect. Control Hosp. Epidemiol.* **36**, 742-744, doi:10.1017/ice.2015.39 (2015).

427 Snelling, A. M., Beggs, C. B., Kerr, K. G. & Shepherd, S. J. Spores of *Clostridium difficile* in Hospital Air. *Clin. Infect. Dis.* **51**, 1104-1105; author reply 1105, doi:10.1086/656686 (2010).

428 Tarrant, J., Jenkins, R. O. & Laird, K. T. From ward to washer: The survival of *Clostridium difficile* spores on hospital bed sheets through a commercial UK NHS healthcare laundry process. *Infect. Control Hosp. Epidemiol.*, 1-6, doi:10.1017/ice.2018.255 (2018).

429 Cahill, O. J. *et al.* Cold air plasma to decontaminate inanimate surfaces of the hospital environment. *Appl. Environ. Microbiol.* **80**, 2004-2010, doi:10.1128/AEM.03480-13 (2014).

430 Klampfl, T. G. *et al.* Cold atmospheric air plasma sterilization against spores and other microorganisms of clinical interest. *Appl. Environ. Microbiol.* **78**, 5077-5082, doi:10.1128/AEM.00583-12 (2012).

431 Sorg, J. A. & Sonenshein, A. L. Inhibiting the initiation of *Clostridium difficile* spore germination using analogs of chenodeoxycholic acid, a bile acid. *J. Bacteriol.* **192**, 4983-4990, doi:10.1128/JB.00610-10 (2010).

432 Halpern, M. D. *et al.* Bile acids induce ileal damage during experimental necrotizing enterocolitis. *Gastroenterology* **130**, 359-372, doi:10.1053/j.gastro.2005.10.023 (2006).

433 de Belle, R. C. *et al.* Intestinal absorption of bile salts: immature development in the neonate. *J Pediatr* **94**, 472-476 (1979).

434 Halpern, M. D. & Dvorak, B. Does abnormal bile acid metabolism contribute to NEC? *Semin Perinatol* **32**, 114-121, doi:10.1053/j.semperi.2008.01.005 (2008).

435 Hulzebos, C. V. *et al.* Fecal Bile Salts and the Development of Necrotizing Enterocolitis in Preterm Infants. *PLoS One* **12**, e0168633, doi:10.1371/journal.pone.0168633 (2017).

436 Inoue, T., Kimura, A., Aoki, K., Tohma, M. & Kato, H. Developmental pattern of 3-oxo-delta 4 bile acids in neonatal bile acid metabolism. *Arch Dis Child Fetal Neonatal Ed* **77**, F52-56 (1997).

437 Sorg, J. A. & Sonenshein, A. L. Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination. *J. Bacteriol.* **191**, 1115-1117, doi:10.1128/JB.01260-08 (2009).

438 Ramirez-Perez, O., Cruz-Ramon, V., Chinchilla-Lopez, P. & Mendez-Sanchez, N. The Role of the Gut Microbiota in Bile Acid Metabolism. *Ann Hepatol* **16**, s15-s20, doi:10.5604/01.3001.0010.5494 (2017).

439 Jarocki, P., Podlesny, M., Glibowski, P. & Targonski, Z. A new insight into the physiological role of bile salt hydrolase among intestinal bacteria from the genus *Bifidobacterium*. *PLoS One* **9**, e114379, doi:10.1371/journal.pone.0114379 (2014).

440 Wu, W. *et al.* Bifidobacterium adolescentis protects against necrotizing enterocolitis and upregulates TOLLIP and SIGIRR in premature neonatal rats. *BMC Pediatr* **17**, 1, doi:10.1186/s12887-016-0759-7 (2017).

441 Rood, J. I., Scott, V. N. & Duncan, C. L. Identification of a transferable tetracycline resistance plasmid (pCW3) from *Clostridium perfringens*. *Plasmid* **1**, 563-570 (1978).

442 Traub, W. H. & Raymond, E. A. In vitro resistance of *Clostridium perfringens* type A to gentamicin sulfate, and reduced activity of the antibiotic under anaerobic atmospheric conditions. *Chemotherapy* **16**, 162-165 (1971).

443 Vaishnavi, C. & Kaur, S. *Clostridium perfringens* enterotoxin in antibiotic-associated diarrhea. *Indian J Pathol Microbiol* **51**, 198-199 (2008).

444 McDonel, J. L. & Duncan, C. L. Histopathological effect of *Clostridium perfringens* enterotoxin in the rabbit ileum. *Infect. Immun.* **12**, 1214-1218 (1975).

445 Sugimoto, N., Chen, Y. M., Lee, S. Y., Matsuda, M. & Lee, C. Y. Pathodynamics of intoxication in rats and mice by enterotoxin of *Clostridium perfringens* type A. *Toxicon* **29**, 751-759 (1991).

446 Caserta, J. A. *et al.* Development and application of a mouse intestinal loop model to study the *in vivo* action of *Clostridium perfringens* enterotoxin. *Infect. Immun.* **79**, 3020-3027, doi:10.1128/IAI.01342-10 (2011).

447 Fisher, D. J. *et al.* Dissecting the contributions of *Clostridium perfringens* type C toxins to lethality in the mouse intravenous injection model. *Infect. Immun.* **74**, 5200-5210, doi:10.1128/IAI.00534-06 (2006).

448 Chen, X. *et al.* A mouse model of *Clostridium difficile*-associated disease. *Gastroenterology* **135**, 1984-1992, doi:10.1053/j.gastro.2008.09.002 (2008).

449 Giallourou, N. *et al.* A novel mouse model of *Campylobacter jejuni* enteropathy and diarrhea. *PLoS Pathog* **14**, e1007083, doi:10.1371/journal.ppat.1007083 (2018).

450 Williams, S. H. *et al.* New York City House Mice (Mus musculus) as Potential Reservoirs for Pathogenic Bacteria and Antimicrobial Resistance Determinants. *MBio* **9**, doi:10.1128/mBio.00624-18 (2018).

451 Fernandez-Miyakawa, M. E. *et al.* Development and application of an oral challenge mouse model for studying *Clostridium perfringens* type D infection. *Infect. Immun.* **75**, 4282-4288, doi:10.1128/IAI.00562-07 (2007).

452 Sun, X. *et al.* Mouse relapse model of *Clostridium difficile* infection. *Infect. Immun.* **79**, 2856-2864, doi:10.1128/IAI.01336-10 (2011).

453 Hutton, M. L., Mackin, K. E., Chakravorty, A. & Lyras, D. Small animal models for the study of *Clostridium difficile* disease pathogenesis. *FEMS Microbiol. Lett.* **352**, 140-149, doi:10.1111/1574-6968.12367 (2014).

454 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, doi:10.1128/IAI.00953-12 (2013).

455 McGuckin, M. A., Linden, S. K., Sutton, P. & Florin, T. H. Mucin dynamics and enteric pathogens. *Nat. Rev. Microbiol.* **9**, 265-278, doi:10.1038/nrmicro2538 (2011).

456 Bosman, E. S., Chan, J. M., Bhullar, K. & Vallance, B. A. Investigation of Host and Pathogen Contributions to Infectious Colitis Using the *Citrobacter rodentium* Mouse Model of Infection. *Methods Mol Biol* **1422**, 225-241, doi:10.1007/978-1-4939-3603-8_21 (2016).

457 Crepin, V. F., Collins, J. W., Habibzay, M. & Frankel, G. *Citrobacter rodentium* mouse model of bacterial infection. *Nat Protoc* **11**, 1851-1876, doi:10.1038/nprot.2016.100 (2016).

458 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, doi:10.1002/cpim.34 (2017).

459 Bergstrom, K. S. *et al.* Modulation of intestinal goblet cell function during infection by an attaching and effacing bacterial pathogen. *Infect. Immun.* **76**, 796-811, doi:10.1128/IAI.00093-07 (2008).

460 Sayeed, S. *et al.* Beta toxin is essential for the intestinal virulence of *Clostridium perfringens* type C disease isolate CN3685 in a rabbit ileal loop model. *Mol. Microbiol.* **67**, 15-30, doi:10.1111/j.1365-2958.2007.06007.x (2008).

461 Koenigsknecht, M. J. *et al.* Dynamics and Establishment of *Clostridium difficile* Infection in the Murine Gastrointestinal Tract. *Infect. Immun.* **83**, 934-941, doi:10.1128/Iai.02768-14 (2015).

462 Wang, S. *et al.* Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clin. Microbiol. Infect.*, doi:10.1016/j.cmi.2018.03.026 (2018).
463 Weinstein, L. & Barza, M. A. Gas gangrene. *New Engl. J. Med.* **289**, 1129-1131 (1973).
464 Engstrom, B. E. *et al.* Molecular typing of isolates of *Clostridium perfringens* from healthy and diseased poultry. *Vet. Microbiol.* **94**, 225-235 (2003).
465 Schlegel, B. J., Van Dreumel, T., Slavic, D. & Prescott, J. F. *Clostridium perfringens* type A fatal acute hemorrhagic gastroenteritis in a dog. *Can. Vet. J.* **53**, 555-557 (2012).
466 De Plaen, I. G. Inflammatory signaling in necrotizing enterocolitis. *Clin Perinatol* **40**, 109-124, doi:10.1016/j.clp.2012.12.008 (2013).
467 Yibar, A., Cetin, E., Ata, Z., Erkose, E. & Tayar, M. *Clostridium perfringens* Contamination in Retail Meat and Meat-Based Products in Bursa, Turkey. *Foodborne Pathog Dis* **15**, 239-245, doi:10.1089/fpd.2017.2350 (2018).
468 Lee, C. A. & Labbe, R. Distribution of Enterotoxin- and Epsilon-Positive *Clostridium perfringens* Spores in U.S. Retail Spices. *J. Food Prot.* **81**, 394-399, doi:10.4315/0362-028X.JFP-17-352 (2018).
469 Philipson, C. W. *et al.* Characterizing Phage Genomes for Therapeutic Applications. *Viruses* **10**, doi:10.3390/v10040188 (2018).
470 Stratton, C. W. Phages, Fitness, Virulence, and Synergy: A Novel Approach for the Therapy of Infections Caused by *Pseudomonas aeruginosa*. *J. Infect. Dis.* **215**, 668-670, doi:10.1093/infdis/jiw634 (2017).
471 Fortier, L. C. & Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **4**, 354-365, doi:10.4161/viru.24498 (2013).
472 Miller, R. W., Skinner, E. J., Sulakvelidze, A., Mathis, G. F. & Hofacre, C. L. Bacteriophage therapy for control of necrotic enteritis of broiler chickens experimentally infected with *Clostridium perfringens*. *Avian Dis.* **54**, 33-40, doi:10.1637/8953-060509-Reg.1 (2010).

Appendix 1 Supplementary data for Chapter 3

This appendix contains all the supplementary materials for Chapter 3.

Table S3.1: Clinical data from preterm neonates of 88 *C. perfringens* isolates in this study.

Table S3.2: Assembly statistics of isolates (n=90) in this study including reference genomes ATCC13124 and NCTC8239.

Table S3.3: Genes specific to NEC-associated isolates (Scoary simplified outputs).

Table S3.4: Genes specific to lineage II and III isolates (Scoary simplified outputs).

Table S3.5: Pair-wise SNP distances and ANI value (%) of lineage IIIc isolates.

Table S3.1: Clinical data from preterm neonates of 88 *C. perfringens* isolates in this study. DOL: days of life.

	Infant ID	Isolate ID	DOL	NEC/ Control	Admission Hospital
1	M009	M009.1C	15	Control	SMH
2	M010	M010.1C	16	NEC (Bell I)	SMH
3		M010.CN1	15	NEC (Bell I)	SMH
4	M011	M011.1C	16	NEC (Bell I)	SMH
5	M030	M030.1C	82	NEC (Bell I)	SMH
6		M030.N1C	71	NEC (Bell I)	SMH
7	M087	M087.10C	62	Control	SMH
8		M087.13C	73	Control	SMH
9		M087.1C	31	Control	SMH
10		M087.4C	44	Control	SMH
11		M087.7C	52	Control	SMH
12	M088	M088.10C	45	Control	SMH
13		M088.1C	17	Control	SMH
14		M088.24C	79	Control	SMH
15		M088.4C	31	Control	SMH
16	M089	M089.1C	22	Control	SMH
17		M089.3C	30	Control	SMH
18	M090	M090.1C	15	Control	SMH
19		M090.3C	24	Control	SMH
20	M091	M091.1C	39	Control	SMH
21	M092	M092.1C	35	Control	SMH
22	M094	M094.1C	33	Control	SMH
23	M096	M096.1C	18	Control	SMH
24		M096.3C	29	Control	SMH
25	M111	M111.1C	23	Control	SMH
26	M112	M112.1C	16	Control	SMH
27		M112.4C	23	Control	SMH
28	M124	M124.1C	30	Control	SMH
29	M125	M125.1C	23	Control	SMH
30		M125.4C	29	Control	SMH
31	Q020	Q020.CN1	19.5	NEC (Bell II/III)	QCCH
32		Q020.CN4	19	NEC (Bell II/III)	QCCH
33	Q087	Q087.1C	37	Control	QCCH
34	Q088	Q088.CN1	13	NEC (Bell II/III)	QCCH
35		Q088.CN12	24	NEC (Bell II/III)	QCCH
36		Q088.CN3	16	NEC (Bell II/III)	QCCH
37		Q088.CN6	17	NEC (Bell II/III)	QCCH
38		Q088.CN9	22	NEC (Bell II/III)	QCCH
39	Q090	Q090.5C	25	Control	QCCH
40		Q090.8C	32	Control	QCCH
41	Q096	Q096.10C	51	Control	QCCH

Continued on next page

	Infant ID	Isolate ID	DOL	NEC/ Control	Admission Hospital
42		Q096.13C	59	Control	QCCH
43		Q096.16C	67	Control	QCCH
44		Q096.1C	15	Control	QCCH
45		Q096.4C	36	Control	QCCH
46	Q097	Q097.1C	25	Control	QCCH
47		Q097.4C	32	Control	QCCH
48		Q097.7C	38	Control	QCCH
49	Q100	Q100.CN1	29	NEC (Bell II/III)	QCCH
50		Q100.CN10	38	NEC (Bell II/III)	QCCH
51		Q100.CN12	39	NEC (Bell II/III)	QCCH
52		Q100.CN4	32	NEC (Bell II/III)	QCCH
53		Q100.CN7	35	NEC (Bell II/III)	QCCH
54	Q101	Q101.1C	47	Control	QCCH
55	Q105	Q105.1C	41	Control	QCCH
56		Q105.4C	49	Control	QCCH
57	Q106	Q106.10C	46	Control	QCCH
58		Q106.1C	12	Control	QCCH
59		Q106.4C	33	Control	QCCH
60		Q106.7C	39	Control	QCCH
61	Q107	Q107.12C	67	Control	QCCH
62		Q107.1C	22	Control	QCCH
63		Q107.4C	37	Control	QCCH
64		Q107.7C	45	Control	QCCH
65		Q107.9C	59	Control	QCCH
66	Q118	Q118.1C	54	Control	QCCH
67		Q118.4C	60	Control	QCCH
68	Q126	Q126.1C	42	Control	QCCH
69	Q135	Q135.CN1	29	NEC (Bell II/III)	QCCH
70		Q135.CN4	34	NEC (Bell II/III)	QCCH
71	Q137	Q137.1C	19	Control	QCCH
72		Q137.4C	25	Control	QCCH
73	Q138	Q138.1C	12	Control	QCCH
74		Q138.4C	21	Control	QCCH
75	Q142	Q142.1C	53	Control	QCCH
76	Q143	Q143.1C	16	NEC (Bell II/III)	QCCH
77		Q143.CN1	13	NEC (Bell II/III)	QCCH
78		Q143.CN4	17	NEC (Bell II/III)	QCCH
79		Q143.CN8	19	NEC (Bell II/III)	QCCH
80	Q164	Q164.1C	67	Control	QCCH
81	Q167	Q167.10C	43	Control	QCCH
82		Q167.1C	13	Control	QCCH
83		Q167.4C	27	Control	QCCH

Continued on next page

	Infant ID	Isolate ID	DOL	NEC/ Control	Admission Hospital
84		Q167.7C	34	Control	QCCH
85	Q168	Q168.1C	30	Control	QCCH
86		Q168.4C	41	Control	QCCH
87		Q168.7C	50	Control	QCCH
88	Q215	Q215.CN1	40	NEC (Bell II/III)	QCCH

Table S3.2: Assembly statistics of isolates (n=90) in this study including reference genomes ATCC13124 and NCTC8239

Average Nucleotide Identities (ANI) generated by mapping with reference genome ATCC13124. ANI≥95% computationally confirms species designation [217]. Statistics were based on genome annotation by Prokka v1.13 [218]. Reference genomes are shaded in grey.

	Isolate	Genome (bp)	Reads	Depth	Contigs	N50 (bp)	G+C (%)	Gene	ANI (%)
1	ATCC13124	3,256,683	n/a	n/a	1	3,256,683	28.38	2,879	100.00
2	NCTC8239	3,008,497	n/a	n/a	2	2,940,812	28.19	2,803	97.18
3	M009.1C	3,228,292	1,329,895	102	22	351,119	28.05	2,971	98.3
4	M010.CN1	3,344,094	1,714,993	128	17	2,122,759	28.08	3,060	98.65
5	M010.1C	3,344,966	2,293,122	171	18	2,122,622	28.08	3,060	98.65
6	M011.1C	3,345,019	2,617,284	195	15	1,000,076	28.08	3,062	98.66
7	M030.1C	3,205,090	1,991,040	155	19	2,037,290	28.23	2,938	98.6
8	M030.N1C	3,207,141	1,894,558	147	16	2,036,992	28.22	2,946	98.61
9	M087.10C	3,462,641	2,325,014	167	27	323,469	28.05	3,231	98.53
10	M087.13C	3,455,084	1,967,057	142	23	405,742	28.02	3,222	98.54
11	M087.1C	3,284,487	1,889,985	143	35	205,249	28.11	3,057	98.34
12	M087.4C	3,460,509	2,131,041	153	28	405,776	28.04	3,234	98.54
13	M087.7C	3,456,430	2,159,055	156	27	405,758	28.04	3,219	98.53
14	M088.10C	3,144,583	2,377,740	189	45	159,921	28.10	2,885	98.34
15	M088.1C	3,162,957	1,990,062	157	38	148,763	28.09	2,900	98.33
16	M088.24C	3,147,539	2,126,494	168	39	159,840	28.10	2,887	98.34
17	M088.4C	3,177,376	2,439,554	191	35	243,556	28.14	2,924	98.34
18	M089.1C	3,462,801	2,122,381	153	27	405,750	28.05	3,228	98.54
19	M089.3C	3,458,059	2,189,234	158	30	405,751	28.03	3,224	98.54
20	M090.1C	3,453,643	2,391,039	173	27	405,601	28.01	3,221	98.54
21	M090.3C	3,418,228	2,106,126	154	26	405,787	28.01	3,172	98.54
22	M091.1C	3,107,914	2,024,086	162	26	256,827	28.09	2,827	98.35
23	M092.1C	3,184,687	2,471,758	194	37	225,186	28.22	2,900	98.41
24	M094.1C	3,123,459	2,482,864	198	34	173,532	28.08	2,840	98.31
25	M096.1C	3,402,370	2,118,403	155	20	981,998	28.11	3,165	99.02
26	M096.3C	3,403,754	2,134,587	156	26	381,709	28.11	3,164	99.01
27	M111.1C	3,170,976	2,081,022	164	46	185,786	28.13	2,901	98.38
28	M112.1C	3,171,762	2,227,274	175	37	177,430	28.14	2,901	98.37
29	M112.4C	3,171,262	2,168,563	170	40	196,214	28.13	2,901	98.37
30	M124.1C	3,166,878	1,727,552	136	31	187,477	28.12	2,914	98.33
31	M125.1C	3,167,259	2,151,147	169	32	266,888	28.13	2,913	98.32
32	M125.4C	3,165,453	2,400,356	189	31	268,068	28.12	2,911	98.32
33	Q020.CN4	3,240,291	2,419,945	186	19	2,028,783	28.11	2,956	98.57
34	Q020.CN1	3,242,179	2,528,412	194	16	2,027,560	28.12	2,963	98.57
35	Q087.1C	3,280,619	2,014,395	153	25	370,656	28.10	2,972	98.38
36	Q088.CN1	3,062,159	1,403,417	114	11	1,934,220	28.22	2,876	96.15
37	Q088.CN3	3,065,157	1,327,748	108	18	1,747,693	28.23	2,882	96.14
38	Q088.CN6	3,066,288	1,679,544	136	17	1,934,223	28.24	2,878	96.15
39	Q088.CN9	3,287,606	1,300,904	98	21	371,237	28.13	2,976	98.38
40	Q088.CN12	3,286,041	2,711,369	206	21	370,678	28.12	2,975	98.38

Continued on next page

	Isolate	Genome (bp)	Reads	Depth	Contigs	N50 (bp)	G+C (%)	Gene	ANI (%)
41	Q090.5C	3,194,717	1,387,128	108	14	700,330	28.14	2,900	98.6
42	Q090.8C	3,202,127	1,584,605	123	16	700,360	28.19	2,906	98.61
43	Q096.10C	3,194,624	2,480,928	194	29	247,015	28.18	2,918	98.34
44	Q096.13C	3,244,272	1,873,957	144	28	182,180	28.14	2,977	98.33
45	Q096.16C	3,190,020	1,912,783	149	27	1,650,471	28.17	2,954	96.14
46	Q096.1C	4,147,421	1,605,746	96	46	220,673	28.10	3,998	98.31
47	Q096.4C	3,309,035	2,186,855	165	21	2,087,340	28.17	3,055	98.59
48	Q097.1C	3,160,474	1,813,042	143	25	440,486	28.13	2,913	98.28
49	Q097.4C	3,154,351	2,460,896	195	21	275,007	28.12	2,914	98.27
50	Q097.7C	3,196,530	2,063,679	161	27	385,797	28.10	2,965	98.27
51	Q100.CN10	3,177,242	1,577,556	124	21	355,350	28.11	2,911	98.29
52	Q100.CN1	3,209,378	2,178,043	169	26	239,497	28.11	2,963	98.25
53	Q100.CN4	3,211,356	1,994,849	155	23	239,759	28.12	2,966	98.26
54	Q100.CN7	3,212,734	2,016,704	156	22	195,946	28.11	2,968	98.25
55	Q100.CN12	3,212,162	1,652,692	128	21	282,818	28.11	2,964	98.24
56	Q101.1C	3,231,229	1,427,677	110	22	231,485	28.17	2,981	98.4
57	Q105.1C	3,138,208	1,136,678	90	28	241,908	28.17	2,880	98.33
58	Q105.4C	3,136,273	1,385,214	110	33	231,285	28.09	2,879	98.33
59	Q106.10C	3,494,367	1,618,543	115	24	782,807	28.07	3,251	99.12
60	Q106.1C	3,125,382	1,387,392	110	39	188,844	28.12	2,838	98.33
61	Q106.4C	3,100,264	1,439,868	116	20	1,962,756	28.28	2,934	96.15
62	Q106.7C	3,092,582	1,314,111	106	15	1,966,081	28.27	2,925	96.14
63	Q107.12C	3,295,802	1,243,651	94	16	2,107,591	28.11	2,996	98.62
64	Q107.1C	3,143,128	1,281,583	101	27	243,728	28.14	2,878	98.3
65	Q107.4C	3,270,656	1,445,861	110	29	321,580	28.13	2,990	98.4
66	Q107.7C	3,133,904	1,426,136	113	21	352,768	28.16	2,869	98.35
67	Q107.9C	3,299,946	1,771,584	134	15	2,108,040	28.12	2,999	98.61
68	Q118.1C	3,333,932	1,483,468	111	24	378,475	28.04	3,056	98.41
69	Q118.4C	3,374,700	1,403,074	103	30	378,470	28.03	3,094	98.41
70	Q126.1C	3,203,802	1,827,538	142	14	1,945,226	28.09	3,029	96.09
71	Q135.CN1	3,156,272	1,509,457	119	32	188,322	28.13	2,900	98.3
72	Q135.CN4	3,154,909	1,035,124	82	38	188,305	28.13	2,901	98.29
73	Q137.1C	3,459,480	1,613,628	116	26	406,919	28.03	3,223	98.54
74	Q137.4C	3,456,126	1,603,842	116	24	405,596	28.03	3,226	98.54
75	Q138.1C	3,460,229	1,687,249	121	23	405,744	28.04	3,233	98.53
76	Q138.4C	3,459,336	1,555,756	112	21	405,744	28.03	3,225	98.54
77	Q142.1C	3,146,959	1,528,544	121	36	215,934	28.16	2,862	98.3
78	Q143.1C	3,459,999	1,804,361	130	21	405,770	28.04	3,227	98.53
79	Q143.CN1	3,460,998	1,725,528	124	25	405,749	28.04	3,230	98.54
80	Q143.CN4	3,458,736	1,618,971	117	28	405,767	28.03	3,224	98.53
81	Q143.CN8	3,460,283	1,607,717	116	24	405,740	28.04	3,228	98.53
82	Q164.1C	3,360,445	1,713,473	127	19	2,084,408	28.11	3,114	98.94
83	Q167.10C	3,193,775	1,778,950	139	21	582,989	28.14	2,934	98.29
84	Q167.1C	3,191,399	1,952,892	152	17	582,691	28.15	2,930	98.31
85	Q167.4C	3,192,114	1,912,392	149	18	582,697	28.15	2,932	98.3
86	Q167.7C	3,187,987	1,770,156	138	18	2,030,903	28.14	2,927	98.3
87	Q168.1C	3,172,313	1,763,212	138	59	115,338	28.09	2,856	98.3

Continued on next page

	Isolate	Genome (bp)	Reads	Depth	Contigs	N50 (bp)	G+C (%)	Gene	ANI (%)
88	Q168.4C	3,183,576	1,933,586	151	16	769,066	28.13	2,921	98.3
89	Q168.7C	3,195,069	1,772,728	138	13	871,716	28.16	2,939	98.3
90	Q215.CN1	3,282,615	1,824,997	138	24	572,833	28.09	3,003	98.69
Stats	**Mean**	3,263,705	1,858,221	142	25	752,353	28.12	3,009	98.27
	Max	4,147,421	2,711,369	206	59	3,256,683	28.38	3,998	100.00
	Min	3,008,497	1,035,124	82	1	115,338	28.01	2,803	96.09

Table S3.3: Genes specific to NEC-associated isolates (Scoary simplified outputs)

Gene	Annotation	Sensitivity	Specificity	Odds ratio
group_3210	hypothetical protein	75.00	80.77	12.60
group_434	membrane protein	91.67	73.08	29.86
group_578	hypothetical protein	91.67	73.08	29.86
group_1083	N-acetyltransferase	83.33	71.79	12.73
group_2242	hypothetical protein	91.67	67.95	23.32
group_583	glycosyl hydrolase family 25	75.00	67.95	6.36
group_1523	hypothetical protein	75.00	66.67	6.00
group_2185	hypothetical protein	75.00	66.67	6.00
group_412	hypothetical protein	75.00	66.67	6.00
group_1493	hypothetical protein	75.00	65.38	5.67
group_1495	hypothetical protein	75.00	64.10	5.36
group_2947	hypothetical protein	75.00	64.10	5.36
group_2186	hypothetical protein	75.00	64.10	5.36
group_1050	helix-turn-helix domain-containing protein	75.00	64.10	5.36
group_1053	membrane protein	75.00	64.10	5.36
group_2943	ParA family protein	75.00	64.10	5.36
group_1522	hypothetical protein	75.00	64.10	5.36
group_1052	CAP domain-containing protein	75.00	64.10	5.36
group_1499	hypothetical protein	75.00	64.10	5.36
group_2946	hypothetical protein	75.00	64.10	5.36
group_428	hypothetical protein	75.00	64.10	5.36
group_1049	chromosome partitioning protein ParB	75.00	64.10	5.36
group_338	thioester-forming surface-anchored protein	75.00	64.10	5.36
group_2926	pilus biosynthesis protein HicB	75.00	64.10	5.36
group_163	recombinase	75.00	64.10	5.36
group_264	ImmA/IrrE family metallo-endopeptidase	75.00	64.10	5.36
group_1498	class D sortase	75.00	64.10	5.36
group_1497	type II toxin-antitoxin system PemK/MazF family toxin	75.00	64.10	5.36
group_2179	hypothetical protein	75.00	64.10	5.36
group_431	hypothetical protein	75.00	64.10	5.36
group_2178	hypothetical protein	75.00	64.10	5.36
group_2949	hypothetical protein	75.00	64.10	5.36
group_1501	type VI secretion protein	75.00	64.10	5.36
group_2184	DNA-binding response regulator	75.00	64.10	5.36
group_2927	addiction module toxin, HicA family	75.00	64.10	5.36
group_2920	cell wall-binding protein	75.00	64.10	5.36
group_2948	DNA topoisomerase III	75.00	64.10	5.36
spo0A	sporulation transcription factor Spo0A	75.00	64.10	5.36

Continued on next page

Gene	Annotation	Sensitivity	Specificity	Odds ratio
group_1494	hypothetical protein	75.00	64.10	5.36
group_1048	hypothetical protein	75.00	64.10	5.36
group_2714	sugar ABC transporter substrate-binding protein	75.00	62.82	5.07
group_3131	HAMP domain-containing protein	75.00	62.82	5.07
group_219	hypothetical protein	75.00	62.82	5.07
group_1830	DNA-binding response regulator	75.00	62.82	5.07
group_2715	sugar ABC transporter permease	75.00	62.82	5.07
group_1831	hypothetical protein	75.00	62.82	5.07
group_576	hypothetical protein	75.00	62.82	5.07
group_1262	PIG-L family deacetylase	75.00	62.82	5.07
group_1059	glycoside hydrolase family 65 protein	75.00	62.82	5.07
group_2717	endo-beta-galactosidase	75.00	62.82	5.07
group_2716	carbohydrate ABC transporter permease	75.00	62.82	5.07
group_2773	NusG domain II-containing protein	100.00	61.54	inf
group_2772	heptaprenyl diphosphate synthase	100.00	61.54	inf
group_552	S-layer homology domain-containing protein	100.00	61.54	inf
group_2054	FAD:protein FMN transferase	100.00	61.54	inf
group_1703	M protein trans-acting positive regulator (MGA)	100.00	60.26	inf

Table S3.4: Genes specific to lineage II and III isolates (Scoary simplified outputs)

Gene	Annotation	Sensitivity	Specificity	Odds ratio
group_412	hypothetical protein	92.11	100.00	inf
group_1174	hypothetical protein	94.74	98.08	918.00
group_2717	endo-beta-galactosidase	92.11	94.23	190.56
group_1830	DNA-binding response regulator	92.11	94.23	190.56
group_3131	HAMP domain-containing protein	92.11	94.23	190.56
group_2715	sugar ABC transporter permease	92.11	94.23	190.56
group_1059	glycoside hydrolase family 65 protein	92.11	94.23	190.56
group_1262	PIG-L family deacetylase	92.11	94.23	190.56
group_2714	sugar ABC transporter substrate-binding protein	92.11	94.23	190.56
group_1831	hypothetical protein	92.11	94.23	190.56
group_2716	carbohydrate ABC transporter permease	92.11	94.23	190.56
group_431	hypothetical protein	89.47	94.23	138.83
group_1523	hypothetical protein	84.21	94.23	87.11
group_2584	phosphoserine phosphatase	97.37	92.31	444.00
group_2596	Crp/Fnr family transcriptional regulator	97.37	92.31	444.00
group_2594	purine-nucleoside phosphorylase	97.37	92.31	444.00
group_2595	nicotinamide riboside transporter PnuC	97.37	92.31	444.00
group_145	hypothetical protein	97.37	90.38	347.80
group_2947	hypothetical protein	84.21	90.38	50.13
group_1499	hypothetical protein	84.21	90.38	50.13
group_2920	cell wall-binding protein	84.21	90.38	50.13
group_338	thioester-forming surface-anchored protein	84.21	90.38	50.13
group_1495	hypothetical protein	84.21	90.38	50.13
group_2179	hypothetical protein	84.21	90.38	50.13
group_2178	hypothetical protein	84.21	90.38	50.13
group_1049	chromosome partitioning protein ParB	84.21	90.38	50.13
group_2184	DNA-binding response regulator	84.21	90.38	50.13
group_2946	hypothetical protein	84.21	90.38	50.13
group_1494	hypothetical protein	84.21	90.38	50.13
group_163	recombinase	84.21	90.38	50.13
group_2949	hypothetical protein	84.21	90.38	50.13
group_1501	type VI secretion protein	84.21	90.38	50.13
group_2948	DNA topoisomerase III	84.21	90.38	50.13
group_1048	hypothetical protein	84.21	90.38	50.13
group_1497	type II toxin-antitoxin system PemK/MazF family toxin	84.21	90.38	50.13
group_428	hypothetical protein	84.21	90.38	50.13

Continued on next page

Gene	Annotation	Sensitivity	Specificity	Odds ratio
group_1050	helix-turn-helix domain-containing protein	84.21	90.38	50.13
group_264	ImmA/IrrE family metallo-endopeptidase	84.21	90.38	50.13
group_2186	hypothetical protein	84.21	90.38	50.13
spo0A	sporulation transcription factor Spo0A	84.21	90.38	50.13
group_2943	ParA family protein	84.21	90.38	50.13
group_1522	hypothetical protein	84.21	90.38	50.13
group_1052	CAP domain-containing protein	84.21	90.38	50.13
group_2926	pilus biosynthesis protein HicB	84.21	90.38	50.13
group_1498	class D sortase	84.21	90.38	50.13
group_1053	membrane protein	84.21	90.38	50.13
group_2927	addiction module toxin, HicA family	84.21	90.38	50.13
group_1493	hypothetical protein	81.58	90.38	41.63
group_2765	ABC transporter ATP-binding protein	97.37	86.54	237.86
group_2764	ABC transporter ATP-binding protein	97.37	86.54	237.86
group_840	hypothetical protein	94.74	86.54	115.71
group_2637	RNA polymerase subunit sigma-24	94.74	86.54	115.71
group_552	S-layer homology domain-containing protein	92.11	86.54	75.00
group_2773	NusG domain II-containing protein	92.11	86.54	75.00
group_2772	heptaprenyl diphosphate synthase	92.11	86.54	75.00
group_2054	FAD:protein FMN transferase	92.11	86.54	75.00
group_841	zinc transporter ZupT	97.37	84.62	203.50
group_1703	M protein trans-acting positive regulator (MGA)	92.11	84.62	64.17
group_654	CotH protein	97.37	80.77	155.40
group_1845	response regulator	97.37	80.77	155.40
group_1844	potassium-transporting ATPase A chain	97.37	80.77	155.40
group_2726	DNA-binding response regulator	97.37	80.77	155.40
group_2718	VTC domain-containing protein	97.37	80.77	155.40
group_2725	potassium-transporting ATPase subunit B	97.37	80.77	155.40
group_877	sensor histidine kinase	97.37	80.77	155.40
group_2719	hypothetical protein	97.37	80.77	155.40
group_662	potassium-transporting ATPase subunit KdpC	97.37	80.77	155.40
group_881	sensor histidine kinase KdpD	97.37	80.77	155.40
group_1835	DNA-binding response regulator	97.37	80.77	155.40
group_1836	hypothetical protein	94.74	80.77	75.60
group_575	hypothetical protein	86.84	80.77	27.72
group_2575	thiol-activated toxin perfringolysin O	92.11	78.85	43.48

Continued on next page

Gene	Annotation	Sensitivity	Specificity	Odds ratio
group_1228	DNA-binding response regulator	100.00	76.92	inf
group_506	diacylglyceryl transferase	97.37	76.92	123.33
group_882	FUSC family protein	97.37	76.92	123.33
group_836	hypothetical protein	94.74	76.92	60.00
group_347	3-phosphoglycerate dehydrogenase	100.00	75.00	inf
group_165	hypothetical protein	100.00	75.00	inf
group_2557	alanine--glyoxylate aminotransferase family protein	100.00	75.00	inf
group_729	hypothetical protein	97.37	75.00	111.00
group_370	FAD-dependent oxidoreductase	81.58	75.00	13.29
group_236	hypothetical protein	84.21	73.08	14.48

Table S3.5: Pair-wise SNP distances and ANI value (%) of lineage IIIc isolates

Pair-wise ANI comparison was performed via mummer alignment in relation to isolate Q143.CN8.

Isolate ID	M087.10C	M087.13C	M087.4C	M087.7C	M089.1C	M089.3C	M090.1C	M090.3C	Q137.1C	Q137.4C	Q138.1C	Q138.4C	Q143.1C	Q143.CN1	Q143.CN4	Q143.CN8	ANI (%)
M087.10C	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
M087.13C	2	0	2	3	2	2	4	2	2	2	2	3	4	2	2	3	99.99
M087.4C	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
M087.7C	1	3	1	0	1	1	3	1	1	1	1	2	3	1	1	2	99.99
M089.1C	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
M089.3C	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
M090.1C	2	4	2	3	2	2	0	2	2	2	2	3	4	2	2	3	99.99
M090.3C	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
Q137.1C	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
Q137.4C	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
Q138.1C	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
Q138.4C	1	3	1	2	1	1	3	1	1	1	1	0	3	1	1	2	99.99
Q143.1C	2	4	2	3	2	2	4	2	2	2	2	3	0	2	2	3	99.99
Q143.CN1	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
Q143.CN4	0	2	0	1	0	0	2	0	0	0	0	1	2	0	0	1	99.99
Q143.CN8	1	3	1	2	1	1	3	1	1	1	1	2	3	1	1	0	100.00

Appendix 2 Supplementary data for Chapter 4

This appendix contains all the supplementary materials for Chapter 4.

Table S4.1: Clinical and genomic (assembly and annotation) data of 109 outbreak-associated *C. perfringens* newly-sequenced isolates in this study.

Table S4.2: Pair-wise SNP distance (from core-gene alignment) among 18 CH isolates in lineage IVc recovered from 9 outbreaks between 2013-2017.

Table S4.3: Pair-wise SNP distance among clonal FP isolates within lineage VII.

Table S4.4: Comparison of subset-specific accessory genes based on EggNog annotations.

Table S4.1: Clinical and genomic (assembly and annotation) data of 109 outbreak-associated *C. perfringens* newly-sequenced isolates in this study

ANI (Average Nucleotide Identity) was compared with reference genome (RefSeq) *C. perfringens* NCTC8239. Cohort: CH represents care home-associated; FP denotes food poisoning isolates.

