Aus dem
Friedrich-Loeffler-Institut,
Bundesforschungsinstitut für Tiergesundheit,
Institut für Bakterielle Infektionen und Zoonosen, Jena

eingereicht über das
Institut für Mikrobiologie und Tierseuchen
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

In silico genome analysis and molecular typing of *Clostridium perfringens*

Inaugural-Dissertation
zur Erlangung des Grades eines
Doktors der Veterinärmedizin
an der
Freien Universität Berlin

vorgelegt von
Mostafa Y. Abdel-Glil
Tierarzt aus Dakahliya, Ägypten

Berlin 2019
Journal-Nr.: 4166

mbvberlin | mensch und buch verlag

Aus dem

Friedrich-Loeffler-Institut,
Bundesforschungsinstitut für Tiergesundheit,
Institut für Bakterielle Infektionen und Zoonosen,
Jena

eingereicht über das
Institut für Mikrobiologie und Tierseuchen
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

In silico genome analysis and molecular typing of *Clostridium perfringens*

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors der Veterinärmedizin
an der Freien Universität Berlin

vorgelegt von
Mostafa Y. Abdel-Glil
Tierarzt aus Dakahliya, Ägypten

Berlin 2019
Journal-Nr.: 4166

Gedruckt mit Genehmigung des Fachbereichs Veterinärmedizin

der Freien Universität Berlin

Dekan:	Univ.-Prof. Dr. Jürgen Zentek
Erster Gutachter:	Prof. Dr. Lothar H. Wieler
Zweiter Gutachter:	Prof. Dr. Heinrich Neubauer
Dritter Gutachter:	Univ.-Prof. Dr. Thomas Alter

Deskriptoren (nach CAB-Thesaurus):
Clostridium perfringens; genomes; genome analysis; phylogenetics; polymerase chain reaction

Tag der Promotion: 03.12.2019

Bibliografische Information der *Deutschen Nationalbibliothek*
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://dnb.de> abrufbar.

ISBN: 978-3-96729-030-1
Zugl.: Berlin, Freie Univ., Diss., 2019
Dissertation, Freie Universität Berlin
D188

Choriner Str. 85 - 10119 Berlin
verlag@menschundbuch.de – www.menschundbuch.de

Printed with the support of the German Academic Exchange Service

Table of contents

Tables

Figures

Abbreviations

AFLP	Amplified Fragment Length Polymorphism
AGP	Antimicrobial Growth Promoter
ANI	Average Nucleotide Identity
AP-PCR	Arbitrarily Primed PCR
BEC	Binary Enterotoxin of *C. perfringens*
BLAST	Basic Local Alignment Search Tool
C.	*Clostridium*
CDC	Cholesterol-Dependent Cytolysin
CDS	Coding Sequence
cgMLST	Core Genome-based Multi-locus Sequence Typing
CH	CPE Hexamer
COG	Clusters of Orthologous Group
CPA	*C. perfringens* Alpha toxin
CPB	*C. perfringens* Beta toxin
CPB2	*C. perfringens* Beta2 toxin
CPE	*C. perfringens* Enterotoxin
CPILE	*C. perfringens* Iota-like toxin
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CT	Cluster Type
DNA	Deoxyribonucleic Acid
DLV	Double Locus Variant
DP	Discriminatory Power
dsDNA	Double Stranded DNA
ECMM	Extracellular Matrix Molecule
ELISA	Enzyme-Linked Immunosorbent Assay
ETX	Epsilon Toxin
GC	Guanosine – Cytosine
GI	Genomic Island
GO	Gene Ontology
GTR	General Time Reversible
HGAP	Hierarchical Genome Assembly Process
INDEL	Insertion or Deletion
IS	Insertion Sequence
ITX	Iota Toxin
Kbp	Kilo Basepair
LCB	Locally Collinear Block
Mbp	Mega Basepair
MGE	Mobile genetic element
ML	Maximum Likelihood
MLST	Multi-locus Sequence Typing
MLVA	Multiple-locus Variable Number Tandem Repeat Analysis
MMPs	Matrix Metalloproteinases
MST	Minimum Spanning Tree
NCTC	National Collection of Type Cultures, UK