	Clinical metadata					Sequencing information				Genome assembly and annotation statistics								PCR
	Isolate	Source	Year	Region	Cohort	Out-break	Illumina sequencing platform	Read length (bp)	Reads (bp)	Sequencing coverage (X)	GC (%)	Contig number	Genome Size (bp)	tRNA	rRNA	CDS	Gene	ANI	PCR *cpe*
1	PH015	Individual	2013	North east	CH	1	HiSeq	151	3854793	377	28.01	100	3081476	84	13	2731	2829	98.70	positive
2	PH096	Individual 1	2013	North east	CH	1	MiSeq	101	1267411	74	28.08	239	3422714	58	7	3105	3171	97.32	positive
3	PH016	Individual 1	2014	North east	CH	2	HiSeq	151	2853518	244	28.07	66	3518998	84	12	3206	3303	97.29	positive
4	PH017	Individual 2	2014	North east	CH	2	HiSeq	151	2625370	240	28.18	23	3291535	72	13	2935	3021	97.15	negative
5	PH018	Individual 2	2014	North east	CH	2	HiSeq	151	2325775	202	28.1	26	3475247	79	11	3164	3255	97.30	positive
6	PH019	Individual 3	2014	North east	CH	2	HiSeq	151	2615637	226	28.14	25	3487876	87	11	3170	3269	97.30	positive
7	PH020	Individual 1	2014	North east	CH	3	HiSeq	151	2483432	220	28.05	33	3397222	75	11	3075	3162	97.29	positive
8	PH021	Individual 2	2014	North east	CH	3	HiSeq	151	3350806	293	28	31	3445998	74	11	3144	3230	97.28	positive
9	PH097	Individual 3	2014	North east	CH	3	MiSeq	101	1138038	68	28.12	398	3371290	29	3	3049	3082	97.36	positive
10	PH098	Individual 1	2014	North east	CH	4	MiSeq	101	1500399	87	28.08	245	3454553	58	6	3137	3202	97.34	positive
11	PH100	Individual 1	2014	North east	CH	5	MiSeq	101	1457550	87	28.15	184	3354426	56	8	3010	3075	97.32	positive
12	PH101	Individual 2	2014	North east	CH	5	MiSeq	101	2004838	116	28.04	87	3481641	60	7	3179	3247	97.30	positive
13	PH102	Individual 1	2014	North east	CH	6	MiSeq	101	1476804	87	28.04	311	3420587	60	7	3110	3178	97.32	positive
14	PH103	Individual 2	2014	North east	CH	6	MiSeq	101	860585	50	28.07	274	3448541	42	5	3128	3176	97.33	positive
15	PH030	Individual 1	2014	North east	CH	7	HiSeq	151	2567442	226	28.07	27	3427777	67	8	3092	3168	97.30	positive
16	PH031	Individual 1	2014	North east	CH	7	HiSeq	151	2606506	229	28.08	21	3430445	74	10	3093	3178	97.31	positive
17	PH032	Individual 2	2014	North east	CH	7	HiSeq	151	2889707	277	27.94	109	3139693	78	13	2827	2919	98.70	positive
18	PH033	Individual 2	2014	North east	CH	7	HiSeq	151	2287790	218	27.99	105	3163229	70	9	2850	2930	98.70	positive
19	PH051	Individual 1	2015	North east	CH	8	HiSeq	151	2218172	192	28.09	27	3480399	92	11	3155	3259	97.21	positive
20	PH052	Individual 2	2015	North east	CH	8	HiSeq	151	2665002	231	28.07	34	3481680	74	8	3188	3271	97.30	positive
21	PH053	Individual 3	2015	North east	CH	8	HiSeq	151	2587606	224	28.09	25	3484752	95	8	3188	3292	97.31	positive
22	PH072	Individual 2	2016	North east	CH	9	HiSeq	151	2510560	217	28.01	59	3492601	86	12	3153	3252	97.27	positive
23	PH073	Individual 3	2016	North east	CH	9	HiSeq	151	2426321	207	28.07	35	3537414	92	10	3246	3349	97.30	positive
24	PH074	Individual 4	2016	North east	CH	9	HiSeq	151	2486225	213	28.07	26	3522802	93	11	3229	3334	97.31	positive
25	PH075	Individual 5	2016	North east	CH	9	HiSeq	151	2353413	200	28.08	29	3538037	95	13	3250	3359	97.30	positive
26	PH076	Individual 6	2016	North east	CH	9	HiSeq	151	1995409	174	28.05	23	3450640	83	11	3121	3216	97.31	positive
27	PH077	Individual 7	2016	North east	CH	9	HiSeq	151	2455240	212	28.11	23	3491708	89	10	3178	3278	97.30	positive
28	PH138	Individual 1	2016	Wales	CH	10	MiSeq	101	1559472	92	28.06	222	3401096	58	9	3056	3124	97.33	positive
29	PH139	Individual 2	2016	Wales	CH	10	MiSeq	101	1111879	66	28.04	97	3402225	60	11	3069	3141	97.30	positive
30	PH140	Individual 3	2016	Wales	CH	10	MiSeq	101	1264793	75	28.06	257	3386184	34	5	3043	3083	97.33	positive
31	PH082	Individual 1	2017	North east	CH	11	HiSeq	151	2961546	263	28.11	30	3394345	90	10	3065	3166	97.18	positive
32	PH083	Individual 2	2017	North east	CH	11	HiSeq	151	1865948	165	28.07	21	3410579	89	12	3088	3190	97.18	positive
33	PH084	Individual 3	2017	North east	CH	11	HiSeq	151	1964261	174	28.1	22	3392793	95	11	3067	3174	97.18	positive

Continued on next page

	Clinical metadata					Sequencing information				Genome assembly and annotation statistics								PCR
	Isolate	Source	Year	Region	Cohort	Out-break	Illumina sequencing platform	Read length (bp)	Reads (bp)	Sequencing coverage (X)	GC (%)	Contig number	Genome Size (bp)	tRNA	rRNA	CDS	Gene	ANI	PCR *cpe*
34	PH085	Individual 4	2017	North east	CH	11	HiSeq	151	2764341	245	28.11	23	3394191	88	11	3064	3164	97.18	positive
35	PH079	Individual	2017	North east	CH	Sporadic	HiSeq	151	2568032	221	28.1	26	3496302	91	10	3184	3286	97.3	positive
36	PH007	Curry goat	2011	South East	FP	1	HiSeq	151	2776491	284	27.93	81	2946980	74	13	2660	2748	99.42	positive
37	PH004	Individual 1	2011	South East	FP	1	HiSeq	151	2522471	260	27.89	95	2923443	89	12	2633	2735	99.43	positive
38	PH006	Individual 3	2011	South East	FP	1	HiSeq	151	2763075	284	27.87	95	2928275	91	13	2635	2740	99.42	positive
39	PH005	Individual 2	2011	South East	FP	1	HiSeq	151	2827331	291	27.89	96	2934006	87	14	2640	2742	99.42	positive
40	PH014	Individual 2	2013	London	FP	2	HiSeq	151	2781207	245	28.11	42	3423911	89	11	3074	3175	97.19	positive
41	PH092	Individual 2	2013	London	FP	2	MiSeq	101	702080	41	28.12	378	3378829	38	6	3006	3051	97.22	positive
42	PH095	Individual 5	2013	London	FP	2	MiSeq	101	1024205	61	28.17	446	3348605	34	3	2999	3037	97.26	positive
43	PH091	Individual 1	2013	London	FP	2	MiSeq	101	881394	53	28.17	491	3337668	35	4	2968	3008	97.25	positive
44	PH024	Individual 3	2014	North east	FP	3	HiSeq	151	3024896	274	28.03	28	3331444	90	12	3016	3119	97.23	positive
45	PH025	Individual 4	2014	North east	FP	3	HiSeq	151	2413851	219	28	34	3320821	72	7	3002	3082	97.23	positive
46	PH023	Individual 2	2014	North east	FP	3	HiSeq	151	2909655	264	28.01	35	3325484	88	7	3002	3098	97.23	positive
47	PH022	Individual 1	2014	North east	FP	3	HiSeq	151	2319050	205	28.01	36	3401212	74	12	3077	3164	97.24	positive
48	PH026	Cooked sliced beef	2014	North east	FP	3	HiSeq	151	2769966	250	28.07	37	3335204	91	12	3009	3113	97.24	positive
49	PH028	Individual 4	2014	London	FP	4	HiSeq	151	3031314	305	27.96	82	2994736	75	10	2731	2817	99.64	positive
50	PH029	Individual 5	2014	London	FP	4	HiSeq	151	3106630	317	27.92	97	2954967	87	12	2694	2794	99.65	positive
51	PH104	Individual 1	2014	London	FP	4	MiSeq	101	1072977	73	27.9	214	2929035	43	4	2668	2716	99.64	positive
52	PH107	Individual 6	2014	London	FP	4	MiSeq	101	877265	60	27.91	246	2914363	38	5	2650	2694	99.66	positive
53	PH105	Individual 3	2014	North East	FP	4	MiSeq	101	1022558	70	27.93	379	2944409	57	7	2672	2737	99.65	positive
54	PH115	Individual 2	2014	North east	FP	5	MiSeq	101	1537158	104	27.87	317	2977232	68	6	2694	2769	99.51	positive
55	PH118	Individual 2	2015	North East	FP	6	MiSeq	101	1632880	100	28.01	214	3282569	62	7	2917	2987	97.21	positive
56	PH119	Individual 3	2015	North East	FP	6	MiSeq	101	1287449	79	28	215	3289115	66	6	2913	2986	97.20	positive
57	PH117	Individual 1	2015	North East	FP	6	MiSeq	101	1569491	96	28.02	246	3285816	74	10	2917	3002	97.22	positive
58	PH039	Individual 4	2015	North East	FP	7	HiSeq	151	2709205	238	28.13	23	3432249	86	11	3094	3192	97.22	positive
59	PH040	Individual 5	2015	North East	FP	7	HiSeq	151	2211747	194	28.12	25	3426487	87	11	3089	3188	97.21	positive
60	PH037	Individual 3	2015	North East	FP	7	HiSeq	151	2113659	186	28.12	26	3426315	71	10	3088	3170	97.21	positive
61	PH044	Individual 8	2015	North East	FP	7	HiSeq	151	1993382	175	28.12	29	3426828	83	9	3088	3181	97.21	positive
62	PH036	Individual 2	2015	North East	FP	7	HiSeq	151	2238843	197	28.14	30	3430269	88	13	3090	3192	97.21	positive
63	PH041	Individual 6	2015	North East	FP	7	HiSeq	151	1997998	175	28.13	31	3433511	88	12	3099	3200	97.21	positive
64	PH035	Individual 1	2015	North East	FP	7	HiSeq	151	1947365	171	28.11	32	3422738	86	10	3081	3178	97.21	positive
65	PH045	Individual 8	2015	North East	FP	7	HiSeq	151	2313617	218	28.07	32	3191860	90	9	2864	2964	97.23	negative
66	PH042	Individual 11	2015	North East	FP	7	HiSeq	151	1766545	155	28.12	34	3427997	87	13	3087	3188	97.21	positive

Continued on next page

	Clinical metadata						Sequencing information				Genome assembly and annotation statistics								PCR
	Isolate	Source	Year	Region	Cohort	Out-break	Illumina sequencing platform	Read length (bp)	Reads (bp)	Sequencing coverage (X)	GC (%)	Contig number	Genome Size (bp)	tRNA	rRNA	CDS	Gene	ANI	PCR *cpe*
67	PH048	Individual 12	2015	North East	FP	7	HiSeq	151	2875352	253	28.13	34	3428504	92	8	3087	3188	97.21	positive
68	PH046	Individual 9	2015	North East	FP	7	HiSeq	151	2091933	184	28.11	35	3425020	93	9	3080	3183	97.20	positive
69	PH043	Individual 7	2015	North East	FP	7	HiSeq	151	2248776	198	28.1	43	3421670	69	8	3081	3159	97.20	positive
70	PH047	Individual 10	2015	North East	FP	7	HiSeq	151	2214491	195	28.11	43	3422937	71	9	3084	3165	97.21	positive
71	PH154	Individual	2015	North East	FP	7	MiSeq	101	1800754	106	28.11	177	3413080	73	8	3065	3147	97.22	positive
72	PH155	Individual 8	2015	North East	FP	7	MiSeq	101	1230813	72	28.11	209	3411087	58	7	3069	3135	97.23	positive
73	PH156	Individual 8	2015	North East	FP	7	MiSeq	101	1206348	76	28.07	239	3172681	51	5	2846	2903	97.25	negative
74	PH050	Individual 1	2015	South West	FP	8	HiSeq	151	2468557	224	28.13	59	3315627	93	12	2944	3050	97.21	positive
75	PH120	Individual 2	2015	South West	FP	8	MiSeq	101	1102939	68	28.05	185	3274008	40	4	2894	2939	97.22	positive
76	PH121	Lamb	2015	South West	FP	8	MiSeq	101	1382732	86	28.08	289	3233465	38	8	2854	2901	97.22	positive
77	PH122	Lamb	2015	South West	FP	8	MiSeq	101	1008241	62	28.11	371	3282469	70	7	2901	2979	97.24	positive
78	PH055	Individual 3	2015	North east	FP	9	HiSeq	151	3085261	270	28	27	3441127	76	9	3130	3216	97.28	positive
79	PH054	Chicken curry	2015	North east	FP	9	HiSeq	151	2270852	197	28.03	57	3476853	87	12	3153	3253	97.27	positive
80	PH124	Individual 2	2015	North east	FP	9	MiSeq	101	1394099	82	27.99	132	3431905	46	5	3120	3172	97.29	positive
81	PH123	Individual 1	2015	North east	FP	9	MiSeq	101	1445464	85	28	156	3429751	67	5	3113	3186	97.29	positive
82	PH057	Individual 2	2015	North east	FP	10	HiSeq	151	2899190	257	28.12	18	3398027	92	10	3037	3140	97.20	positive
83	PH066	3 birds roast	2015	North east	FP	10	HiSeq	151	2252903	200	28.13	18	3398355	90	11	3037	3139	97.20	positive
84	PH063	Individual 8	2015	North east	FP	10	HiSeq	151	2652135	235	28.11	19	3395429	75	10	3037	3123	97.19	positive
85	PH060	Individual 5	2015	North east	FP	10	HiSeq	151	2421609	215	28.13	20	3399028	89	12	3038	3140	97.19	positive
86	PH061	Individual 6	2015	North east	FP	10	HiSeq	151	2774093	246	28.12	20	3396723	89	10	3037	3137	97.19	positive
87	PH065	Individual 10	2015	North east	FP	10	HiSeq	151	2302879	204	28.1	20	3392964	93	8	3037	3139	97.19	positive
88	PH062	Individual 7	2015	North east	FP	10	HiSeq	151	2623765	233	28.12	21	3397544	87	9	3037	3134	97.20	positive
89	PH067	Individual 11	2015	North east	FP	10	HiSeq	151	2428595	216	28.1	21	3392663	75	8	3036	3120	97.18	positive
90	PH058	Individual 3	2015	North east	FP	10	HiSeq	151	2571555	228	28.12	22	3396864	83	12	3036	3132	97.19	positive
91	PH056	Individual 1	2015	North east	FP	10	HiSeq	151	2422080	215	28.11	23	3395441	80	10	3037	3128	97.19	positive
92	PH068	Individual 12	2015	North east	FP	10	HiSeq	151	2060812	183	28.13	23	3399957	87	10	3037	3135	97.20	positive
93	PH059	Individual 4	2015	North east	FP	10	HiSeq	151	2864609	254	28.11	24	3395953	89	9	3037	3136	97.19	positive
94	PH127	Individual 3	2016	Newcastle	FP	11	MiSeq	101	1014192	70	27.93	262	2924057	50	7	2649	2707	99.47	positive
95	PH125	Individual 1	2015	Newcastle	FP	11	MiSeq	101	1011178	59	28.09	363	3406444	27	3	3089	3120	97.31	positive
96	PH126	Individual 2	2016	Newcastle	FP	11	MiSeq	101	1153871	68	28.09	418	3408431	32	4	3077	3114	97.31	positive
97	PH134	Cooked turkey	2016	West Midlands	FP	12	MiSeq	101	1617315	94	28.08	192	3446865	76	8	3077	3162	97.17	positive
98	PH130	Individual 3	2016	West Midlands	FP	12	MiSeq	101	1374433	80	28.09	250	3446961	68	7	3078	3154	97.16	positive
99	PH129	Individual 2	2016	West Midlands	FP	12	MiSeq	101	1599870	93	28.09	266	3441457	57	6	3070	3134	97.17	positive

Continued on next page

	Clinical metadata						Sequencing information				Genome assembly and annotation statistics								PCR
	Isolate	Source	Year	Region	Cohort	Out-break	Illumina sequencing platform	Read length (bp)	Reads (bp)	Sequencing coverage (X)	GC (%)	Contig number	Genome Size (bp)	tRNA	rRNA	CDS	Gene	ANI	PCR *cpe*
100	PH133	Cooked meat	2016	West Midlands	FP	12	MiSeq	101	1833668	107	28.11	347	3432746	54	6	3055	3116	97.18	positive
101	PH128	Individual 1	2016	West Midlands	FP	12	MiSeq	101	1316831	77	28.09	352	3436609	45	8	3064	3118	97.20	positive
102	PH132	Individual 6	2016	West Midlands	FP	12	MiSeq	101	1569588	92	28.14	466	3418685	57	8	3029	3095	97.21	positive
103	PH136	Individual 2	2016	London	FP	13	MiSeq	101	1317728	91	27.91	312	2905546	45	5	2638	2689	99.40	positive
104	PH135	Individual 1	2016	London	FP	13	MiSeq	101	1152631	79	27.9	356	2912738	40	6	2644	2691	99.40	positive
105	PH081	Individual 3	2017	North West	FP	14	HiSeq	151	2085111	211	27.89	67	2984289	91	11	2729	2832	99.47	positive
106	PH078	Individual 1	2017	North West	FP	14	HiSeq	151	2152453	217	27.92	70	2987740	89	13	2727	2830	99.47	positive
107	PH080	Individual 2	2017	North West	FP	14	HiSeq	151	2501153	253	27.9	73	2983246	89	11	2718	2819	99.47	positive
108	PH143	Individual 1	2016	South West	FP	sporadic	MiSeq	101	1333457	76	28.05	256	3528864	56	6	3176	3239	97.24	positive
109	PH147	Individual 1	2013	North East	FP	sporadic	MiSeq	101	1204345	74	28.03	351	3267085	53	6	2918	2978	97.28	positive

Table S4.2: Pair-wise SNP distance (from core-gene alignment) among 18 CH isolates in lineage IVc recovered from 9 outbreaks between 2013-2017

Isolate ID	PH016	PH018	PH019	PH030	PH031	PH052	PH053	PH073	PH074	PH075	PH076	PH077	PH079	PH096	PH097	PH098	PH101	PH103
PH016	0	9	6	19	17	21	17	29	23	25	23	25	61	39	54	22	25	52
PH018	9	0	7	20	18	22	18	30	24	26	24	26	62	40	55	23	26	53
PH019	6	7	0	17	15	19	15	27	21	23	21	23	59	37	52	20	23	50
PH030	19	20	17	0	2	20	16	16	10	12	10	12	50	28	43	11	14	41
PH031	17	18	15	2	0	18	14	14	8	10	8	10	48	26	41	9	12	39
PH052	21	22	19	20	18	0	6	30	24	26	24	26	62	40	55	23	26	53
PH053	17	18	15	16	14	6	0	26	20	22	20	22	58	36	51	19	22	49
PH073	29	30	27	16	14	30	26	0	20	4	20	6	56	38	53	21	24	51
PH074	23	24	21	10	8	24	20	20	0	16	6	16	54	32	47	15	18	45
PH075	25	26	23	12	10	26	22	4	16	0	16	2	52	34	49	17	20	47
PH076	23	24	21	10	8	24	20	20	6	16	0	16	54	32	47	15	18	45
PH077	25	26	23	12	10	26	22	6	16	2	16	0	52	34	49	17	20	47
PH079	61	62	59	50	48	62	58	56	54	52	54	52	0	54	77	53	54	45
PH096	39	40	37	28	26	40	36	38	32	34	32	34	54	0	55	31	32	45
PH097	54	55	52	43	41	55	51	53	47	49	47	49	77	55	0	46	47	68
PH098	22	23	20	11	9	23	19	21	15	17	15	17	53	31	46	0	17	44
PH101	25	26	23	14	12	26	22	24	18	20	18	20	54	32	47	17	0	45
PH103	52	53	50	41	39	53	49	51	45	47	45	47	45	45	68	44	45	0

Table S4.3: Pair-wise SNP distance among clonal FP isolates within lineage VII

Colour-coded cells represent individual outbreaks. Empty cells represent SNP pair-wise distances range 900-1700 SNPs.

	PH014	PH091	PH092	PH095	PH035	PH036	PH037	PH039	PH040	PH041	PH042	PH043	PH044	PH046	PH047	PH048	PH154	PH155	PH056	PH057	PH058	PH059	PH060	PH061	PH062	PH063	PH065	PH066	PH067	PH068	PH143
PH014	0	14	6	10																											
PH091	14	0	12	16																											
PH092	6	12	0	8																											
PH095	10	16	8	0																											
PH035					0	1	1	3	2	1	4	3	4	2	1	0	1	4													
PH036					1	0	0	2	1	2	3	2	3	3	2	1	0	3													
PH037					1	0	0	2	1	2	3	2	3	3	2	1	0	3													
PH039					3	2	2	0	3	4	5	4	5	5	4	3	2	5													
PH040					2	1	1	3	0	3	4	3	4	4	3	2	1	4													
PH041					1	2	2	4	3	0	5	4	5	3	2	1	2	5													
PH042					4	3	3	5	4	5	0	5	6	6	5	4	3	6													
PH043					3	2	2	4	3	4	5	0	5	5	4	3	2	5													
PH044					4	3	3	5	4	5	6	5	0	6	5	4	3	2													
PH046					2	3	3	5	4	3	6	5	6	0	3	2	3	6													
PH047					1	2	2	4	3	2	5	4	5	3	0	1	2	5													
PH048					0	1	1	3	2	1	4	3	4	2	1	0	1	4													
PH154					1	0	0	2	1	2	3	2	3	3	2	1	0	3													
PH155					4	3	3	5	4	5	6	5	2	6	5	4	3	0													
PH056																			0	2	3	3	3	1	1	1	1	2	1	2	
PH057																			2	0	3	3	3	1	1	1	1	2	1	2	
PH058																			3	3	0	4	4	2	2	2	2	3	2	3	
PH059																			3	3	4	0	4	2	2	2	2	3	2	3	
PH060																			3	3	4	4	0	2	2	2	2	3	2	3	
PH061																			1	1	2	2	2	0	0	0	0	1	0	1	
PH062																			1	1	2	2	2	0	0	0	0	1	0	1	
PH063																			1	1	2	2	2	0	0	0	0	1	0	1	
PH065																			1	1	2	2	2	0	0	0	0	1	0	1	
PH066																			2	2	3	3	3	1	1	1	1	0	1	2	
PH067																			1	1	2	2	2	0	0	0	0	1	0	1	
PH068																			2	2	3	3	3	1	1	1	1	2	1	0	
PH143																															0

Table S4.4: Comparison of subset-specific accessory genes based on EggNog annotations

FP-specific accessory genes	COG category	CH-specific accessory genes	COG category
Regulatory DnaK co-chaperone. Direct interaction between DnaK and DjlA is needed for the induction of the wcaABCDE operon, involved in the synthesis of a colanic acid polysaccharide capsule, possibly through activation of the RcsB RcsC phosphotransfer signaling pathway. The colanic acid capsule may help the bacterium survive conditions outside the host (By similarity)	O	restriction modification system DNA specificity	V
hydrolase, family 25	M	NA	S
toxin secretion phage lysis holin	S	Inherit from COG: peptidase'	M
HNH endonuclease	V	NA	S
Cell wall anchor domain protein	M	NA	S
serine threonine protein kinase	T	TRANSCRIPTIONal	K
Plays a major role in protein secretion by helping the post-translocational extracellular folding of several secreted proteins (By similarity)	O	PTS System	G
EA59 protein	S	PTS System	G
NA	S	PTS system, galactitol-specific IIc component	G
NA	S	PTS system, galactitol-specific IIc component	G
response regulator	T	NA	S
Histidine kinase	T	NA	S
NA	S	ribonuclease	S
Histidine kinase	T	Toxic component of a toxin-antitoxin (TA) module. A stable (half-life 27.6 minutes) endoribonuclease that in the absence of cognate antitoxin RnlB causes generalized RNA degradation. Degrades late enterobacteria phage T4 mRNAs, protecting the host against T4 reproduction. Activity is inhibited by cognate antitoxin RnlB and by enterobacteria phage T4 protein Dmd. Targets cyaA mRNA	A
cell division protein FtsK	D	NA	S
transposase	L	phage protein	S
Mor transcription activator family	S		

Appendix 3 Supplementary data for Chapter 5

This appendix comprises all supplementary materials for Chapter 5.

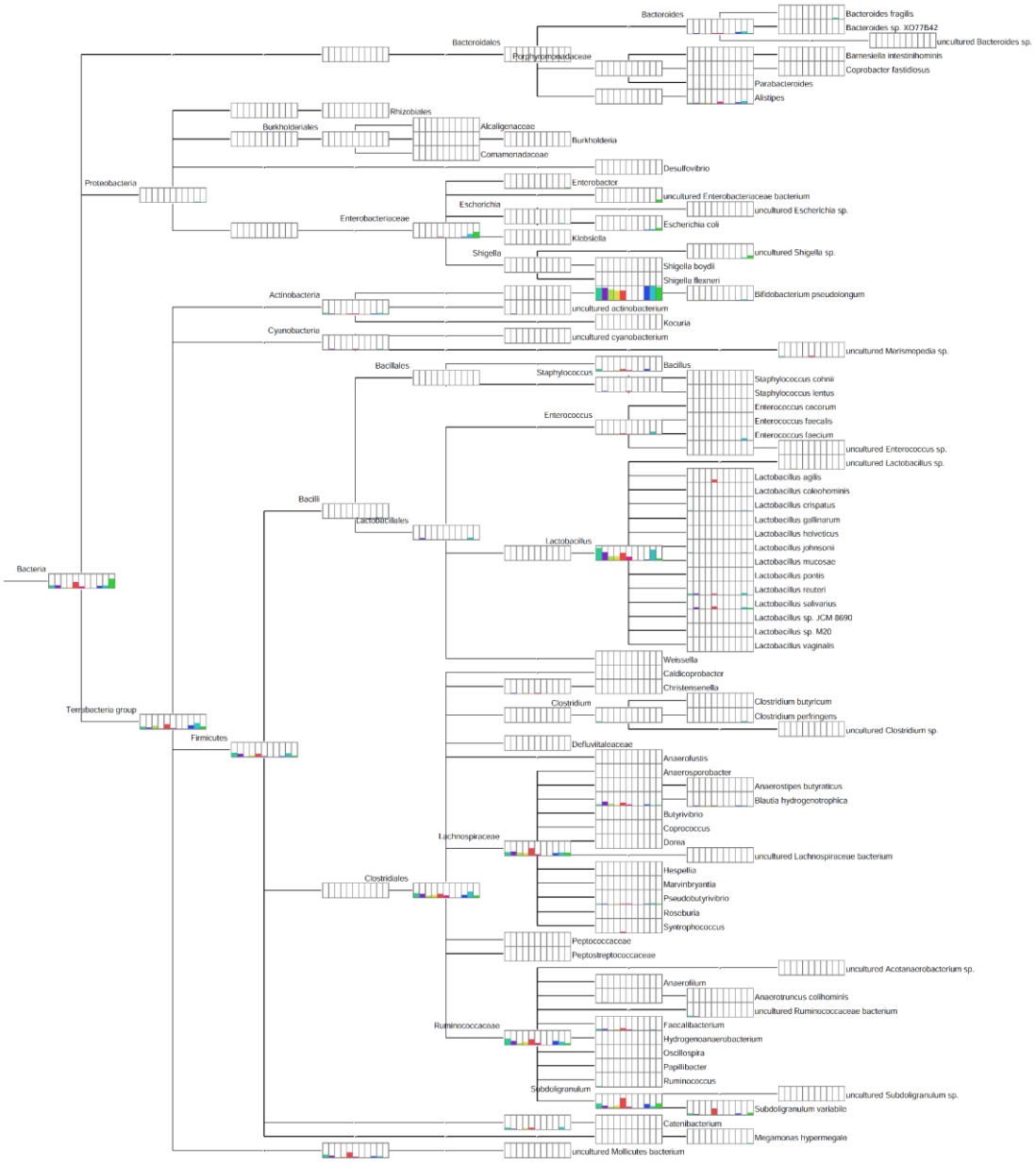
Table S5.1: Summary statistics of 16S rRNA sequences analysed including OTU numbers detected.

Figure S5.1: Reads assigned to 16S rRNA sequences including species level using paired-end BLASTN approach.

Table S5.1: Summary statistics of 16S rRNA sequences analysed including OTU numbers detected

Sample	Host condition	Sequencing platform	No. of OTUs	Raw reads
29	Healthy control	Illumina MiSeq	2,749	83,117
2	Healthy control	Illumina MiSeq	2,772	89,344
25	Healthy control	Illumina MiSeq	4,027	143,001
14	NE	Illumina MiSeq	3,050	126,582
33	NE	Illumina MiSeq	3,857	164,027
36	NE	Illumina MiSeq	3,236	135,808
21	SNE	Illumina MiSeq	3,440	150,198
22	SNE	Illumina MiSeq	1,780	60,036
26	SNE	Illumina MiSeq	4,972	112,820
34	SNE	Illumina MiSeq	1,792	79,227
38	SNE	Illumina MiSeq	2,641	123,828
		Mean ± SD	3,120 ± 902	115,272 ± 31,696

Figure S5.1: Reads assigned to 16S rRNA sequences including species level using paired-end BLASTN approach



Appendix 4 Supplementary data for Chapter 6

This appendix comprises all supplementary materials for Chapter 6.

Table S6.1: Details of all *C. perfingens* isolates (n=552) included in this thesis with basic assembly statistics, origin data and public database accessions (NCBI/ENA).

Table S6.1: Details of all *C. perfingens* isolates (n=552) included in this thesis with basic assembly statistics, origin data and public database accessions (NCBI/ENA)

A total of 90 genome assemblies were downloaded from NCBI assembly databases whereas 22 genomes of NCTC isolates were re-assembled in this study from raw-read sequencing data. The remaining isolates are newly sequenced in my PhD work. The data are arranged in the alphabetical order of isolate strain names. ANI (%) were compared with type strain ATCC13124.

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
11	3197244	35	Birds	Belgium	Public database	98.62	PRJNA422745
13	3085740	2	Undetermined	Undetermined: from NCBI	Public database	98.50	PRJNA79
37	3481068	74	Birds	Belgium	Public database	98.60	PRJNA422745
48	3250386	33	Birds	Belgium	Public database	98.60	PRJNA422745
67	3448290	48	Birds	Belgium	Public database	98.64	PRJNA422745
1207	3207161	95	Undetermined	Undetermined: from NCBI	Public database	98.56	PRJNA267549
2789STDY5608889	3321076	48	Adults	Undetermined: from NCBI	Public database	98.78	PRJEB10915
96-7415	3383570	48	Birds	USA	Public database	98.94	PRJNA422745
98.78718-2	3608029	55	Birds	Denmark	Public database	98.97	PRJNA422745
An185	3254381	71	Birds	Undetermined: from NCBI	Public database	98.54	PRJNA377666
An68	3263985	47	Birds	Undetermined: from NCBI	Public database	98.53	PRJNA377666
ATCC13124	3256683	1	Undetermined	Undetermined: from NCBI	Public database	100.00	PRJNA304
B_ATCC3626	3896305	98	Sheep	Undetermined: from NCBI	Public database	98.96	PRJNA20027
BER-NE33	3428087	14	Birds	Australia	Public database	98.68	PRJNA422745
BzA	3282838	64	Undetermined	Undetermined: from NCBI	Public database	98.59	PRJNA451295
C_JGS1495	3661329	84	Pigs	Undetermined: from NCBI	Public database	98.30	PRJNA20025
C033	3198143	41	Birds	Midlands, England	Lincolnshire	98.41	PRJEB25766
C036	3213346	14	Birds	Midlands, England	Lincolnshire	99.05	PRJEB25766
C056	3125087	35	Birds	Midlands, England	Lincolnshire	98.32	PRJEB25766
C55A	3375555	20	Birds	Midlands, England	Lincolnshire	97.15	PRJEB25766
CLP2R01	3470089	42	Adults	Scotland	Scotland	98.55	PRJEB25766
CLP2R02	3472015	47	Adults	Scotland	Scotland	98.54	PRJEB25766
CLP2R03	3468756	43	Adults	Scotland	Scotland	98.54	PRJEB25766
CLP3R02	3471259	40	Adults	Scotland	Scotland	98.54	PRJEB25766
CLP3R03	3471795	51	Adults	Scotland	Scotland	98.54	PRJEB25766
CP4	3642209	98	Birds	Canada	Public database	98.99	PRJNA294154
CPE_F4969	3510272	74	Adults	Undetermined: from NCBI	Public database	98.61	PRJNA20031
CR034	3257012	14	Birds	Midlands, England	Lincolnshire	97.11	PRJEB25766
D_JGS1721	4045016	221	Sheep	Undetermined: from NCBI	Public database	98.62	PRJNA28587
E_JGS1987	4127102	101	Cows	Undetermined: from NCBI	Public database	98.41	PRJNA20029
EHE-NE18	3663932	4	Birds	Australia	Public database	98.75	PRJNA341531
EHE-NE7	3430497	88	Birds	Australia	Public database	98.46	PRJNA422745

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
EM082-T6-5	3294625	20	Older adults	Ireland	ELDERMET	98.66	PRJEB28308
EM095-T0-2	3167702	35	Older adults	Ireland	ELDERMET	98.30	PRJEB28308
EM124-T3-6	3195277	30	Older adults	Ireland	ELDERMET	98.60	PRJEB28308
EM151-T0-4	3219100	24	Older adults	Ireland	ELDERMET	98.56	PRJEB28308
EM159-T0-8	3394504	49	Older adults	Ireland	ELDERMET	98.43	PRJEB28308
EM159-T6-2	3273806	41	Older adults	Ireland	ELDERMET	98.33	PRJEB28308
EM165-T0-4	3249888	24	Older adults	Ireland	ELDERMET	98.56	PRJEB28308
EM169-T0-2	3268354	24	Older adults	Ireland	ELDERMET	98.56	PRJEB28308
EM177-T0-6	3297346	20	Older adults	Ireland	ELDERMET	98.63	PRJEB28308
EM186-T0-5	3203537	24	Older adults	Ireland	ELDERMET	98.61	PRJEB28308
EM200-T6	3489434	45	Older adults	Ireland	ELDERMET	98.61	PRJEB28308
EM297-T0	3312580	18	Older adults	Ireland	ELDERMET	98.52	PRJEB28308
EM313-T6	3260252	19	Older adults	Ireland	ELDERMET	98.55	PRJEB28308
EM322-T0	3298048	20	Older adults	Ireland	ELDERMET	98.65	PRJEB28308
EM325-T0	3310070	30	Older adults	Ireland	ELDERMET	98.60	PRJEB28308
EM331-T0	3313180	37	Older adults	Ireland	ELDERMET	98.43	PRJEB28308
EM342-T0	3420719	31	Older adults	Ireland	ELDERMET	98.60	PRJEB28308
EUR-NE15	3588449	62	Birds	Australia	Public database	98.76	PRJNA422745
F262	3464306	53	Cows	Undetermined: from NCBI	Public database	98.59	PRJNA51969
FC2	3486930	81	Birds	USA	Public database	98.95	PRJNA422745
FORC003	3395109	2	Undetermined	Undetermined: from NCBI	Public database	98.46	PRJNA259939
FORC025	3343822	1	Undetermined	Undetermined: from NCBI	Public database	98.53	PRJNA299533
FR002	3325832	13	Soil	Midlands, England	Lincolnshire	99.09	PRJEB25766
FR006	3259735	15	Soil	Midlands, England	Lincolnshire	98.57	PRJEB25766
FR011	3388890	23	Soil	Midlands, England	Lincolnshire	98.65	PRJEB25766
FR013	3432631	26	Soil	Midlands, England	Lincolnshire	98.54	PRJEB25766
FR025	3253037	12	Soil	Midlands, England	Lincolnshire	98.59	PRJEB25766
FR063	3480913	47	Soil	Midlands, England	Lincolnshire	98.58	PRJEB25766
G035	3429313	22	Birds	Midlands, England	Lincolnshire	97.19	PRJEB25766
G049	3698052	46	Birds	Midlands, England	Lincolnshire	98.62	PRJEB25766
G050	3228163	26	Birds	Midlands, England	Lincolnshire	98.39	PRJEB25766
GNP-1	3725979	128	Birds	USA	Public database	99.00	PRJNA422745
I055	3385548	19	Birds	Midlands, England	Lincolnshire	99.11	PRJEB25766
I056	3136168	47	Birds	Midlands, England	Lincolnshire	98.32	PRJEB25766
I058	3648990	90	Birds	Midlands, England	Lincolnshire	98.62	PRJEB25766
I060	3688521	57	Birds	Midlands, England	Lincolnshire	98.61	PRJEB25766
ITX1105-12MP	3651954	252	Birds	USA	Public database	99.04	PRJNA422745
J060	3134537	57	Birds	Midlands, England	Lincolnshire	98.33	PRJEB25766
J36	3694265	46	Birds	Midlands, England	Lincolnshire	98.62	PRJEB25766

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
J46	3322108	14	Birds	Midlands, England	Lincolnshire	97.15	PRJEB25766
J55	3385936	22	Birds	Midlands, England	Lincolnshire	99.11	PRJEB25766
JFP718	3652220	56	Dogs	Canada	Public database	97.11	PRJNA327316
JFP727	3624033	47	Horses	Canada	Public database	97.11	PRJNA327317
JFP728	3579626	85	Horses	Canada	Public database	97.11	PRJNA327319
JFP771	3495308	81	Dogs	Canada	Public database	98.96	PRJNA327398
JFP774	3548595	56	Dogs	Canada	Public database	97.12	PRJNA327402
JFP795	3578912	67	Dogs	Canada	Public database	97.12	PRJNA327404
JFP796	3601148	114	Dogs	Canada	Public database	97.11	PRJNA327405
JFP801	3580610	59	Horses	USA	Public database	97.11	PRJNA327406
JFP804	3625968	117	Horses	USA	Public database	97.11	PRJNA327409
JFP810	3830406	95	Dogs	Canada	Public database	97.11	PRJNA327410
JFP826	3660408	70	Dogs	USA	Public database	97.11	PRJNA327411
JFP828	3564370	54	Horses	USA	Public database	97.12	PRJNA327440
JFP829	3639686	108	Horses	USA	Public database	97.11	PRJNA327441
JFP833	3521770	96	Horses	USA	Public database	98.97	PRJNA327442
JFP834	3549140	121	Horses	USA	Public database	98.96	PRJNA327443
JFP836	3599532	51	Dogs	USA	Public database	97.11	PRJNA327444
JFP914	3819232	79	Dogs	Switzerland	Public database	97.12	PRJNA327445
JFP916	3660641	67	Dogs	Switzerland	Public database	97.12	PRJNA327446
JFP921	3601767	78	Dogs	USA	Public database	97.12	PRJNA327447
JFP922	3612099	76	Dogs	USA	Public database	97.12	PRJNA327448
JFP923	3589715	48	Dogs	USA	Public database	97.13	PRJNA327449
JFP941	3594640	64	Dogs	Canada	Public database	97.11	PRJNA327450
JFP961	3617219	66	Dogs	Canada	Public database	97.12	PRJNA327451
JFP978	3585277	68	Horses	USA	Public database	98.96	PRJNA327452
JFP980	3654588	67	Horses	USA	Public database	97.11	PRJNA327453
JFP981	3580367	68	Horses	USA	Public database	98.96	PRJNA327455
JFP982	3670557	62	Horses	USA	Public database	97.12	PRJNA327456
JFP983	3657522	46	Horses	USA	Public database	97.11	PRJNA327457
JFP986	3669049	58	Horses	USA	Public database	97.10	PRJNA327458
JFP992	3657868	70	Horses	USA	Public database	97.11	PRJNA327459
JJC	3259329	69	Undetermined	Undetermined: from NCBI	Public database	98.63	PRJNA215269
JP55	3347300	1	Horses	Canada	Public database	98.98	PRJNA276414
JP838	3530414	1	Dogs	USA	Public database	97.15	PRJNA276513
K473	3411764	55	Birds	Australia	Public database	98.55	PRJNA422745
LH019	3605600	28	Infants	East of England	BAMBI	96.54	PRJEB25762
LH020	3572073	22	Infants	East of England	BAMBI	96.57	PRJEB25762
LH023	3607194	26	Infants	East of England	BAMBI	96.54	PRJEB25762

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
LH024	3605555	29	Infants	East of England	BAMBI	96.55	PRJEB25762
LH031	3293275	16	Infants	East of England	BAMBI	98.57	PRJEB25762
LH032	3289023	15	Infants	East of England	BAMBI	98.58	PRJEB25762
LH033	3294377	13	Infants	East of England	BAMBI	98.58	PRJEB25762
LH035	3290682	15	Infants	East of England	BAMBI	98.57	PRJEB25762
LH043	3294719	18	Infants	East of England	BAMBI	98.57	PRJEB25762
LH044	3292797	18	Infants	East of England	BAMBI	98.57	PRJEB25762
LH045	3287528	15	Infants	East of England	BAMBI	98.57	PRJEB25762
LH046	3231068	24	Infants	East of England	BAMBI	98.36	PRJEB25762
LH047	3229224	25	Infants	East of England	BAMBI	98.36	PRJEB25762
LH048	3236291	25	Infants	East of England	BAMBI	98.36	PRJEB25762
LH051	3350033	16	Infants	East of England	BAMBI	98.62	PRJEB25762
LH055	3253813	14	Infants	East of England	BAMBI	98.56	PRJEB25762
LH056	3349772	18	Infants	East of England	BAMBI	98.61	PRJEB25762
LH057	3349425	16	Infants	East of England	BAMBI	98.61	PRJEB25762
LH059	3347675	15	Infants	East of England	BAMBI	98.61	PRJEB25762
LH060	3176811	27	Infants	East of England	BAMBI	98.31	PRJEB25762
LH062	3313187	53	Infants	East of England	BAMBI	98.32	PRJEB25762
LH063	3311517	52	Infants	East of England	BAMBI	98.32	PRJEB25762
LH064	3313206	47	Infants	East of England	BAMBI	98.32	PRJEB25762
LH065	3255191	15	Infants	East of England	BAMBI	98.57	PRJEB25762
LH066	3358561	34	Infants	East of England	BAMBI	98.42	PRJEB25762
LH067	3363409	36	Infants	East of England	BAMBI	98.43	PRJEB25762
LH070	3364587	40	Infants	East of England	BAMBI	98.42	PRJEB25762
LH071	3362921	38	Infants	East of England	BAMBI	98.43	PRJEB25762
LH080	3173380	30	Infants	East of England	BAMBI	98.27	PRJEB25762
LH083	3718165	58	Infants	East of England	BAMBI	98.96	PRJEB25762
LH084	3758696	53	Infants	East of England	BAMBI	98.96	PRJEB25762
LH086	3721164	66	Infants	East of England	BAMBI	98.96	PRJEB25762
LH101	3273117	17	Infants	East of England	BAMBI	98.61	PRJEB25762
LH102	3270536	15	Infants	East of England	BAMBI	98.61	PRJEB25762
LH103	3274777	21	Infants	East of England	BAMBI	98.61	PRJEB25762
LH105	3273978	20	Infants	East of England	BAMBI	98.61	PRJEB25762
LH107	3303134	31	Infants	East of England	BAMBI	98.41	PRJEB25762
LH108	3303663	30	Infants	East of England	BAMBI	98.41	PRJEB25762
LH109	3306644	27	Infants	East of England	BAMBI	98.41	PRJEB25762
LH111	3303073	30	Infants	East of England	BAMBI	98.41	PRJEB25762
LH112	3364130	32	Infants	East of England	BAMBI	98.42	PRJEB25762
LH113	3365350	42	Infants	East of England	BAMBI	98.43	PRJEB25762

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
LH114	3364314	38	Infants	East of England	BAMBI	98.42	PRJEB25762
LH115	3364031	39	Infants	East of England	BAMBI	98.42	PRJEB25762
LH116	3365880	37	Infants	East of England	BAMBI	98.42	PRJEB25762
LH120	3272004	15	Infants	East of England	BAMBI	98.61	PRJEB25762
LH121	3275022	20	Infants	East of England	BAMBI	98.61	PRJEB25762
LH122	3272157	14	Infants	East of England	BAMBI	98.60	PRJEB25762
LH135	3292320	18	Infants	East of England	BAMBI	98.57	PRJEB25762
LH136	3292069	19	Infants	East of England	BAMBI	98.57	PRJEB25762
LH137	3287845	14	Infants	East of England	BAMBI	98.57	PRJEB25762
LH139	3291122	16	Infants	East of England	BAMBI	98.57	PRJEB25762
LHC1	3313944	54	Infants	East of England	BAMBI	98.32	PRJEB25762
LHC13	3252214	16	Infants	East of England	BAMBI	98.57	PRJEB25762
LHC14	3350685	17	Infants	East of England	BAMBI	98.61	PRJEB25762
LHC5	3231824	26	Infants	East of England	BAMBI	98.36	PRJEB25762
LHZ154	3225350	26	Zoo animals: bear	Wales	Zoo Microbiome	98.40	PRJEB28309
LHZ155	3195467	21	Zoo animals: bear	Wales	Zoo Microbiome	98.41	PRJEB28309
LHZ157	3336624	18	Zoo animals: bear	Wales	Zoo Microbiome	97.35	PRJEB28309
LHZ158	3208816	18	Zoo animals: bear	Wales	Zoo Microbiome	98.42	PRJEB28309
LHZ161	3310316	59	Zoo animals: lion	East of England	Zoo Microbiome	98.47	PRJEB28309
LHZ162	3326006	19	Zoo animals: lion	East of England	Zoo Microbiome	99.05	PRJEB28309
LHZ163	3281094	13	Zoo animals: leopard	East of England	Zoo Microbiome	98.59	PRJEB28309
LHZ164	3101486	17	Zoo animals: leopard	East of England	Zoo Microbiome	96.17	PRJEB28309
LHZ170	3306129	22	Zoo animals: monkey	East of England	Zoo Microbiome	98.66	PRJEB28309
LHZ171	3303919	19	Zoo animals: monkey	East of England	Zoo Microbiome	98.66	PRJEB28309
LHZ459	3169911	21	Zoo animals: seriema	East of England	Zoo Microbiome	96.13	PRJEB28309
LHZ460	3448674	11	Zoo animals: seriema	East of England	Zoo Microbiome	98.61	PRJEB28309
LHZ472	3352988	20	Zoo animals: aardvark	East of England	Zoo Microbiome	99.02	PRJEB28309
LHZ474	3335111	22	Zoo animals: aardvark	East of England	Zoo Microbiome	97.23	PRJEB28309
LHZ476	3363510	13	Zoo animals: mongoose	East of England	Zoo Microbiome	98.71	PRJEB28309
LHZ513	3366540	13	Zoo animals: goose	East of England	Zoo Microbiome	98.71	PRJEB28309
LHZ574	3469057	18	Zoo animals: monkey	East of England	Zoo Microbiome	96.71	PRJEB28309
LLY_N11	3446916	4	Birds	Undetermined: from NCBI	Public database	98.91	PRJNA403823
M009.1C	3228292	22	Infants	London, England	NEOM	98.30	PRJEB25762
M009.2.1C	3261764	26	Children	London, England	NEOM	98.58	PRJEB28307
M009.2C	3227979	24	Infants	London, England	NEOM	98.30	PRJEB25762
M010.1C	3344966	18	Infants	London, England	NEOM	98.65	PRJEB25762
M010.2C	3347565	18	Infants	London, England	NEOM	98.66	PRJEB25762
M010.CN1	3344094	17	Infants	London, England	NEOM	98.65	PRJEB25762
M010.CN2	3346940	17	Infants	London, England	NEOM	98.65	PRJEB25762

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
M010f.1C	3345584	18	Infants	London, England	NEOM	98.65	PRJEB25762
M011.1C	3345019	15	Infants	London, England	NEOM	98.66	PRJEB25762
M011f.1C	3347080	21	Infants	London, England	NEOM	98.66	PRJEB25762
M030.1C	3205090	19	Infants	London, England	NEOM	98.60	PRJEB25762
M030.2C	3206308	16	Infants	London, England	NEOM	98.61	PRJEB25762
M030.N1C	3207141	16	Infants	London, England	NEOM	98.61	PRJEB25762
M030.N2C	3209453	16	Infants	London, England	NEOM	98.60	PRJEB25762
M030.N3C	3205909	14	Infants	London, England	NEOM	98.60	PRJEB25762
M030.N4C	3202137	13	Infants	London, England	NEOM	98.60	PRJEB25762
M040.2.1C	3143104	35	Children	London, England	NEOM	98.30	PRJEB28307
M087.10C	3462641	27	Infants	London, England	NEOM	98.53	PRJEB25762
M087.11C	3456083	28	Infants	London, England	NEOM	98.53	PRJEB25762
M087.13C	3455084	23	Infants	London, England	NEOM	98.54	PRJEB25762
M087.1C	3284487	35	Infants	London, England	NEOM	98.34	PRJEB25762
M087.2C	3283083	33	Infants	London, England	NEOM	98.34	PRJEB25762
M087.4C	3460509	28	Infants	London, England	NEOM	98.54	PRJEB25762
M087.5C	3462398	29	Infants	London, England	NEOM	98.53	PRJEB25762
M087.7C	3456430	27	Infants	London, England	NEOM	98.53	PRJEB25762
M087.8C	3456813	24	Infants	London, England	NEOM	98.54	PRJEB25762
M088.10C	3144583	45	Infants	London, England	NEOM	98.34	PRJEB25762
M088.11C	3146349	44	Infants	London, England	NEOM	98.34	PRJEB25762
M088.1C	3162957	38	Infants	London, England	NEOM	98.33	PRJEB25762
M088.24C	3147539	39	Infants	London, England	NEOM	98.34	PRJEB25762
M088.2C	3173224	37	Infants	London, England	NEOM	98.33	PRJEB25762
M088.4C	3177376	35	Infants	London, England	NEOM	98.34	PRJEB25762
M088.5C	3179386	37	Infants	London, England	NEOM	98.34	PRJEB25762
M089.1C	3462801	27	Infants	London, England	NEOM	98.54	PRJEB25762
M089.2C	3452649	28	Infants	London, England	NEOM	98.54	PRJEB25762
M089.3C	3458059	30	Infants	London, England	NEOM	98.54	PRJEB25762
M089.4C	3454935	31	Infants	London, England	NEOM	98.54	PRJEB25762
M090.1C	3453643	27	Infants	London, England	NEOM	98.54	PRJEB25762
M090.2C	3456978	28	Infants	London, England	NEOM	98.54	PRJEB25762
M090.3C	3418228	26	Infants	London, England	NEOM	98.54	PRJEB25762
M090.4C	3453548	28	Infants	London, England	NEOM	98.54	PRJEB25762
M091.1C	3107914	26	Infants	London, England	NEOM	98.35	PRJEB25762
M091.2C	3108502	27	Infants	London, England	NEOM	98.35	PRJEB25762
M092.1C	3184687	37	Infants	London, England	NEOM	98.41	PRJEB25762
M092.2C	3180571	34	Infants	London, England	NEOM	98.41	PRJEB25762
M094.1C	3123459	34	Infants	London, England	NEOM	98.31	PRJEB25762

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
M094.2C	3134877	36	Infants	London, England	NEOM	98.32	PRJEB25762
M096.1C	3402370	20	Infants	London, England	NEOM	99.02	PRJEB25762
M096.2C	3402836	23	Infants	London, England	NEOM	99.02	PRJEB25762
M096.3C	3403754	26	Infants	London, England	NEOM	99.01	PRJEB25762
M096.4C	3345030	20	Infants	London, England	NEOM	99.02	PRJEB25762
M111.1C	3170976	46	Infants	London, England	NEOM	98.38	PRJEB25762
M111.2C	3167862	46	Infants	London, England	NEOM	98.38	PRJEB25762
M112.1C	3171762	37	Infants	London, England	NEOM	98.37	PRJEB25762
M112.2C	3172334	45	Infants	London, England	NEOM	98.37	PRJEB25762
M112.4C	3171262	40	Infants	London, England	NEOM	98.37	PRJEB25762
M112.5C	3166211	44	Infants	London, England	NEOM	98.38	PRJEB25762
M124.1C	3166878	31	Infants	London, England	NEOM	98.33	PRJEB25762
M124.2C	3161455	33	Infants	London, England	NEOM	98.33	PRJEB25762
M125.1C	3167259	32	Infants	London, England	NEOM	98.32	PRJEB25762
M125.2.4C	3496522	30	Children	London, England	NEOM	98.53	PRJEB28307
M125.2C	3168973	32	Infants	London, England	NEOM	98.32	PRJEB25762
M125.4C	3165453	31	Infants	London, England	NEOM	98.32	PRJEB25762
M125.5C	3164915	32	Infants	London, England	NEOM	98.31	PRJEB25762
M128.2.3C	3425328	48	Children	London, England	NEOM	98.40	PRJEB28307
M135.2.3C	3327279	31	Children	London, England	NEOM	98.72	PRJEB28307
M137.2.1C	3142190	46	Children	London, England	NEOM	98.31	PRJEB28307
MJR7757A	3585666	274	Undetermined	Undetermined: from NCBI	Public database	96.62	PRJNA272114
NAG-NE1	3334514	128	Birds	Australia	Public database	98.71	PRJNA422745
NAG-NE31	3469324	116	Birds	Australia	Public database	99.01	PRJNA422745
NCTC10239	3081789	7	Undetermined	Undetermined: from NCTC	Public database	97.16	PRJEB28310
NCTC10240	3035468	4	Birds	Undetermined: from NCTC	Public database	97.19	PRJEB28310
NCTC10578	3353304	8	Undetermined	Undetermined: from NCTC	Public database	98.57	PRJEB28310
NCTC10612	3065426	5	Undetermined	Undetermined: from NCTC	Public database	97.14	PRJEB28310
NCTC10613	3215051	35	Undetermined	Undetermined: from NCTC	Public database	97.20	PRJEB28310
NCTC10614	3106776	17	Undetermined	Undetermined: from NCTC	Public database	97.19	PRJEB28310
NCTC10719	3657605	25	Pigs	Undetermined: from NCTC	Public database	98.31	PRJEB6403
NCTC13170	3310238	1	Undetermined	Undetermined: from NCTC	Public database	99.03	PRJEB28310
NCTC2544	3195093	1	Undetermined	Undetermined: from NCTC	Public database	98.60	PRJEB28310
NCTC2837	3306216	3	Undetermined	Undetermined: from NCTC	Public database	98.61	PRJEB28310
NCTC3182	3877250	34	Sheep	Undetermined: from NCTC	Public database	98.60	PRJEB28310
NCTC8081	3665176	21	Undetermined	Undetermined: from NCTC	Public database	97.18	PRJEB28310
NCTC8238	3051538	2	Undetermined	Undetermined: from NCTC	Public database	97.16	PRJEB28310
NCTC8239	3008497	2	Undetermined	Undetermined: from NCTC	Public database	97.18	PRJEB28310
NCTC8246	3391864	3	Undetermined	Undetermined: from NCTC	Public database	99.07	PRJEB28310

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
NCTC8247	3011897	2	Undetermined	Undetermined: from NCTC	Public database	97.15	PRJEB28310
NCTC8359	3202545	21	Undetermined	Undetermined: from NCTC	Public database	97.21	PRJEB28310
NCTC8449	3096490	29	Undetermined	Undetermined: from NCTC	Public database	97.89	PRJEB6403
NCTC8503	3577234	9	Sheep	Undetermined: from NCTC	Public database	98.61	PRJEB28310
NCTC8678	3005443	6	Undetermined	Undetermined: from NCTC	Public database	97.17	PRJEB28310
NCTC8679	3151371	11	Undetermined	Undetermined: from NCTC	Public database	97.20	PRJEB28310
NCTC8797	3021559	5	Undetermined	Undetermined: from NCTC	Public database	97.16	PRJEB28310
NCTC8798	3084171	14	Undetermined	Undetermined: from NCTC	Public database	97.18	PRJEB6403
NCTC8799	3108291	7	Undetermined	Undetermined: from NCTC	Public database	97.17	PRJEB28310
NCTC9851	3035210	6	Undetermined	Undetermined: from NCTC	Public database	97.20	PRJEB28310
NobL1	3508314	241	Birds	USA	Public database	98.29	PRJNA422745
PBD1	3428087	14	Birds	Australia	Public database	98.68	PRJNA422745
PBS5	3110163	10	Birds	Australia	Public database	95.95	PRJNA422745
PC5	3099804	11	Birds	Australia	Public database	95.95	PRJNA422745
Pennington	3749339	148	Birds	USA	Public database	99.02	PRJNA422745
PH004	2923443	95	Adults	South East England	PHE	97.13	PRJEB25764
PH005	2934006	96	Adults	South East England	PHE	97.13	PRJEB25764
PH006	2928275	95	Adults	South East England	PHE	97.13	PRJEB25764
PH007	2946980	81	Adults	South East England	PHE	97.12	PRJEB25764
PH014	3423911	42	Adults	London, England	PHE	98.55	PRJEB25764
PH015	3081476	100	Older adults	North East England	PHE	97.21	PRJEB25764
PH016	3518998	66	Older adults	North East England	PHE	98.47	PRJEB25764
PH017	3291535	23	Older adults	North East England	PHE	98.61	PRJEB25764
PH018	3475247	26	Older adults	North East England	PHE	98.47	PRJEB25764
PH019	3487876	25	Older adults	North East England	PHE	98.47	PRJEB25764
PH020	3397222	33	Older adults	North East England	PHE	98.51	PRJEB25764
PH021	3445998	31	Older adults	North East England	PHE	98.50	PRJEB25764
PH022	3401212	36	Adults	North East England	PHE	98.33	PRJEB25764
PH023	3325484	35	Adults	North East England	PHE	98.32	PRJEB25764
PH024	3331444	28	Adults	North East England	PHE	98.33	PRJEB25764
PH025	3320821	34	Adults	North East England	PHE	98.32	PRJEB25764
PH026	3335204	37	Adults	North East England	PHE	98.33	PRJEB25764
PH028	2994736	82	Adults	London, England	PHE	97.12	PRJEB25764
PH029	2954967	97	Adults	London, England	PHE	97.12	PRJEB25764
PH030	3427777	27	Older adults	North East England	PHE	98.47	PRJEB25764
PH031	3430445	21	Older adults	North East England	PHE	98.47	PRJEB25764
PH032	3139693	109	Older adults	North East England	PHE	97.16	PRJEB25764
PH033	3163229	105	Older adults	North East England	PHE	97.17	PRJEB25764
PH035	3422738	32	Adults	North East England	PHE	98.54	PRJEB25764

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information					Study Accession
	Genome (bp)	Contigs	Source	Region/Country		Study/Site	ANI	
PH036	3430269	30	Adults	North England	East	PHE	98.54	PRJEB25764
PH037	3426315	26	Adults	North England	East	PHE	98.54	PRJEB25764
PH039	3432249	23	Adults	North England	East	PHE	98.54	PRJEB25764
PH040	3426487	25	Adults	North England	East	PHE	98.54	PRJEB25764
PH041	3433511	31	Adults	North England	East	PHE	98.54	PRJEB25764
PH042	3427997	34	Adults	North England	East	PHE	98.54	PRJEB25764
PH043	3421670	43	Adults	North England	East	PHE	98.54	PRJEB25764
PH044	3426828	29	Adults	North England	East	PHE	98.54	PRJEB25764
PH045	3191860	32	Adults	North England	East	PHE	98.30	PRJEB25764
PH046	3425020	35	Adults	North England	East	PHE	98.54	PRJEB25764
PH047	3422937	43	Adults	North England	East	PHE	98.54	PRJEB25764
PH048	3428504	34	Adults	North England	East	PHE	98.54	PRJEB25764
PH050	3315627	59	Adults	South England	West	PHE	98.30	PRJEB25764
PH051	3480399	27	Older adults	North England	East	PHE	98.57	PRJEB25764
PH052	3481680	34	Older adults	North England	East	PHE	98.47	PRJEB25764
PH053	3484752	25	Older adults	North England	East	PHE	98.47	PRJEB25764
PH054	3476853	57	Adults	North England	East	PHE	98.52	PRJEB25764
PH055	3441127	27	Adults	North England	East	PHE	98.52	PRJEB25764
PH056	3395441	23	Adults	North England	East	PHE	98.56	PRJEB25764
PH057	3398027	18	Adults	North England	East	PHE	98.56	PRJEB25764
PH058	3396864	22	Adults	North England	East	PHE	98.56	PRJEB25764
PH059	3395953	24	Adults	North England	East	PHE	98.56	PRJEB25764
PH060	3399028	20	Adults	North England	East	PHE	98.56	PRJEB25764
PH061	3396723	20	Adults	North England	East	PHE	98.55	PRJEB25764
PH062	3397544	21	Adults	North England	East	PHE	98.56	PRJEB25764
PH063	3395429	19	Adults	North England	East	PHE	98.56	PRJEB25764
PH065	3392964	20	Adults	North England	East	PHE	98.56	PRJEB25764
PH066	3398355	18	Adults	North England	East	PHE	98.56	PRJEB25764
PH067	3392663	21	Adults	North England	East	PHE	98.55	PRJEB25764
PH068	3399957	23	Adults	North England	East	PHE	98.56	PRJEB25764
PH072	3492601	59	Older adults	North England	East	PHE	98.56	PRJEB25764
PH073	3537414	35	Older adults	North England	East	PHE	98.46	PRJEB25764
PH074	3522802	26	Older adults	North England	East	PHE	98.46	PRJEB25764
PH075	3538037	29	Older adults	North England	East	PHE	98.46	PRJEB25764
PH076	3450640	23	Older adults	North England	East	PHE	98.46	PRJEB25764
PH077	3491708	23	Older adults	North England	East	PHE	98.47	PRJEB25764
PH078	2987740	70	Older adults	North England	East	PHE	97.12	PRJEB25764
PH079	3496302	26	Older adults	North England	East	PHE	98.46	PRJEB25764
PH080	2983246	73	Adults	North England	West	PHE	97.13	PRJEB25764

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
PH081	2984289	67	Adults	North West England	PHE	97.12	PRJEB25764
PH082	3394345	30	Older adults	North East England	PHE	98.65	PRJEB25764
PH083	3410579	21	Older adults	North East England	PHE	98.64	PRJEB25764
PH084	3392793	22	Older adults	North East England	PHE	98.65	PRJEB25764
PH085	3394191	23	Older adults	North East England	PHE	98.64	PRJEB25764
PH091	3337668	491	Adults	London, England	PHE	98.58	PRJEB25764
PH092	3378829	378	Adults	London, England	PHE	98.56	PRJEB25764
PH095	3348605	446	Adults	London, England	PHE	98.58	PRJEB25764
PH096	3422714	239	Older adults	North East England	PHE	98.48	PRJEB25764
PH097	3371290	398	Older adults	North East England	PHE	98.50	PRJEB25764
PH098	3454553	245	Older adults	North East England	PHE	98.48	PRJEB25764
PH100	3354426	184	Older adults	North East England	PHE	98.52	PRJEB25764
PH101	3481641	87	Older adults	North East England	PHE	98.47	PRJEB25764
PH102	3420587	311	Older adults	North East England	PHE	98.49	PRJEB25764
PH103	3448541	274	Older adults	North East England	PHE	98.49	PRJEB25764
PH104	2929035	214	Adults	London, England	PHE	97.14	PRJEB25764
PH105	2944409	379	Adults	North East England	PHE	97.16	PRJEB25764
PH107	2914363	246	Adults	London, England	PHE	97.15	PRJEB25764
PH115	2977232	317	Adults	North East England	PHE	97.17	PRJEB25764
PH117	3285816	246	Adults	North East England	PHE	98.35	PRJEB25764
PH118	3282569	214	Adults	North East England	PHE	98.35	PRJEB25764
PH119	3289115	215	Adults	North East England	PHE	98.35	PRJEB25764
PH120	3274008	185	Adults	South West England	PHE	98.31	PRJEB25764
PH121	3233465	289	Adults	South West England	PHE	98.32	PRJEB25764
PH122	3282469	371	Adults	South West England	PHE	98.32	PRJEB25764
PH123	3429751	156	Adults	North East England	PHE	98.52	PRJEB25764
PH124	3431905	132	Adults	North East England	PHE	98.52	PRJEB25764
PH125	3406444	363	Adults	North East England	PHE	98.63	PRJEB25764
PH126	3408431	418	Adults	North East England	PHE	98.62	PRJEB25764
PH127	2924057	262	Adults	North East England	PHE	97.13	PRJEB25764
PH128	3436609	352	Adults	Midlands, England	PHE	98.97	PRJEB25764
PH129	3441457	266	Adults	Midlands, England	PHE	98.97	PRJEB25764
PH130	3446961	250	Adults	Midlands, England	PHE	98.96	PRJEB25764
PH132	3418685	466	Adults	Midlands, England	PHE	98.98	PRJEB25764
PH133	3432746	347	Adults	Midlands, England	PHE	98.98	PRJEB25764
PH134	3446865	192	Adults	Midlands, England	PHE	98.96	PRJEB25764
PH135	2912738	356	Adults	London, England	PHE	97.15	PRJEB25764
PH136	2905546	312	Adults	London, England	PHE	97.16	PRJEB25764
PH138	3401096	222	Older adults	North East England	PHE	98.52	PRJEB25764

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
PH139	3402225	97	Older adults	North East England	PHE	98.52	PRJEB25764
PH140	3386184	257	Older adults	North East England	PHE	98.53	PRJEB25764
PH143	3528864	256	Adults	South West England	PHE	98.54	PRJEB25764
PH147	3267085	351	Adults	North East England	PHE	98.35	PRJEB25764
PH154	3413080	177	Adults	North East England	PHE	98.55	PRJEB25764
PH155	3411087	209	Adults	North East England	PHE	98.55	PRJEB25764
PH156	3172681	239	Adults	North East England	PHE	98.31	PRJEB25764
Q012.2.1C	3135616	46	Children	London, England	NEOM	98.32	PRJEB28307
Q020.CN1	3242179	16	Infants	London, England	NEOM	98.57	PRJEB25762
Q020.CN2	3243904	21	Infants	London, England	NEOM	98.57	PRJEB25762
Q020.CN4	3240291	19	Infants	London, England	NEOM	98.57	PRJEB25762
Q020.CN5	3241785	21	Infants	London, England	NEOM	98.58	PRJEB25762
Q020f.CN1	3244276	22	Infants	London, England	NEOM	98.57	PRJEB25762
Q034.2.1C	3186736	28	Children	London, England	NEOM	96.16	PRJEB28307
Q036.2.1C	3427754	28	Children	London, England	NEOM	98.61	PRJEB28307
Q041.2.4C	3573312	61	Children	London, England	NEOM	98.59	PRJEB28307
Q048.2.1C	3232498	44	Children	London, England	NEOM	98.31	PRJEB28307
Q049.2.3C	3347476	40	Children	London, England	NEOM	99.20	PRJEB28307
Q061.2.1C	3442469	40	Children	London, England	NEOM	97.15	PRJEB28307
Q063.2.1C	3317138	30	Children	London, England	NEOM	98.62	PRJEB28307
Q087.1C	3280619	25	Infants	London, England	NEOM	98.38	PRJEB25762
Q087.2C	3283123	29	Infants	London, England	NEOM	98.38	PRJEB25762
Q088.CN1	3062159	11	Infants	London, England	NEOM	96.15	PRJEB25762
Q088.CN10	3403885	27	Infants	London, England	NEOM	98.30	PRJEB25762
Q088.CN12	3286041	21	Infants	London, England	NEOM	98.38	PRJEB25762
Q088.CN13	3065620	16	Infants	London, England	NEOM	96.15	PRJEB25762
Q088.CN2	3069367	19	Infants	London, England	NEOM	96.15	PRJEB25762
Q088.CN3	3065157	18	Infants	London, England	NEOM	96.14	PRJEB25762
Q088.CN4	3068997	16	Infants	London, England	NEOM	96.15	PRJEB25762
Q088.CN6	3066288	17	Infants	London, England	NEOM	96.15	PRJEB25762
Q088.CN7	3063712	13	Infants	London, England	NEOM	96.16	PRJEB25762
Q088.CN9	3287606	21	Infants	London, England	NEOM	98.38	PRJEB25762
Q090.5C	3194717	14	Infants	London, England	NEOM	98.60	PRJEB25762
Q090.6C	3200511	24	Infants	London, England	NEOM	98.61	PRJEB25762
Q090.8C	3202127	16	Infants	London, England	NEOM	98.61	PRJEB25762
Q090.9C	3201249	23	Infants	London, England	NEOM	98.61	PRJEB25762
Q096.10C	3194624	29	Infants	London, England	NEOM	98.34	PRJEB25762
Q096.11C	3187873	30	Infants	London, England	NEOM	98.32	PRJEB25762
Q096.13C	3244272	28	Infants	London, England	NEOM	98.33	PRJEB25762

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
Q096.14C	3239882	33	Infants	London, England	NEOM	98.33	PRJEB25762
Q096.16C	3190020	27	Infants	London, England	NEOM	96.14	PRJEB25762
Q096.17C	3184400	32	Infants	London, England	NEOM	98.33	PRJEB25762
Q096.1C	4147421	46	Infants	London, England	NEOM	98.31	PRJEB25762
Q096.2C	4086994	39	Infants	London, England	NEOM	98.32	PRJEB25762
Q096.4C	3309035	21	Infants	London, England	NEOM	98.59	PRJEB25762
Q096.5C	3305014	17	Infants	London, England	NEOM	98.58	PRJEB25762
Q097.1C	3160474	25	Infants	London, England	NEOM	98.28	PRJEB25762
Q097.2C	3152325	26	Infants	London, England	NEOM	98.27	PRJEB25762
Q097.4C	3154351	21	Infants	London, England	NEOM	98.27	PRJEB25762
Q097.5C	3151690	25	Infants	London, England	NEOM	98.28	PRJEB25762
Q097.7C	3196530	27	Infants	London, England	NEOM	98.27	PRJEB25762
Q097.8C	3152354	26	Infants	London, England	NEOM	98.27	PRJEB25762
Q100.CN1	3209378	26	Infants	London, England	NEOM	98.25	PRJEB25762
Q100.CN10	3177242	21	Infants	London, England	NEOM	98.29	PRJEB25762
Q100.CN11	3177983	21	Infants	London, England	NEOM	98.29	PRJEB25762
Q100.CN12	3212162	21	Infants	London, England	NEOM	98.24	PRJEB25762
Q100.CN13	3210861	19	Infants	London, England	NEOM	98.25	PRJEB25762
Q100.CN2	3214133	27	Infants	London, England	NEOM	98.26	PRJEB25762
Q100.CN4	3211356	23	Infants	London, England	NEOM	98.26	PRJEB25762
Q100.CN5	3212371	20	Infants	London, England	NEOM	98.26	PRJEB25762
Q100.CN7	3212734	22	Infants	London, England	NEOM	98.25	PRJEB25762
Q100.CN8	3213979	18	Infants	London, England	NEOM	98.25	PRJEB25762
Q101.1C	3231229	22	Infants	London, England	NEOM	98.40	PRJEB25762
Q101.2C	3238879	21	Infants	London, England	NEOM	98.40	PRJEB25762
Q105.1C	3138208	28	Infants	London, England	NEOM	98.33	PRJEB25762
Q105.2C	3137322	27	Infants	London, England	NEOM	98.33	PRJEB25762
Q105.4C	3136273	33	Infants	London, England	NEOM	98.33	PRJEB25762
Q105.5C	3134159	35	Infants	London, England	NEOM	98.33	PRJEB25762
Q106.10C	3494367	24	Infants	London, England	NEOM	99.12	PRJEB25762
Q106.11C	3475235	26	Infants	London, England	NEOM	99.12	PRJEB25762
Q106.1C	3125382	39	Infants	London, England	NEOM	98.33	PRJEB25762
Q106.2C	3137937	28	Infants	London, England	NEOM	98.32	PRJEB25762
Q106.4C	3100264	20	Infants	London, England	NEOM	96.15	PRJEB25762
Q106.5C	3091941	14	Infants	London, England	NEOM	96.14	PRJEB25762
Q106.7C	3092582	15	Infants	London, England	NEOM	96.14	PRJEB25762
Q106.8C	3267326	31	Infants	London, England	NEOM	98.39	PRJEB25762
Q107.10C	3297761	20	Infants	London, England	NEOM	98.61	PRJEB25762
Q107.12C	3295802	16	Infants	London, England	NEOM	98.62	PRJEB25762

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
Q107.13C	3296932	15	Infants	London, England	NEOM	98.61	PRJEB25762
Q107.1C	3143128	27	Infants	London, England	NEOM	98.30	PRJEB25762
Q107.2C	3143673	26	Infants	London, England	NEOM	98.30	PRJEB25762
Q107.4C	3270656	29	Infants	London, England	NEOM	98.40	PRJEB25762
Q107.5C	3268477	27	Infants	London, England	NEOM	98.39	PRJEB25762
Q107.7C	3133904	21	Infants	London, England	NEOM	98.35	PRJEB25762
Q107.8C	3134078	23	Infants	London, England	NEOM	98.36	PRJEB25762
Q107.9C	3299946	15	Infants	London, England	NEOM	98.61	PRJEB25762
Q118.1C	3333932	24	Infants	London, England	NEOM	98.41	PRJEB25762
Q118.2C	3376789	26	Infants	London, England	NEOM	98.41	PRJEB25762
Q118.4C	3374700	30	Infants	London, England	NEOM	98.41	PRJEB25762
Q118.5C	3375850	22	Infants	London, England	NEOM	98.41	PRJEB25762
Q124.1C	3591545	17	Infants	London, England	NEOM	97.20	PRJEB25762
Q124.2C	3592345	19	Infants	London, England	NEOM	97.20	PRJEB25762
Q124.4C	3591925	17	Infants	London, England	NEOM	97.20	PRJEB25762
Q124.5C	3591587	18	Infants	London, England	NEOM	97.21	PRJEB25762
Q125.1C	3592250	16	Infants	London, England	NEOM	97.21	PRJEB25762
Q125.2C	3592243	17	Infants	London, England	NEOM	97.20	PRJEB25762
Q125.4C	3592144	17	Infants	London, England	NEOM	97.21	PRJEB25762
Q125.5C	3591863	17	Infants	London, England	NEOM	97.20	PRJEB25762
Q126.1C	3203802	14	Infants	London, England	NEOM	96.09	PRJEB25762
Q126.2C	3202647	15	Infants	London, England	NEOM	96.09	PRJEB25762
Q126.3C	3204069	16	Infants	London, England	NEOM	96.09	PRJEB25762
Q126.4C	3203065	14	Infants	London, England	NEOM	96.09	PRJEB25762
Q135.2.1C	3527791	31	Children	London, England	NEOM	97.24	PRJEB28307
Q135.CN1	3156272	32	Infants	London, England	NEOM	98.30	PRJEB25762
Q135.CN2	3158560	29	Infants	London, England	NEOM	98.30	PRJEB25762
Q135.CN4	3154909	38	Infants	London, England	NEOM	98.29	PRJEB25762
Q135.CN5	3157891	34	Infants	London, England	NEOM	98.31	PRJEB25762
Q137.1C	3459480	26	Infants	London, England	NEOM	98.54	PRJEB25762
Q137.2.2C	3333854	21	Children	London, England	NEOM	99.11	PRJEB28307
Q137.2.3C	3372422	90	Children	London, England	NEOM	98.90	PRJEB28307
Q137.2C	3460184	26	Infants	London, England	NEOM	98.53	PRJEB25762
Q137.4C	3456126	24	Infants	London, England	NEOM	98.54	PRJEB25762
Q137.5C	3460469	25	Infants	London, England	NEOM	98.54	PRJEB25762
Q138.1C	3460229	23	Infants	London, England	NEOM	98.53	PRJEB25762
Q138.2.5C	3283796	25	Children	London, England	NEOM	98.62	PRJEB28307
Q138.2.7C	3282195	28	Children	London, England	NEOM	98.61	PRJEB28307
Q138.2C	3458864	24	Infants	London, England	NEOM	98.53	PRJEB25762

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
Q138.4C	3459336	21	Infants	London, England	NEOM	98.54	PRJEB25762
Q138.5C	3459391	23	Infants	London, England	NEOM	98.54	PRJEB25762
Q142.1C	3146959	36	Infants	London, England	NEOM	98.30	PRJEB25762
Q142.2C	3142942	36	Infants	London, England	NEOM	98.31	PRJEB25762
Q143.1C	3459999	21	Infants	London, England	NEOM	98.53	PRJEB25762
Q143.2C	3458899	23	Infants	London, England	NEOM	98.53	PRJEB25762
Q143.CN1	3460998	25	Infants	London, England	NEOM	98.54	PRJEB25762
Q143.CN2	3459787	24	Infants	London, England	NEOM	98.53	PRJEB25762
Q143.CN4	3458736	28	Infants	London, England	NEOM	98.53	PRJEB25762
Q143.CN5	3459306	24	Infants	London, England	NEOM	98.53	PRJEB25762
Q143.CN7	3460400	24	Infants	London, England	NEOM	98.53	PRJEB25762
Q143.CN8	3460283	24	Infants	London, England	NEOM	98.53	PRJEB25762
Q143f.CN7	3454358	26	Infants	London, England	NEOM	98.54	PRJEB25762
Q146.2.5C	3026392	97	Children	London, England	NEOM	97.11	PRJEB28307
Q154.2.6C	3320860	63	Children	London, England	NEOM	98.32	PRJEB28307
Q161.2.3C	3312339	21	Children	London, England	NEOM	98.52	PRJEB28307
Q164.1C	3360445	19	Infants	London, England	NEOM	98.94	PRJEB25762
Q164.2C	3361328	21	Infants	London, England	NEOM	98.94	PRJEB25762
Q167.11C	3191571	17	Infants	London, England	NEOM	98.31	PRJEB25762
Q167.1C	3191399	17	Infants	London, England	NEOM	98.31	PRJEB25762
Q167.20C	3193775	21	Infants	London, England	NEOM	98.29	PRJEB25762
Q167.2C	3191805	20	Infants	London, England	NEOM	98.30	PRJEB25762
Q167.4C	3192114	18	Infants	London, England	NEOM	98.30	PRJEB25762
Q167.5C	3191783	20	Infants	London, England	NEOM	98.31	PRJEB25762
Q167.7C	3187987	18	Infants	London, England	NEOM	98.30	PRJEB25762
Q167.8C	3192625	16	Infants	London, England	NEOM	98.29	PRJEB25762
Q168.1C	3172313	59	Infants	London, England	NEOM	98.30	PRJEB25762
Q168.2C	3172634	64	Infants	London, England	NEOM	98.30	PRJEB25762
Q168.4C	3183576	16	Infants	London, England	NEOM	98.30	PRJEB25762
Q168.5C	3193363	16	Infants	London, England	NEOM	98.30	PRJEB25762
Q168.7C	3195069	13	Infants	London, England	NEOM	98.30	PRJEB25762
Q168.8C	3184875	17	Infants	London, England	NEOM	98.30	PRJEB25762
Q182.2.2C	3346383	39	Children	London, England	NEOM	98.56	PRJEB28307
Q210.2.1C	3283184	24	Children	London, England	NEOM	98.61	PRJEB28307
Q213.2.1C	3109696	15	Children	London, England	NEOM	96.15	PRJEB28307
Q215.CN1	3282615	24	Infants	London, England	NEOM	98.69	PRJEB25762
Q215.CN2	3281952	17	Infants	London, England	NEOM	98.69	PRJEB25762
SAF-1	3300478	93	Birds	Australia	Public database	98.61	PRJNA422745
SM101	2921996	3	Undetermined	Undetermined: from NCBI	Public database	97.17	PRJNA12521

Continued on next page

Strain	Assembly Statistics		Clinical/Origin Information				Study Accession
	Genome (bp)	Contigs	Source	Region/Country	Study/Site	ANI	
SOM-NE34	3418437	2	Birds	Australia	Public database	98.45	PRJNA422745
SOM-NE35	3567592	59	Birds	Australia	Public database	98.46	PRJNA422745
SYD-NE41	3738529	31	Birds	Australia	Public database	98.64	PRJNA422745
TAM-NE38	3646969	78	Birds	Australia	Public database	99.00	PRJNA422745
TAM-NE40	3645901	78	Birds	Australia	Public database	99.00	PRJNA422745
TAM-NE42	3363281	109	Birds	Australia	Public database	98.91	PRJNA422745
TAM-NE43	3275545	28	Birds	Australia	Public database	98.64	PRJNA422745
TAM-NE46	3677993	216	Birds	Australia	Public database	99.00	PRJNA422745
UDE95-1372	3644497	124	Birds	USA	Public database	99.00	PRJNA422745
W1319	3308237	12	Birds	Australia	Public database	98.75	PRJNA422745
WAL_14572	3462156	35	Undetermined	Undetermined: from NCBI	Public database	98.57	PRJNA46397
Warren	3776328	184	Birds	USA	Public database	99.03	PRJNA422745
WER-NE36	3411304	156	Birds	Australia	Public database	98.39	PRJNA422745

Appendix 5 Supplementary data for Chapter 8

This appendix contains all supplementary materials for Chapter 8.

Table S8.1: Raw pathological scoring of colonic sections for each mouse in infection model (1).

Table S8.2: Raw pathological scoring of colonic sections for each mouse in each individual group for infection model (2) - tetracycline model.

Colonic sections were scored single-blinded by Dr Lindsay J Hall. Values given for single criteria are not additive but represent the final score value that include all relevant criteria (as below):

1. Inflammatory infiltration (0-4)
2. Epithelial hyperplasia and goblet cell loss (0-5)
3. Mucosal architecture (0-5)

Table S8.1: Raw pathological scoring of colonic sections for each mouse in each individual group for infection model 1.

		Inflammatory infiltration (0-4)	Epithelial hyperplasia and goblet cell loss (0-5)	Mucosal architecture (0-5)	Total Score (out of 14)
Control - untreated	Mouse 1	0.5	0.5	0.5	1.5
	Mouse 2	1	0	0.5	1.5
	Mouse 3	0.5	0	0.5	1
	Mouse 4	1	1	2	4
	Mouse 5	0	0	0.5	0.5
Abx	Mouse 1	1.5	0.5	1.5	3.5
	Mouse 2	0.5	0.5	0.5	1.5
	Mouse 3	1.5	1.5	1.5	4.5
Abx + Q215	Mouse 1	2	1.5	3.5	7
	Mouse 2	2	3	3	8
	Mouse 3	2.5	3	3.5	9
	Mouse 4	2.5	4	4	10.5
	Mouse 5	3.5	4	3.5	11
	Mouse 6	2.5	4	4	10.5

Table S8.2: Raw pathological scoring of colonic sections for each mouse in each individual group for infection model 2 (tetracycline model).

		Inflammatory infiltration (0-4)	Epithelial hyperplasia and goblet cell loss (0-5)	Mucosal architecture (0-5)	Total Score (out of 14)
Abx + *tet*	Mouse 1	1	0.5	1.5	3
	Mouse 2	0.5	0.5	0.5	1.5
	Mouse 3	1	0.5	1	2.5
	Mouse 4	0.5	0	0.5	1
	Mouse 5	1	1.5	1.5	4
Abx + *tet* + Q215	Mouse 1	1.5	2	3	6.5
	Mouse 2	2	1.5	2	5.5
	Mouse 3	2	1.5	2.5	6
	Mouse 4	3	1.5	2	6.5
	Mouse 5	2.5	1.5	2.5	6.5
Abx + *tet* + Q215-GFP	Mouse 1	3	3	4	10
	Mouse 2	2	2	3	7
	Mouse 3	2	3.5	3	8.5
	Mouse 4	2	2.5	2.5	7
	Mouse 5	2.5	2.5	4	9
Abx + *tet* + Q143	Mouse 1	2	2.5	2.5	7
	Mouse 2	2.5	2.5	2	7
	Mouse 3	3	4	4	11
	Mouse 4	3	4	4	11
	Mouse 5	3	4	4	11

Appendix 6 Presentations and awards

Seminars/ oral presentations

1. Kiu R., Caim S., Amar C., Hall L. (2018) Phylogenomic analysis of food-poisoning associated *Clostridium perfringens* identifies isogenic strains in multiple outbreaks, and novel virulence-related features. *Quadram Institute Bioscience (QIB). Hall laboratory seminar. 4 September 2018.*
2. Kiu R., Hall L. (2018) Understanding the role of *Clostridium perfringens* in intestinal diseases. *Quadram Institute Bioscience (QIB). Hall laboratory seminar. 23 May 2018.*
3. Kiu R., Caim S., Hall L. (2018) Probing genomic aspects of the multi-host pathogen *Clostridium perfringens. The Institute for Microbiology and Infection (IMI) at the University of Birmingham. 2018 Junior Awards for Microbiology (J.A.M.) March Seminar. 9 March 2018.*
4. Kiu R. (2018) The role of *Clostridium perfringens* in intestinal diseases. *Quadram Institute Bioscience (QIB). Director's luncheon. 15 February 2018.*
5. Kiu R., Hall L. (2018) The role of *Clostridium perfringens* in intestinal diseases. *Quadram Institute Bioscience (QIB). Hall laboratory seminar. 25 November 2017.*
6. Kiu R., Caim S., Hall L. (2017) Probing genomic evolutionary & virulence traits in the multi-host pathogen *Clostridium perfringens. Quadram Institute Bioscience (QIB). Gut Health & Food Safety ISP Research-In-Progress Seminar. 4 September 2017.*
7. Kiu R., Hall L. (2018) The role of *Clostridium perfringens* in intestinal diseases. *Public Health England, Colindale. Collaboration and knowledge exchange meeting. 24 July 2017.*
8. Kiu R. (2017) The role of *Clostridium perfringens* in intestinal diseases. *Quadram Institute Bioscience (QIB). Hall laboratory seminar. 9 May 2017.*
9. Kiu R. (2016) The role of *Clostridium perfringens* in early life intestinal diseases. *Institute of Food Research (IFR). Hall laboratory seminar. 19 October 2016.*

10. Kiu R., Hall L. (2016) The role of *Clostridium perfringens* in early life intestinal diseases. *Institute of Food Research (IFR). IFR Student Seminar. 6 October 2016.*
11. Kiu R., Hall L. (2016) The role of *Clostridium perfringens* in early life intestinal diseases. *Trinity College Dublin, The University of Dublin, Ireland. Dublin Knowledge Exchange Programme oral presentation. 26 September 2016.*
12. Kiu R., Hall L. (2016) Impact of early life gut microbiota on infection risk and resistance. *Institute of Food Research (IFR). Annual Progress Meeting. 12 August 2016.*
13. Kiu R., Hall L. (2016) Impact of early life gut microbiota on infection risk and resistance. *Institute of Food Research (IFR). Hall laboratory seminar. 1 June 2016.*
14. Kiu R. (2016) Impact of early life gut microbiota on infection risk and resistance. *Institute of Food Research (IFR). Collaboration and knowledge exchange meeting. 18 May 2016.*
15. Kiu R., Hall L. (2016) Impact of early life gut microbiota on infection risk and resistance. *Institute of Food Research (IFR). PhD Transfer Panel Meeting. 14 March 2016.*
16. Kiu R., Hall L. (2016) Pre-term gut wars: Villain vs Hero. *Institute of Food Research (IFR). IFR Coffee Break Science. 11 March 2016.*
17. Kiu R., Hall L. (2016) How early life microbiota impacts on gut infection resistance. *Institute of Food Research (IFR). Hall laboratory seminar. 19 February 2016.*
18. Kiu R., Hall L. (2015) How early life microbiota impacts on infection resistance. *Institute of Food Research (IFR). Hall laboratory seminar. 16 September 2015.*

Poster presentations

1. Kiu R., Caim S., Alexander S., Pachori P., Hall L. (2018) Probing Genomic Aspects of the Multi-Host Pathogen *Clostridium perfringens.* Reveals Significant Pangenome Diversity, and a Diverse Array of Virulence Factors *Microbiology Society Annual Conference 2018.*
2. Kiu R., Caim S., Pickard D., Bedwell H., Alcon C., Ketskemety J., Shaw A., Sim K., Brown J., Belteki G., Clarke P., Dixon R., Dougan G., Kroll S., Hall L. J. (2017) Probing evolutionary relationship and virulence traits of *Clostridium perfringens. Microbiology Society Annual Conference 2017.*
3. Kiu R., Caim S., Pickard D., Bedwell H., Alcon C., Ketskemety J., Shaw A., Sim K., Brown J., Belteki G., Clarke P., Dixon R., Gordon G., Kroll J. S., Hall L. J. (2017) Probing evolutionary relationship and virulence traits of *Clostridium perfringens. Microbes in Norwich 2017.*
4. Leclaire C., Kiu R., Caim S., Flegg Z., O'Neill I., Hall L. J. (2017) Bile salt hydrolase activity in Bifidobacteria, a probiotic trait? *Microbes in Norwich*
5. Kiu R., Caim S., Pickard D., Bedwell H., Alcon C., Ketskemety J., Shaw A., Sim K., Brown J., Belteki G., Clarke P., Dixon R., Dougan G., Kroll S., Hall L. (2017) Probing evolutionary relationship and virulence traits of *Clostridium perfringens UEA FMH Postgraduate Research Conference.*
6. Kiu R., Caim S., Pickard D., Bedwell H., Alcon C., Ketskemety J., Shaw A., Sim K., Brown J., Belteki G., Clarke P., Dixon R., Dougan G., Kroll S., Hall L. (2017) Probing evolutionary relationship and virulence traits of *Clostridium perfringens. IFR Student Science Showcase.*
7. Kiu R., Lawson M. L., Hall L. H. (2015) How early life microbiota impacts immune function and infection resistance. *IFR Science Symposium.*
8. Kiu R., Lawson M., Alcon C., Walker A., Carding S., Hall L. (2015) How early life microbiota impacts infection resistance *Wageningen, the Netherlands. IFR-Wageningen University Student Symposium.*

Awards and scholarships

1. Kiu R. (2018) J.A.M. Talks Invited Speaker Travel Grant by Microbiology Society. March 2018.
2. Kiu R. (2018) Microbiology Society Annual Conference (Birmingham, UK) Travel Award. April 2018.
3. Kiu R. (2017) Conference on the Molecular Biology and Pathogenesis of the Clostridia (CLOSTPATH) Travel Award. July 2017.
4. Kiu R. (2017) Microbiology Society Annual Conference (Edinburgh, UK) Travel Award. April 2017.
5. Kiu R. (2015) UEA Faculty of Medical and Health Sciences Postgraduate Research International Bursary (2015-2017). January 2015.

Appendix 7 Peer-reviewed publications

All published peer-reviewed journal articles (including a genome report, a major research article, a Review article and a rebuttal article) during my PhD (2015-2018) were reproduced in this section (in chronological order):

1. Kiu, R., Caim, S., Alcon-Giner, C., Belteki, G., Clarke, P., Pickard, D., Dougan, G., Hall, L J. Preterm infant-associated *Clostridium tertium*, *Clostridium cadaveris*, and *Clostridium paraputrificum* strains: genomic and evolutionary insights. *Genome Biol Evol* **9**, 2707-2714, doi:10.1093/gbe/evx210 (2017).
    - My contribution in this work as first author:
        - Isolation of bacterial strains and DNA extraction.
        - Performed bioinformatics analysis and graphing.
        - Wrote up the entire article.

2. 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, doi:10.3389/fmicb.2017.02485 (2017).
    - My contribution in this work as joint-first author:
        - Performed bioinformatics analysis independently/with major support from Mr Shabhonam Caim (also joint-first author).
        - Graphed figures and wrote up the entire article (including results and discussion).

3. Kiu, R. & Hall, L. J. An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerging Microbes & Infections* **7**, 141, doi:10.1038/s41426-018-0144-8 (2018).
    - My contribution in this work as first author:
        - Researched existing literature and wrote up the Review article.
        - Researched and graphed new figures.

4. Kiu, R. & Hall, L. J. Response: Commentary: Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant pangenome diversity, and a diverse array of virulence factors. *Frontiers in Microbiology* **9**, doi:10.3389/fmicb.2018.01857 (2018).
    - My contribution in this work as first author:
        - Wrote up the rebuttal article.
        - Performed bioinformatics analysis.

GBE

Preterm Infant-Associated *Clostridium tertium*, *Clostridium cadaveris*, and *Clostridium paraputrificum* Strains: Genomic and Evolutionary Insights

Raymond Kiu[1,2], Shabhonam Caim[1], Cristina Alcon-Giner[1], Gusztav Belteki[3], Paul Clarke[4], Derek Pickard[5], Gordon Dougan[5], and Lindsay J. Hall[1,*]

[1]The Gut Health and Food Safety Programme, Quadram Institute Bioscience, Norwich Research Park, Norwich, United Kingdom

[2]Norwich Medical School, Norwich Research Park, University of East Anglia, Norwich, United Kingdom

[3]Neonatal Intensive Care Unit, The Rosie Hospital, Cambridge University Hospitals NHS Foundation Trust, United Kingdom

[4]Neonatal Intensive Care Unit, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, United Kingdom

[5]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom

*Corresponding author: E-mail: lindsay.hall@quadram.ac.uk.

Accepted: September 28, 2017

Data deposition: This project has been deposited at European Nucleotide Archive (EMBL-EBI) under the accession PRJEB22142. Bacterial strain deposition: Newly sequenced strains are deposited at National Collection of Type Cultures (NCTC; a culture depository of Public Health England).

Abstract

Clostridium species (particularly *Clostridium difficile*, *Clostridium botulinum*, *Clostridium tetani* and *Clostridium perfringens*) are associated with a range of human and animal diseases. Several other species including *Clostridium tertium*, *Clostridium cadaveris*, and *Clostridium paraputrificum* have also been linked with sporadic human infections, however there is very limited, or in some cases, no genomic information publicly available. Thus, we isolated one *C. tertium* strain, one *C. cadaveris* strain and three *C. paraputrificum* strains from preterm infants residing within neonatal intensive care units and performed Whole Genome Sequencing (WGS) using Illumina HiSeq. In this report, we announce the open availability of the draft genomes: *C. tertium* LH009, *C. cadaveris* LH052, *C. paraputrificum* LH025, *C. paraputrificum* LH058, and *C. paraputrificum* LH141. These genomes were checked for contamination in silico to ensure purity, and we confirmed species identity and phylogeny using both 16S rRNA gene sequences (from PCR and in silico) and WGS-based approaches. Average Nucleotide Identity (ANI) was used to differentiate genomes from their closest relatives to further confirm speciation boundaries. We also analysed the genomes for virulence-related factors and antimicrobial resistance genes, and detected presence of tetracycline and methicillin resistance, and potentially harmful enzymes, including multiple phospholipases and toxins. The availability of genomic data in open databases, in tandem with our initial insights into the genomic content and virulence traits of these pathogenic *Clostridium* species, should enable the scientific community to further investigate the disease-causing mechanisms of these bacteria with a view to enhancing clinical diagnosis and treatment.