NE	Necrotic Enteritis
NetB	Necrotic Enteritis Toxin Beta-like
NJ	Neighbour Joining
PacBio	Pacific Biosciences
PC	Phosphatidylcholine
PCR	Polymerase Chain Reaction
PFGE	Pulse Field Gel Electrophoresis
PFO	Perfringolysin O
PLA	Phospholipase
QUAST	QUality ASsessment Tool
RAPD	Randomly Amplified Polymorphic DNA
RAST	Rapid Annotation using the Subsystem Technology
REP-PCR	Repetitive Sequencing-based PCR
rRNA	Ribosomal RNA
SLV	Single Locus Variant
SM	Sphingomyelin
SNP	Single Nucleotide Polymorphism
ST	Sequence Type
TCP	transfer of clostridial plasmid
TE	Typing Efficiency
TLV	Triple-Locus Variant
TSC	Tryptose-Sulphite-Cycloserine
UV	Ultraviolet
VFDB	Virulence Factor Database

Chapter 1: Introduction

1.1. *Clostridium perfringens*: a pathogen of humans and animals

1.1.1 *Clostridium perfringens* (history, characteristics and types)

Clostridium (C.) perfringens was first isolated in 1892 by Welch and Nuttall at the Johns Hopkins Hospital (Baltimore, MD, USA) from a fatal aortic aneurism (Rood et al., 2018; Welch and Nuttall, 1892). It was initially named by Welch and Nuttall as *Bacillus aerogenes capsulatus* (also known as *Bacillus welchii* and *Bacterium welchii*) (Rood et al., 2018). In 1898, Veillon and Zuber reported the isolation of the bacterium as *Bacillus perfringens* (Rainey et al., 2009; Rood et al., 2018). However, by 1929 the name was changed to *Clostridium welchii* (Lucey and Hutchins, 2004) and later to *Clostridium perfringens*, where *Clostridium* means spindle-like and *perfringens* means breaking through (Rainey et al., 2009).

C. perfringens is a Gram-positive, rod-shaped, spore forming anaerobe, which is known for its toxin-production capability. It is widely distributed in the environment i.e. sewage and soil, but can be also part of the intestinal flora of humans and animals (Kiu and Hall, 2018). *C. perfringens* is a member of the genus *Clostridium* cluster I (*Clostridium* sensu stricto*),* based on the 16S rRNA gene sequence (Rainey et al., 2009). Species (~ 77 species) within this cluster are considered true representatives of the genus *Clostridium* that are phylogenetically distinct from other clostridiae and maintain unique molecular characteristics (Gupta and Gao, 2009). *C. perfringens* is a highly versatile bacterium that is able to inhabit different ecosystems and to grow over a wide temperature range. The spores of some *C. perfringens* strains such as human food-poisoning- and human enteritis necroticans-causing strains are known for their high thermal resistance (Robertson et al., 2014; Zeissler et al., 1949). The optimum temperature for the growth of food poisoning *C. perfringens* strains is between 43°C and 45°C, at which the organism has one of the fastest known growth rates with an extremely short generation time (Labbé, 2003; Li and McClane, 2006b; Robertson et al., 2014). *C. perfringens* is described as a non-motile bacterium due to the absence of flagella. However, a type IV pili-dependent gliding motility was described in some strains (Liu et al., 2014; Varga et al., 2006). Under anaerobic conditions, *C. perfringens* grows well on blood agar with smooth, circular, grayish colonies which are usually surrounded by a double zone of hemolysis: an inner zone of complete hemolysis due to the perfringolysin toxin and an outer zone of incomplete hemolysis due to the alpha toxin (Goossens et al., 2016; Vidal et al., 2015). However, non-hemolytic *C. perfringens* were also described (Schoepe et al., 2001).

Genome sequencing of *C. perfringens* revealed its lack of many enzyme genes, which are required for aminoacid biosynthesis (Myers et al., 2006; Shimizu et al., 2002). The bacterium secretes numerous hydrolytic enzymes and toxins that act extracellularly on organic materials,

mobilising essential nutrients and energy sources, which are required for its growth (Shimizu et al., 2002). *C. perfringens* produces more than 20 extracellular enzymes and toxins, which are considered virulence elements (Kiu and Hall, 2018; Popoff, 2014). Most of these toxins disrupt the host cellular membranes liberating the cellular nutrient sources. Others degrade complex molecules into simple forms for utilisation (Popoff and Bouvet, 2009).