Key words: Clostridium, functional annotation, whole genome sequencing, virulence.

Medical Relevance

Clostridium, which means "a small spindle" in Greek (due to its rod-shaped morphology), is classified as a genus under the phylum Firmicutes and class Clostridia, and comprises 221 species to date (September 2017) (Parte 2014). *Clostridium* spp. are Gram-positive spore-forming anaerobes found ubiquitously in the environment (soil and water) and the intestinal tract of humans and animals (Yamagishi et al. 1964; Miwa 1975; de Vos et al. 1982). There are several significant human and animal disease causing *Clostridium* species including *Clostridium difficile* (pseudomembranous colitis), *Clostridium botulinum* (infant botulism), *Clostridium tetani* (tetanus), and *Clostridium perfringens* (acute watery diarrhea/necrotising enterocolitis [NEC]), with associated pathology ascribed to the

Table 1
Genome Description, Assembly Statistics, and Clinical Information of Isolates Used in This Study

	C. tertium LH009	*C. cadaveris* LH052	*C. paraputrificum* LH025	*C. paraputrificum* LH058	*C. paraputrificum* LH141
Genome size (bp)	3,970,462	3,460,249	3,797,748	3,776,795	3,630,606
No. of contigs	49	46	40	84	29
Genes	3,910	3,395	3,896	3,823	3,655
CDS	3,821	3,310	3,813	3,745	3,565
N_{50} (bp)	258,765	118,391	479,233	101,241	390,404
tRNAs	89	84	83	77	90
GC content (%)	27.8	31.2	29.6	29.9	29.7
Origin of isolates	29-week preterm infant	32-week preterm infant	29-week preterm infant	32-week preterm infant	27-week preterm infant
Hospital	RH	NNUH	RH	NNUH	RH

*RH: Rosie Hospital, Cambridge, UK; NNUH: Norfolk and Norwich Hospital, Norwich, UK.

toxins they produce (Bruggemann and Gottschalk 2004; Awad et al. 2014; Carter and Peck 2015; Sim et al. 2015). There are also several less well-studied species including *Clostridium tertium*, *Clostridium paraputrificum*, and *Clostridium cadaveris*, which have been sporadically reported in the literature to be associated with human infection.

C. cadaveris (formerly *Clostridium capitovale*), is thought to be a key tissue-decomposing bacterium in dead carcasses, and is generally not considered pathogenic in living individuals (Poduval et al. 1999). However, this bacterium has infrequently been associated with human systemic diseases, including intraperitoneal infection (Leung et al. 2009) and bacteremia (Poduval et al. 1999; Schade et al. 2006).

C. tertium, is an aerotolerant and nontoxin-producing species. During The First World War, it was the third most frequently isolated bacteria from war wounds, after *C. perfringens* and *C. sporogenes* (Henry 1917). This organism was officially recognized as a pathogen in 1963, when the first *C. tertium*-associated septicemia case was reported (King et al. 1963). *C. tertium* has also been associated with infections including peritonitis (Butler and Pitt 1982) and pneumonia (Johnson and Tenover 1988). Importantly, *C. tertium* is also linked with cattle enteritis (Silvera et al. 2003), preterm NEC (Cheah et al. 2001) and adult enterocolitis (Coleman et al. 1993).

C. paraputrificum has previously been isolated from formula-fed infants within their first weeks of life (Stark and Lee 1982). This pathogen has been associated with paediatric infection (sepsis) (Brook 1995), adult necrotizing enterocolitis (Shandera et al. 1988), bacteremia (Shinha and Hadi 2015), and preterm NEC (Smith et al. 2011). Interestingly, this organism was shown to produce NEC-like lesions, including gas cysts, in an animal model and thus supports their disease-causing link (Waligora-Dupriet et al. 2005).

Whole genome sequencing (WGS) has contributed significantly to biomedical and veterinary research through our increased understanding of pathogens at a genomic level. Despite the medical importance of these three pathogenic *Clostridium* species, there is currently no sequenced genomes of *C. tertium* or *C. cadaveris* available to the research community (apart from 16S rRNA gene sequences) and only four genomes of *C. paraputrificum* accessible on NCBI databases as of September 2017 (Geer et al. 2010). In this study, we sequenced one *C. cadaveris* isolate, one *C. tertium* isolate and three *C. paraputrificum* isolates from preterm infant faecal samples obtained from two neonatal intensive care units (NICUs) units in England. We identified these using their 16S rRNA gene sequences (both full-length PCR and in silico) and WGS-based k-mer phylogenetic assignment, thus contributing new genomic data on these pathogenic bacteria. We also verified their phylogenetic positions using WGS data, measured genetic distances via Average Nucleotide Identity (ANI), and performed genome-wide functional annotation (COG classification). These genomic data and analyses increases our understanding of the virulence potentials and functionalities of these pathogenic bacteria, with a future view to unraveling disease-causing mechanisms.

Genome Description

Here, we report the release of draft genomes sequenced on Illumina HiSeq 2500 platform as stated in table 1. *C. paraputrificum* isolates have a genome size between 3.6 and 3.7 million bases and a stable GC content from 29.6 to 29.9%, which is in line with the four public genomes (Geer et al. 2010). *C. tertium* has a larger genome (3.9 million bases) and relatively lower GC content of 27.8%, whilst *C. cadaveris* has a smaller genome (3.4 million bases) compared with *C. paraputrificum*, and a significantly higher GC content of 31.2%. All draft genomes were assembled using Prokka de novo assembler and 80% (four out of five) of the genomes analyzed were <50 contigs, except for *C. paraputrificum* LH058 with 84 contigs.

These five strains were isolated from preterm infants residing at two different NICUs (table 1), which is in line with previous findings that report frequent detection of *C. paraputrificum* (16–22%) and *C. tertium* (4–9%) in infant cohorts (Tonooka et al. 2005; Ferraris et al. 2012). However, to date there are no reports of *C. cadaveris* isolation from infants.

Phylogenetic Positions

To assign phylogenetic position, and identify these isolates, we computationally extracted 16S rRNA sequences from genomes to construct a *Clostridium* 16S rRNA phylogeny (based on 19 isolates in the NCBI nucleotide database) as in figure 1*A*. Here, we coupled three genomic approaches to confirm taxonomic position of these newly released genomes. We firstly, performed a PCR targeting almost the full length of the 16S rRNA gene, and predicted the whole 16S rRNA gene sequence in silico. Secondly, we employed Average Nucleotide Identity (ANI) to confirm species boundaries; ANI cut-offs for species discrimination is known to be approximately 95%, and this value has been reported to mirror the traditional taxonomic gold standard method DNA–DNA hybridization (DDH) to define species (Richter and Rossello 2009). Lastly, we performed CVTree—an alignment-free whole genome-based phylogenetic construction method, which is known for speed and accuracy for taxonomic assignment (Xu and Hao 2009).

At a 16S rRNA level, LH058, LH141, and LH025 fall in the same lineage as *C. paraputrificum* DSM2630, indicating species-level relatedness (fig. 1*A*), with LH052 clustering with *C. cadaveris* JCM1392, and LH009 within the same lineage as *C. tertium* ATCC14573 and *Clostridium sartagoforme* KAR69. CVTree phylogenetic analysis, providing greater resolution based on sequence comparison, showed similar relationships. (fig. 1*B*); all *C. paraputrificum* isolates grouped within the same lineage as *C. paraputrificum* DSM2630, when compared with other *Clostridium* species, indicating correct species assignment for isolates LH058, LH141, and L025. *C. cadaveris* LH052 is most closely related to *C. perfringens* ATCC13124, and LH009 (*C. tertium* as assigned according to 16S data) is closely related to *C. sartagoforme* AAU1 (fig. 1*B*).

We next used ANI analysis to provide higher phylogenetic resolution (fig. 2*A*). *C. paraputrificum* AGR2156 are identical to LH025, LH141, and LH058 in terms of nucleotide sequences, sharing ANI of >95.7%, thus determined to be the same species. Although LH009 is closely related to *C. sartagoforme* AAU1, the ANI calculation does not allocate these two within the same species (ANI = 83.6%, < 95% as species cut-off), which indicates LH009 is distinct from its closest relatives, and may be identified as the species *C. tertium*. LH052 is also evolutionarily distant (based on ANI calculation, 68.5%) from other *Clostridium*, indicating this isolate is a separate species, *C. cadaveris*.

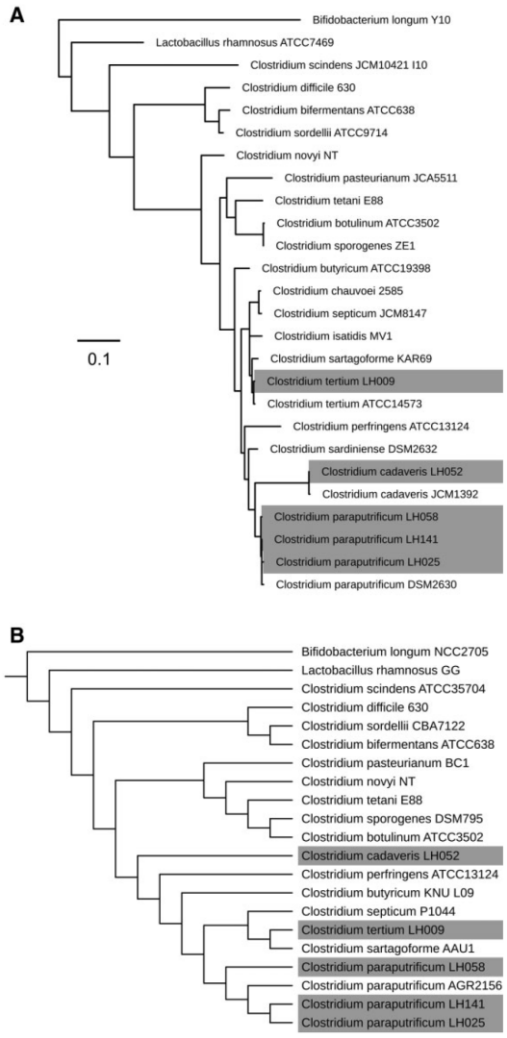


Fig. 1.—(*A*) 16S rRNA maximum-likelihood (ML) phylogenetic tree of 19 species of *Clostridium*. (*B*) WGS-based alignment-free cladogram of representative *Clostridium* species. *Lactobacillus rhanmosus* and *Bifidobacterium longum* have been used as outgroups. Grey labels indicate newly sequenced isolates in this study.

Virulence Traits and Genome-Wide Functional Analyses

Using genome annotations, we performed a thorough search on virulence-related terms including "phospholipase," "hemolysin," "resistance," "lactamase," "drug," and "toxin" to provide initial insights into the potential

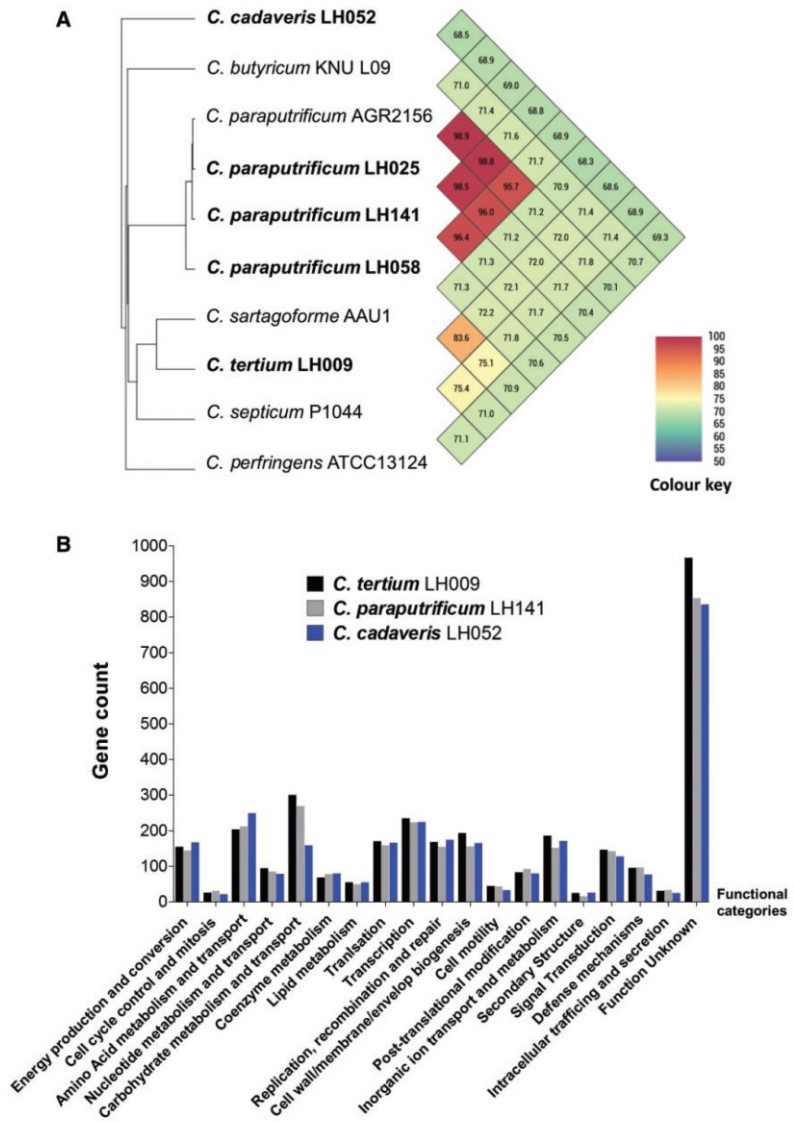


FIG. 2.—(*A*) Average Nucleotide Identity (ANI) values in representative *Clostridium* genomes. (*B*) Comparison of functional annotations based on COG classifications on three representative genomes.

virulence-linked genes encoded within these genomes (table 2).

C. tertium LH009, *C. cadaveris* LH052, and *C. paraputrificum* LH141 harbour phospholipase genes (*ytpA*) that are homologous to phospholipases encoded in other pathogen genomes including *Bacillus subtilis*, *Pseudomonas aeruginosa*, and *Streptococcus pneumoniae*. Phospholipases are known to possess hydrolytic activity against eukaryotic cell membranes, and are thus considered key virulence factors. *C. perfringens* produces homologous phospholipase C (also known as alpha toxin) that has previously been reported to damage epithelial cells (Verherstraeten et al. 2013), and which shares >58% protein sequence identity with the phospholipase encoded by gene *ytpA*. Importantly, LH052 and LH141 also

Table 2
Virulence-Related Genes Detected in Selected *Clostridium* Genomes

Isolate	Gene Names	Gene Description and Functions
C. tertium LH009	*ytpA*	Phospholipase
	vanW	Vancomycin B-type resistance
	stp	Multidrug resistance
	mdtK	Multidrug resistance
	tetM	Tetracycline resistance
	marA	Multiple antibiotic resistance
	mecR1	Methicillin resistance
	norM	Multidrug resistance
	mepA	Multidrug export protein
	hcpC	Beta-lactamase precursor
	sme-1	Carbapenem-hydrolyzing beta-lactamase precursor
C. cadaveris LH052	*ytpA*	Phospholipase
	(n/a)	Patatin-like phospholipase
	(n/a)	Phospholipase C precursor
	toxA	Toxin A
	marA	Multiple antibiotic resistance
	vanW	Vancomycin B-type resistance
	norm	Multidrug resistance protein
	tetM	Tetracycline resistance
	marR	Multiple antibiotic resistance
	mdtA	Multidrug resistance
	mepA	Multidrug export protein
	(n/a)	Beta-lactamase precursor
C. paraputrificum LH141	*ytpA*	Phospholipase
	toxA	Toxin A
	marR	Multiple antibiotic resistance
	tetM	Tetracycline resistance
	norM	Multidrug resistance
	vanW	Vancomycin B-type resistance
	mdtK	Multidrug resistance
	mepA	Multidrug export protein
	(n/a)	Beta-lactamase precursor
	sme-1	Carbapenem-hydrolyzing beta-lactamase precursor

possess *toxA* gene, which encodes *C. difficile*-associated Toxin A, known to be one of the main virulence factors during infection having cytotoxic and proinflammatory activities (Awad et al. 2014).

Notably, antimicrobial resistance genes are encoded in all three genomes, including vancomycin (*vanW*) and tetracycline resistance (*tetM*) (Evers and Courvalin 1996; Donhofer et al. 2012). Other resistance traits include multidrug efflux pumps; that is, those encoded by *mdtK* and *norM* (fluoroquinolones) (Horiyama et al. 2011; Golparian et al. 2014), *mdtA* (aminocoumarin) (Guerrero et al. 2012), and efflux pump transcriptional regulators *marA* and *marR* (Maira-Litran et al. 2000). In addition, methicillin resistance gene *mecR1* was detected in LH009 (Shore et al. 2011), whilst beta-lactamase (penicillins and carbapenems) precursor (inactive protein sequence that could potentially be activated via posttranslational modification) was encoded in all genomes (Marciano et al. 2007). The prevalence of multiple antimicrobial resistance genes in these clinical strains may correspond to the environment in which they were isolated; preterm infants residing in NICUs where antimicrobial usage is extensive (Albrich et al. 2004).

From the COG-based genome-wide annotation, most genes (>40% in each genome) did not map with any known functional orthologs, which highlights the limitation of genomic tools and current databases, for understanding these bacteria at a functional level. Gene counts in most categories of these three genomes did not differ significantly from one another (fig. 2*B*). However, the number of genes involved in carbohydrate metabolism and transport is lower in *C. cadaveris* LH052 ($n = 159$), than encoded in *C. tertium* LH009 ($n = 300$) and *C. paraputrificum* LH141 ($n = 269$), whereas LH052 possesses more genes ($n = 249$) involved in amino acid metabolism and transport as compared with LH009 ($n = 203$) and LH141 ($n = 212$). These functional differences may correspond to divergent modes of metabolism and nutritional substrates for *C. cadaveris*, which is distinct from *C. tertium* and *C. paraputrificum* (correlates to WGS phylogeny positions), and may link to previous isolations of this species from additional environmental niches, i.e. dead carcasses. Therefore, we conclude that these three *Clostridium* species are similar in terms of genomic functionalities, however due to the high number of function-unknown genes, this somewhat reduces in-depth comparison between genomes and will require further experimental work.

Materials and Methods

Faecal Sample Collection

Fecal sample collection was performed under an on-going preterm infant study (BAMBI) which is approved by University of East Anglia (UEA) Faculty of Medical and Health Sciences (FMH) Ethics Committee. Sample collection was done in accordance with the procedures outlined by National Research Ethics Service (NRES) approved UEA Biorepository (Licence no.: 11208). Participating infants were given written consent by their parents for fecal sample collection at Norfolk and Norwich University Hospital (Norwich, UK) and Rosie Hospital (Cambridge, UK). Fecal samples were routinely collected from infant nappies in the NICUs into sterile stool containers and stored at 4 °C.

Bacterial Isolates and Preliminary 16S rRNA PCR Identification

A total of five *Clostridium* isolates (including *C. tertium*, *C. cadaveris*, and *C. paraputrificum*) were analyzed in this study. Isolates were preliminarily identified using 16S rRNA full-length PCR (Weisburg et al. 1991). Primers used as in table 3.

Table 3
Sequence of Primers Used for PCR Amplification of 16S rRNA Gene

Primers	Sequence
fD1	5′-AGA GTT TGA TCC TGG CTC AG-3′
fD2	5′-AGA GTT TGA TCA TGG CTC AG-3′
rP1	5′-ACG GTT ACC TTG TTA CGA CTT-3′

Near 1kbp PCR products were subsequently sequenced (Eurofins, Luxembourg) and compared with 16S rRNA bacteria sequence database on NCBI using BLASTn (optimized for megablast) search algorithm (Camacho et al. 2009).

Genomic DNA Extraction

Overnight 10 ml pure cultures in BHI were harvested for phenol-chloroform DNA extraction. Briefly, bacterial pellets were resuspended in 2 ml 25% sucrose in 10 mM Tris and 1 mM EDTA at pH 8.0. Cells were lysed using 50 µl 100 mg/ml lysozyme (Roche). 100 µl 20 mg/ml Proteinase K (Roche), 30 µl 10 mg/ml RNase A (Roche), 400 µl 0.5 M EDTA (pH 8.0) and 250 µl 10% Sarkosyl NL30 (Fisher) were added subsequently into the lysed bacterial suspension. This follows by 1-h ice incubation and 50 °C overnight water bath.

Second-day protocol comprises three rounds of phenol-chloroform-isoamyl alcohol (Sigma) extraction using 15 ml gel-lock tubes (Qiagen). Chloroform-Isoamyl alcohol (Sigma) extraction was performed to remove residual phenol, followed by ethanol precipitation and 70% ethanol wash. DNA pellets were finally resuspended in 200–300 µl of 10 mM Tris (pH 8.0). DNA concentration was quantified using Qubit dsDNA BR assay kit (Invitrogen) and DNA quality assessed by Nanodrop spectrophotometer.

Whole Genome Sequencing, Genome Assembly and Annotation

Isolated DNA of pure cultures was subjected to multiplex standard Illumina library preparation protocol followed by sequencing via Illumina HiSeq 2500 platform with read length 2 × 125 bp (paired-end reads) and an average sequencing coverage of 60×. Draft genome assemblies were generated using an assembly and annotation pipeline as described previously (Page et al. 2016). All genomes were annotated using *Prokka* v1.11 (Seemann 2014).

Contamination Estimation

Webtool *ContEst16S* was used to check for potential contamination of the draft genomes based on Genbank database (Lee et al. 2017).

16S rRNA Phylogeny

Publicly available16S rRNA genes were retrieved from NCBI nucleotide database (Geer et al. 2010). 16S rRNA genes from our isolates were predicted using *Barrnap* v0.7 (https://github.com/Victorian-Bioinformatics-Consortium/barrnap, last accessed September 20, 2017) and extracted using *BEDTools* getfasta utility (Quinlan and Hall 2010). All 16S rRNA sequences were subsequently concatenated as a multisequence fasta, and sequences were aligned with *MUSCLE* (Edgar 2004). Neighbor-joining (NJ) tree was generated in 1000 bootstrap replicates using Juke-Cantor distance (Gouy et al. 2010). Maximum-likelihood (ML) tree was produced by *PhyML* GTR model with 1000 bootstrap replicates (Guindon et al. 2010). Trees were edited using *iTOL* (Letunic and Bork 2016).

Alignment-Free WGS Phylogeny

Selected *Clostridium* genome sequences were retrieved from NCBI genome database. Annotated multiple protein sequences were used as input for *CVTree* v5.0 to generate alignment-free WGS-based phylogeny using the optimized six as the k-tuple length (Xu and Hao 2009). Tree was edited using *iTOL* as described in previous section.

Average Nucleotide Identity (ANI)

OrthoANI Tool v.093 (OAT) was employed to calculate the ANI (both directions) between genomes (Lee et al. 2016). Identity >95% was used as cut-off for species delineation.

Genome-Wide Functional Assignment (COG)

Functional assignments were implemented using *eggNOG-mapper* v0.99.3 (Huerta-Cepas et al. 2017), based on *eggNOG* orthology data (Huerta-Cepas et al. 2016). Sequence searches were performed using HMMER3 (Eddy 2011). Data were extracted using Shell scripts (https://github.com/raymondkiu/eggnog-mapper_COGextraction, last accessed September 20, 2017) and visualized in *GraphPad PRISM* v5.04.

Ethics Approval and Consent for Participation

This study was approved by the University of East Anglia (UEA) Faculty of Medical and Health Sciences (FMH) Ethics Committee. Sample collection follows the protocols outlined by NRES approved UEA Biorepository (Licence no.: 11208). Written consent was given by the parents for their infants for participation in this study.

Acknowledgments

This work was supported by a Wellcome Trust Investigator Award to LJH (grant number 100974/C/13/Z). RK is a PhD student at the Norwich Medical School of University of East Anglia (UEA), partially funded by UEA international bursary.

Literature Cited

Albrich WC, Monnet DL, Harbarth S. 2004. Antibiotic selection pressure and resistance in *Streptococcus pneumoniae* and *Streptococcus pyogenes*. Emerg Infect Dis. 10(3): 514–517.

Awad MM, Johanesen PA, Carter GP, Rose E, Lyras D. 2014. *Clostridium difficile* virulence factors: insights into an anaerobic spore-forming pathogen. Gut Microbes 5(5): 579–593.

Brook I. 1995. Clostridial infection in children. J Med Microbiol. 42(2): 78–82.

Bruggemann H, Gottschalk G. 2004. Insights in metabolism and toxin production from the complete genome sequence of *Clostridium tetani*. Anaerobe 10(2): 53–68.

Butler T, Pitt S. 1982. Spontaneous bacterial peritonitis due to *Clostridium tertium*. Gastroenterology 82(1): 133–134.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Carter AT, Peck MW. 2015. Genomes, neurotoxins and biology of *Clostridium botulinum* Group I and Group II. Res Microbiol. 166(4): 303–317.

Cheah FC, Lim KE, Boo NY. 2001. *Clostridium tertium* in cerebrospinal fluid of a premature neonate with necrotizing enterocolitis: contamination or real? Acta Paediatr. 90(6): 704–705.

Coleman N, et al. 1993. Neutropenic enterocolitis associated with *Clostridium tertium*. J Clin Pathol. 46(2): 180–183.

de Vos N, Mevissen-Verhage E, van Amerongen WH, Marcelis J. 1982. A new selective medium for the culture of Clostridia from human faeces. Eur J Clin Microbiol. 1(5): 267–271.

Donhofer A, et al. 2012. Structural basis for TetM-mediated tetracycline resistance. Proc Natl Acad Sci U S A. 109(42): 16900–16905.

Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol. 7(10): e1002195.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5): 1792–1797.

Evers S, Courvalin P. 1996. Regulation of VanB-type vancomycin resistance gene expression by the VanS(B)-VanR (B) two-component regulatory system in *Enterococcus faecalis* V583. J Bacteriol. 178(5): 1302–1309.

Ferraris L, et al. 2012. Clostridia in premature neonates' gut: incidence, antibiotic susceptibility, and perinatal determinants influencing colonization. PLoS One 7(1): e30594.

Geer LY, et al. 2010. The NCBI BioSystems database. Nucleic Acids Res. 38(Database issue): D492–D496.

Golparian D, Shafer WM, Ohnishi M, Unemo M. 2014. Importance of multidrug efflux pumps in the antimicrobial resistance property of clinical multidrug-resistant isolates of *Neisseria gonorrhoeae*. Antimicrob Agents Chemother. 58(6): 3556–3559.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 27(2): 221–224.

Guerrero P, et al. 2012. Characterization of the BaeSR two-component system from *Salmonella* Typhimurium and its role in ciprofloxacin-induced mdtA expression. Arch Microbiol. 194(6): 453–460.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3): 307–321.

Henry H. 1917. An investigation of the cultural reactions of certain anaerobes found in wounds. J Pathol Bacteriol. 21: 344–385.

Horiyama T, Nikaido E, Yamaguchi A, Nishino K. 2011. Roles of *Salmonella* multidrug efflux pumps in tigecycline resistance. J Antimicrob Chemother. 66(1): 105–110.

Huerta-Cepas J, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. doi: 10.1093/molbev/msx148.

Huerta-Cepas J, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44(D1): D286–D293.

Johnson JR, Tenover FC. 1988. *Clostridium tertium* bacteremia in a patient with aspiration pneumonia: an elusive diagnosis. J Infect Dis. 157(4): 854–855.

King BM, Ranck BA, Daugherty FD, Rau CA. 1963. *Clostridium tertium* septicemia. N Engl J Med. 269: 467–469.

Lee I, et al. 2017. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. Int J Syst Evol Microbiol. 67(6): 2053–2057.

Lee I, Kim YO, Park SC, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 66(2): 1100–1103.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44(W1): W242–W245.

Leung J, Sasson M, Patel SR, Viveiros K. 2009. *Clostrium cadaveris* intraperitoneal abscess. Am J Gastroenterol. 104(10): 2635–2636.

Maira-Litran T, Allison DG, Gilbert P. 2000. An evaluation of the potential of the multiple antibiotic resistance operon (mar) and the multidrug efflux pump acrAB to moderate resistance towards ciprofloxacin in *Escherichia coli* biofilms. J Antimicrob Chemother. 45(6): 789–795.

Marciano DC, Karkouti OY, Palzkill T. 2007. A fitness cost associated with the antibiotic resistance enzyme SME-1 beta-lactamase. Genetics 176(4): 2381–2392.

Miwa T. 1975. Clostridia in soil of the Antarctica. Jpn J Med Sci Biol. 28(4): 201–213.

Page AJ, et al. 2016. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microb Genom. 2(8): e000083.

Parte AC. 2014. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res. 42(Database issue): D613–D616. doi: 10.1093/nar/gkt1111

Poduval RD, Mohandas R, Unnikrishnan D, Corpuz M. 1999. *Clostridium cadaveris* bacteremia in an immunocompetent host. Clin Infect Dis. 29(5): 1354–1355. doi: 10.1086/313491

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6): 841–842.

Richter M, Rossello MR. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 106(45): 19126–19131.

Schade RP, et al. 2006. *Clostridium cadaveris* bacteraemia: two cases and review. Scand J Infect Dis. 38(1): 59–62.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14): 2068–2069.

Shandera WX, Humphrey RL, Stratton LB. 1988. Necrotizing enterocolitis associated with *Clostridium paraputrificum* septicemia. South Med J. 81(2): 283–284.

Shinha T, Hadi C. 2015. *Clostridium paraputrificum* bacteremia associated with colonic necrosis in a patient with AIDS. Case Rep Infect Dis. 2015: 312919. doi: 10.1155/2015/312919

Shore AC, et al. 2011. Detection of staphylococcal cassette chromosome mec type XI carrying highly divergent mecA, mecI, mecR1, blaZ, and ccr genes in human clinical isolates of clonal complex 130 methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother. 55(8): 3765–3773. doi: 10.1128/AAC.00187-11

Silvera M, Finn B, Reynolds KM, Taylor DJ. 2003. *Clostridium tertium* as a cause of enteritis in cattle. Vet Rec. 153(2): 60.

Sim K, et al. 2015. Dysbiosis anticipating necrotizing enterocolitis in very premature infants. Clin Infect Dis. 60(3): 389–397.

Smith B, et al. 2011. Community analysis of bacteria colonizing intestinal tissue of neonates with necrotizing enterocolitis. BMC Microbiol. 11: 73. doi: 10.1186/1471-2180-11-73

Stark PL, Lee A. 1982. Clostridia isolated from the feces of infants during the first year of life. J Pediatr. 100(3): 362–365.

Tonooka T, et al. 2005. Detection and quantification of four species of the genus *Clostridium* in infant feces. Microbiol Immunol. 49(11): 987–992.

Verherstraeten S, et al. 2013. The synergistics necrohemorrhagic action of *Clostridium perfringens* perfringolysin and alpha toxin in the bovine intestine and against bovine endothelial cells. Vet Res. 44(1): 45.

Waligora-Dupriet AJ, Dugay A, Auzeil N, Huerre M, Butel MJ. 2005. Evidence of clostidial implication in necrotizing enterocolitis through bacterial fermentation in a gnotobiotic quail model. Pediatr Res. 58(4): 629–635.

Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 173(2): 697–703.

Xu Z, Hao B. 2009. CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. Nucleic Acids Res. 37(Web Server issue): W174–W178.

Yamagishi T, Ishida S, Nishida S. 1964. Isolation of toxigenic strains of *Clostridium perfringens* from soil. J Bacteriol. 88: 646–652.

Associate editor: Howard Ochman

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 December 2017
doi: 10.3389/fmicb.2017.02485

Probing Genomic Aspects of the Multi-Host Pathogen *Clostridium perfringens* Reveals Significant Pangenome Diversity, and a Diverse Array of Virulence Factors

Raymond Kiu[1,2†], Shabhonam Caim[1†], Sarah Alexander[3], Purnima Pachori[4] and Lindsay J. Hall[1]*

[1] Gut Health and Food Safety, Quadram Institute Bioscience, Norwich Research Park, Norwich, United Kingdom, [2] Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, United Kingdom, [3] Public Health England, London, United Kingdom, [4] Earlham Institute, Norwich Research Park, Norwich, United Kingdom

OPEN ACCESS

Edited by:
Frank T. Robb,
University of Maryland, Baltimore, United States

Reviewed by:
Nikolai Ravin,
Research Center for Biotechnology of Russian Academy of Sciences, Russia
Karl Hassan,
University of Newcastle, Australia

***Correspondence:**
Lindsay J. Hall
lindsay.hall@quadram.ac.uk

[†] These authors have contributed equally to this work and shared first authorship.

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

Received: 24 August 2017
Accepted: 29 November 2017
Published: 12 December 2017

Citation:
Kiu R, Caim S, Alexander S, Pachori P and Hall LJ (2017) Probing Genomic Aspects of the Multi-Host Pathogen Clostridium perfringens Reveals Significant Pangenome Diversity, and a Diverse Array of Virulence Factors. Front. Microbiol. 8:2485. doi: 10.3389/fmicb.2017.02485

Clostridium perfringens is an important cause of animal and human infections, however information about the genetic makeup of this pathogenic bacterium is currently limited. In this study, we sought to understand and characterise the genomic variation, pangenomic diversity, and key virulence traits of 56 *C. perfringens* strains which included 51 public, and 5 newly sequenced and annotated genomes using Whole Genome Sequencing. Our investigation revealed that *C. perfringens* has an "open" pangenome comprising 11667 genes and 12.6% of core genes, identified as the most divergent single-species Gram-positive bacterial pangenome currently reported. Our computational analyses also defined *C. perfringens* phylogeny (16S rRNA gene) in relation to some 25 *Clostridium* species, with *C. baratii* and *C. sardiniense* determined to be the closest relatives. Profiling virulence-associated factors confirmed presence of well-characterised *C. perfringens*-associated exotoxins genes including α-toxin (*plc*), enterotoxin (*cpe*), and Perfringolysin O (*pfo* or *pfoA*), although interestingly there did not appear to be a close correlation with encoded toxin type and disease phenotype. Furthermore, genomic analysis indicated significant horizontal gene transfer events as defined by presence of prophage genomes, and notably absence of CRISPR defence systems in >70% (40/56) of the strains. In relation to antimicrobial resistance mechanisms, tetracycline resistance genes (*tet*) and anti-defensins genes (*mprF*) were consistently detected *in silico* (*tet*: 75%; *mprF*: 100%). However, pre-antibiotic era strain genomes did not encode for *tet*, thus implying antimicrobial selective pressures in *C. perfringens* evolutionary history over the past 80 years. This study provides new genomic understanding of this genetically divergent multi-host bacterium, and further expands our knowledge on this medically and veterinary important pathogen.

Keywords: *Clostridium perfringens*, pangenome, antimicrobial resistance, genomics, whole genome sequencing, Clostridial infection, exotoxins

INTRODUCTION

Clostridium perfringens is a Gram-positive spore-forming anaerobe, best known as the causative agent for the tissue necrotic disease gas gangrene (also known as Clostridial myonecrosis). Notably, *C. perfringens* has also been widely associated with various intestinal diseases across human and animal species including: broiler necrotic enteritis (Keyburn et al., 2008), human food poisoning (Scallan et al., 2011), and preterm necrotising enterocolitis (Sim et al., 2015; Heida et al., 2016). Remarkably, it is reported to secrete >20 degradative toxins which constitute its primary arsenal to initiate histotoxic pathogenesis in both humans and animals (Uzal et al., 2014; Revitt-Mills et al., 2015). Intestinal-associated disease aetiology is characterised by rapid anaerobic proliferation in host tissue accompanied by *in vivo* production of several key pore-forming toxins including α-toxin, β-toxin, and β2-toxin, that consequently disrupt epithelial barrier function and induce histotoxic infections or tissue necrosis (Li and Mcclane, 2006). Common disease symptoms include pronounced diarrhoea (human adult food poisoning) (Mcclung, 1945), gaseous tissue necrosis (neonatal humans and animals) (Kosloske et al., 1978), and a haemorrhagic gut (neonatal horses) (Mehdizadeh Gohari et al., 2015).

C. perfringens was first isolated in the 1890s and named *Clostridium welchii* after its discoverer William Henry Welch (Welch and Nuttall, 1891). This gut pathogen is serotyped (or, specifically toxinotyped) clinically from A to E based upon the combination of "major toxins" (namely, α-toxin, β-toxin, ι-toxin, and ε-toxin) it encodes (Hassan et al., 2015). Multiplex PCR on the key toxin genes and Matrix Assisted Laser Desorption/Ionization time-of-flight mass (MALDI-TOF) spectrometry are the most common typing/identification methods in medical laboratories alongside conventional culture-based biochemical identification (Van Asten et al., 2009; Nagy et al., 2012).

Although this widespread gut pathogen has been studied and experimentally characterised over the past century, these studies have primarily focused on *C. perfringens*-associated toxins (Nagler, 1939; Ohno-Iwashita et al., 1986; Gibert et al., 1997; Stevens and Bryant, 2002). Notably, information on the entire pangenome [or, "the entire collection of gene families that are found in a given species" (Mcinerney et al., 2017)] of this pathogen is limited; to the date of analysis (April 2017) only 51 strains have been sequenced and made available publicly (Geer et al., 2010).

Previous comparative genomic studies on *C. perfringens* (up to 12 strains) indicated substantial genome variation within the pangenome (Hassan et al., 2015). Hence, we sought to construct the latest and largest pangenome of *C. perfringens* to probe evolutionary relationships and understand the genomic makeup of this bacterium. The *C. perfringens* genomes publicly available represent strains isolated from various diseased hosts including gas gangrene (human), food poisoning (human), necrotic enteritis (poultry), enterotoxaemia (sheep), and haemorrhagic enteritis (dogs and foals) from the past 100 years. This collection also represents the full range of *C. perfringens* toxinotypes (A–E). In this study, we also sequenced an additional 5 *C. perfringens* strains (as part of the NCTC3000 project[1]), for their historical significance (mostly isolated before 1960s), and thus included 56 genomes (including 51 genomes in NCBI databases) in our *in silico* investigation. Here we characterise genomic aspects of this medically and veterinary important pathogen, and through bioinformatic analysis of Whole Genome Sequencing (WGS) data identify virulence traits, define evolutionary relationships and functional annotations of the pangenome. Our analysis indicates that *C. perfringens* has a surprisingly diverse pangenome (core genes ∼12.6%), potentially driven by horizontal gene transfer (HGT).

RESULTS

Probing Evolutionary Relationships

Initially we sought to determine taxonomy and phylogenetic relatedness between *C. perfringens* and a subset of pathogenic/environmental *Clostridium* species (25) using 16S rRNA approach (Supplementary Table 1) (Woese et al., 1990) as previous earlier studies had reported that *Clostridium pasteurianum*, *Clostridium baratii,* and *Clostridium absonum* were the closest relatives (Canard et al., 1992; Collins et al., 1994). Using the 16S rRNA gene predictor Barrnap (>800 bp sequences used, 38/56 strains) we determined that the 16S rRNA regions appear to be highly conserved (identity >99.1%; Supplementary Table 2) across all selected 38 *C. perfringens* strains as they form a monophyletic lineage in the 16S *Clostridium* phylogenetic tree. *C. perfringens* cluster was clearly separated from other *Clostridium* species, with closest relatives identified as *C. baratii* (toxin producer associated with infant botulism) and *Clostridium sardiniense* (α-toxin producers isolated from gas gangrene cases) (**Figure 1**) (Masaki et al., 1988; Harvey et al., 2002). Clinically important toxin-generating bacteria *C. botulinum*, potentially pathogenic *Clostridium paraputrificum*, *Clostridium tertium*, scavenger *Clostridium cadaveris,* and deadly-toxin producer *Clostridium tetani* all fall in the same sub-lineage (cluster 2) as *C. perfringens*, with *Clostridium difficile* (known for nosocomial antibiotic-associated diarrhea), *Clostridium sordellii* and *C. bifermentans* (uncommon environmental species infrequently linked to human diseases) diverging earlier from their common ancestor (cluster 1), suggesting an ancient divergence in evolutionary history (Edagiz et al., 2015).

Evolutionary distances based on near full-length 16S rRNA gene sequence (**Table 1**) also indicates that *C. baratii* and *C. sardiniense* are the closest relatives of *C. perfringens* in this dataset with sequence identities of ∼95%, followed by *C. absonum* (94%), *C. paraputrificum* (93%), *C. septicum* (93%), *C. butyricum* (93%), *C. tetani* (91%), and *C. pasteurianum* (90%). These data support the 16S rRNA ML tree (**Figure 1**) as strains with higher sequence identities all appear to be more closely related to *C. perfringens* as indicated by tree topology.

[1]NCTC 3000 Project website: www.sanger.ac.uk/resources/downloads/bacteria/nctc/

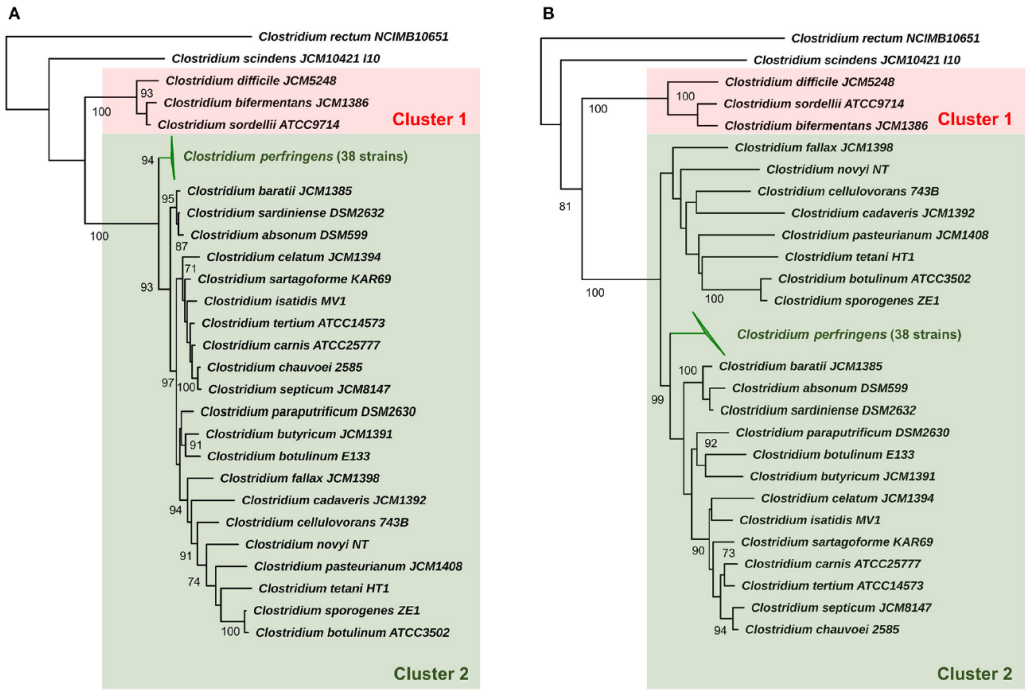


FIGURE 1 | 16S rRNA phylogeny of representative *Clostridium* species (25) and *C. perfringens*. **(A)** Maximum-likelihood phylogenetic tree of *C. perfringens* strains and representative *Clostridium* species (25) based on 16S rRNA genes with aLRT branch support values displayed. **(B)** Neighbour-joining phylogeny based on 16S rRNA genes supported by 1000 bootstrap replicates based on 1585 sites. Clusters 1 and 2 were assigned for description purposes. Branch support values >70 are shown on the nodes.

Genome Description

A total of 51 genomes were retrieved in preassembled nucleotide FASTA files from NCBI database for further analysis (March 2017) and 5 NCTC strains were subjected to optimised DNA extraction procedures and sequenced (PacBio) as part of NCTC 3000 project (www.sanger.ac.uk/resources/downloads/bacteria/nctc/) with genomes annotated using Prokka. Whilst, contig number was variable (1–274) for NCBI-associated genomes (sequenced by different platforms and assembled using different methods), NCTC13170 and NCTC2544 reads assembled into 1 unitig, and NCTC8503, NCTC8797, and NCTC8678 genomes <10 contigs. Genome size of *C. perfringens* isolates ranged from 2.9 to 4.1 million bases, with an average GC content between 27.7 and 28.7% (**Table 2**). Predicted genes range from 2,600 to 3,800 with a median of 3304 genes.

Pangenome Construction and Analysis

To define key genomic components, we next constructed the pangenome of *C. perfringens* encompassing 56 isolates, which represents the largest analysis of this type to date. This pangenome comprises 11,667 genes; 1470 core genes and 10,197 accessory genes (**Figure 2A**). Remarkably, only 12.6% represented core genes and a surprisingly high 44% (5139/11667) of unique genes (defined as genes only present in 1 strain in this pangenome; **Figures 2B–E**). This analysis implies very high genome plasticity in this pathogen, more than any other known prokaryotes currently published and studied (Mcinerney et al., 2017).

Phylogeny and Genome Comparison

A core-genome phylogenetic approach was used for inference of *C. perfringens* genetic relatedness of these 56 isolates (**Figure 3**). The generated phylogenetic tree indicated four main clades, which correlated with pertinent metadata (**Table 3**). Food poisoning isolates SM101, NCTC8797, and NCTC8678 (collected between1940–60 from diseased human adults) clustered in a single lineage (clade 1), whereas human, sheep, chicken, dog, horse, and soil isolates intermingled in clade 2 or clade 3, suggesting potential spread and transmission between hosts. In clade 4 (clearly split from clade 1–3), most isolates were collected from a North American dog and horse *C. perfringens* haemorrhagic enteritis study (JFP isolates) indicating these

TABLE 1 | 16S rRNA sequence pair-wise comparison between five representative *C. perfringens* strains and 20 selected *Clostridium* species.

Strains	*Clostridium perfringens*									
	ATCC13124 (1,511 bp)		Str13 (1,513 bp)		FORC003 (1,511 bp)		JP838 (1,511 bp)		NCTC2544 (1,511 bp)	
	Identity (%)	Coverage (%)	Identity (%)	Coverage (%)	Identity (%)	Coverage (%)	Identity (%)	Coverage (%)	Identity (%)	Coverage (%)
Clostridium baratii JCM1385	95.33	100	95.06	100	95.33	100	95.26	100	95.33	100
Clostridium sardiniense DSM2632	95.18	99.67	94.92	99.80	95.05	99.67	95.11	99.67	95.18	99.67
Clostridium absonum DSM599	94.35	99.79	94.08	99.93	94.21	99.79	94.16	100	94.35	99.79
Clostridium paraputrificum DSM2630	93.77	99.93	93.52	100	93.77	99.93	93.91	100	93.64	99.93
Clostridium sartagoforme KAR69	93.73	99.86	93.60	100	93.66	99.86	93.73	99.86	93.59	99.86
Clostridium septicum JCM8147	93.70	100	93.57	100	93.77	100	93.63	100	93.56	100
Clostridium isatidis MV1	93.58	99.93	93.32	100	93.71	99.93	93.91	99.93	93.44	99.93
Clostridium fallax JCM1398	93.57	100	93.45	100	93.50	100	93.52	100	93.44	100
Clostridium tertium ATCC14573	93.53	100	93.25	100	93.46	100	93.53	100	93.38	100
Clostridium botulinum E133	93.45	100	91.09	100	93.52	100	93.67	100	93.59	100
Clostridium butyricum JCM1391	93.42	100	93.30	100	93.49	100	93.63	100	93.29	100
Clostridium botulinum Eklund17B	93.32	100	93.10	100	93.32	100	93.48	100	93.47	100
Clostridium celatum JCM1394	92.82	100	92.56	100	92.89	100	92.82	100	92.69	100
Clostridium carnis ATCC25777	92.57	99.86	92.30	100	92.50	99.86	92.57	99.86	92.43	99.86
Clostridium cellulovorans 743B	91.64	100	91.39	100	91.71	100	91.91	100	91.51	100
Clostridium tetani HT1	91.21	99.93	91.09	100	91.14	99.93	91.35	100	91.35	99.93
Clostridium cadaveris JCM1392	90.58	100	90.32	100	90.58	100	90.79	100	90.51	100
Clostridium pasteurianum JCM1408	90.22	100	90.10	100	90.29	100	90.28	100	90.09	100
Clostridium difficile JCM5248	86.25	100	86.14	100	86.32	100	86.39	100	86.25	100
Clostridium bifermentans JCM1386	85.82	100	85.68	100	85.89	100	85.89	100	85.82	100
Clostridium rectum NCIMB10651	79.66	100	79.60	100	79.66	100	79.63	100	79.57	100

isolates might be host- and disease-specific. This supports previous work where these JFP *netF*-positive isolates were found to cluster closely in a cgMLST tree (Mehdizadeh Gohari et al., 2017). Furthermore, the presence of 2 JFP isolates of Swiss origin also clustered in clade 4, which again suggests disease- and host- specificity, but exclusion of geographical linkages. We also observed that different toxinotypes grouped within the same clade (**Figure 3** and **Table 3**). These findings are similar to those previously reported by Hassan et al. and Mehdizadeh Gohari et al. and suggest that certain toxin genes are encoded within the accessory genome (Li et al., 2013; Hassan et al., 2015; Park et al., 2016; Mehdizadeh Gohari et al., 2017). Our analysis also demonstrates the presence of several major toxins encoded within *C. perfringens* plasmids (sequences individually retrieved from NCBI nucleotide database), highlighting the potential for toxin gene transfer via HGT (Supplementary Figure 1). Interestingly, whole-genome alignment-free phylogeny (Supplementary Figure 2) shows the clustering of toxinotypes (B-E in 1 lineage), in contrast to the core-genome based phylogeny (**Figure 3**), which further supports the hypothesis that several specific major toxins obtained via HGT are present in accessory genome.

Multi-dimensional scaling (MDS) investigation (Supplementary Figure 3) based on whole-genome gene-absence-presence phylogeny (Supplementary Figure 4) also mirrors the core-genome phylogeny findings (**Figure 3**); food-poisoning isolates cluster tightly, JFP isolates nest in 2 distinct lineages (clades 2 and 4) and remaining isolates form another cluster. Using JP838 as our reference genome (as it contains a large 3.5 Mbp chromosome for nucleotide comparison) we generated a circular genome comparison figure, based on 1–3 strains from each of the 4 core-genome clades (Supplementary Figure 5). This comparison supports our results from the pangenomic analysis highlighting significant variation; faded areas (<80% sequence similarity) exist in several regions as well as regions (>10) with differences in GC content, which may indicate genomic island insertion sites via HGT. Interestingly, an additional whole-genome search on strain JP838 indicated that most coding sequences predicted in divergent GC regions encode for hypothetical protein genes (via BLAST NR database) of unknown function.

Functional Annotation of Core and Accessory Genomes

We next analysed both the core and accessory genomes at a functional level using "Clusters of Orthologous Groups of proteins" (COG) and "Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups" (eggNOG) databases (Galperin et al., 2015; Huerta-Cepas et al., 2016), to probe

TABLE 2 | Genome description of 56 *C. perfringens* isolates included in this study.

	Strain	Genome size (bp)	Contigs	Predicted genes	G+C (%)
1	1207	3207161	95	2838	28.11
2	13	3085740	2	2827	28.51
3	2789STDY5608889	3321076	48	3000	28.08
4	ATCC13124	3256683	1	2973	28.38
5	B_ATCC3626	3896305	98	3563	28.36
6	C_JGS1495	3661329	84	3354	28.58
7	CP4	3642209	98	3469	27.80
8	CPE_F4969	3510272	74	3172	28.57
9	D_JGS1721	4045016	221	3768	28.26
10	E_JGS1987	4127102	101	3896	28.08
11	F262	3464306	53	3213	28.00
12	FORC003	3395109	2	3091	28.41
13	FORC025	3343822	1	3082	28.49
14	JFP718	3652220	56	3387	27.99
15	JFP727	3624033	47	3355	28.01
16	JFP728	3579626	85	3313	27.98
17	JFP771	3495308	81	3245	27.99
18	JFP774	3548595	56	3252	28.03
19	JFP795	3578912	67	3286	28.01
20	JFP796	3601148	114	3330	28.01
21	JFP801	3580610	59	3298	28.01
22	JFP804	3625968	117	3331	28.02
23	JFP810	3830406	95	3651	28.04
24	JFP826	3660408	70	3398	27.99
25	JFP828	3564370	54	3303	27.99
26	JFP829	3639686	108	3354	27.96
27	JFP833	3521770	96	3277	27.94
28	JFP834	3549140	121	3305	27.88
29	JFP836	3599532	51	3341	28.03
30	JFP914	3819232	79	3607	27.96
31	JFP916	3660641	67	3400	28.02
32	JFP921	3601767	78	3310	27.95
33	JFP922	3612099	76	3334	27.98
34	JFP923	3589715	48	3291	28.02
35	JFP941	3594640	64	3330	28.00
36	JFP961	3617219	66	3330	28.01
37	JFP978	3585277	68	3383	27.90
38	JFP980	3654588	67	3385	28.00
39	JFP981	3580367	68	3388	27.90
40	JFP982	3670557	62	3406	28.01
41	JFP983	3657522	46	3388	28.00
42	JFP986	3669049	58	3406	27.99
43	JFP992	3657868	70	3376	28.01
44	JJC	3259329	69	2970	28.12
45	JP55	3347300	1	3088	28.38
46	JP838	3530414	1	3244	28.38
47	MJR7757A	3585666	274	3265	27.79
48	NA	3417203	31	3157	28.33
49	NCTC8239	3324319	55	2942	28.66
50	SM101	2921996	3	2696	28.23

(Continued)

TABLE 2 | Continued

	Strain	Genome size (bp)	Contigs	Predicted genes	G+C (%)
51	WAL14572	3462156	35	3264	28.10
52	NCTC13170	3310238	1	2993	28.36
53	NCTC2544	3195093	1	2854	28.46
54	NCTC8503	3577234	9	3294	28.20
55	NCTC8678	3005443	6	2785	28.12
56	NCTC8797	3021559	5	2813	28.21
Summary: Median(range)		3.58(2.92-4.12) Mb	65(1-274)	3304(2696-3896)	28.02(27.79-28.66)

potential *C. perfringens* host adaption and/or pathogenesis traits (**Figure 4**).

Interestingly, 849 genes were assigned to protein families associated with "replication, recombination, and repair" comprising mainly transposases ($n = 254$), integrases ($n = 31$), and phage proteins ($n = 7$) within the accessory genome. Whereas, only 79 genes encoded within the core genome were assigned to this functional class, suggesting a key difference between both genomes at a functional level (849 vs. 79; 16.2 vs. 5.6%). Genes associated with "defense mechanisms" were higher within the accessory genome ($n = 202$), when compared to the core genome ($n = 48$), which includes genes encoding efflux pumps, restriction enzymes and ABC transporters (linked with iron uptake systems). Genes involved in metabolism (including carbohydrate, amino acid, and lipid) are proportionately more represented within the core genome (16.8%) vs. the accessory genome (8.3%), whilst recombination and cell-wall biogenesis related genes dominate the functionality of the accessory genome (27.1% vs. core genome: 12.2%). Notably, as indicated above there is a high proportion of unknown-function genes observed in both genomes (core: 24% vs. accessory: 38.7%; **Figure 4**).

In Silico Profiling of Virulence Traits

Toxins secreted by *C. perfringens* have long been considered key virulence factors and are implicated in multiple clinical disease phenotypes (Welch and Nuttall, 1891; Rood, 1998). Primary toxins produced include α-toxin, β-toxin, ι-toxin, and ε-toxins, which are used to classify this pathogen into different toxinotypes (Petit et al., 1999). These toxins, as well as additional toxins β2-toxin and perfringolysin O, potentially initiate pathogenesis through pore-forming and cytolysis (Supplementary Table 4). Therefore, we analysed the toxin profiles of the 56 isolates, and additional virulence-associated factors such as antimicrobial resistance (AMR) and prophage content (**Figure 5**).

Several toxins, including α-toxin (*plc*), α-clostripain (*ccp*), and microbial collagenase (*colA*) genes are conserved in all isolates (**Figure 5**). Other toxins, such as enterotoxin (*cpe*), a food-poisoning-linked toxin, was found to be encoded in food-poisoning isolates SM101 and NCTC8797, however this gene is also present in all JFP isolates (haemorrhagic enteritis isolates). Perfringolysin O (*pfoA* or *pfo*), a well-characterised pore-forming toxin, is encoded in most genomes except food-poisoning and JJC environmental isolates and presence of sialidase genes (*nan* and *nag*) vary across the isolates while *netB*, *netE*, *netF*, and *netG* genes were found to be uniquely encoded within JFP isolates (associated with haemorrhagic enteritis in dogs and foals), as reported previously (Mehdizadeh Gohari et al., 2016).

Antimicrobial-peptide (AMP) resistance gene multi-peptide resistance factor (*mprF*) was detected in all isolates. This gene is highly-conserved in certain Gram-positive and Gram-negative bacteria (**Figure 5**) and is known to confer resistance through modifying surface molecules of AMPs, including host defensins, and potentially also aminoglycoside antibiotics (Ernst and Peschel, 2011; Cole and Nizet, 2016). Tetracycline-resistant efflux protein *tetA(P)* gene (Bannam et al., 2004) were detected in 75% of the isolates ($n = 42$), and 23% of the isolates ($n = 14$) harboured double *tet*-resistance genes—*tetA(P)* and *tetB(P)*. The only other AMR-associated gene detected was specific aminoglycosides resistance gene (*ANT(6)-Ib*), which was encoded within 1 toxinotype C porcine isolate JGS1495.

Finally, genome scanning for prophage elements indicated that 19 in total were encoded, potentially derived from a variety of bacterial genera including *Aeromonas, Bacillus, Clostridium, Escherichia,* and *Staphylococcus* (**Figure 5**). Notably, JFP isolates also displayed unique prophage profiles including *Clostridium* phage ϕvB_cpes_CP51 ($n = 48$), *Aeromonas* phage ϕphiAS4 ($n = 47$), ϕaes508 and ϕ25 ($n = 45$), and *Clostridium* phage ϕphiS63 ($n = 44$). Interestingly, 40/56 (71.5%) isolates do not encode CRISPR system at all (*in silico* prediction = 0; **Figure 5**).

DISCUSSION

Although *C. perfringens* is an important causative agent of animal and human infections; our knowledge on the genome content and phylogenic characteristics of this pathogenic bacterium is currently limited. Our current study, analysing the largest number of *C. perfringens* genomes to date, highlights significant genomic plasticity, and an arsenal of key virulence factors, including toxins and AMR genes, that are encoded by this pathogen.

Using the 16S rRNA gene to initially probe *C. perfringens* phylogenic relationships (Koch et al., 1994), we observed clear deviation of 16S alignment from other *Clostridium* species (**Figure 1**), with *C. baratii* and *C. sardiniense* as the closest relatives of *C. perfringens* (16S sequence similarity ∼95.3 and 95.1%), which agrees with previous 16S phylogenetic studies (Collins et al., 1994; Vos et al., 2009).We also performed alignment-free whole-genome-based phylogenetics, which has

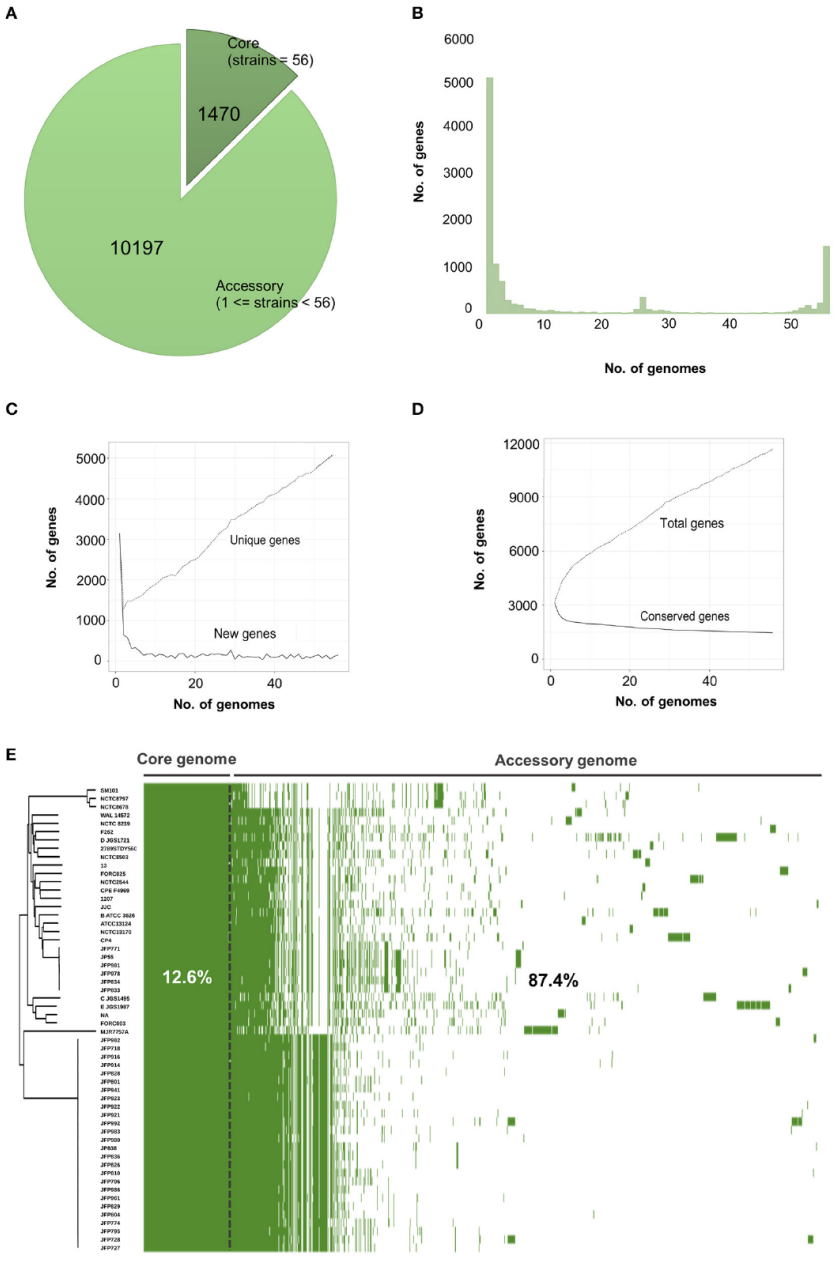


FIGURE 2 | Visualisation and statistics of *C. perfringens* pangenome. **(A)** Core and accessory genes statistics. **(B)** Frequency bar graphs of number of identical genes against number of genomes. **(C)** Number of new genes and unique genes along pangenome computation. **(D)** Number of conserved genes and total genes along pangenome calculation. **(E)** Linearised pangenomic view of 56 *C. perfringens* isolates. Green cells represent syntenies aligned next to Neighbour-Joining core-genome based phylogenetic tree.

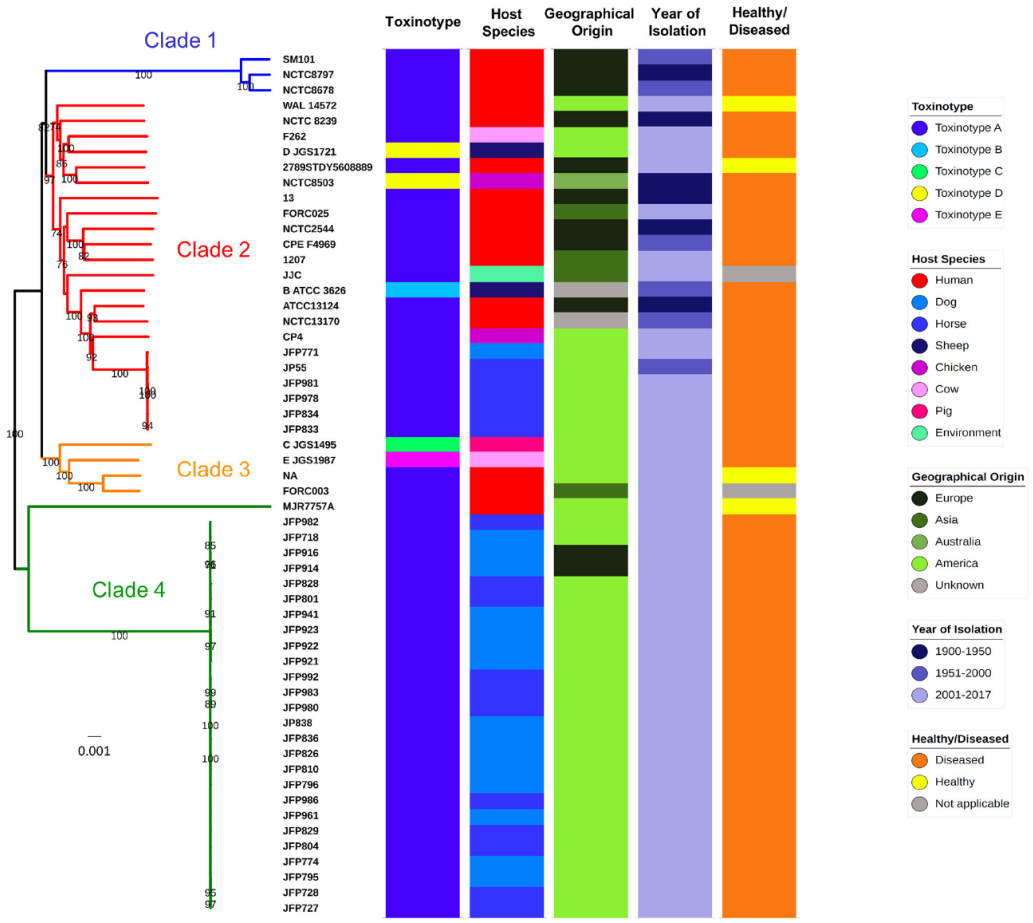


FIGURE 3 | Genome comparison and core-genome phylogeny of *C. perfringens*. Core-genome based Neighbour-Joining (NJ) phylogeny of 56 isolates. Branch support values >70 (based on 1,000 bootstrap replicates) are shown on the nodes. Categories of metadata were displayed for comparison. MJR7757A is assigned to Clade 4 although it is distant from the rest of JFP isolates.

been reported to represent a more comprehensive and robust method for clustering genetically variable prokaryotes including *Helicobacter pylori* (Van Vliet and Kusters, 2015; Bernard et al., 2016). Our analysis indicates that this approach can also be used for *C. perfringens*, as we observed near-identical phylogeny when compared to the core-genome based method (Supplementary Figure 2), and may prove useful in epidemiology and outbreak management.

Previous limited genomic studies have indicated that *C. perfringens* is a genetically diverse organism, and our larger scale study probing 56 genomes also supports this notion. We determined that *C. perfringens* has a highly variable open/infinite pangenome, with new genes potentially being added at an average of 106 genes for each new genome sequenced (Supplementary Figure 6), which is in comparison to other pathogens, including *Yersinia pestis* (Eppinger et al., 2010) and *Bacillus anthracis* (Rouli et al., 2015), which have closed pangenomes. The "lifestyle" of the organism is proposed to reflect the nature of the pangenome, and *C. perfringens*' ability to thrive in multiple host species and environmental niches, including the human and animal gut and soil, allows active interaction with other bacterial species, and links to our open pangenome data (Rouli et al., 2015). To our knowledge, only a handful of bacterial pangenome studies have reported extreme pan-genomic variation. This trait (12.5% core genes) is an uncommon observation in other bacteria and is significantly lower than other reported

TABLE 3 | Metadata of the *C. perfringens* genomes used for analysis in this study.

	Isolate	Alternative name	Toxino-type	Year of isolation*	Origin of isolates	Host diseased states	NCBI accession	References
1	ATCC13124	NCTC8237	A	1941	Human	Gas gangrene	SAMN02604008	Möllby et al., 1976
2	SM101	NCTC8798	A	1953	Meat rissole/Human	Food poisoning	SAMN02604026	Hobbs et al., 1953; Myers et al., 2006
3	Strain 13	Lechlen	A	1939	Human	Gas gangrene	SAMD00061119	Paquette and Fredette, 1967; Mahony and Moore, 1976; Shimizu et al., 2002
4	FORC_003	FORC3	A	2015	Contaminated food	n/a	SAMN03140316	n/a
5	JP55	JFP55	A	1999	Foal	Necrotizing enteritis	SAMN03372134	Mehdizadeh Gohari et al., 2016
6	JP838	JFP838	A	2009	Dog	Haemorrhagic gastroenteritis	SAMN03377063	Mehdizadeh Gohari et al., 2016
7	FORC_025	FORC25	A	2016	Human	Food poisoning	SAMN04209542	n/a
8	F262		A	2011	Calf	Bovine Clostridial abomasitis	SAMN00199225	Nowell et al., 2012
9	WAL-14572		A	2012	Human (gut of an autistic child)	Human Microbiome Project	SAMN02463856	Ribeiro et al., 2012
10	CP4		A	2006	Chicken	Necrotic enteritis	SAMN04017578	Lepp et al., 2010
11	MJR7757A		A	2016	Human (vagina)	Human Microbiome Project	SAMN03842618	Ribeiro et al., 2012
12	JGS1495		C	2007*	Pig	Diarrhoea	SAMN02436294	Lepp et al., 2010
13	ATCC3626	NCTC13110, NCIMB 10691	B	1955	Lamb	Dysentery	SAMN02436295	Lepp et al., 2010
14	JGS1987		E	2007*	Calf	Haemorrhagic Enteritis	SAMN02436167	Lepp et al., 2010
15	F4969		A	1995	Human	Diarrhoea	SAMN02436168	Cornillot et al., 1995
16	NCTC8239	ATCC12917	A	1949	Salt beef/ Human	Food poisoning	SAMN02436239	Hassan et al., 2015
17	JGS1721		D	2008*	Sheep	Enterotoxaemia	SAMN02436277	Hassan et al., 2015
18	JJC		A	2013	Landfill sludge	n/a	SAMN02317206	Wong et al., 2014
19	2789STDY5608889		A	2015*	Human	Healthy donor	SAMEA3545297	n/a
20	JFP718		A	2011	Dog	Haemorrhagic gastroenteritis	SAMN05323879	Mehdizadeh Gohari et al., 2017
21	JFP774		A	2011	Dog	Haemorrhagic gastroenteritis	SAMN05323883	Mehdizadeh Gohari et al., 2017
22	JFP728		A	2011	Foal	Necrotizing enteritis	SAMN05323881	Mehdizadeh Gohari et al., 2017
23	JFP727		A	2011	Foal	Necrotizing enteritis	SAMN05323880	Mehdizadeh Gohari et al., 2017
24	JFP795		A	2012	Dog	Haemorrhagic gastroenteritis	SAMN05323884	Mehdizadeh Gohari et al., 2017
25	JFP801		A	2002	Foal	Necrotizing enteritis	SAMN05323886	Mehdizadeh Gohari et al., 2017
26	JFP804		A	2010	Foal	Necrotizing enteritis	SAMN05323887	Mehdizadeh Gohari et al., 2017
27	JFP796		A	2012	Dog	Haemorrhagic gastroenteritis	SAMN05323885	Mehdizadeh Gohari et al., 2017
28	JFP771		A	2011	Dog	Haemorrhagic gastroenteritis	SAMN05323882	Mehdizadeh Gohari et al., 2017
29	JFP810		A	2012	Dog	Haemorrhagic gastroenteritis	SAMN05323888	Mehdizadeh Gohari et al., 2017
30	JFP826		A	2012	Dog	Haemorrhagic gastroenteritis	SAMN05323889	Mehdizadeh Gohari et al., 2017
31	JFP828		A	2011	Foal	Necrotizing enteritis	SAMN05323890	Mehdizadeh Gohari et al., 2017
32	JFP829		A	2010	Foal	Necrotizing enteritis	SAMN05323891	Mehdizadeh Gohari et al., 2017
33	JFP833		A	2000	Foal	Necrotizing enteritis	SAMN05323892	Mehdizadeh Gohari et al., 2017
34	JFP834		A	2002	Foal	Necrotizing enteritis	SAMN05323893	Mehdizadeh Gohari et al., 2017
35	JFP836		A	2008	Dog	Haemorrhagic gastroenteritis	SAMN05323894	Mehdizadeh Gohari et al., 2017

(Continued)

TABLE 3 | Continued

	Isolate	Alternative name	Toxino-type	Year of isolation*	Origin of isolates	Host diseased states	NCBI accession	References
36	JFP914		A	2009	Dog	Haemorrhagic gastroenteritis	SAMN05323895	Mehdizadeh Gohari et al., 2017
37	JFP916		A	2009	Dog	Haemorrhagic gastroenteritis	SAMN05323896	Mehdizadeh Gohari et al., 2017
38	JFP921		A	2007	Dog	Haemorrhagic gastroenteritis	SAMN05323897	Mehdizadeh Gohari et al., 2017
39	JFP922		A	2006	Dog	Haemorrhagic gastroenteritis	SAMN05323898	Mehdizadeh Gohari et al., 2017
40	JFP923		A	2006	Dog	Haemorrhagic gastroenteritis	SAMN05323899	Mehdizadeh Gohari et al., 2017
41	JFP941		A	2013	Dog	Haemorrhagic gastroenteritis	SAMN05323900	Mehdizadeh Gohari et al., 2017
42	JFP961		A	2013	Dog	Haemorrhagic gastroenteritis	SAMN05323901	Mehdizadeh Gohari et al., 2017
43	JFP978		A	2011	Foal	Necrotizing enteritis	SAMN05323902	Mehdizadeh Gohari et al., 2017
44	JFP980		A	2006	Foal	Necrotizing enteritis	SAMN05323903	Mehdizadeh Gohari et al., 2017
45	JFP981		A	2004	Foal	Necrotizing enteritis	SAMN05323904	Mehdizadeh Gohari et al., 2017
46	JFP982		A	2001	Foal	Necrotizing enteritis	SAMN05323905	Mehdizadeh Gohari et al., 2017
47	JFP983		A	2004	Foal	Necrotizing enteritis	SAMN05323906	Mehdizadeh Gohari et al., 2017
48	JFP992		A	2008	Foal	Necrotizing enteritis	SAMN05323908	Mehdizadeh Gohari et al., 2017
49	JFP986		A	2011	Foal	Necrotizing enteritis	SAMN05323907	Mehdizadeh Gohari et al., 2017
50	1207_CPER		A	2012	Human (blood)	ICU patient	SAMN03197169	Roach et al., 2015
51	NA (not available)		A	2013	Human/infant	Premature infant	SAMN05509317	Raveh-Sadka et al., 2015
52	NCTC8678	ATCC12919	A	1951	Human	Food poisoning	SAMEA3867459	Hobbs et al., 1953
53	NCTC8797	HF2985/50	A	1950	Salt beef	Food poisoning	SAMEA3867461	Hobbs et al., 1953
54	NCTC13170	DSMZ 100947/ D10	A	1993	Human	Food poisoning	SAMEA3867463	Mooijman et al., 2003
55	NCTC8503	CN366	D	1930	Chicken	Necrotic enteritis	SAMEA3879480	Bennetts, 1932
56	NCTC2544	STEELE COL 5	A	1928	Human	Infected gall bladder	SAMEA3919787	n/a

*Year of isolation here is predicted based on sequence submission date on public genome databases due to non-traceability in literature.

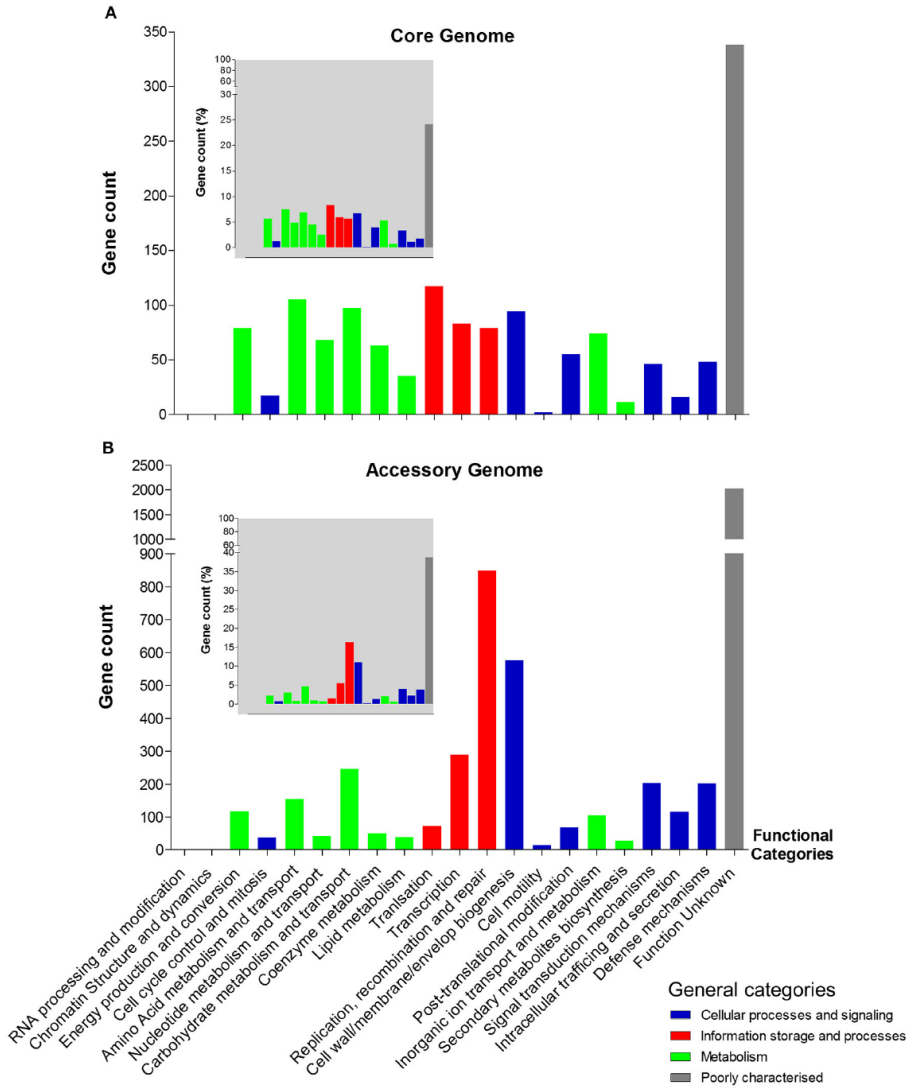


FIGURE 4 | Functional annotation (COG) of *C. perfringens* core and accessory genomes. **(A)** Core genome functional annotation. **(B)** Accessory genome functional annotation. The smaller-scale grey-background charts show the percentage of gene count of each functional class in both core and accessory genomes corresponding to the functional categories. Non-matching genes (unclassified) are not included in the statistics.

prokaryotic species including *C. difficile* (30.3%) (Knight et al., 2016), *Streptococcus pneumoniae* (46.5%) (Hiller et al., 2007), *Salmonella enterica* (16%), and *Klebsiella pneumoniae* (26%) (Mcinerney et al., 2017). It should be noted that this pangenome analysis may be limited in terms of capturing total *C. perfringens* diversity due to the fact that 30 isolates in our analysis are from a single study (Mehdizadeh Gohari et al., 2017). These *netF*-positive isolates were sequenced and assembled for plasmid pathogenicity analysis, and were reported to have tight clonal relationship using a MLST approach that could potentially bias the analysis of this dataset. Thus, we expect that due to *C. perfringens*' wide host and environmental range, additional

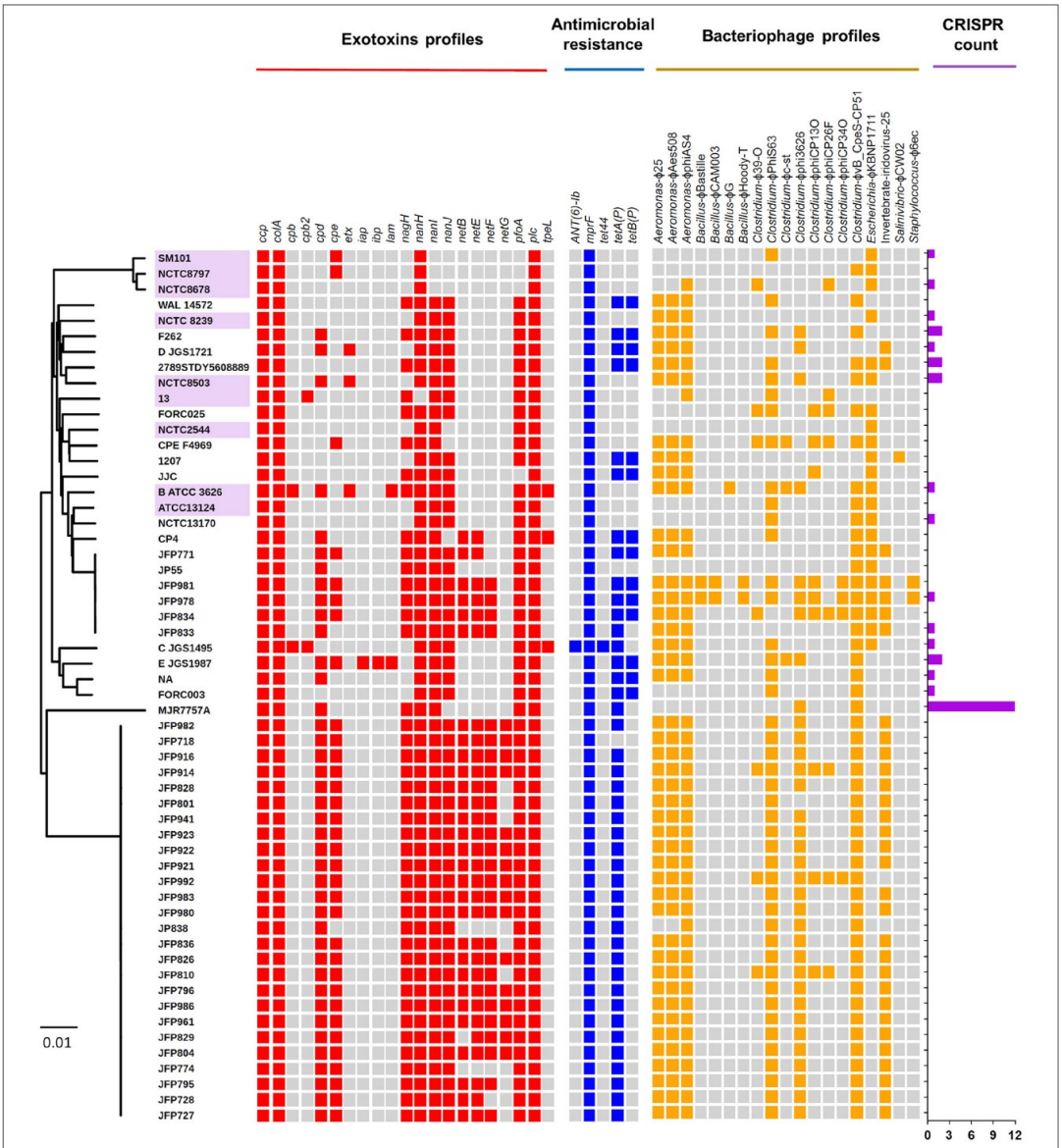


FIGURE 5 | Virulence-associated potentials in *C. perfringens* genomes and plasmids. Heat-map visualising exotoxins, antimicrobial resistance, prophage profiles, and CRISPR counts of 56 *C. perfringens* genomes. Presence of genes are indicated by cell colours: grey (absence); red, blue, and yellow (presence). Purple-labelled isolates indicate historical isolates (pre-antibiotic era).

genetic variation potentially exists, and that a significantly larger-scale study, encompassing a greater variety and number of isolates from multiple ecological niches, would be required to obtain higher-resolution evolutionary insights.

C. perfringens is recognised to be the most prolific toxin-producing organism currently characterised, and is reported to produce >20 exotoxins. However, not all strains produce all toxins, and certain toxins are encoded on plasmids

rather than chromosomes (Supplementary Figure 1). In our analysis, we showed that individual strains produce a diverse profile of toxins, with the exception of the JFP isolates, which display a unique pattern of toxin gene profiles (netE, netF, and netG), inferring that these encoded toxins may be associated with specific hosts (i.e., dogs and horses), and/or disease (i.e., haemorrhagic enteritis) (Mehdizadeh Gohari et al., 2016). Although poultry necrotic enteritis (NE) is thought to be mediated through the *netB* toxin, the NE isolate NCTC8503 was not found to encode this gene, which confirms a previous study by Brady et al. who determined that *netB* (via PCR testing) was present in both healthy and diseased chickens (Brady et al., 2010). Interestingly, *netB* was detected in 87.5% JFP isolates (28/32), suggesting that netB toxin might be involved in haemorrhagic enteritis in dogs and horses. Whist our analysis provides interesting linkages with toxin-associated genes with disease and host phenotypes, determination of specific associations will require further WGS data from a wider variety of isolates (including from healthy hosts), and confirmation from experimental studies.

As *C. perfringens* produces this vast array of toxins, a combination of several "key toxins" are traditionally used to type this species. However, our understanding behind the disease-toxinotype link is limited at present, in part due to lack of in-depth WGS studies. We observed that most of the disease-causing isolates in this dataset are toxinotype A (88%; 44/50), and are associated with diverse diseases in different hosts, suggesting the need to review current toxinotyping methods to reflect *C. perfringens* genetic and pathogenesis relationships more closely. From a public health standpoint, a reliable typing method is also necessary to track the spread of bacterial pathogens, and outbreaks. Currently, there is a move from traditional serotyping and MLST approaches to more recent WGS core-genome based methods, which are employed by national-level health department including Public Health England in the UK (Ashton et al., 2016). Notably, phylogenomic clustering (using CVTree) resembles the core-genome tree, indicating the possibility of employing phylogenomic approaches for large dataset epidemic analysis, as it uses significantly less computational resources (Bernard et al., 2016).

The extreme pan-genomic variation of *C. perfringens*, appears to be driven by HGT events, as indicated by the high number of genes associated with transposases, integrases, and phages, which infers the high potential for gene gain or loss events (**Figure 4**). Notably, *C. perfringens* is found in multiple ecological niches, which are associated with diverse microbial (including viral) communities. Genomic integration of prophages seemingly represents a key source of genetic diversity in *C. perfringens,* and prophage diversity in these isolates surpass the phage contents of *C. difficile*; 7 prophages were found in a total of 44 isolates (Knight et al., 2016). Notably, phages in the human gut are reported to outnumber bacteria by at least a factor of 10, and their lysogenic lifestyle has a powerful biological impact on its hosts such as modifying the fitness and metabolism of bacterial hosts (Brüssow and Hendrix, 2002). Bacteriophages were reported to boost the sporulation capacity of *C. perfringens* in several studies, which reflects the survival strategy employed by these viral particles (Stewart and Johnson, 1977). Some prophage-like genes were also associated with toxin secretion in *C. difficile* (Goh et al., 2005), *Salmonella enterica* (Thomson et al., 2004), and *E. coli* (Zhang et al., 2000). Thus, diverse prophage-associated gene integration in *C. perfringens* genomes may be central for our understanding of the pathogenicity of *C. perfringens* in terms of human and animal health. This high prophage genome plasticity is further supported by our CRISPR (Clustered regularly interspaced short palindromic repeats) array investigation on these 56 isolates as 71.5% (40/56) are predicted to possess no CRISPR system (count = 0), which encodes systems to defend against viral/phage invasion (**Figure 5**). Interestingly, genes involved in "defense mechanisms" are encoded at higher numbers within the accessory genome (n = 202), than the core genome (n = 48), again suggesting constant genomic adaptation to environmental change. These include genes encoding multidrug efflux pumps (associated with antibiotic resistance; n = 11) and ABC transporters (linked with iron-uptake mechanisms and multi-drug resistance; n = 83) (Köster, 2001; Andersen et al., 2015). Notably, genes encoding functions expected to be conserved, such as metabolism (i.e., carbohydrate, amino acid, and lipid metabolism) are more represented within the core (16.8%) rather than the accessory genome (8.3%), indicating their critical functionality in terms of colonisation and growth in different nutritional environments.

There is an ever-increasing incidence of multi-drug resistance pathogens that now pose a critical risk to human and animal health. AMR within *C. perfringens* strains is a serious concern due to its ability to encode multiple virulence associated factors e.g. toxins, that are linked to severe disease. Previous phenotyping-based studies have reported *C. perfringens* as a multidrug resistant pathogen (resist against >1 antimicrobial drug) (Tansuphasiri et al., 2005). However, our genomic analysis indicates that tetracycline resistance genes are the only AMR-specific genes that are wide-spread in this pathogen, which corresponds to previous findings of phenotypically observed tetracycline resistance in *C. perfringens* strains isolated from pigs, human, and environments (Rood et al., 1978). Resistance genes *tetA(P)* and *tetB(P)* are both tetracycline efflux proteins most commonly known to be encoded in *C. perfringens* since the 1980s (Lyras and Rood, 1996). Notably, our data indicates that strains isolated prior to the 1950s do not encode any *tet* genes, which may correlate with the fact that tetracycline did not come into commercial use until 1978, supporting the emergence of antimicrobial resistance under selective pressure in the post-antibiotic era. Although *C. perfringens* does encode tetracycline resistance-associated genes, this is a limited AMR profile when compared to other Gram-positive enteric pathogens like *C. difficile* (Knight et al., 2016). Novel/uncharacterised AMR genes and transmission route (directly from environments such as soil) may account for these findings, however a larger dataset from hospital and/or farming environments and phenotypic testing is required to confirm this hypothesis. Nevertheless, it is suggestive that efflux pumps families including efflux pumps (e.g., *mepA*) and ATP-Binding Cassette (ABC) transporters were *in silico* detected as these could potentially play a significant role in multidrug resistance as observed in previous clinical

cases (Andersen et al., 2015). Furthermore, the environment in which *C. perfringens* resides (e.g., the gut) could also potentially play a significant role in the acquisition of new AMR genes via the "resistome," and as such warrants further investigations (Browne et al., 2017). The host also has specific mechanisms for directly killing gut-associated bacteria including via AMPs. Interestingly, the *mprF* gene (known to be involved in resistance to human defensins) was found to be conserved in all *C. perfringens* genomes, which may link to its ability to evade host-associated defenses for gut colonisation. Indeed, previous studies in *Staphylococcus aureus* have indicated that *mprF* mutant strains are more prone to mammalian AMP killing including neutrophil defensins (Peschel et al., 2001), and mutants are less virulent in animal infection models (Kristian et al., 2003). Importantly, homologues of *mprF* are also reported to dampen the effectiveness of certain clinical antimicrobials (aminoglycosides) such as vancomycin and gentamicin (Nishi et al., 2004), and this may link to previous studies showing resistance of *C. perfringens* to these antibiotics (Traub and Raymond, 1971; Tyrrell et al., 2006; Osman and Elhariri, 2013).

CONCLUSION

High-throughput WGS has paved the way for in-depth pangenomic studies into bacterial species, however previous studies on *C. perfringens* have utilised a maximum of 12 strains for analysis, and thus genomic information on this pathogen is currently limited. Our analysis of 56 strains indicates a surprisingly diverse pangenome, and *in silico* functional analysis reveals frequent exchange of genes in its evolutionary history, associated with prophages, toxins and AMR genes, that potentially influence pathogenicity. Interestingly, this study did not correlate any toxinotype to a specific disease in phylogenetic clustering, however a larger dataset may be needed to establish any causative relationships. As most of the isolates in our study are from diseased patients/animals, inclusion of isolates from healthy samples may enable us to understand the genomic signatures that separate commensal from pathogenic *C. perfringens* (Nagpal et al., 2015). Whilst these genomic studies are of central importance it is vital to also determine the phenotypic disease-causing mechanisms, thus experimental studies are essential to probe our understanding of *C. perfringens* pathogenesis in tandem with *in silico* exploration.

MATERIALS AND METHODS

Bacterial Isolates and Metadata

C. perfringens isolates previously sequenced and deposited in NCBI were used in this study ($n = 51$; retrieved in March 2017), a further 5 *C. perfringens* isolates (NCTC8678, NCTC8797, NCTC13170, NCTC2544, and NCTC8503) were sequenced as part of the NCTC 3000 sequencing project to be included in this study. Metadata of all 56 isolates were retrieved manually online (March 2017) either from published literature, National Center of Biological Information (NCBI) databases or European Bioinformatics Institute (EBI) databases. Year of isolation and geographical origins of isolates were taken from published research papers, otherwise estimated using dates/locations on public databases as indicated. WGS accessions, references and metadata for all genomes are provided in the **Table 3**.

Genomic DNA Extraction and Whole Genome Sequencing

For genomic DNA extraction, *C. perfringens* strains (NCTC8678, NCTC8797, NCTC13170, NCTC2544, and NCTC8503) were propagated on blood agar at 37°C for 36 h under anaerobic conditions. Bacterial lysates were prepared in buffer B1 (containing additional RNase A, protease, and Lysozyme) and the lysates were incubated at 80°C for 24 h. Post-incubation high Molecular weight DNA was isolated from *C. perfringens* strains using the Qiagen midi kit. DNA quality (>60 kb) and quantity (>3 μg) was assessed using the Agilent 2200 TapeStation and Qubit® dsDNA BR Assay Kit respectively. WGS was performed using the PacBio SMRT® DNA Sequencing technology. Appropriate biosafety procedures were in place for both bacterial culturing and DNA extraction.

De Novo Assembly and Annotation

Assembled genomes (51) previously sequenced were retrieved online on NCBI Genomes database (Geer et al., 2010). All 56 genomes were annotated using Prokka v1.10 (Seemann, 2014). HGAP.3 from SMRTpipe-2.3.0 was used to perform *de novo* assembly of the 5 strains of *C. perfringens* sequenced on PacBio RS I obtained from ENA, with one SMRT cell per strain assembled into unitigs. The minlongreadlength (assembly seed length) parameter was set such that there was 30X genome coverage above the seed length. Two strains (NCTC13170 and NCTC2544) were assembled into one single unitig of 3,310,238 bp and the rest three strains (NCTC8678, NCTC8503, and NCTC8797) were assembled into 5–9 unitigs with an N50 of over 2,969,557 bp. These assemblies were run on a High-Performance Computing Cluster using SLURM, utilizing 50 Gb of memory and running parallel on four computing nodes taking ∼634.93 min to finish.

Pangenome Assembly and Visualisation

The annotation files (general feature format, GFF) of 56 *C. perfringens* genomes were submitted to Roary pangenome pipeline v3.6.1 to perform pangenomic analysis (Page et al., 2015). Roary was run with default parameters. A gene-absence-presence data matrix was subsequently derived and visualised in Phandango genome visualizer (v0.87) (Hadfield and Harris, 2017). Fripan[2] was used to visualised MDS clustering of isolates (Supplementary Figure 3) and gene-presence-absence tree (Supplementary Figure 4). Recommended R scripts from Roary package were used to generate pangenome stats using R studio(R Development Core Team, 2010; Rstudio Team, 2015).

Core-Genome Alignment and Construction of Phylogenetic Trees

Core-genome alignment generated using MAFFT v7.305b (Katoh and Standley, 2013) was used as input for Seaview v4.0

[2]Fripan website: drpowell.github.io/FriPan/

(Gouy et al., 2010) to construct core-genome phylogeny. Neighbour-Joining tree (NJ) was built using Jukes-Cantor model of DNA evolution on a set of 1,000 bootstrap replicates; Maximum-Likelihood (ML) tree was constructed using PhyML v3.0 (Guindon et al., 2010), GTR model and supported by Approximate Likelihood Ratio Test (aLRT). Interactive Tree of Life (iTOL) was used as a tree editor to assign clade and annotate metadata (Letunic and Bork, 2016).

Alignment-Free Whole-Genome Phylogeny

CVTree standalone v5.0 was used for constructing alignment-free whole-genome phylogeny with parameters $k = 6$ on amino acid sequences as recommended for prokaryotes and ATCC13124 was selected as the reference sequence (Qi et al., 2004). All genomes were subject to Prokka annotation prior to building phylogeny to obtain annotated FASTA files (genes in amino acid sequences) (Seemann, 2014). Webtool iTOL was used for phylogeny annotation and editing (Letunic and Bork, 2016).

16S rRNA Gene Alignment, Phylogeny, and Pair-Wise Blast

This study employs BAsic Rapid Ribosomal RNA Predictor (Barrnap) v0.5 as 16S rRNA gene predictor (Seemann, 2013). In-house Perl scripts were utilised for extracting the maximum-length 16S rRNA sequence on the positive strands predicted by Barrnap on the WGS data in each isolate for optimal consistency and comparison (Supplementary Table 1). Sequence were aligned using MUSCLE v3.8.31 (Edgar, 2004). Phylogenetic trees were generated as described in previous section. *Clostridium* species 16S rRNA gene sequences were retrieved from NCBI databases (Supplementary Table 3). Pair-wise BLASTn was performed using BLAST+ v2.2.30 and in-house Perl scripts (Camacho et al., 2009).

Functional Annotation of Genomes (COG)

Core and accessory genes were extracted using in-house Perl scripts based on pangenome data. The Clusters of Orthologous Groups (COG) classification was done on both the extracted datasets using eggnog-mapper (v0.99.3) based on COG and EggNOG databases, and HMMER was used for sequence homology search and COG class assignment on default parameters (Huerta-Cepas et al., 2016, 2017). COG classes assigned to each gene were then extracted using Shell script[3] and were plotted using GraphPad Prism v5.04. In-house bash scripts and Linux commands were used to explore annotated data on different functional classes.

In Silico Analysis of Virulence Potentials

Exotoxin profiles: 23 presently identified toxins and virulent enzymes sequence data from NCBI database were built into database for BLAST query (Supplementary Table 4). BLAST+ v2.2.30 was employed for sequence similarity search (BLASTn) using in-house Perl and bash scripts (Camacho et al., 2009). Heat maps were generated using R heatmap2 package (R Development Core Team, 2010). Individual *C. perfringens* plasmids were retrieved from NCBI database for toxin profiling (Supplementary Table 5). **Antimicrobial resistance**: sequence similarity search was performed as described above using Comprehensive Antimicrobial Resistance Database (CARD) (Jia et al., 2017). **Prophage profiles**: sequence similarity search was performed as described above using viral databases retrieved from NCBI (March 2017). Percentage identity of 90% was applied for secondary BLASTn filter. **BLAST parameters**: Double-filter parameters were implemented for screening—expect value of 1e-10 and percentage identity of >80% for sequence homology inference unless otherwise indicated.

[3]Public repository: github.com/raymondkiu/eggnog-mapper_COGextraction

CRISPR Array Detection

CRISPR were predicted using MinCED v0.1.6, a tool derived from CRT (Bland et al., 2007; Angly and Skennerton, 2011). Default parameters were implemented as follows: minimum number of repeats a CRISPR must contain–3, minimum length of CRISPR repeats–23, maximum CRISPR spacers–26, maximum length of CRISPR spaces–50. Counts were parsed in custom bash scripts and visualised using GraphPad Prism v5.04.

Multiple Circular Genome Comparison

Circular whole genome comparisons were generated and visualised using BLAST Ring Image Generator (BRIG) v0.95 (Alikhan et al., 2011). Percentage identity cut-offs were indicated in the graph legends.

AUTHOR CONTRIBUTIONS

LH conceived the study and RK, SC, and LH conceptualised the study. RK and SC designed and performed all bioinformatics analysis, visualised all data, and co-wrote the manuscript. PP and SA contributed to the sequencing and assembly work, and co-wrote the manuscript. LH co-wrote and revised the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work was supported by a Wellcome Trust Investigator Award to LH (100974/C/13/Z) and support of the Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Programme grant for Gut Health and Food Safety BB/J004529/1 (LH). RK is a PhD student at the Norwich Medical School of University of East Anglia (UEA), partially funded by UEA international bursary (Faculty of Medical and Health Sciences). NCTC3000 project is funded by the Wellcome Trust (101503/Z/13/Z). Funders had no input into the design of the study, analysis, or interpretation of the data, or in writing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02485/full#supplementary-material

REFERENCES

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Andersen, J. L., He, G. X., Kakarla, P., K., C., R., Kumar, S., Lakra, W. S., et al. (2015). Multidrug efflux pumps from Enterobacteriaceae, *Vibrio cholerae* and *Staphylococcus aureus* bacterial food pathogens. *Int. J. Environ. Res. Public Health* 12, 1487–1547. doi: 10.3390/ijerph120201487

Angly, F., and Skennerton, C. (2011). *Mining CRISPRs in Environmental Datasets [Online].* Available online at: https://github.com/ctSkennerton/minced/tree/master (Accessed on June, 1 2017).

Ashton, P. M., Nair, S., Peters, T. M., Bale, J. A., Powell, D. G., Painset, A., et al. (2016). Identification of Salmonella for public health surveillance using whole genome sequencing. *PeerJ* 4:e1752. doi: 10.7717/peerj.1752

Bannam, T. L., Johanesen, P. A., Salvado, C. L., Pidot, S. J., Farrow, K. A., and Rood, J. I. (2004). The *Clostridium perfringens* TetA(P) efflux protein contains a functional variant of the Motif A region found in major facilitator superfamily transport proteins. *Microbiology* 150, 127–134. doi: 10.1099/mic.0.26614-0

Bennetts, H. W. (1932). Enzootic ataxia of lambs in Western Australia. *Aust. Vet. J.* 11, 84–92.

Bernard, G., Chan, C. X., and Ragan, M. A. (2016). Alignment-free microbial phylogenomics under scenarios of sequence divergence, genome rearrangement and lateral genetic transfer. *Sci. Rep.* 6:28970. doi: 10.1038/srep28970

Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., et al. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. doi: 10.1186/1471-2105-8-209

Brady, J., Hernandez-Doria, J. D., Bennett, C., Guenter, W., House, J. D., and Rodriguez-Lecompte, J. C. (2010). Toxinotyping of necrotic enteritis-producing and commensal isolates of *Clostridium perfringens* from chickens fed organic diets. *Avian Pathol.* 39, 475–481. doi: 10.1080/03079457.2010.521141

Browne, H. P., Neville, B. A., Forster, S. C., and Lawley, T. D. (2017). Transmission of the gut microbiota: spreading of health. *Nat. Rev. Microbiol.* 15, 531–543. doi: 10.1038/nrmicro.2017.50

Brüssow, H., and Hendrix, R. W. (2002). Phage genomics: small is beautiful. *Cell* 108, 13–16. doi: 10.1016/S0092-8674(01)00637-7

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Canard, B., Garnier, T., Lafay, B., Christen, R., and Cole, S. T. (1992). Phylogenetic analysis of the pathogenic anaerobe *Clostridium perfringens* using the 16S rRNA nucleotide sequence. *Int. J. Syst. Bacteriol.* 42, 312–314. doi: 10.1099/00207713-42-2-312

Cole, J. N., and Nizet, V. (2016). "Bacterial evasion of host antimicrobial peptide defenses," in *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition,* eds I. Kudva, N. Cornick, P. Plummer, Q . Zhang, T. Nicholson, J. Bannantine, and B. Bellaire (Washington, DC: ASM Press), 413–443. doi: 10.1128/microbiolspec.VMBF-0006-2015

Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., et al. (1994). The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. *Int. J. Syst. Bacteriol.* 44, 812–826. doi: 10.1099/00207713-44-4-812

Cornillot, E., Saint-Joanis, B., Daube, G., Katayama, S., Granum, P. E., Canard, B., et al. (1995). The enterotoxin gene (cpe) of *Clostridium perfringens* can be chromosomal or plasmid-borne. *Mol. Microbiol.* 15, 639–647. doi: 10.1111/j.1365-2958.1995.tb02373.x

Edagiz, S., Lagace-Wiens, P., Embil, J., Karlowsky, J., and Walkty, A. (2015). Empyema caused by *Clostridium bifermentans*: a case report. *Can. J. Infect. Dis. Med. Microbiol.* 26, 105–107.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Eppinger, M., Worsham, P. L., Nikolich, M. P., Riley, D. R., Sebastian, Y., Mou, S., et al. (2010). Genome sequence of the deep-rooted *Yersinia pestis* strain angola reveals new insights into the evolution and pangenome of the plague bacterium. *J. Bacteriol.* 192, 1685–1699. doi: 10.1128/JB.01518-09

Ernst, C. M., and Peschel, A. (2011). Broad-spectrum antimicrobial peptide resistance by MprF-mediated aminoacylation and flipping of phospholipids. *Mol. Microbiol.* 80, 290–299. doi: 10.1111/j.1365-2958.2011.07576.x

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., et al. (2010). The NCBI BioSystems database. *Nucleic Acids Res.* 38, D492–D496. doi: 10.1093/nar/gkp858

Gibert, M., Jolivet-Reynaud, C., and Popoff, M. R. (1997). Beta2 toxin, a novel toxin produced by *Clostridium perfringens*. *Gene* 203, 65–73. doi: 10.1016/S0378-1119(97)00493-9

Goh, S., Chang, B. J., and Riley, T. V. (2005). Effect of phage infection on toxin production by *Clostridium difficile*. *J. Med. Microbiol.* 54, 129–135. doi: 10.1099/jmm.0.45821-0

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Hadfield, J., and Harris, S. (2017). *Interactive Visualization of Genome Phylogenies - Phandango [Online].* Wellcome Trust Sange Institute. Available online at: https://jameshadfield.github.io/phandango/ (Accessed on April, 5 2017).

Harvey, S. M., Sturgeon, J., and Dassey, D. E. (2002). Botulism due to *Clostridium baratii* type F toxin. *J. Clin. Microbiol.* 40, 2260–2262. doi: 10.1128/JCM.40.6.2260-2262.2002

Hassan, K. A., Elbourne, L. D., Tetu, S. G., Melville, S. B., Rood, J. I., and Paulsen, I. T. (2015). Genomic analyses of *Clostridium perfringens* isolates from five toxinotypes. *Res. Microbiol.* 166, 255–263. doi: 10.1016/j.resmic.2014.10.003

Heida, F. H., van Zoonen, A. G. J. F., Hulscher, J. B. F., Te Kiefte, B. J. C., Wessels, R., Kooi, E. M. W., et al. (2016). A necrotizing enterocolitis-associated gut microbiota is present in the meconium: results of a prospective study. *Clin. Infect. Dis.* 62, 863-870. doi: 10.1093/cid/ciw016

Hiller, N. L., Janto, B., Hogg, J. S., Boissy, R., Yu, S., Powell, E., et al. (2007). Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. *J. Bacteriol.* 189, 8186–8195. doi: 10.1128/JB.00690-07

Hobbs, B. C., Smith, M. E., Oakley, C. L., Warrack, G. H., and Cruickshank, J. C. (1953). *Clostridium welchii* food poisoning. *J. Hyg.* 51, 75–101. doi: 10.1017/S0022172400015515

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115-2122. doi: 10.1093/molbev/msx148

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Keyburn, A. L., Boyce, J. D., Vaz, P., Bannam, T. L., Ford, M. E., Parker, D., et al. (2008). NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. *PLoS Pathog.* 4:e26. doi: 10.1371/journal.ppat.0040026

Knight, D. R., Squire, M. M., Collins, D. A., and Riley, T. V. (2016). Genome analysis of *Clostridium difficile* pcr ribotype 014 lineage in australian pigs and humans reveals a diverse genetic repertoire and signatures of long-range interspecies transmission. *Front. Microbiol.* 7:2138.doi: 10.3389/fmicb.2016.02138

Koch, C., Rainey, F. A., and Stackebrandt, E. (1994). 16s Rdna studies on members of Arthrobacter and Micrococcus - an aid for their future taxonomic restructuring. *FEMS Microbiol. Lett.* 123, 167–171. doi: 10.1111/j.1574-6968.1994.tb07217.x

Kosloske, A. M., Ulrich, J. A., and Hoffman, H. (1978). Fulminant necrotising enterocolitis associated with Clostridia. *Lancet* 312, 1014–1016. doi: 10.1016/S0140-6736(78)92337-1

Köster, W. (2001). ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Res. Microbiol.* 152, 291–301. doi: 10.1016/S0923-2508(01)01200-1

Kristian, S. A., Lauth, X., Nizet, V., Goetz, F., Neumeister, B., Peschel, A., et al. (2003). Alanylation of teichoic acids protects *Staphylococcus aureus* against Toll-like receptor 2-dependent host defense in a mouse tissue cage infection model. *J. Infect. Dis.* 188, 414–423. doi: 10.1086/376533

Lepp, D., Roxas, B., Parreira, V. R., Marri, P. R., Rosey, E. L., Gong, J., et al. (2010). Identification of novel pathogenicity loci in *Clostridium perfringens* strains that cause avian necrotic enteritis. *PLoS ONE* 5:e10795. doi: 10.1371/annotation/501e5656-71ac-420f-a194-80141f6381e5

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Li, J., Adams, V., Bannam, T. L., Miyamoto, K., Garcia, J. P., Uzal, F. A., et al. (2013). Toxin plasmids of *Clostridium perfringens*. *Microbiol. Mol. Biol. Rev.* 77, 208–233. doi: 10.1128/MMBR.00062-12

Li, J., and Mcclane, B. A. (2006). Further comparison of temperature effects on growth and survival of *Clostridium perfringens* type A isolates carrying a chromosomal or plasmid-borne enterotoxin gene. *Appl. Environ. Microbiol.* 72, 4561–4568. doi: 10.1128/AEM.00177-06

Lyras, D., and Rood, J. I. (1996). Genetic organization and distribution of tetracycline resistance determinants in *Clostridium perfringens*. *Antimicrob. Agents Chemother.* 40, 2500–2504.

Mahony, D. E., and Moore, T. I. (1976). Stable L-forms of *Clostridium perfringens* and their growth on glass surfaces. *Can. J. Microbiol.* 22, 953–959. doi: 10.1139/m76-138

Masaki, T., Umehashi, H., Miyazaki, H., Takano, M., Yamakawa, K., and Nakamura, S. (1988). *Clostridium absonum* from gas gangrene. *Jpn. J. Med. Sci. Biol.* 41, 27–30. doi: 10.7883/yoken1952.41.27

Mcclung, L. (1945). Human food poisoning due to growth of *Clostridium perfringens* (*C. welchii*) in freshly cooked chicken. *J. Bacteriol.* 50, 229–231.

Mcinerney, J. O., Mcnally, A., and O'connell, M. J. (2017). Why prokaryotes have pangenomes. *Nat Microbiol* 2:17040. doi: 10.1038/nmicrobiol.2017.40

Mehdizadeh Gohari, I., Parreira, V. R., Nowell, V. J., Nicholson, V. M., Oliphant, K., and Prescott, J. F. (2015). A novel pore-forming toxin in type A *Clostridium perfringens* is associated with both fatal canine hemorrhagic gastroenteritis and fatal foal necrotizing enterocolitis. *PLoS ONE* 10:e0122684. doi: 10.1371/journal.pone.0122684

Mehdizadeh Gohari, I., Kropinski, A. M., Weese, S. J., Parreira, V. R., Whitehead, A. E., Boerlin, P., et al. (2016). Plasmid characterization and chromosome analysis of two netF+ *Clostridium perfringens* isolates associated with foal and canine necrotizing enteritis. *PLoS ONE* 11:e0148344. doi: 10.1371/journal.pone.0148344

Mehdizadeh Gohari, I., Kropinski, A. M., Weese, S. J., Whitehead, A. E., Parreira, V. R., Boerlin, P., et al. (2017). NetF-producing *Clostridium perfringens*: clonality and plasmid pathogenicity loci analysis. *Infect. Genet. Evol.* 49, 32–38. doi: 10.1016/j.meegid.2016.12.028

Möllby, R., Holme, T., Nord, C. E., Smyth, C. J., and Wadström, T. (1976). Production of phospholipase C (alpha-toxin), haemolysins and lethal toxins by *Clostridium perfringens* types A to D. *J. Gen. Microbiol.* 96, 137–144. doi: 10.1099/00221287-96-1-137

Mooijman, K. A., During, M., and Nagelkerke, N. J. D. (2003). *MICROCRM: Preparation and Control of Batches of Microbiological Reference Materials Consisting of Capsules*. The Netherlands National Institute for Public Health and the Environment (RIVM).

Myers, G. S., Rasko, D. A., Cheung, J. K., Ravel, J., Seshadri, R., Deboy, R. T., et al. (2006). Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. *Genome Res.* 16, 1031–1040. doi: 10.1101/gr.5238106

Nagler, F. P. O. (1939). Observations on a reaction between the lethal toxin of *Cl. welchii* (Type A) and human serum. *Br. J. Exp. Pathol.* 20, 473–485.

Nagpal, R., Ogata, K., Tsuji, H., Matsuda, K., Takahashi, T., Nomoto, K., et al. (2015). Sensitive quantification of *Clostridium perfringens* in human feces by quantitative real-time PCR targeting alpha-toxin and enterotoxin genes. *BMC Microbiol.* 15:219. doi: 10.1186/s12866-015-0561-y

Nagy, E., Becker, S., Kostrzewa, M., Barta, N., and Urbán, E. (2012). The value of MALDI-TOF MS for the identification of clinically relevant anaerobic bacteria in routine laboratories. *J. Med. Microbiol.* 61, 1393–1400. doi: 10.1099/jmm.0.043927-0

Nishi, H., Komatsuzawa, H., Fujiwara, T., Mccallum, N., and Sugai, M. (2004). Reduced content of lysyl-phosphatidylglycerol in the cytoplasmic membrane affects susceptibility to moenomycin, as well as vancomycin, gentamicin, and antimicrobial peptides, in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 48, 4800–4807. doi: 10.1128/AAC.48.12.4800-4807.2004

Nowell, V. J., Kropinski, A. M., Songer, J. G., Macinnes, J. I., Parreira, V. R., and Prescott, J. F. (2012). Genome sequencing and analysis of a type A *Clostridium perfringens* isolate from a case of bovine clostridial abomasitis. *PLoS ONE* 7:e32271. doi: 10.1371/journal.pone.0032271

Ohno-Iwashita, Y., Iwamoto, M., Mitsui, K., Kawasaki, H., and Ando, S. (1986). Cold-labile hemolysin produced by limited proteolysis of theta-toxin from *Clostridium perfringens*. *Biochemistry* 25, 6048–6053. doi: 10.1021/bi00368a032

Osman, K. M., and Elhariri, M. (2013). Antibiotic resistance of *Clostridium perfringens* isolates from broiler chickens in Egypt. *Rev. Sci. Technol.* 32, 841–850. doi: 10.20506/rst.32.2.2212

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Paquette, G., and Fredette, V. (1967). Avirulent *Clostridium perfringens* strains obtained by euflavine treatment. *J. Bacteriol.* 94, 1437–1442.

Park, M., Deck, J., Foley, S. L., Nayak, R., Songer, J. G., Seibel, J. R., et al. (2016). Diversity of *Clostridium perfringens* isolates from various sources and prevalence of conjugative plasmids. *Anaerobe* 38, 25–35. doi: 10.1016/j.anaerobe.2015.11.003

Peschel, A., Jack, R. W., Otto, M., Collins, L. V., Staubitz, P., Nicholson, G., et al. (2001). *Staphylococcus aureus* resistance to human defensins and evasion of neutrophil killing via the novel virulence factor MprF is based on modification of membrane lipids with l-lysine. *J. Exp. Med.* 193, 1067–1076. doi: 10.1084/jem.193.9.1067

Petit, L., Gibert, M., and Popoff, M. (1999). *Clostridium perfringens*: toxinotype and genotype. *Trends Microbiol.* 7, 104–110. doi: 10.1016/S0966-842X(98)01430-9

Qi, J., Luo, H., and Hao, B. (2004). CVTree: a phylogenetic tree reconstruction tool based on whole genomes. *Nucleic Acids Res.* 32, W45–W47. doi: 10.1093/nar/gkh362

Raveh-Sadka, T., Thomas, B. C., Singh, A., Firek, B., Brooks, B., Castelle, C. J., et al. (2015). Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. *Elife* 4:e05477. doi: 10.7554/eLife.05477

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Revitt-Mills, S. A., Rood, J. I., and Adams, V. (2015). *Clostridium perfringens* extracellular toxins and enzymes: 20 and counting. *Microbiol. Aust.* 36, 114–117. doi: 10.1071/MA15039

Ribeiro, F. J., Przybylski, D., Yin, S., Sharpe, T., Gnerre, S., Abouelleil, A., et al. (2012). Finished bacterial genomes from shotgun sequence data. *Genome Res.* 22, 2270–2277. doi: 10.1101/gr.141515.112

Roach, D. J., Burton, J. N., Lee, C., Stackhouse, B., Butler-Wu, S. M., Cookson, B. T., et al. (2015). A year of infection in the intensive care unit: prospective whole genome sequencing of bacterial clinical isolates reveals cryptic transmissions and novel microbiota. *PLoS Genet.* 11:e1005413. doi: 10.1371/journal.pgen.1005413

Rood, J. I. (1998). Virulence genes of *Clostridium perfringens*. *Annu. Rev. Microbiol.* 52, 333–360. doi: 10.1146/annurev.micro.52.1.333

Rood, J. I., Scott, V. N., and Duncan, C. L. (1978). Identification of a transferable tetracycline resistance plasmid (pCW3) from *Clostridium perfringens*. *Plasmid* 1, 563–570. doi: 10.1016/0147-619X(78)90013-6

Rouli, L., Merhej, V., Fournier, P. E., and Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* 7, 72–85. doi: 10.1016/j.nmni.2015.06.005

Rstudio Team (2015). *RStudio: Integrated Development for R.* Boston, MA: RStudio, Inc.

Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M. A., Roy, S. L., et al. (2011). Foodborne illness acquired in the United States–major pathogens. *Emerg Infect Dis* 17, 7–15. doi: 10.3201/eid1701.P11101

Seemann, T. (2013). *barrnap 0.7: Rapid Ribosomal RNA Prediction [Online].* Available online at: https://github.com/Victorian-Bioinformatics-Consortium/barrnap (Accessed on April, 6 2017).

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., et al. (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc. Natl. Acad. Sci. U.S.A.* 99, 996–1001. doi: 10.1073/pnas.022493799

Sim, K., Shaw, A. G., Randell, P., Cox, M. J., Mcclure, Z. E., Li, M. S., et al. (2015). Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clin. Infect. Dis.* 60, 389–397. doi: 10.1093/cid/ciu822

Stevens, D. L., and Bryant, A. E. (2002). The role of clostridial toxins in the pathogenesis of gas gangrene. *Clin. Infect. Dis.* 35, S93–S100. doi: 10.1086/341928

Stewart, A. W., and Johnson, M. G. (1977). Increased numbers of heat-resistnat spores produced by two strains of *Clostridium perfringens* bearing temperate phage ϕ9. *J. Gen. Microbiol.* 103, 45–50. doi: 10.1099/00221287-103-1-45

Tansuphasiri, U., Matra, W., and Sangsuk, L. (2005). Antimicrobial resistance among *Clostridium perfringens* isolated from various sources in Thailand. *Southeast Asian J. Trop. Med. Public Health* 36, 954–961.

Thomson, N., Baker, S., Pickard, D., Fookes, M., Anjum, M., Hamlin, N., et al. (2004). The role of prophage-like elements in the diversity of *Salmonella enterica* serovars. *J. Mol. Biol.* 339, 279–300. doi: 10.1016/j.jmb.2004.03.058

Traub, W. H., and Raymond, E. A. (1971). *In vitro* resistance of *Clostridium perfringens* type A to gentamicin sulfate, and reduced activity of the antibiotic under anaerobic atmospheric conditions. *Chemotherapy* 16, 162–165. doi: 10.1159/000220723

Tyrrell, K. L., Citron, D. M., Warren, Y. A., Fernandez, H. T., Merriam, C. V., and Goldstein, E. J. C. (2006). *In vitro* activities of daptomycin, vancomycin, and penicillin against *Clostridium* difficile, *C. perfringens*, *Finegoldia magna*, and *Propionibacterium acnes*. *Antimicrob. Agents Chemother.* 50, 2728–2731. doi: 10.1128/Aac.00357-06

Uzal, F. A., Freedman, J. C., Shrestha, A., Theoret, J. R., Garcia, J., Awad, M. M., et al. (2014). Towards an understanding of the role of *Clostridium perfringens* toxins in human and animal disease. *Future Microbiol.* 9, 361–377. doi: 10.2217/fmb.13.168

Van Asten, A. J., Van Der Wiel, C. W., Nikolaou, G., Houwers, D. J., and Gröne, A. (2009). A multiplex PCR for toxin typing of *Clostridium perfringens* isolates. *Vet. Microbiol.* 136, 411–412. doi: 10.1016/j.vetmic.2008.11.024

Van Vliet, A. H., and Kusters, J. G. (2015). Use of alignment-free phylogenetics for rapid genome sequence-based typing of *Helicobacter pylori* virulence markers and antibiotic susceptibility. *J. Clin. Microbiol.* 53, 2877–2888. doi: 10.1128/JCM.01357-15

Vos, P., Garrity, G., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A., et al. (2009). *Bergey's Manual of Systematic Bacteriology.* New York, NY: Springer-Verlag.

Welch, W. H., and Nuttall, G. H. F. (1891). A gas-producing bacillus (*Bacillus aerogenes capsulatus*, Nov, Spec.) capable of rapid development in the body after death. *Bull. John Hopkins Hosp.* 3, 81–91.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576

Wong, Y. M., Juan, J. C., Gan, H. M., and Austin, C. M. (2014). Draft Genome Sequence of *Clostridium perfringens* Strain, J. J. C., a Highly Efficient Hydrogen Producer Isolated from Landfill Leachate Sludge. *Genome Announc* 2:e00064-14. doi: 10.1128/genomeA.00064-14

Zhang, X., Mcdaniel, A. D., Wolf, L. E., Keusch, G. T., Waldor, M. K., and Acheson, D. W. (2000). Quinolone antibiotics induce shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J.f Infect. Dis.* 181, 664–670. doi: 10.1086/315239

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

Kiu and Hall *Emerging Microbes & Infections* (2018)7:141
DOI 10.1038/s41426-018-0144-8

Emerging Microbes & Infections
www.nature.com/emi

REVIEW ARTICLE Open Access

An update on the human and animal enteric pathogen *Clostridium perfringens*

Raymond Kiu[1,2] and Lindsay J. Hall[1]

Abstract

Clostridium perfringens, a rapid-growing pathogen known to secrete an arsenal of >20 virulent toxins, has been associated with intestinal diseases in both animals and humans throughout the past century. Recent advances in genomic analysis and experimental systems make it timely to re-visit this clinically and veterinary important pathogen. This Review will summarise our understanding of the genomics and virulence-linked factors, including antimicrobial potentials and secreted toxins of this gut pathogen, and then its up-to-date clinical epidemiology and biological role in the pathogenesis of several important human and animal-associated intestinal diseases, including pre-term necrotising enterocolitis. Finally, we highlight some of the important unresolved questions in relation to *C. perfringens*-mediated infections, and implications for future research directions.

Introduction

Clostridium perfringens (formerly known as *Bacillus aerogenes capsulatus, Bacillus perfringens, Bacillus welchii* or *Clostridium welchii*) is a Gram positive, spore-forming, anaerobic, rod-shaped bacterium[1]. It was first isolated and identified as a novel bacterium in 1891 by William H. Welch from the autopsy of a 38-year-old man, where gas bubbles were observed within infected blood vessels. This gas-forming trait (the original name *Bacillus aerogenes capsulatus*, 'aerogenes' literally means 'air-producing' in Latin) was later linked to gas gangrene symptoms seen in British and French soldiers during World War I[2].

C. perfringens is associated with diverse environments including soils, food, sewage, and as a member of the gastrointestinal (GI) tract microbial community (i.e., microbiota) of both diseased, and non-diseased humans and animals. Notably it has been associated with humans for thousands of years as evidenced by the recent identification of *C. perfringens* using next generation sequencing (NGS) technology in the mummified GI tract of a >5000-year-old Neolithic 'Tyrolean Iceman' (also known as Ötzi), found in an Alpine glacier in 1991[3].

Clinically, *C. perfringens* has been constantly associated with various significant systemic and enteric diseases, in both humans and animals, including gas gangrene (Clostridial myonecrosis), food poisoning and non-foodborne diarrhoea, and enterocolitis[4,5]. Importantly, *C. perfringens* strains are known to secrete >20 identified toxins or enzymes that could potentially be the principal virulence factors involved in pathophysiology[6].

In this Review, we explore phenotypic and genomic features of this important and re-emerging pathogen, including virulence factors, and antimicrobial resistance (AMR) profiles, and the clinical impact of this bacterium in relation to numerous medically important intestinal diseases, across several host species. Finally, we highlight some of the important unresolved questions in relation to *C. perfringens*-associated infections, and implications for future research directions.

Isolation, identification and typing methods

There are numerous methods for isolation of *C. perfringens*, and biochemical and molecular methods for identification as detailed in Table 1.

Correspondence: Lindsay J. Hall (Lindsay.Hall@quadram.ac.uk)
[1]Gut Microbes and Health Programme, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK
[2]Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK

Table 1 Common isolation, identification and typing methods used in *C. perfringens* research

Method	Method details in brief	Refs
Isolation		
Direct plating	Direct plating on TSC-EYA + 18–24 h anaerobic incubation at 37 °C. Pitch black colonies with opaque halos are presumptively *C. perfringens*.	146
Ethanol pre-treatment	Ethanol pre-treatment (50% ethanol) for 30 mins + plating on Fastidious Anaerobe Agar (sometimes supplemented with 0.1% taurocholate). Colonies that exhibit beta-haemolysis are preliminarily identified as *C. perfringens*.	147
Biochemical identification		
Nagler's reaction	Known as lecithinase (alpha-toxin) test. Egg-yolk agar plates are split into two halves, where one half contains anti-alpha-toxin, and following anaerobic incubation, positivity is defined as turbidity around colonies on the anti-alpha-toxin-free side which confirms *C. perfringens*.	148
Reverse CAMP test	*Streptococcus agalactiae* is streaked on blood agar, and *C. perfringens* streaked perpendicular to *S. agalactiae*. After 24–48 h of anaerobic incubation, a 'bow-tie' zone will form due to the synergistic haemolysis, this confirms *C. perfringens*.	149
Molecular identification		
16S ribosomal RNA PCR	Appropriating full-length 16S rRNA universal primers/smaller region of 16S rRNA gene in PCR to amplify 16S rRNA gene + sequencing, and identify using informatics approach (based on sequence similarity >97% to assign taxonomy).	30
MALDI-TOF Mass Spectrometry	Rapid and inexpensive identification method based upon the mass spectrum analysis of highly-conserved ribosomal proteins from whole bacterial cells—apply bacterial colonies straight onto the MALDI-TOF metal target and followed by appropriate treatments and analysis.	4
Typing		
Multiplex PCR	Multiplex PCR approach is commonly used to amplify key toxins genes to classify *C. perfringens* into 7 (A to G) different toxinotypes according to the toxin genes combination.	150
Genome-wide sequence search	Genome-wide search on relevant toxin genes using sequence similarity search program e.g. BLAST for toxinotyping on *C. perfringens* genomes.	10

As *C. perfringens* is a pathogen, typing methods are currently used to differentiate between strains that may be associated with serious infection (Table 1). *C. perfringens* strains, clinically well-known for toxin production, are typed (recently updated for seven toxinotypes: A–G; previously only types A–E[7]) according to the combination of typing toxins (Fig. 1) they produce, i.e., α-toxin, β-toxin, ε-toxin and ι-toxin, enterotoxin (CPE) and NetB[8]. Historically this method indicated that certain toxins were associated with certain hosts and diseases e.g., type B (particularly the β-toxin it harbours), is exclusively linked to dysentery in sheep, and rarely seen in other hosts[9]. Food-poisoning associated CPE was genotyped typically in type F strains (previously named as CPE-positive type A; not to be confused with heat-resistant type C strains) although CPE can also be produced by certain type C, D and E strains, whereas β2-toxin and θ-toxin (perfringolysin O; PFO) could be found in any toxinotypes[1,10,11]. However, it is important to note that no single strain is known to produce the entire panoply of toxins[10].

Genomic insights

C. perfringens has a relatively low GC content, between 27–28%, and genome sizes range from 3.0–4.1 Mb, with 2500–3600 predicted genes present in each individual genome[10]. The first ever complete genome sequence of *C. perfringens* (strain 13) was published in early 2002; a historical gas gangrene-associated strain from 1939 (this also represents the first Gram-positive anaerobic pathogen to be sequenced)[12]. Virulence-linked genes within the sequence data include phospholipase C (α-toxin), hyaluronidase (μ-toxin), collagenase (κ-toxin), PFO (θ-toxin), and β2 toxin. Moreover, 61 sporulation-related and germination-related genes were encoded, which supports the fact that *C. perfringens* is a spore-former. Notably, most virulence genes in this type A strain were not found on genomic islands, or near insertion/transposase sequences, and GC content of those genes were similar to the average GC percentage, implying that horizontal gene transfer (HGT) of these toxin genes was an ancient evolutionary event. This study also highlighted differences in environmental adaption genes, including glycolytic

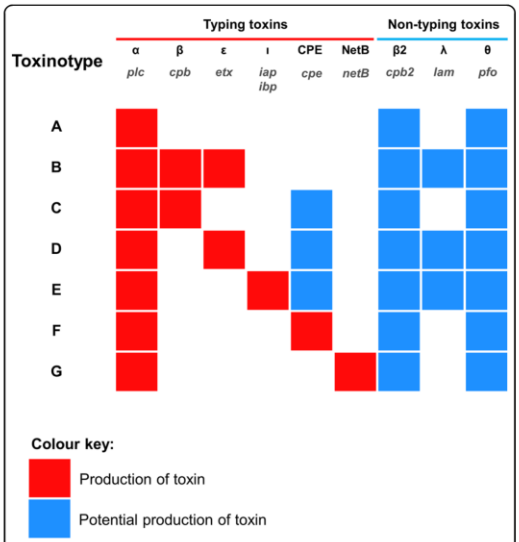


Fig. 1 *C. perfringens* current toxinotyping system. The names of toxin genes are printed in grey

enzymes, and sporulation-linked genes, whereas recent studies indicated that phylogenetic relationships between *C. perfringens* (based on core genome analysis) does not relate to clone origin or toxinotypes[1,10]

A recent large-scale genomic study that investigated 56 strains *C. perfringens* strains, representing the largest *C. perfringens* genomic study to date, revealed a diverse pangenome (a repertoire of genes in a defined number of genomes), with only 12.6% core genes (common genes that are present in each genome) (Fig. 2)[10]. A wide array of toxin genes was profiled in this study: α-toxin (*plc*), α-clostripain (*ccp*) and microbial collagenase (*colA*) were conserved in all examined genomes, the pore-forming toxin PFO gene *pfo* was found to be encoded in most genomes (>90%), whereas the food-poisoning causative toxin CPE gene *cpe* was consistently detected in historical food poisoning isolates (isolated before 1950s). Importantly, prevalence of aminoglycoside-resistant/anti-defensin gene *mprF* (100% detection), and tetracycline-resistant efflux protein genes *tetA (P)* (>75% detection) encoded within the genomes underlined the potential antimicrobial resistant threat of *C. perfringens*-associated infection. This study also suggested that these diverse genetic variations may have been driven by HGT, especially prophage insertion, within the 'CRISPR-free' genomes (no single CRISPR prophage defence system detected in >70% of the genomes).

Box 1

Glossary

- 16S rRNA gene: a housekeeping gene (of ribosomes) that is conserved across all species therefore is widely used as a chronological/biological identity marker in evolutionary genetics and genomics.
- Accessory gene: a gene that is absent in one or more strains in an analysed pangenome.
- Antimicrobial resistance: a biological characteristic that enable a microorganism to survive and thrive in the presence of antibiotics.
- Core gene: a gene that is present in all strains in an analysed pangenome.
- Genotyping: a process to differentiate/identify characteristics of genes (genotypes).
- Germination: the process which microorganisms grow from a dormant state (usually from endospores).
- Necrotising Enterocolitis: severe gut inflammation (often accompanied by necrosis of the gut) which occurs mostly in pre-term infants with high mortality.
- Pangenome: The repertoire of collective genes in closely related organisms.
- Phenotyping: a process to observe the expression/traits (phenotypes) of an organism's genotype.
- Sporulation: a process certain species/genus of bacteria (e.g., Firmicutes) undergo to produce a tough external structure (spore/endospore) to withstand unfavourable conditions in order to survive through.

The substantial genetic divergence in the *C. perfringens* pangenome suggests there may be additional novel virulence-related genes encoded within the 'accessory genome', in addition to the plasmid-borne toxins, known to be primarily responsible for specific disease pathologies. Thus, it is important to analyse plasmid-encoded genes/toxins, in tandem to chromosomally-encoded genes in the accessory genome, to provide a more complete genomic picture of *C. perfringens*. Potentially, with the aid of bioinformatics tools, more novel toxins or virulence genes will be discovered and identified for investigation in relation to *C. perfringens* pathogenicity, and disease transmission tracking in hospital and environmental settings, as has been performed for the related gut pathogen *C. difficile*[13]. To fully determine and research the pangenome, and discover potentially novel virulence genes, more isolate sequencing will be required, including addition of isolates from a range of hosts and environmental conditions, which may pave the way for translational research into new treatment strategies against *C. perfringens* infections.

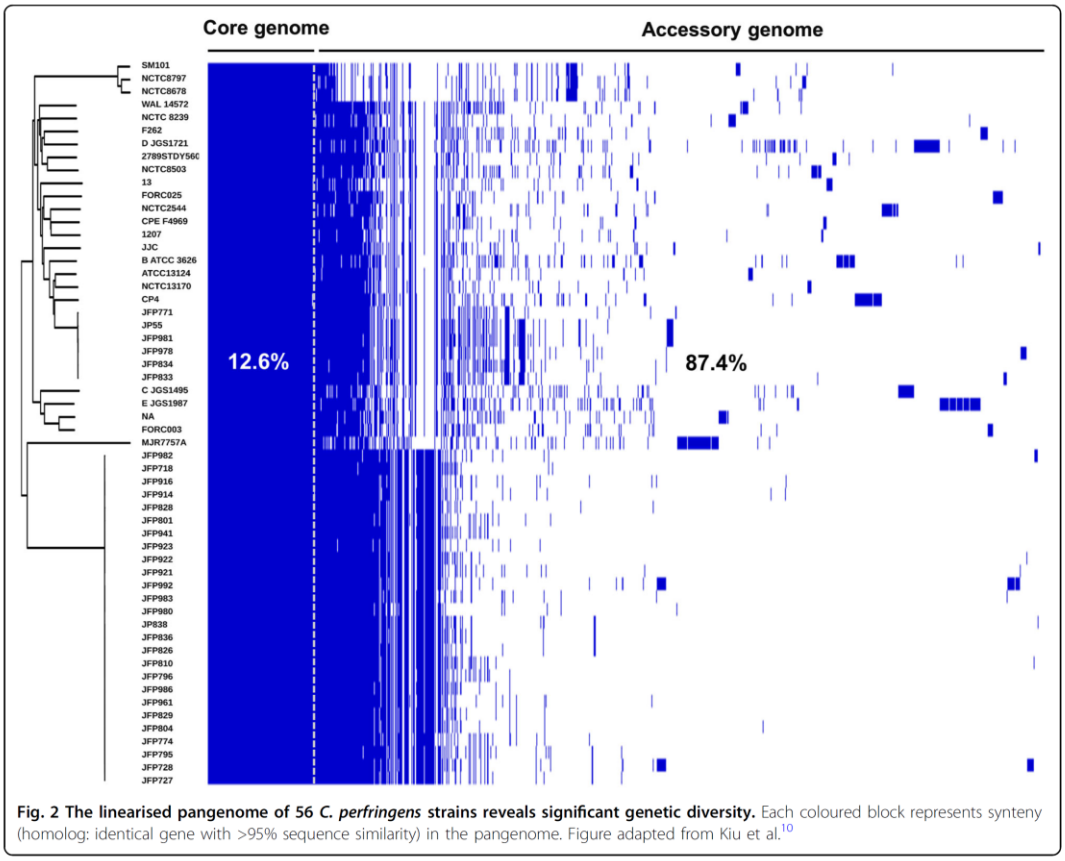


Fig. 2 The linearised pangenome of 56 *C. perfringens* strains reveals significant genetic diversity. Each coloured block represents synteny (homolog: identical gene with >95% sequence similarity) in the pangenome. Figure adapted from Kiu et al.[10]

Plasmids

Importantly, *C. perfringens* strains are known to carry plasmids, which often encode virulence-associated proteins including toxins and antimicrobials[14]. Disease-associated toxins including; CPE, ε-toxin, ι-toxin, NetB, β2-toxin and binary enterotoxin (BEC) have all been detected on *C. perfringens* plasmids[10,15]. HGT of plasmids between strains is reported to be via the *tcp* conjugation locus that exists in most plasmids[16]. Important toxin-carrying plasmids will be briefly reviewed in order of toxinotypes. For comprehensive reviews on *C. perfringens* plasmids, and other related insertion sequences see[14,15].

CPE is known to be involved in food poisoning and non-foodborne gastroenteritis. It was reported that up to 70% of food poisoning cases, causative *C. perfringens* strains (mainly type F, previously known as CPE-positive type A) were shown to encode chromosomal-*cpe*, rather than plasmid-*cpe*, while the latter was associated with non-foodborne gastroenteritis (responsible for 5–15% of all cases)[17]. Some *cpe*-associated plasmids have also been shown to encode *cpb2* genes or *iab/ibp* genes (e.g., pCPF5603)[14]. Type B strains carry plasmids which encode one of the deadliest toxins on earth, ε-toxin, and β-toxin. Plasmid pCP8533etx encodes *etx* and *cpb2* genes, but not *cpb*[18]. Most plasmids in type B strains also possess other virulence genes encoding λ-toxin and urease[19]. Similarly, plasmids in type D strains also carry *cpe* and/or *cpb2* genes, and plasmids range in size from 45 to 110 kb[20]. *C. perfringens* type C strains possess plasmid-borne β-toxin gene *cpb*, and other plasmids reported to carry other toxin genes including *cpb2*, *tpeL*, or *cpe*[21]. In type E strains, two major families of plasmids have been identified: (1) plasmids carry *iap/ibp* genes, λ-toxin gene and *cpe* gene, however, *cpe* contains nonsense and missense mutations in the *cpe* ORF and is thus non-functional[22]. (2) These plasmids, including pCPPB-1, encode both *iap/ibp* genes and *cpe* gene but not the λ-toxin gene[23]. NetB is a chicken NE-associated toxin, which is plasmid encoded. These

netB-encoded plasmids can potentially carry other virulence genes including tetracycline resistance genes (e.g., pJIR3537 and pCW3) and *cpb2* gene (e.g., pJIR3844)[24].

Notably, many toxin plasmids of *C. perfringens* are highly similar, sharing up to 35–40 kb of identical sequences[15]. In addition, some *C. perfringens* isolates can carry multiple near-identical plasmids that encode different toxin and AMR genes[24]. However, no one isolate has ever been found to possess plasmids that encode all key toxins; only certain plasmid combination could can be maintained, which suggests plasmid incompatibilities, due to presence of distinct plasmid segregation systems, and warrants further research[14,25,26]. Furthermore, the universal existence of *tcp* locus indicates that conjugative plasmid transfer could be a key HGT event for increasing virulence of *C. perfringens* strains.

Virulence-related factors

C. perfringens can generate a complement of extracellular toxins and hydrolytic enzymes (>20), survive in aerobic environments (i.e., oxygen tolerance), produce toxic gases, and rapidly grow. Therefore, it possesses the capacity to effect various histotoxic (i.e., toxigenic to tissues) infections in humans, including gas gangrene in contaminated wounds, gastroenteritis (including foodborne and non-foodborne diarrhoea) in human adults, necrotic enteritis (NE) in animals, and recent links to NEC in pre-term infants[27]. Relevant virulence factors of *C. perfringens* are represented and summarised in Fig. 3.

Transmission and colonisation

Oxygen sensitivity

C. perfringens is commonly known as a strict anaerobe; however, this bacterium can also survive in the presence of oxygen, and/or low concentrations of superoxide or hydroxyl-radical-generating compounds[28]. Notably, as an aero-tolerant anaerobe, *C. perfringens* could potentially survive through aerobic environments (such as on surfaces in hospital wards[29]) and trigger disease development in aerophillic environments (i.e., adult/pre-term infant gut), and additionally oxygen-exposed tissues (such as gas gangrene), which may facilitate bacterial host-to-host transmission[30].

Sporulation

C. perfringens is a spore-former, which enables this bacterium to survive in extreme or nutrient-depleted conditions. This spore-forming characteristic plays a vital role in the transmission of this Gram-positive bacterium from diverse environments to hosts, leading to infection, including food poisoning in human adults[31]. Some spores of *C. perfringens* (especially food-poisoning associated strains) had been demonstrated to resist extreme temperature conditions, which may contribute to *C. perfringens* survival and subsequent disease pathology[32,33]. Molecular regulation of sporulation in *C. perfringens* is less well studied compared to the model bacterium *Bacillus subtilis*. Previous studies have indicated that the transcriptional regulator CcpA (encoded by *ccpA*) is required for effectual sporulation[34] and Spo0A (*spo0A*), a transcriptional factor known to be present in *Clostridium* species including *C. perfringens*, is associated with sporulation initiation[35]. Sigma factors, including SigF, SigE, SigG and SigK (genes *sigF*, *sigE*, *sigG* and *sigK* respectively) are known to regulate the sporulation process of *C. perfringens*[36,37]. In addition to this, it was also reported that *C. perfringens* sporulation processes could be regulated by an Agr-like quorum-sensing system, potentially promoted by the *codY* gene in type A food-poisoning strains[38,39]. Importantly, CPE is known to be produced by *C. perfringens* during sporulation, which correlates with the pathogenesis of food poisoning-associated diarrhoea[31].

Germination

Germination of *C. perfringens* can be triggered by small molecules termed germinants—commonly known as sugars, nucleosides, amino acids, salts and purines[40]. Importantly, primary bile acids (including glycocholate, cholate and taurocholate) in the human GI tract are also known to act as potent germinants in *Clostridium* species, whereas secondary bile acids, such as deoxycholate (derivatives generated by certain gut microbes), can inhibit the in vitro proliferation of *C. perfringens*[41]. This proposed 'colonising germination' survival modus operandi of spore-formers, including *C. perfringens*, potentially acts as a signal to initiate growth, and to persist in the intestinal environment[30]. This persistence, and long-term gut colonisation may correlate with the ongoing symptoms (up to several weeks) of nonfoodborne diarrhoea in patients[42]. From a colonisation resistance perspective, inhibition of germination, by secondary bile products produced by resident commensal bacteria, implies a long-standing competition for limited nutrients and niches within the gut.

Rapid proliferation

Notably, *C. perfringens* represents one of the fastest-growing organisms currently known, and is reported to have a very short 8–12 min generation time when cultured at 43 °C in optimal media, and 12–17 min at 37 °C[43]. Rapid growth of this bacterium potentially predisposes the host to tissue infection, as in the cases of avian NE, bovine necro-haemorrhagic enteritis (could be <5 h), and gas gangrene. Also, as *C. perfringens* has a two-fold quicker generation time, when compared to intestinal commensals like wild-type *Escherichia coli* (typically 20–30 min in Luria-Bertani broth), this could represent a

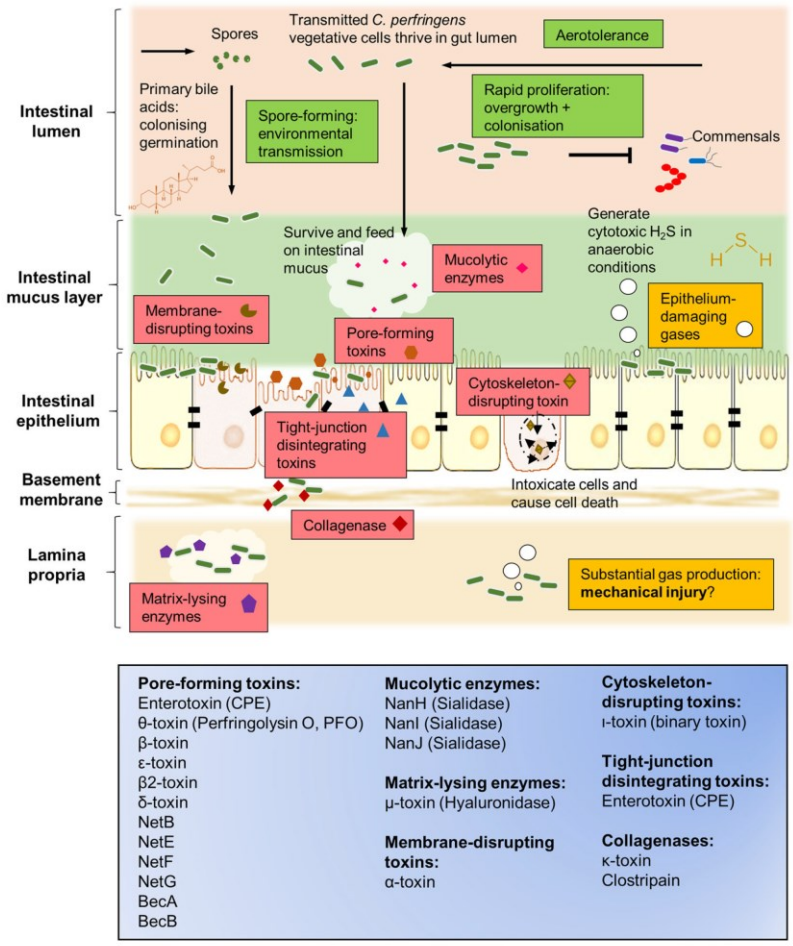


Fig. 3 Graphical representation of disease-linked virulence factors of *C. perfringens* in the context of intestinal infections. Classification of toxins according to general mechanisms of action are shown in the bottom blue box

potent mechanism for outgrowing other resident bacteria, leading to efficient gut colonisation.

Disease initiations

Histotoxic gas production

Clostridial myonecrosis (better known as gas gangrene), is accompanied by profuse gas production, which could in theory be the mechanical determinant of tissue injury, alongside rapid cell proliferation of *C. perfringens*. Importantly, *C. perfringens* is known to generate hydrogen sulphide[44] (using ubiquitous sulphuric sources from the environment), a readily-permeable toxic substance to human cells, which when produced in excess is associated with other intestinal inflammatory diseases (e.g., Ulcerative Colitis)[45]. Thus, *C. perfringens*-associated gas production represents an understudied virulence trait in the pathogenesis of gut infection, such as infant NEC (discussed in more detail below), which symptoms significantly mimic those of gas gangrene; necrosis of tissue accompanied by abundant gas production[46,47].

Toxins and virulent enzymes

Presently, 23 virulence genes that encode toxins and virulent enzymes have been identified in *C. perfringens*

(summarised in Table S1); the most 'prolific' toxin-producing pathogen currently known. Some key toxins will be briefly discussed and summarised in this subsection.

Alpha-toxin

Alpha-toxin (encoded by gene *plc* or *cpa*), which is produced by all strains of *C. perfringens*, has been shown to hydrolyse cell membrane phospholipids, which eventually leads to cell necrosis; a key characteristic in gas gangrene. Indeed, this toxin is essential for gas gangrene pathology, as shown in an in vivo myonecrosis model[48]. Mechanistically, α-toxin may play three major roles in gas gangrene pathology; firstly, it is able to impact the transferring of immune cells such as neutrophils to infected tissues (mechanism currently unknown), hence potentially reducing pathogen clearance at infected sites. Secondly, it can cause constriction of blood vessels, which may reduce the blood supply to tissues, thus creating a micro-aerophillic environment conducive to *C. perfringens* overgrowth. Thirdly, this toxin can activate inflammation cascades in host cell metabolism (arachidonic acid and protein kinase C), which may lead to direct immune-mediated pathology of tissues[49].

Beta-toxin

Beta-toxin, encoded by gene *cpb*, is a plasmid-encoded pore-forming toxin that is thought to be intestinal-necrotic, and also important for systemic enterotoxaemia, in humans and neonatal animals, as shown in in vivo studies[50,51]. *C. perfringens* type C strains that possess the *cpb* gene were associated with historical Clostridial gut infections; Darmbrand (post-world war II in Germany), and Pig Bel (Papua New Guinea)[52]. Previous intestinal loop studies have indicated a synergistic effect with CPE (encoded by *cpe*)[53]. Furthermore, this toxin shares sequence similarity with several pore-forming toxins produced by *Staphylococcus aureus* – ɣ toxin (A: 22 and B:28%), α-toxin (28%) and leucocidin (S:17% and F:28%)[54].

Perfringolysin O

PFO (also θ-toxin, encoded by gene *pfoA* or *pfo*), is also a pore-forming toxin that acts on cholesterol-comprising cell membranes. This toxin has been shown to be involved in the pathogenesis of gas gangrene, and haemorrhagic enteritis in calves[55,56]. Notably, PFO shares structural homology with similar pore-forming toxins identified in *Streptococcus*, *Bacillus*, *Listeria* and many other genera[57]. θ-toxin is also known for its capacity to induce tumour necrosis factor alpha (TNF-α) and interleukin 6 (IL-6) expression in the host, and it could activate apoptosis through p38 MAPK (mitogen-activated protein kinase) pathway as demonstrated in in vitro models[58]. Interestingly, it has synergistic cytotoxic effects with α-toxin on bovine epithelial cells, highlighting the significant role of this toxin in disease development[56,59].

Beta2-toxin

Beta2-toxin (encoded by *cpb2* plasmid gene), a pore-forming cytolytic toxin shares <15% sequence homology with β-toxin, distinguished itself as a novel toxin produced by *C. perfringens*[54]. β2-toxin-producing *C. perfringens* strains have been associated with gut diseases such as NE in piglets, and enterocolitis in foals[60]. Importantly, this pore-forming toxin has been suggested as an accessory toxin in *C. perfringens*-associated non-foodborne diarrhoea[61], and proposed to play an important role in pre-term NEC in potential synergistic effects with antibiotic gentamicin[62].

Epsilon-toxin

This deadly toxin, is plasmid-encoded by *etx* gene *C. perfringens* type B and D strains, and is essential for pathogenesis. Epsilon toxin is involved in animal (goat and sheep) enterotoxaemia[63]. ε-toxin is currently thought to be the most potent of all toxins produced by *C. perfringens*, evidenced by the LD_{50} of 70 ng/kg body weight, ranked only behind *C. botulinum* and *C. tetani* neurotoxins. It has been demonstrated to affect various organs, such as kidneys and liver, through unknown mechanisms, that allow this toxin to enter systemic circulation[64]. Due to its potential use as a biological weapon, ε-toxin-producing *C. perfringens* strains are on the export control list in a number of countries including USA and the UK.

Iota-toxin

This cytoskeleton-damaging toxin (also considered as a major toxin produced by *C. perfringens*), consists of two cell-binding monomers Ia (enzymatic component) and Ib (binding component; encoded on plasmids by *iap* and *ibp* gene respectively), which act synergistically to first translocate the toxin into host cells, then exert enzymatic activity on ADP-ribosylating actin to disassemble the cytoskeleton, which eventually leads to apoptosis and cell death[65].

Microbial collagenase

Microbial collagenase (also known as κ-toxin), encoded by chromosomal gene *colA*, is a key toxin produced by *C. perfringens* that degrades collagen. This enzyme might contribute to intestinal infection, as collagen, the substrate for collagenase, is a primary component of intestinal connective tissues/basal membrane of human and animal hosts, thus disruption may damage basal integrity, which may lead to eventual tissue necrosis[56,66]. However, Awad et al.[67] indicated that κ-toxin is not a major determinant in a Clostridial myonecrosis mouse model (gas gangrene), despite its capacity to hydrolyse collagen.

Enterotoxin (CPE)

CPE (encoding gene *cpe*) is recognised as the key toxin to cause food-poisoning and non-foodborne diarrhoea, it has also been demonstrated to disrupt the intercellular claudin tight junctions in gut epithelial cells[68,69]. Importantly, this pore-forming toxin was demonstrated to bind and necrotise human ileal and colonic epithelium in vitro, and can induce cell death i.e., apoptosis, via the caspase-3 pathway[70]. Hence, the potential pathogenesis mechanisms underlying food poisoning could result from CPE-induced tight junction rearrangements or pore-formation.

Sialidase

Three sialidases, encoded by genes *nanH, nanI* and *nanJ* in *C. perfringens*, are also named as neuraminidases or exo-sialidases (NanI and NanJ). This group of enzymes represent important virulence factors during *C. perfringens*-mediated tissue infection; they catalyse hydrolysis of terminal sialic acids from glycoprotein, glycolipids and polysaccharides of cell membranes that aids in bacterial attachment to host cells[71]. This mucolytic potential suggests that *C. perfringens* may utilise intestinal mucus as a nutrient source, and thereby potentiates intestinal colonisation, which has been modelled using the in vitro Caco-2 cell line[72]. Importantly, in vitro studies have also demonstrated that α-toxin associated with NanI (exo-alpha-sialidase) increased the virulence of *C. perfringens*[73]. Furthermore, NanI was shown to potentiate the virulence of ε-toxin, β-toxin and CPE, via binding-enhancing (ε-toxin) and proteolytic activation (β-toxin and CPE) mechanisms, potentially enhancing *C. perfringens* pathogenesis[74]. Notably, in a gas gangrene mouse model, NanI and NanJ are not essential for virulence[75].

NetB

NetB, a pore-forming toxin (encoded by *netB*), shares limited amino acid sequence similarity with beta-toxin (38% sequence identity), α-toxin from *S. aureus* (31% sequence identity) and δ-toxin (40% sequence identity) from *C. perfringens*[76]. Less than a decade ago, this chicken-NE associated toxin (plasmid-borne), was discovered and shown to be essential (instead of α-toxin) for lesion formation both in cell line models, and avian in vivo models via genetic mutant strains, and thus fulfilled molecular Koch's postulates as a disease determinant[77]. This toxin is important in avian agriculture as *netB*-positive *C. perfringens* strains recovered from broiler chickens (healthy birds) can be as high as 61%, however expression of NetB was shown to be lower in healthy birds, when compared with NE-associated chickens (92% vs. 29%)[78]. In addition, *netB* was later identified as part of a plasmid-encoded pathogenicity locus named NELoc-1 capable of being transferred via conjugation, indicating the potential spread in relation to chicken NE epidemiology[24,79]. Please refer to Review by Rood et al.[80] for a comprehensive description on NetB and poultry NE.

These extracellular toxins represent potent virulence factors central to intestinal disease development. However, not all toxins are secreted by all strains of *C. perfringens* (excluding α-toxin), hence, in silico identification of virulence genes using NGS techniques and bioinformatics tools are essential for rapid and comprehensive geno-toxinotyping of *C. perfringens*, compared with conventional molecular tools.

Antimicrobial resistance (AMR)

AMR traits in *C. perfringens* pose a serious clinical treatment concern, due to its capacity to generate an array of lethal toxins. Presently there are several phenotyping studies published (based on in vitro susceptibility testing of minimal inhibitory concentration; MIC, Table S2) on the AMR profile of *C. perfringens*, however, there is currently limited genetic/WGS-based AMR genotyping studies.

Tetracycline resistance in *C. perfringens* was first described in 1968, when 11 strains of *C. perfringens* were tested, and found to possess 'some degree' of resistance to tetracycline, thus penicillin G (or, benzylpenicillin) was recommended as the drug of choice for treating Clostridial infection[81]. Plasmid-carrying tetracycline resistance (*tet* components) in *C. perfringens* was genetically shown for the first time in porcine samples (three strains)[82]. Tetracycline resistance elements *tetA*(P), was then detected in 81 tetracycline-resistance *C. perfringens* strains (100%), with 93% of these strains carrying a secondary resistance gene, either *tet*B(P) or *tet*(M), notably, no single strain possesses all *tet* genes[83]. Multidrug-resistant *C. perfringens* strains were first reported back in 1977, multiple strains ($n = 7$) isolated from porcine faeces were demonstrated to be resistant against tetracycline (MIC > 32 μg/ml), erythromycin (MIC > 128 μg/ml), clindamycin (MIC > 64 μg/ml), and lincomycin (MIC > 128 μg/ml)[84]. A PCR-based AMR study on 160 environmental strains (water, soil and sewage) revealed encoded macrolide resistance genes *erm*(B)(26%), *erm*(Q)(1%), and *mef*(A)(18%), in addition to the commonly found *tetA*(P)(53%), *tetB*(P)(22%), and *tet*(M)(8%)[85].

Macrolide and Lincosamide resistance (mainly erythromycin and lincomycin) appears widespread[86], and therefore is considered ineffective in treating *C. perfringens* infections. This is supported by a recent multidrug-resistance study of 260 strains of *C. perfringens* isolated from diarrheal neonatal piglets in Thailand, where higher resistance was observed for erythromycin, lincomycin and tylosin[87].

Recent WGS studies on AMR genes have detected *mepA* (multidrug-resistance gene), using various public AMR databases on *C. perfringens* strains[88], and also

tetracycline resistant genes *tetA*(P), *tetB*(P) and *tet38*. The genes, *mprF* and *rpoB* (rifampin-resistant) have also been reported to be encoded[88]. Anti-defensin gene *mprF* (possibly involved in multidrug-resistant, including resistance against gentamicin) was recently reported in a large-scale genomic study of *C. perfringens* ($n = 56$ strains) to be present in 100% of the genomes[10]. In this latest genomic study, *tetA*(P) was detected in 75% of the 56 strains, which is more prevalent than *tetB*(P)(42%). Interestingly, *ANT(6)-Ib*, an aminoglycoside resistance gene, was also reported to be encoded in a *C. perfringens* toxinotype C strain. Although mainly anaerobic bacteria like *C. perfringens* may have reduced transport of aminoglycosides intracellularly, there are also findings that *C. perfringens* are sensitive to aminoglycosides like Gentamicin at higher concentration, which indicates that *C. perfringens* might also have acquired additional resistance to aminoglycosides[89,90].

WGS will be a key tool in the fight against *C. perfringens* AMR, particularly in rapid diagnostics, however gold-standard phenotypic MIC tests will also be required to clinically determine the antimicrobial susceptibilities, and facilitate antibiotic management.

Clinical associations in humans and animals

C. perfringens has been constantly associated with gut diseases across both animal and human hosts (summarised in Table S3) as described in this section.

Animal hosts

Poultry NE

Poultry NE was first documented in England in 1961[91]. Since then, NE has been consistently reported in every continent around the globe. Importantly, *C. perfringens* is unequivocally identified as the key aetiological organism of NE in broiler chickens[92]. Global financial impact of NE is significant, with an estimated economic loss of 6 billion US $ in 2015, projecting profit loss per bird at >US$0.062[93].

Classic NE pathologies are characterised by gaseous lesions and mucosa necrosis in gas-filled distended small intestine, however, these can also involve kidney and liver as secondary infection sites[94]. Proposed key biological factors that contribute to NE are hydrolytic enzymes (e.g., collagenase)[95], toxin production (traditionally understood as α-toxin, and more recently NetB and TpeL, both are pore-forming toxins, were linked to the onset of NE[96]. Other NE-predisposing factors include: a high-protein diet that favours the growth of *C. perfringens*, and environmental stress (e.g., high stocking density), which alters gut microbiota/host immunity that eventually increases risk of infection[97].

Equine acute necrotising enterocolitis

Acute necrotising enterocolitis (ANEC) in foals/horses (caused by *C. perfringens*) is a severe intestinal disease that resembles the classic clinical signs of *C. perfringens*-infections; rapid disease development and necrotic intestines (mainly the colon)[98]. ANEC symptoms are characterised by bloody diarrhoea, followed by a haemorrhagic and necrotic gut[99]. Type F *C. perfringens* strains (previously type A) harbouring both CPE and β2-toxin are often associated with this deadly condition[98]. More recently, NetE, NetF and NetG toxins (pore-forming toxins) were proposed to underlie the pathogenesis of foal NEC[100].

Swine enterocolitis

C. perfringens type A strains (also less frequently, CPE-harbouring type F strains) are widely considered as the invasive agent of non-haemorrhagic EC in piglets, although the actual pathogenesis remains undefined. Similar to other *C. perfringens*-intestinal infection, this disease commonly affects one-week-old piglets, suspected to inherit (i.e., via vertical transmission during birth) the bacterium from the microbiota of mother sows[101]. Symptoms involve severe diarrhoea (non-haemorrhagic), accompanied by necrotic mucosa and atrophy of intestinal villi. β2-toxin was initially believed to drive the development of this disease (backed by epidemiological studies in 1990s), and as such was used as a diagnostic biomarker, although this has become controversial in recent years[60]. Type C strains that carry β-toxin are associated with haemorrhagic EC in piglets and largely affect 1 to 4 days old neonates[101]. In contrast to type A-infection (to a lesser degree, type F-infection), type C-infection is characterised by haemorrhagic NE, which is proposed to be driven by the presence of type C strains and low trypsin secretion (trypsin can inactivate β-toxin) in the immature host gut[102].

Human hosts

Darmbrand

Darmbrand, which literally means 'burning (fire) bowels' in German, was used to describe a particular type of necrotic inflammatory gut disease (also known as enteritis necroticans, EN) associated with *C. perfringens*, that occurred epidemically post-second World War (1944–1949) in north-west Germany[103]. Darmbrand causative strains were later classified as type C strains that carry β-toxin, and these strains are highly resistant to heat[104]. Many of these type C strains also harbour CPE, which has been shown to act synergistically with β-toxin in an intestinal loop model and therefore CPE was suggested to play a role in some cases of EN[53]. Notably, the heat-resistance of this strain might be attributed to the production of a small acid-soluble protein (Ssp4) that could potentially play a central role in dissemination of *C. perfringens* strains[33]. Darmbrand was believed to be facilitated by poor post-war sanitary (bacterial contamination), and malnutrition (protein-malnourished)

conditions, as this disease disappeared within a few years after its first recognition.

Pigbel

Enteritis necroticans (EN), or commonly known as pigbel in the highlands of Papua New Guinea (PNG), is a form of inflammatory gut disease that was associated with pork-feasting activities that took place among PNG Highlanders in outbreaks first documented back in 1966[105]. The classical description of the symptoms is 'spontaneous gangrene of small intestine, without obvious vascular or mechanical cause' which resembles Darmbrand, and occurred particularly in children[106]. The aetiology of this fatal infection was thought to be large co-consumption of *C. perfringens* type C-contaminated pork (due to poor hygiene), and large amount of sweet potatoes which contain trypsin inhibitor (trypsin secreted in the gut could break down β-toxin secreted by *C. perfringens*)[107]. Similar cases of EN were also reported in Uganda[108] and Indonesia[109]. Notably, EN-like cases were also reported (rare cases) in developed countries, including USA, in exclusively diabetic subjects, who also have attenuated trypsin production[110].

Acute watery diarrhoea (food poisoning)

C. perfringens has been associated with food poisoning since it was first documented in both the UK and the USA in the 1940s[111]. *C. perfringens*-type A food poisoning, is ranked the second most prevalent bacterial food poisoning in the US, estimated at 1 million cases per annum, after *Salmonella*[112]. In the European Union (EU) member countries, *C. perfringens*-linked food-poisoning outbreaks were projected at approximately 5 million cases per year (2011)[113]. *C. perfringens* has also been reported to affect elderly communities, especially in care homes (North East of England, 2012–2014; 83% of the outbreaks reported from care homes)[114]. *C. perfringens*-linked foodborne cases are suggested to be under-reported due to its self-limiting symptoms, thus the statistics published based on laboratory-confirmed cases may be lower than the actual numbers, suggesting a higher actual epidemiological impact.

Hallmark symptoms (self-limiting, lasting 12–24 h; mortality is uncommon) include, rapid appearance within 8–14 h after food ingestion, intestinal cramp, watery diarrhoea without fever or vomiting[115]. Currently, food poisoning *C. perfringens* cases are thought to be caused by CPE (encoded by *cpe* gene). This secreted pore-forming toxin is also demonstrated to disrupt intestinal tight junction barriers and initiate disease development[68].

Non-foodborne diarrhoea

C. perfringens has also been associated with non-foodborne diarrhoea (a distinct disease entity from food poisoning, characterised primarily by more severe symptoms and longer duration), which includes antibiotic-associated diarrhoea (AAD), and sporadic diarrhoea (SD)[116,117]. AAD typically occurs in 5–25% of patients after administration of broad-spectrum antibiotics[118]. Non-foodborne diarrhoea typically affects older adults that are >60 years of age (although SD is also associated with younger age groups)[119]. Clinical symptoms are abdominal pain, prolonged diarrhoea (>3 days to several weeks), and bloody stools[120,121]. Patients suffering from non-foodborne diarrhoea, particularly AAD, can become seriously ill due to diarrhoea-induced dehydration, and may progress to develop colitis, although full recovery is common[122].

Although other pathogenic bacteria including *C. difficile* (most common AAD pathogen; ~25% of cases[123]) and *S. aureus* have also been implicated in AAD, enterotoxigenic *C. perfringens* type F strains (producing CPE) are estimated to be responsible for up to 15% of all AAD cases[116,124,125]. CPE, produced by *C. perfringens* type F strains (encoded by plasmid-borne *cpe* gene), has been reported to be the aetiological agent for *C. perfringens*-associated non-foodborne diarrhoea, as evidenced by the high prevalence in diarrhoea patients, but not healthy individuals. Importantly, type F strains have also been linked to SD, although to a lesser extent[116]. AAD-associated *C. perfringens* strains are also described to be more adherent to Caco-2 intestinal cells, when compared to other food-poisoning strains, which is attributed to the production of sialidase NanI[126]. The spore-forming nature of *C. perfringens* could also potentially contribute to the commonly observed disease persistence and relapse[127]. Molecular detection of faecal CPE or PCR confirmation of *cpe* gene represents the current clinical diagnostic method for *C. perfringens*-AAD[128,129].

Pre-term infant NEC

Pre-term infant NEC has been clinically linked with *C. perfringens* since the 1970s, as type A-*C. perfringens* were isolated from necrotic tissues in many NEC cases. Notably, it was even called 'gas gangrene of the bowel' owing to its reflection of the highly similar diseased histology with the infamous tissue myonecrosis[47].

NEC has ~14% prevalence in pre-term infants less than 1000 g birth weight (i.e., extremely low birth weight, ELBW), a high mortality of 30% with surgical NEC, and mortality up to 50%. NEC-related deaths account for 10% of all mortality causes in neonatal intensive care units (NICUs) according to a UK-nationwide study[130], and is the most severe and lethal neonatal GI emergency worldwide[131]. The additional annual cost for treating NEC in pre-term infants is conservatively estimated at £13.5 million in the UK (excluding long-term post-surgical treatment), whilst in the US $1 billion is estimated to be spent annually by neonatal departments on NEC[132].

The involvement of bacteria in the pathogenesis of NEC has been strongly implicated since the disease was first described[133]. Both *Klebsiella* spp., and *C. perfringens*, have been linked to clinical NEC in recent years[4]. Importantly, *C. perfringens* was isolated from most neonates (70%), as early as the fifth day of life, which supports the theory of pathogen colonisation that then leads to microbial invasion. Classic NEC symptoms such as pneumatosis intestinalis, suggests gas-producing Gram-positive bacteria including *Clostridium* spp. (also implicated consistently in recent metagenomics studies), as the key causative agent[134,135]. Most recently, the emergence of two extensive pre-term infant NEC cohort studies that profiled metagenomic faecal samples, have both indicated that *C. perfringens* is significantly abundant in the infant gut microbiota prior to NEC development[4,5]. These studies are supported by the substantial number of NEC studies (14 studies to date since 1970s, more than any individual bacterial agents, as summarised in Table S4) that have reported *C. perfringens* as a potential pathogenic agent, in addition to its established pathogenesis in many other neonatal animal gut diseases, and further supported by in vivo studies[136]. Thus, *C. perfringens* appears to be a pathogenic agent of microbial-NEC, i.e., *C. perfringens*-associated NEC, and may link to specific disease symptoms[137].

Being universal and resilient in habitable niches, spore former *C. perfringens* may readily pass on to in-hospital neonates through either environmental or oral transmission, and then initiate diseases in the intestine[138]. Proposed underlying mechanisms for *C. perfringens*-associated NEC include reduced bowel peristalsis in pre-term infants (increased pathogen retention time), universal administration of broad-spectrum antibiotics[139] (that leads to reduced microbiota diversity, and the potential rapid overgrowth of resistant spores), reduced gut barrier integrity (non-existent mucus layer), lack of protective bacteria[4,5,131] (including *Bifidobacterium* species, which may antagonise the growth of *C. perfringens*), and immature/underdeveloped immune system that is ineffective at fighting off pathogenic bacteria[46] (Fig. 4). Bile salt factors may also be linked to disease onset. Reduced diversity of the microbiota in pre-term infants, including bacteria that can de-conjugate bile salts to secondary bile acids, may allow germination of *C. perfringens*, which has also been directly indicated in the pathogenesis of the well-studied *C. difficile* colitis (both species form spores and secrete colitis-related toxins)[140].

To date, no particular toxin(s) has been specifically linked with pre-term NEC, although β2-toxin has been suggested[4]. Notably, *C. perfringens* can also be part of a 'healthy' human microbiota (as low as 3.5% in pre-term neonates, nevertheless this is controversial[141]), thus suggesting other factors play a role. Antibiotic administration

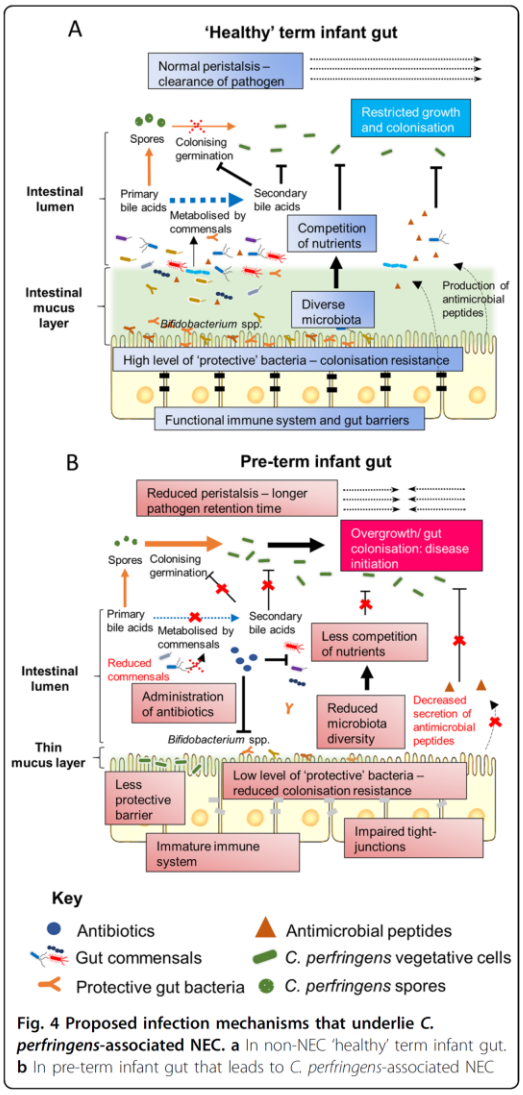


Fig. 4 Proposed infection mechanisms that underlie *C. perfringens*-associated NEC. a In non-NEC 'healthy' term infant gut. **b** In pre-term infant gut that leads to *C. perfringens*-associated NEC

in pre-term babies is very high, with up to 97% of pre-term infants residing in NICU exposed to at least one dosage of broad-spectrum antibiotics[142]. These antibiotics eliminate 'protective' members of the gut microbiota, and in tandem with presence of AMR-*C. perfringens*, may facilitate successful colonisation and infection[143]. An immature gut barrier in neonatal infants may also expose gut niches for overgrowth, and finally the presence of 'hyper-virulent' strains of *C. perfringens* may be implicated. However, to comprehensibly answer these important questions a more global genomic approach, i.e., WGS

Box 2

What are the key knowledge gaps in understanding *C. perfringens* as a pathogen, and what are the potential approaches to investigate/mitigate the virulence of *C. perfringens* in clinical diseases?

Questions

- Compared to other pathogens (e.g., *C. difficile*), there is limited WGS data for *C. perfringens*, what is the genomic diversity of this important pathogen?
- Can we use WGS to 'type' *C. perfringens* isolates, and can WGS be used in diagnostics (including AMR profiling), and to track spread in outbreaks settings?
- *C. perfringens* causes diverse diseases in a wide host range, but what are the microbial virulence factors responsible for pathology, and is this host/strain specific?
- What are the host immune cells/signalling cascades involved in *C. perfringens*-associated clearance and/or pathology?
- *C. perfringens* can reside as a 'normal' member of the gut microbiota, what are the external factors that may lead to overgrowth and disease?
- Can *C. perfringens* be directly linked to NEC pathogenesis?
- What are the therapeutic approaches that could be used to prevent *C. perfringens* infection in multi-host species?

Approaches

- Isolate diverse strains from various environments, i.e., from both case and control samples
- Perform comprehensive bioinformatics analyses to understand evolution, e.g., SNP based phylogeny, and investigate variants and spread clinical settings
- Carry out pan-GWAS/comparative genomics to identify/predict specific virulence genes related to disease development.
- Develop/optimise clinical diagnostics from faecal samples via e.g. ultra-long read real-time sequencing method such as Oxford Nanopore
- Develop and characterise in vivo infection models for different disease types (e.g., NEC and gastroenteritis)
- Pinpoint specific virulence genes involved in infection via bacterial transcriptomics (RNAseq) and knock-out bacterial strains, to fulfil molecular Koch's postulates.
- Perform microbiota profiling (e.g., 16S rRNA metagenomics or shotgun metagenomics) to investigate the impact of *C. perfringens* to microbiota members, and the role of microbiome in *C. perfringens* infections
- Understand host defence mechanisms using immunological approaches, and knock out in vivo models
- Develop therapeutics for *C. perfringens* infections including phage therapy, vaccines, microbiota treatment i.e., probiotics.

and analysis, and validation using suitable in vivo models is required.

Advancement of WGS and genomic analyses, particularly those focussing on other gut pathogens including *C. difficile*, *K. pneumoniae* and *Salmonella enterica* has proved successful in understanding the pathogenic role of these bacteria in disease development and tracking bacterial transmission[144,145]. These enteric pathogens have all been sequenced extensively with genomic information made public on NCBI database (as of May 2018; bracketed digits indicate number of genomes): *C. difficile* (1322), *E. coli* (10541), *C. botulinum* (225), *K. pneumoniae* (3927) and *S. enterica* (8780). Yet, including a batch addition of 30 genomes towards the end of 2016, there are only 59 publicly available genomes of *C. perfringens* (May 2018), highlighting the lack of in-depth understanding of this pathogen with respect to its genomic information. Importantly, no NEC-associated pre-term infant isolates are known to be sequenced at present, and genomic investigation may allow identification of putative/novel NEC-associated virulence factors, that may facilitate intervention strategies and/or therapy development.

Future research directions

The clinical link between intestinal diseases and *C. perfringens* is clear and defined, however the underlying factors responsible for specific aspects of pathology remains uncertain (Box 1, Box 2). Deciphering the genes that are involved in the sporulation, germination, enzyme/toxin production, oxygen tolerance, AMR and other novel virulence factors could lead to more targeted clinical preventions in *C. perfringens*-associated adult and neonatal intestinal diseases, whether it be humans or animals. Important questions may be answered utilising WGS and in silico tools, to delve into the genetic makeup of this notorious, yet under-studied pathogen. In tandem further in vitro and in vivo studies should be carried out to confirm the importance of suspected novel virulence traits, i.e., to fulfil Molecular Koch's postulates. These approaches may help establish future platforms for disease prevention strategies such as vaccines, phage therapy, microbiota-based therapeutics, or implementation of specific antibiotic administration policies. Furthermore, understanding the genomes will potentially enable epidemiological tracking (as is taking place for other pathogenic bacteria), allowing us to pinpoint the origin and route of spread of isolates in the hospital settings, which is vital within clinical contexts.

Acknowledgements

This review was supported by a Wellcome Trust Investigator Award (100974/C/13/Z), and the Biotechnology and Biological Sciences Research Council (BBSRC); Institute Strategic Programme Gut Microbes and Health BB/R012490/1, and Institute Strategic Programme Gut Health and Food Safety BB/J004529/1 to L.J.H. R.K. is a recipient of University of East Anglia International Bursary (Faculty of Medical and Health Sciences).

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41426-018-0144-8).

Received: 2 March 2018 Revised: 28 June 2018 Accepted: 2 July 2018
Published online: 06 August 2018

References

1. Hassan, K. A. et al. Genomic analyses of *Clostridium perfringens* isolates from five toxinotypes. *Res. Microbiol.* **166**, 255–263 (2015).
2. Welch, W. H. & Nuttall, G. H. F. A gas-producing bacillus (*Bacillus aerogenes* capsulatus, Nov, Spec.) capable of rapid development in the body after death. *Bull. John Hopkins Hosp. Baltim.* **3**, 81–91 (1891).
3. Lugli, G. A. et al. Ancient bacteria of the Otzi's microbiome: a genomic tale from the Copper Age. *Microbiome* **5**, 5 (2017).
4. Sim, K. et al. Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clin. Infect. Dis.* **60**, 389–397 (2015).
5. 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).
6. Revitt-Mills, S. A., Rood, J. I. & Adams, V. *Clostridium perfringens* extracellular toxins and enzymes: 20 and counting. *Microbiology Australia*, 114–117 (2015).
7. Rood, J. I. et al. Expansion of the *Clostridium perfringens* toxin-based typing scheme. *Anaerobe* (2018) https://doi.org/10.1016/j.anaerobe.2018.04.011.
8. Petit, L., Gibert, M. & Popoff, M. *Clostridium perfringens*: toxinotype and genotype. *Trends Microbiol.* **7**, 104–110 (1999).
9. Nagahama, M. et al. Recent insights into *Clostridium perfringens* Beta-toxin. *Toxins* **7**, 396–406 (2015).
10. Kiu, R., Caim, S., Alexander, S., Pachori, P. & Hall, L. J. Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant pan-genome diversity, and a diverse array of virulence factors. *Front. Microbiol.* **8**, 2485 (2017).
11. Freedman, J. C., Shrestha, A. & McClane, B. A. *Clostridium perfringens* Enterotoxin: Action, Genetics, and Translational Applications. *Toxins* (Basel) **8**, https://doi.org/10.3390/toxins8030073 (2016).
12. Shimizu, T. et al. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc. Natl Acad. Sci. USA* **99**, 996–1001 (2002).
13. Kumar, N. et al. Genome-based infection tracking reveals dynamics of *Clostridium difficile* transmission and disease recurrence. *Clin. Infect. Dis.*, civ1031, https://doi.org/10.1093/cid/civ1031 (2015).
14. Freedman, J. C. et al. *Clostridium perfringens* type A-E toxin plasmids. *Res. Microbiol.* **166**, 264–279 (2015).
15. Li, J. et al. Toxin plasmids of *Clostridium perfringens*. *Microbiol. Mol. Biol. Rev.* **77**, 208–233 (2013).
16. Wisniewski, J. A. & Rood, J. I. The Tcp conjugation system of *Clostridium perfringens*. *Plasmid* **91**, 28–36 (2017).
17. Grant, K. A. et al. The identification and characterization of *Clostridium perfringens* by real-time PCR, location of enterotoxin gene, and heat resistance. *Foodborne Pathog. Dis.* **5**, 629–639 (2008).
18. Miyamoto, K. et al. Complete sequencing and diversity analysis of the enterotoxin-encoding plasmids in *Clostridium perfringens* type A non-food-borne human gastrointestinal disease isolates. *J. Bacteriol.* **188**, 1585–1598 (2006).
19. Sayeed, S., Li, J. & McClane, B. A. Characterization of virulence plasmid diversity among *Clostridium perfringens* type B isolates. *Infect. Immun.* **78**, 495–504 (2010).
20. Sayeed, S., Li, J. & McClane, B. A. Virulence plasmid diversity in *Clostridium perfringens* type D isolates. *Infect. Immun.* **75**, 2391–2398 (2007).
21. Gurjar, A., Li, J. & McClane, B. A. Characterization of toxin plasmids in *Clostridium perfringens* type C isolates. *Infect. Immun.* **78**, 4860–4869 (2010).
22. Li, J., Miyamoto, K. & McClane, B. A. Comparison of virulence plasmids among *Clostridium perfringens* type E isolates. *Infect. Immun.* **75**, 1811–1819 (2007).
23. Miyamoto, K. et al. Identification of novel *Clostridium perfringens* type E strains that carry an iota toxin plasmid with a functional enterotoxin gene. *PLoS ONE* **6**, e20376 (2011).
24. Bannam, T. L. et al. Necrotic enteritis-derived *Clostridium perfringens* strain with three closely related independently conjugative toxin and antibiotic resistance plasmids. *MBio* **2**, https://doi.org/10.1128/mBio.00190-11 (2011).
25. Ebersbach, G. & Gerdes, K. Plasmid segregation mechanisms. *Annu. Rev. Genet.* **39**, 453–479 (2005).
26. Watts, T. D., Johanesen, P. A., Lyras, D., Rood, J. I. & Adams, V. Evidence that compatibility of closely related replicons in *Clostridium perfringens* depends on linkage to parMRC-like partitioning systems of different subfamilies. *Plasmid* **91**, 68–75 (2017).
27. McClane, B. A., et al. eds 698–752 (Springer, 2006).
28. Briolat, V. & Reysset, G. Identification of the *Clostridium perfringens* genes involved in the adaptive response to oxidative stress. *J. Bacteriol.* **184**, 2333–2343 (2002).
29. Machida, Y. et al. An outbreak of enterocolitis due to *Clostridium perfringens* in a hospital for the severely disabled. *Kansenshogaku Zasshi* **63**, 410–416 (1989).
30. Browne, H. P. et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543–546 (2016).
31. Li, J., Paredes-Sabja, D., Sarker, M. R. & McClane, B. A. *Clostridium perfringens* sporulation and sporulation-associated toxin production. *Microbiol. Spectr.* **4**, https://doi.org/10.1128/microbiolspec.TBS-0022-2015 (2016).
32. Orsburn, B., Melville, S. B. & Popham, D. L. Factors contributing to heat resistance of *Clostridium perfringens* endospores. *Appl. Environ. Microbiol.* **74**, 3328–3335 (2008).
33. Li, J. & McClane, B. A. A novel small acid soluble protein variant is important for spore resistance of most *Clostridium perfringens* food poisoning isolates. *PLoS Pathog.* **4**, e1000056 (2008).
34. Varga, J., Stirewalt, V. L. & Melville, S. B. The CcpA protein is necessary for efficient sporulation and enterotoxin gene (cpe) regulation in *Clostridium perfringens*. *J. Bacteriol.* **186**, 5221–5229 (2004).
35. Myers, G. S. et al. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. *Genome Res.* **16**, 1031–1040 (2006).
36. Li, J. & McClane, B. A. Evaluating the involvement of alternative sigma factors SigF and SigG in *Clostridium perfringens* sporulation and enterotoxin synthesis. *Infect. Immun.* **78**, 4286–4293 (2010).
37. Harry, K. H., Zhou, R., Kroos, L. & Melville, S. B. Sporulation and enterotoxin (CPE) synthesis are controlled by the sporulation-specific sigma factors SigE and SigK in *Clostridium perfringens*. *J. Bacteriol.* **191**, 2728–2742 (2009).
38. Li, J., Chen, J., Vidal, J. E. & McClane, B. A. The Agr-like quorum-sensing system regulates sporulation and production of enterotoxin and beta2 toxin by *Clostridium perfringens* type A non-food-borne human gastrointestinal disease strain F5603. *Infect. Immun.* **79**, 2451–2459 (2011).
39. Li, J., Freedman, J. C., Evans, D. R. & McClane, B. A. CodY promotes sporulation and enterotoxin production by *Clostridium perfringens* type A strain SM101. *Infect. Immun.* **85**, https://doi.org/10.1128/IAI.00855-16 (2017).
40. Setlow, P. Spore germination. *Curr. Opin. Microbiol.* **6**, 550–556 (2003).
41. Sorg, J. A. & Sonenshein, A. L. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J. Bacteriol.* **190**, 2505–2512 (2008).
42. Modi, N. & Wilcox, M. H. Evidence for antibiotic induced *Clostridium perfringens* diarrhoea. *J. Clin. Pathol.* **54**, 748–751 (2001).
43. Li, J. & McClane, B. A. Further comparison of temperature effects on growth and survival of *Clostridium perfringens* type A isolates carrying a chromosomal or plasmid-borne enterotoxin gene. *Appl. Environ. Microbiol.* **72**, 4561–4568 (2006).
44. Fuchs, A. & Bonde, G. J. The availability of sulphur for *Clostridium perfringens* and an examination of hydrogen sulphide production. *J. Gen. Microbiol.* **16**, 330–340 (1957).
45. Rowan, F. E., Docherty, N. G., Coffey, J. C. & O'Connell, P. R. Sulphate-reducing bacteria and hydrogen sulphide in the aetiology of ulcerative colitis. *Br. J. Surg.* **96**, 151–158 (2009).
46. Neu, J. & Walker, A. W. Necrotizing enterocolitis. *N. Engl. J. Med.* **364**, 255–264 (2011).
47. Pedersen, P. V., Hansen, F. H., Halveg, A. B. & Christiansen, E. D. Necrotising enterocolitis of the newborn-is it gas-gangrene of the bowel? *Lancet* **2**, 715–716 (1976).
48. Awad, M. M., Bryant, A. E., Stevens, D. L. & Rood, J. I. Virulence studies on chromosomal alpha-toxin and theta-toxin mutants constructed by allelic exchange provide genetic evidence for the essential role of alpha-toxin in *Clostridium perfringens*-mediated gas gangrene. *Mol. Microbiol.* **15**, 191–202 (1995).

49. Takehara, M. et al. *Clostridium perfringens* alpha-toxin impairs innate immunity via inhibition of neutrophil differentiation. *Sci. Rep.* **6**, 28192 (2016).
50. Timoney, J. F., Gillespie, J. H., Scott, F. W. & Barlough, J. E. *Hagan and Bruner's Microbiology and Infectious Diseases of Domestic Animals*. (Comstock Publishing Associates, New York 1988).
51. Uzal, F. A. et al. Development and application of new mouse models to study the pathogenesis of *Clostridium perfringens* type C Enterotoxemias. *Infect. Immun.* **77**, 5291–5299 (2009).
52. Murrel, T. G. C. Pigbel in Papua New Guinea: an ancient disease rediscovered. *Int. J. Epidemiol.* **12**, 211–214 (1983).
53. Ma, M. et al. Synergistic effects of *Clostridium perfringens* enterotoxin and beta toxin in rabbit small intestinal loops. *Infect. Immun.* **82**, 2958–2970 (2014).
54. Hunter, S. E., Brown, J. E., Oyston, P. C., Sakurai, J. & Titball, R. W. Molecular genetic analysis of beta-toxin of *Clostridium perfringens* reveals sequence homology with alpha-toxin, gamma-toxin, and leukocidin of Staphylococcus aureus. *Infect. Immun.* **61**, 3958–3965 (1993).
55. 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).
56. Goossens, E. et al. Rethinking the role of alpha toxin in *Clostridium perfringens*-associated enteric diseases: a review on bovine necro-haemorrhagic enteritis. *Vet. Res.* **48**, 9 (2017).
57. Verherstraeten, S. et al. Perfringolysin O: the underrated *Clostridium perfringens* toxin? *Toxins* **7**, 1702–1721 (2015).
58. 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).
59. Verherstraeten, S. et al. The synergistics necrohemorrhagic action of *Clostridium perfringens* perfringolysin and alpha toxin in the bovine intestine and against bovine endothelial cells. *Vet. Res.* **44**, 45 (2013).
60. Gibert, M., Jolivet-Reynaud, C. & Popoff, M. R. Beta2 toxin, a novel toxin produced by *Clostridium perfringens*. *Gene* **203**, 65–73 (1997).
61. Fisher, D. J. et al. Association of beta2 toxin production with *Clostridium perfringens* type A human gastrointestinal disease isolates carrying a plasmid enterotoxin gene. *Mol. Microbiol.* **56**, 747–762 (2005).
62. Vilei, E. M. et al. Antibiotic-induced expression of a cryptic cpb2 gene in equine beta2-toxigenic *Clostridium perfringens*. *Mol. Microbiol.* **57**, 1570–1581 (2005).
63. Popoff, M. R. Epsilon toxin: a fascinating pore-forming toxin. *FEBS J.* **278**, 4602–4615 (2011).
64. Tamai, E. et al. Accumulation of *Clostridium perfringens* epsilon-toxin in the mouse kidney and its possible biological significance. *Infect. Immun.* **71**, 5371–5375 (2003).
65. Hilger, H. et al. The long-lived nature of *Clostridium perfringens* iota toxin in mammalian cells induces delayed apoptosis. *Infect. Immun.* **77**, 5593–5601 (2009).
66. Obana, N., Nomura, N. & Nakamura, K. Structural requirement in *Clostridium perfringens* collagenase mRNA 5′ leader sequence for translational induction through small RNA-mRNA base pairing. *J. Bacteriol.* **195**, 2937–2946 (2013).
67. Awad, M. M. et al. Construction and virulence testing of a collagenase mutant of *Clostridium perfringens*. *Microb. Pathog.* **28**, 107–117 (2000).
68. Shinoda, T. et al. Structural basis for disruption of claudin assembly in tight junctions by an enterotoxin. *Sci. Rep.* **6**, 33632 (2016).
69. Eichner, M. et al. In colon epithelia, *Clostridium perfringens* enterotoxin causes focal leaks by targeting claudins which are apically accessible due to tight junction derangement. *J. Infect. Dis.* **217**, 147–157 (2017).
70. Chakrabarti, G. & McClane, B. A. The importance of calcium influx, calpain and calmodulin for the activation of CaCo-2 cell death pathways by *Clostridium perfringens* enterotoxin. *Cell Microbiol.* **7**, 129–146 (2005).
71. Llanco, L. A., Nakano, V. & Avila-Campos, M. J. Sialidase production and genetic diversity in *Clostridium perfringens* type A isolated from chicken with necrotic enteritis in Brazil. *Curr. Microbiol.* **70**, 330–337 (2015).
72. Li, J. & McClane, B. A. NanI sialidase can support the growth and survival of *Clostridium perfringens* strain F4969 using sialyated host macromolecules (Mucin) or Caco-2 cells. *Infect. Immun.*, https://doi.org/10.1128/IAI.00547-17 (2017).
73. Flores-Diaz, M. et al. A cellular deficiency of gangliosides causes hypersensitivity to *Clostridium perfringens* phospholipase C. *J. Biol. Chem.* **280**, 26680–26689 (2005).
74. Theoret, J. R. et al. Native or proteolytically activated NanI sialidase enhances the binding and cytotoxic activity of *Clostridium perfringens* enterotoxin and beta toxin. *Infect. Immun.* **86**, https://doi.org/10.1128/IAI.00730-17 (2018).
75. Chiarezza, M. et al. The NanI and NanJ sialidases of *Clostridium perfringens* are not essential for virulence. *Infect. Immun.* **77**, 4421–4428 (2009).
76. Keyburn, A. L., Bannam, T. L., Moore, R. J. & Rood, J. I. NetB, a pore-forming toxin from necrotic enteritis strains of *Clostridium perfringens*. *Toxins* **2**, 1913–1927 (2010).
77. Keyburn, A. L. et al. NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. *PLoS Pathog.* **4**, e26 (2008).
78. Abildgaard, L., Sondergaard, T. E., Engberg, R. M., Schramm, A. & Hojberg, O. In vitro production of necrotic enteritis toxin B, netB, by netB-positive and netB-negative *Clostridium perfringens* originating from healthy and diseased broiler chickens. *Vet. Microbiol.* **144**, 231–235 (2010).
79. Lepp, D. et al. Identification of novel pathogenicity loci in *Clostridium perfringens* strains that cause avian necrotic enteritis. *PLoS ONE* **5**, e10795 (2010).
80. Rood, J. I., Keyburn, A. L. & Moore, R. J. NetB and necrotic enteritis: the hole movable story. *Avian Pathol.* **45**, 295–301 (2016).
81. Johnstone, F. R. & Cockcroft, W. H. *Clostridium welchii* resistance to tetracycline. *Lancet* **1**, 660–661 (1968).
82. Rood, J. I., Buddle, J. R., Wales, A. J. & Sidhu, R. The occurrence of antibiotic resistance in *Clostridium perfringens* from pigs. *Aust. Vet. J.* **62**, 276–279 (1985).
83. Lyras, D. & Rood, J. I. Genetic organization and distribution of tetracycline resistance determinants in *Clostridium perfringens*. *Antimicrob. Agents Chemother.* **40**, 2500–2504 (1996).
84. Rood, J. I., Maher, E. A., Somers, E. B., Campos, E. & Duncan, C. L. Isolation and characterization of multiply antibiotic-resistant *Clostridium perfringens* strains from porcine feces. *Antimicrob. Agents Chemother.* **13**, 871–880 (1978).
85. Soge, O. O., Tivoli, L. D., Meschke, J. S. & Roberts, M. C. A conjugative macrolide resistance gene, mef(A), in environmental *Clostridium perfringens* carrying multiple macrolide and/or tetracycline resistance genes. *J. Appl. Microbiol.* **106**, 34–40 (2009).
86. Slavic, D. et al. Antimicrobial susceptibility of *Clostridium perfringens* isolates of bovine, chicken, porcine, and turkey origin from Ontario. *Can. J. Vet. Res.* **75**, 89–97 (2011).
87. Ngamwongsatit, B. et al. Multidrug resistance in *Clostridium perfringens* isolated from diarrheal neonatal piglets in Thailand. *Anaerobe* **38**, 88–93 (2016).
88. Li, C., Yan, X. & Lillehoj, H. S. Complete genome sequence of *Clostridium perfringens* LLY_N11, a necrotic enteritis-inducing strain isolated from a healthy chicken intestine. *Genome Announc.* **5**, https://doi.org/10.1128/genomeA.01225-17 (2017).
89. Bryan, L. E., Kowand, S. K. Van Den Elzen, H. M. Mechanism of aminoglycoside antibiotic resistance in anaerobic bacteria: *Clostridium perfringens* and Bacteroides fragilis. *Antimicrob. Agents Chemother.* **15**, 7–13 (1979).
90. Udhayavel, S., Thippichettypalayam Ramasamy, G., Gowthaman, V., Malmarugan, S. & Senthilvel, K. Occurrence of *Clostridium perfringens* contamination in poultry feed ingredients: Isolation, identification and its antibiotic sensitivity pattern. *Anim. Nutr.* **3**, 309–312 (2017).
91. Parish, W. E. Necrotic enteritis in the fowl (Gallus gallus domesticus). I. Histopathology of the disease and isolation of a strain of *Clostridium welchii*. *J. Comp. Pathol.* **71**, 377–393 (1961).
92. Zahoor, I., Ghayas, A. & Basheer, A. Genetics and genomics of susceptibility and immune response to necrotic enteritis in chicken: a review. *Mol. Biol. Rep.* **45**, 31–37 (2017).
93. Wade, B., and Keybun, A. 2015. The true cost of necrotic enteritis. Poultry World [Online]. Available: https://www.poultryworld.net/Meat/Articles/2015/10/The-true-cost-of-necrotic-enteritis-2699819W/ [Accessed 30 July 2018].
94. Timbermont, L., Haesebrouck, F., Ducatelle, R. & Van Immerseel, F. Necrotic enteritis in broilers: an updated review on the pathogenesis. *Avian Pathol.* **40**, 341–347 (2011).
95. Olkowski, A. A., Wojnarowicz, C., Chirino-Trejo, M., Laarveld, B. & Sawicki, G. Sub-clinical necrotic enteritis in broiler chickens: novel etiological consideration based on ultra-structural and molecular changes in the intestinal tissue. *Res. Vet. Sci.* **85**, 543–553 (2008).
96. Coursodon, C. F., Glock, R. D., Moore, K. L., Cooper, K. K. & Songer, J. G. TpeL-producing strains of *Clostridium perfringens* type A are highly virulent for broiler chicks. *Anaerobe* **18**, 117–121 (2012).
97. Drew, M. D., Syed, N. A., Goldade, B. G., Laarveld, B. & Van Kessel, A. G. Effects of dietary protein source and level on intestinal populations of *Clostridium perfringens* in broiler chickens. *Poult. Sci.* **83**, 414–420 (2004).

98. Gohari, I. M. et al. Characterization of *Clostridium perfringens* in the feces of adult horses and foals with acute enterocolitis. *Can. J. Vet. Res.* **78**, 1–7 (2014).
99. Diab, S. S. et al. Pathology of *Clostridium perfringens* type C enterotoxemia in horses. *Vet. Pathol.* **49**, 255–263 (2012).
100. Gohari, I. M. et al. NetF-producing *Clostridium perfringens*: clonality and plasmid pathogenicity loci analysis. *Infect. Genet. Evol.* **49**, 32–38 (2017).
101. Songer, J. G. & Uzal, F. A. Clostridial enteric infections in pigs. *J. Vet. Diagn. Invest.* **17**, 528–536 (2005).
102. Niilo, L. *Clostridium perfringens* Type C Enterotoxemia. *Can. Vet. J.* **29**, 658–664 (1988).
103. Hansen, K. et al. *Darmbrand - Enteritis Necroticans*. (George Thieme, Stuttgart 1949).
104. Sterne, M. & Warrack, G. H. The Types of *Clostridium perfringens*. *J. Pathol. Bacteriol.* **88**, 279–283 (1964).
105. Murrell, T. G., Roth, L., Egerton, J., Samels, J. & Walker, P. D. Pig-bel: enteritis necroticans. A study in diagnosis and management. *Lancet* **1**, 217–222 (1966).
106. Murrell, T. G., Egerton, J. R., Rampling, A., Samels, J. & Walker, P. D. The ecology and epidemiology of the pig-bel syndrome in man in New Guinea. *J. Hyg.* **64**, 375–396 (1966).
107. Sakurai, J. & Duncan, C. L. Some properties of beta-toxin produced by *Clostridium perfringens* type C. *Infect. Immun.* **21**, 678–680 (1978).
108. Foster, W. D. The bacteriology of necrotizing jejunitis in Uganda. *East Afr. Med. J.* **43**, 550–553 (1966).
109. Gan, K. H., Sukono, D., Satari, S., Sujudi, R. W. & Njooo, S. N. First outbreak of necrotising enteritis caused by *Clostridium welchii* in Indonesia. *Nadj Kedkt* **1**, 802–806 (1967).
110. Gui, L., Subramony, C., Fratkin, J. & Hughson, M. D. Fatal enteritis necroticans (pigbel) in a diabetic adult. *Mod. Pathol.* **15**, 66–70 (2002).
111. Know, R. & MacDonald, E. Outbreaks of foodpoisoning in certain Leicester institutions. *Med. Offr.* **69**, 21–22 (1943).
112. Scallan, E. et al. Foodborne illness acquired in the United States-major pathogens. *Emerg. Infect. Dis.* **17**, 7–15 (2011).
113. Felicio, Da. Silva et al. Risk ranking of pathogens in ready-to-eat unprocessed foods of non-animal origin (FoNAO) in the EU: initial evaluation using outbreak data (2007-2011). *Int. J. Food Microbiol.* **195**, 9–19 (2015).
114. Dolan, G. P. et al. An epidemiological review of gastrointestinal outbreaks associated with *Clostridium perfringens*, North East of England, 2012-2014. *Epidemiol. Infect.* **144**, 1386–1393 (2016).
115. DuPont, H. L. Clinical practice. Bacterial diarrhea. *N. Engl. J. Med.* **361**, 1560–1569 (2009).
116. Collie, R. E. & McClane, B. A. Evidence that the enterotoxin gene can be episomal in *Clostridium perfringens* isolates associated with non-food-borne human gastrointestinal diseases. *J. Clin. Microbiol.* **36**, 30–36 (1998).
117. Borriello, S. P. et al. Enterotoxigenic *Clostridium perfringens*: a possible cause of antibiotic-associated diarrhoea. *Lancet* **1**, 305–307 (1984).
118. Hogenauer, C., Hammer, H. F., Krejs, G. J. & Reisinger, E. C. Mechanisms and management of antibiotic-associated diarrhea. *Clin. Infect. Dis.* **27**, 702–710 (1998).
119. Brett, M. M., Rodhouse, J. C., Donovan, T. J., Tebbutt, G. M. & Hutchinson, D. N. Detection of *Clostridium perfringens* and its enterotoxin in cases of sporadic diarrhoea. *J. Clin. Pathol.* **45**, 609–611 (1992).
120. Larson, H. E. & Borriello, S. P. Infectious diarrhea due to *Clostridium perfringens*. *J. Infect. Dis.* **157**, 390–391 (1988).
121. Mpamugo, O., Donovan, T. & Brett, M. M. Enterotoxigenic *Clostridium perfringens* as a cause of sporadic cases of diarrhoea. *J. Med. Microbiol.* **43**, 442–445 (1995).
122. Wong, S. et al. Use of antibiotic and prevalence of antibiotic-associated diarrhoea in-patients with spinal cord injuries: a UK national spinal injury centre experience. *Spinal Cord.* **55**, 583–587 (2017).
123. Ayyagari, A., Agarwal, J. & Garg, A. Antibiotic associated diarrhoea: infectious causes. *Indian. J. Med. Microbiol.* **21**, 6–11 (2003).
124. Asha, N. J. & Wilcox, M. H. Laboratory diagnosis of *Clostridium perfringens* antibiotic-associated diarrhoea. *J. Med. Microbiol.* **51**, 891–894 (2002).
125. Asha, N. J., Tompkins, D. & Wilcox, M. H. Comparative analysis of prevalence, risk factors, and molecular epidemiology of antibiotic-associated diarrhea due to *Clostridium difficile*, *Clostridium perfringens*, and *Staphylococcus aureus*. *J. Clin. Microbiol.* **44**, 2785–2791 (2006).
126. Li, J. & McClane, B. A. Contributions of NanI sialidase to Caco-2 cell adherence by *Clostridium perfringens* type A and C strains causing human intestinal disease. *Infect. Immun.* **82**, 4620–4630 (2014).
127. Larcombe, S., Hutton, M. L. & Lyras, D. Involvement of bacteria other than *Clostridium difficile* in antibiotic-associated diarrhoea. *Trends Microbiol.* **24**, 463–476 (2016).
128. Pituch, H. et al. Laboratory diagnosis of antibiotic-associated diarrhea: a Polish pilot study into the clinical relevance of *Clostridium difficile* and *Clostridium perfringens* toxins. *Diagn. Microbiol. Infect. Dis.* **58**, 71–75 (2007).
129. Kim, Y. J. et al. Prevalence of *Clostridium perfringens* toxin in patients suspected of having antibiotic-associated diarrhea. *Anaerobe* **48**, 34–36 (2017).
130. 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).
131. Lim, J. C., Golden, J. M. & Ford, H. R. Pathogenesis of neonatal necrotizing enterocolitis. *Pediatr. Surg. Int.* **31**, 509–518 (2015).
132. Bisquera, J. A., Cooper, T. R. & Berseth, C. L. Impact of necrotizing enterocolitis on length of stay and hospital charges in very low birth weight infants. *Pediatrics* **109**, 423–428 (2002).
133. Berdon, W. E. et al. Necrotizing enterocolitis in the premature infant. *Radiology* **83**, 879–887 (1964).
134. Zhou, Y. et al. Longitudinal analysis of the premature infant intestinal microbiome prior to necrotizing enterocolitis: a case-control study. *PLoS ONE* **10**, e0118632 (2015).
135. Cassir, N. et al. *Clostridium butyricum* strains and dysbiosis linked to necrotizing enterocolitis in preterm neonates. *Clin. Infect. Dis.* **61**, 1107–1115 (2015).
136. 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).
137. Hanke, C. A. et al. *Clostridium perfringens* intestinal gas gangrene in a preterm newborn. *Eur. J. Pediatr. Surg.* **19**, 257–259 (2009).
138. Kindley, A. D., Rboerts, P. J. & Tulloch, W. H. Neonatal necrotising enterocolitis. *Lancet* **1**, 649 (1977).
139. Silverman, M. A., Konnikova, L. & Gerber, J. S. Impact of antibiotics on necrotizing enterocolitis and antibiotic-associated diarrhea. *Gastroenterol. Clin. North Am.* **46**, 61–76 (2017).
140. Abt, M. C., McKenney, P. T. & Pamer, E. G. *Clostridium difficile* colitis: pathogenesis and host defence. *Nat. Rev. Microbiol.* **14**, 609–620 (2016).
141. Blakey, J. L. et al. Development of gut colonisation in pre-term neonates. *J. Med. Microbiol.* **15**, 519–529 (1982).
142. Soll, R. F. & Edwards, W. H. Antibiotic use in neonatal intensive care. *Pediatrics* **135**, 928–929 (2015).
143. Alexander, V. N., Northrup, V. & Bizzarro, M. J. Antibiotic exposure in the newborn intensive care unit and the risk of necrotizing enterocolitis. *J. Pediatr.* **159**, 392–397 (2011).
144. Izumiya, H. et al. Whole-genome analysis of *Salmonella enterica* serovar Typhimurium T000240 reveals the acquisition of a genomic island involved in multidrug resistance via IS1 derivatives on the chromosome. *Antimicrob. Agents Chemother.* **55**, 623–630 (2011).
145. Ramos, P. I. et al. Comparative analysis of the complete genome of KPC-2-producing *Klebsiella pneumoniae* Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. *BMC Genom.* **15**, 54 (2014).
146. 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).
147. Sim, K. *Defining the gastrointestinal microbiota in premature neonates: Its development and relation to necrotizing enterocolitis* Doctor of Philosophy (PhD) thesis, Imperial College London, (2015).
148. Nagler, F. P. O. Observations on a reaction between the lethal toxin of Cl. welchii (Type A) and human serum. *Br. J. Exp. Pathol.* **20**, 473–485 (1939).
149. Hansen, M. & Elliott, L. New presumptive identification test for *Clostridium perfringens*: reverese CAMP test. *J. Clin. Microbiol.* **12**, 617–619 (1980).
150. van Asten, A. J., van der Wiel, C. W., Nikolaou, G., Houwers, D. J. & Grone, A. A multiplex PCR for toxin typing of *Clostridium perfringens* isolates. *Vet. Microbiol.* **136**, 411–412 (2009).

frontiers in Microbiology

GENERAL COMMENTARY
published: 16 August 2018
doi: 10.3389/fmicb.2018.01857

Response: Commentary: Probing Genomic Aspects of the Multi-Host Pathogen *Clostridium perfringens* Reveals Significant Pangenome Diversity, and a Diverse Array of Virulence Factors

Raymond Kiu[1,2] *and Lindsay J. Hall*[1]*

[1] Gut Microbes and Health Programme, Quadram Institute Bioscience, Norwich Research Park, Norwich, United Kingdom,
[2] Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, United Kingdom

Keywords: pangenome, antimicrobial resistance, *Clostridium perfringens*, genomics, whole genome sequencing, Clostridial infection, exotoxins, pathogen

OPEN ACCESS

Edited by:
John R. Battista,
Louisiana State University,
United States

Reviewed by:
Karl Hassan,
University of Newcastle, Australia

***Correspondence:**
Lindsay J. Hall
Lindsay.Hall@quadram.ac.uk

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

Received: *02 June 2018*
Accepted: *24 July 2018*
Published: *16 August 2018*

Citation:
Kiu R and Hall LJ (2018) Response: Commentary: Probing Genomic Aspects of the Multi-Host Pathogen Clostridium perfringens Reveals Significant Pangenome Diversity, and a Diverse Array of Virulence Factors. Front. Microbiol. 9:1857. doi: 10.3389/fmicb.2018.01857

A Commentary on

Commentary: Probing Genomic Aspects of the Multi-Host Pathogen *Clostridium perfringens* Reveals Significant Pangenome Diversity, and a Diverse Array of Virulence Factors
by Gohari, I. M., and Prescott, J. F. (2018). Front. Microbiol. 9:1856. doi: 10.3389/fmicb.2018.01856

We firstly would like to express our appreciation for the comments and efforts made by Gohari and Prescott regarding their detailed commentary of our *Clostridium perfringens* Whole Genome Sequencing (WGS) based genomic and phylogenetic study we published in December 2017. We are particularly pleased to note that the authors acknowledge that our "work has contributed significantly to understanding of genomic diversity of this bacterium."

The aim of our study was to conduct a robust large-scale WGS study on this important pathogen, and to highlight genomic insights to the global scientific community. We believe that WGS could prove extremely useful in exploring new traits in *C. perfringens*, as has been performed in many different bacterial pathogens. We agree with the authors, and indeed highlight this throughout our original study, that alongside genomic-based studies, clinical metadata, epidemiological studies, and phenotypic testing, will be central to determine the impact of genetic variation in *C. perfringens* in the context of human and animal health.

We would first like to respond to the authors commentary suggesting a "clear mistake" or "incorrect conclusion" in that *netB* gene (a gene that encode NetB toxin, which is associated with avian Necrotising Enteritis) has no role in canine haemorrhagic gastroenteritis and foal necrotising enteritis (Keyburn et al., 2008, 2010; Rood et al., 2016). We appreciate the authors have a track record of working in this particular area of *C. perfringens* virulence, as highlighted by citation of their own work throughout the commentary. They highlight that *netB* and *netE* genes (which encodes proteins NetB and NetE respectively) share high sequence identity of 78% amino acid (according to their published work in 2015, 79% amino acid sequence identity; Mehdizadeh Gohari et al., 2015), which may have potentially contributed to our "understandable misinterpretation" of the data. Based on our informatics filtering parameters, we also determined that *netB* and *netE* are highly identical and not distinctive

A

```
Sequences producing significant alignments:                          Score     E
                                                                     (Bits)  Value

  etx                                                                 1818   0.0


51607  atgaaaaaaaaTCTTGTAAAAAGTTTAGCAATCGCATCAGCGGTGATATCCATCTATTCA  51666
1      ............................................................  60

51667  ATAGTTAATATTGTTTCACCAACTAATGTAATAGCTAAGGAAATATCTAATACAGTATCT  51726
61     ............................................................  120

51727  AATGAAATGTCCAAAAAAGCTTCTTATGATAATGTAGATACATTAATTGAGAAAGGAAGA  51786
121    ............................................................  180

51787  TATAATACAAAATATAATTACTTAAAGAGAATGGAAAAATATTATCCTAATGCTATGGCA  51846
181    ............................................................  240

51847  TATTTTGATAAGGTTACTATAAATCCACAAGGAAATGATTTTTATATTAATAATCCTAAA  51906
241    ............................................................  300

51907  GTTGAATTAGATGGAGAACCATCAATGAATTATCTTGAAGATGTTTATGTTGGAAAAGCT  51966
301    ............................................................  360

51967  CTCTTAACTAATGATACTCAACAAGAACAAAAATTAAAATCACAATCATTCACTTGTAAA  52026
361    ............................................................  420

52027  AATACTGATACAGTAACTGCAACTACTACTCATACTGTGGGAACTTCGATACAAGCAACT  52086
421    ............................................................  480

52087  GCTAAGTTTACTGTTCCTTTTAATGAAACAGGAGTATCATTAACTACTAGTTATAGTTTT  52146
481    ............................................................  540

52147  GCAAATACAAATACAAATACTAATTCAAAAGAAATTACTCATAATGTCCCTTCACAAGAT  52206
541    ............................................................  600

52207  ATACTAGTACCAGCTAATACTACTGTAGAAGTAATAGCATATTTaaaaaaaGTTAATGTT  52266
601    ............................................................  660

52267  AAAGGAAATGTAAAGTTAGTAGGACAAGTAAGTGGAAGTGAATGGGGAGAGATACCTAGT  52326
661    ............................................................  720

52327  TATTTAGCTTTTCCTAGGGATGGTTATAAATTTAGTTTATCGGATACAGTAAATAAGAGT  52386
721    ..........................................A.................  780

52387  GATTTAAATGAAGATGGTACTATTAATATTAATGGAAAAGGAAATTATAGTGCAGTTATG  52446
781    ............................................................  840

52447  GGAGATGAGTTAATAGTTAAGGTTAGAAATTTAAATACAAATAATGTACAAGAATATGTA  52506
841    ............................................................  900

52507  ATACCTGTAGATaaaaaagaaaaaaGTAATGATTCAAATATAGTAAAATATAGGAGTCTT  52566
901    ............................................................  960

52567  TCTATTAAGGCACCAGGAATAAAATAA  52593
961    .........................    987
```

B

CONTIGS	START	END	GENE	COVERAGE	COVERAGE_MAP	GAPS	%COVERAGE	%IDENTITY
Processing:								
Found 8 genes in								
NCTC8503.1	35701	39015	colA	1-3315/3315	==========	0/0	100.00	98.79
NCTC8503.1	51002	52504	pfoA	1-1503/1503	==========	0/0	100.00	99.20
NCTC8503.2	55060	56628	ccp	1-1569/1569	==========	0/0	100.00	97.58
NCTC8503.2	146118	147511	nanH	73-1463/1464	========/======	1/3	95.01	92.68
NCTC8503.4	11085	15851	nanJ	1-4767/4767	==========	0/0	100.00	98.72
NCTC8503.4	1635483	1636679	plc	1-1197/1197	==========	0/0	100.00	98.75
NCTC8503.5	51607	52593	etx	1-987/987	==========	0/0	100.00	99.90
NCTC8503.8	84742	86826	nanI	1-2085/2085	==========	0/0	100.00	98.71

FIGURE 1 | (A) A BLASTn sequence alignment performed on a local machine that aligns epsilon toxin gene *etx* (NCBI accession: M95206.1) with PacBio-sequenced NCTC8503 previously assembled genome (ENA accession: SAMEA3879480; contig 5: nucleotide position 51607-52593). Only a point mutation at position 762 is detected with 100% coverage and > 99% sequence identity. **(B)** Toxin profile of NCTC8503 isolate, performed via sequence similarity search ("best-match" approach) pipeline ABRicate (https://github.com/tseemann/abricate; BLASTn-based tool) that confirmed the presence of *etx* gene in NCTC8503 genome with 99.90% identity (at 100% coverage).

at nucleotide sequence level. This was based on strict double-filtering strategy at 80% identity and E-value of 10^{-20}, (routinely applied in WGS studies to infer identical genes; Pearson, 2013; Kiu et al., 2017), and was used as we were undertaking a global *in silico*-based approach to explore a significant number of *C. perfringens* genomes for virulence-associated traits (not solely focused on these two toxin genes). Consequently, we have reanalysed the data and determined that a higher sequence

identity threshold at 90% (BLASTn) confirms an absence of the *netB* gene in these NetF-associated genomes. We thank the authors for highlighting the high sequence similarly of these toxin genes, and will factor these parameters in for future studies (Camacho et al., 2009). As we carried out a purely bioinformatic-based study, and are aware of the sensitivity/specificity of different computational pipelines and parameters, we were careful not to come to any definitive "conclusions," and "suggested" that NetB toxin "might be involved," thus clarifying our discussion points. Notably, with the advancement in bioinformatics tools, differentiation between highly similar genes may be possible using a "best-match" approach to avoid inaccurate annotations.

To address the question as to whether or not NCTC8503 is an "NE isolate," we have re-traced the source of this isolate and its sequence data. NCTC8503 is a *C. perfringens* strain isolated by Bennetts (1930) according to the history record shown on the Public Health England NCTC official website[1] Nairn and Bamford (1967) state "Bennetts (1930) had earlier reported the isolation of *Bacillus welchii* from a bowel lesion in a Black Orpington pullet with intestinal coccidiosis, and he considered that an enterotoxaemia had contributed to the bird's death"; *B. welchii* was later renamed as *C. perfringens*. Furthermore, Williams et al. states "A frequent, although sporadic, poultry clostridiosis (necrotic enteritis [NE]) was first recorded by Bennetts (1930) in Australia," thus these descriptions indicated a likely link to strain NCTC8503, a *C. perfringens* strain isolated in 1930. We agree that no type D *C. perfringens* strains have been linked with poultry NE. Unfortunately, as NCTC8503 is a historical isolate with no detailed source information recorded, we have been unable to definitively confirm the isolate's origin (Bennetts, 1930; Williams, 2005). However, we can reaffirm that this isolate has not lost the epsilon toxin plasmid during laboratory passage as confirmed by both WGS (**Figure 1**), and multiplex PCR toxinotyping (Baums et al., 2004; Kiu et al., 2017).

The "extreme" level of pangenome (12.6% core genes) in *C. perfringens* has not been widely reported to date (McInerney et al., 2017). However, as stated by the authors a previous study carried out in 2010 in *Escherichia coli,* which analysed 53 genomes, indicated 11% core genes (Lukjancenko et al., 2010). Notably, although these 2 bacteria can colonise the GI tract, they are fundamentally different regarding their oxygen sensitivity and ability to form spores. As described by Lukjanenko et al., analysis of 53 genomes (not 61 genomes as claimed) identified 1,472 core genes out of 13,296 genes in the pangenome (11%), which is lower than the 12.6% reported in our genomic *C. perfringens* analysis, although we note some differences in approaches used. More recently a pangenome study based on 228 *E. coli* genomes identified 23.8% core genes (2,722 core genes in 11,401 gene families) (McNally et al., 2016), and using web-based tool panX pangenome analysis[2] and databases (Ding et al., 2018), based on the same cutoff as our study indicated 13% of core genes, based on analysis of 307 *E. coli* genomes (as of May 2018). We reiterate that our "extreme" observation for spore-forming Gram-positive *C. perfringens* is a rare trait, and is therefore of interest to the wider research community. We agree with their statement "describing a species as having extreme variation depends very much to what it is being compared," and furthermore we would emphasise that prediction of genetic diversity will likely vary due to different factors including sampling bias, number of strains selected, and parameters used during informatics analysis, thus impacting diversity measurements. In this study we compared *C. perfringens* with *Clostridium difficile* (30.3%, which has been reclassified as *Clostridioides difficile*), a closer relative of *C. perfringens*, and *Streptococcus pneumoniae* (46.5%), *Salmonella enterica* (16%), and *Klebsiella pneumoniae* (26%) (Lawson et al., 2016; Kiu et al., 2017).

We agree that there is currently no definitive pathogenicity link with *C. perfringens* prophages, although previous studies have indicated they enhance sporulation, which would be expected to enhance transmission, and can be viewed as a virulence trait (Stewart and Johnson, 1977). Notably, in the closely related pathogen *C. difficile*, bacteriophages are linked to toxin-secretion (Goh et al., 2005). Consequently, we speculate that *C. perfringens* phages contribute to their virulence, which could be confirmed in future studies.

We hope that our commentary provides clarification and context and we look forward to wider discussion with all investigators in the *C. perfringens* field.

AUTHOR CONTRIBUTIONS

RK and LH co-wrote the manuscript. RK performed the bioinformatics analysis.

FUNDING

Funding by a Wellcome Trust Investigator Award (100974/C/13/Z), and the Biotechnology and Biological Sciences Research Council (BBSRC); Institute Strategic Programme Gut Microbes and Health BB/R012490/1, and Institute Strategic Programme Gut Health and Food Safety BB/J004529/1 to LH. RK is a recipient of University of East Anglia International Bursary (Faculty of Medical and Health Sciences).

[1]https://www.phe-culturecollections.org.uk/

[2]http://pangenome.tuebingen.mpg.de/

REFERENCES

Baums, C. G., Schotte, U., Amtsberg, G., and Goethe, R. (2004). Diagnostic multiplex PCR for toxin genotyping of *Clostridium perfringens* isolates. *Vet. Microbiol.* 100, 11–16. doi: 10.1016/S0378-1135(03)00126-3

Bennetts, H. W. (1930). *Bacillus welchii* and bowel lesions. With special reference to a case of coccidiosis. *Aust. Vet. J.* 6, 153–154.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Ding, W., Baumdicker, F., and Neher, R. A. (2018). panX: pan-genome analysis and exploration. *Nucleic Acids Res.* 46, e5. doi: 10.1093/nar/gkx977

Goh, S., Chang, B. J., and Riley, T. V. (2005). Effect of phage infection on toxin production by *Clostridium difficile*. *J. Med. Microbiol.* 54(Pt 2), 129–135. doi: 10.1099/jmm.0.45821-0

Keyburn, A. L., Bannam, T. L., Moore, R. J., and Rood, J. I. (2010). NetB, a pore-forming toxin from necrotic enteritis strains of *Clostridium perfringens*. *Toxins* 2, 1913–1927. doi: 10.3390/toxins2071913

Keyburn, A. L., Boyce, J. D., Vaz, P., Bannam, T. L., Ford, M. E., Parker, D., et al. (2008). NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. *PLoS Pathog.* 4:e26. doi: 10.1371/journal.ppat.0040026

Kiu, R., Caim, S., Alexander, S., Pachori, P., and Hall, L. J. (2017). Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant pangenome diversity, and a diverse array of virulence factors. *Front. Microbiol.* 8:2485. doi: 10.3389/fmicb.2017.02485

Lawson, P. A., Citron, D. M., Tyrrell, K. L., and Finegold, S. M. (2016). Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prevot 1938. *Anaerobe* 40, 95–99. doi: 10.1016/j.anaerobe.2016.06.008

Lukjancenko, O., Wassenaar, T. M., and Ussery, D. W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* 60, 708–720. doi: 10.1007/s00248-010-9717-3

McInerney, J. O., McNally, A., and O'Connell, M. J. (2017). Why prokaryotes have pangenomes. *Nat. Microbiol.* 2:17040. doi: 10.1038/nmicrobiol.2017.40

McNally, A., Oren, Y., Kelly, D., Pascoe, B., Dunn, S., Sreecharan, T., et al. (2016). Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. *PLoS Genet.* 12:e1006280. doi: 10.1371/journal.pgen.1006280

Mehdizadeh Gohari, I., Parreira, V. R., Nowell, V. J., Nicholson, V. M., Oliphant, K., and Prescott, J. F. (2015). A novel pore-forming toxin in type A *Clostridium perfringens* is associated with both fatal canine hemorrhagic gastroenteritis and fatal foal necrotizing enterocolitis. *PLoS ONE* 10:e0122684. doi: 10.1371/journal.pone.0122684

Nairn, M. E., and Bamford, V. W. (1967). Necrotic enteritis of broiler chickens in western Australia. *Aust. Vet. J.* 43, 49–54.

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinformatics* Chapter 3, Unit3 1. doi: 10.1002/0471250953.bi0301s42

Rood, J. I., Keyburn, A. L., and Moore, R. J. (2016). NetB and necrotic enteritis: the hole movable story. *Avian Pathol.* 45, 295–301. doi: 10.1080/03079457.2016.1158781

Stewart, A. W., and Johnson, M. G. (1977). Increased numbers of heat-resistant spores produced by two strains of *Clostridium perfringens* bearing temperate phage s9. *J. Gen. Microbiol.* 103, 45–50. doi: 10.1099/00221287-103-1-45

Williams, R. B. (2005). Intercurrent coccidiosis and necrotic enteritis of chickens: rational, integrated disease management by maintenance of gut integrity. *Avian Pathol.* 34, 159–180. doi: 10.1080/03079450500112195

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