C. perfringens is categorised into five toxinotypes (A to E) according to the presence of four major toxins: alpha, beta, epsilon and iota. The typing scheme was initially developed in 1931 by Wilsdon (as stated in Rood et al., 2018). Recently, an expanded typing scheme was proposed defining the new types F and G to classify isolates that produce the *C. perfringens* enterotoxin (CPE) and the Beta-like necrotic enteritis toxin (NetB), respectively (Table 1) (Rood et al., 2018). ELISAs and more recently molecular identification for toxin genes have been developed and used for easy and fast differentiation (Baums et al., 2004; el Idrissi and Ward, 1992; Nagahama et al., 1991; Yoo et al., 1997). Recent studies report toxin-genotypes of *C. perfringens* based on the presence of the genes encoding the “typing toxins” (Drigo et al., 2010; Fohler et al., 2016; Rood et al., 2018).

Table 1: *C. perfringens* toxin types (Rood et al., 2018).

Typing toxin	Toxin type						
	A	B	C	D	E	F	G
PLC (Phospholipase C or alpha toxin)	+	+	+	+	+	+	+
CPB (*C. perfringens* beta toxin)	-	+	+	-	-	-	-
ETX (Epsilon toxin)	-	+	-	+	-	-	-
ITX (*C. perfringens* iota toxin)	-	-	-	-	+	-	-
CPE (*C. perfringens* enterotoxin)	-	+/-	+/-	+/-	+/-	+	-
NetB (Beta-like necrotic enteritis toxin)	-	-	-	-	-	-	+

1.1.2 *C. perfringens* as a pathogen (types and diseases)

C. perfringens causes various enteric and histotoxic infections in humans and animals. Most of the diseases are mediated by one or more extracellular toxins (Kiu and Hall, 2018) (Table 2). Of the five classical toxin types, *C. perfringens* type A (including newly defined types F and G) is the most abundant and versatile group.

1.1.2.1 *C. perfringens* type A

1.1.2.1.1 Histotoxic infections

Type A strains cause numerous infections in various hosts (Songer, 1996). They are involved in cases of human and also animal histotoxic infections, in which the alpha toxin is believed to be the key virulence factor, while perfringolysin is thought to act in synergism thereby causing progressive (muscle) tissue damage (Awad et al., 1995). Clostridial gas gangrene (histotoxic

myonecrosis) is a disease of low incidence, but high lethality in both humans and animals, where *C. perfringens* is the most commonly recovered bacterium in humans (Uzal et al., 2016c; Weinstein and Barza 1973). A contamination by *C. perfringens* spores or vegetative cells after an extensive traumatic damage is required for the disease development (Bryant et al., 2015). The traumatic damage is essential for the disruption of the blood supply and the induction of an anaerobic niche for clostridial colonization (Bryant et al., 2015). The site of infection rapidly exhibit progressive destruction of the tissue with a foul-smelling necrotic area containing serosanguinous discharge and gas bubbles (Bryant et al., 2015). Systemic absorption of the toxins leads to shock, multi-organ failure and death (Stevens et al., 2012).

1.1.2.1.2 Enteric infections

Enteric diseases in humans and animals are reported in association with type A strains (see also type F [1.1.2.6] and type G [1.1.2.7]). Three new toxins (NetE, NetF and NetG) have recently been described within the type A strain group in association with fatal hemorrhagic gastroenteritis in dogs and fatal necrotizing enterocolitis in horses (Mehdizadeh Gohari et al., 2015). A new toxin called binary enterotoxin of *C. perfringens* (BEC) was reported in few cases of human acute gastroenteritis (Yonogi et al., 2014). Additionally, beta2 toxin is a pore-forming toxin that was reported in association with enteric diseases in different animals such as porcine enteritis and gentamycin-associated equine enteritis. Beta2 toxin can be found in different toxin types (Gibert et al., 1997).

1.1.2.2 *C. perfringens* type B

C. perfringens type B strains essentially carry the genes of two plasmid-borne major toxins (beta and epsilon toxin). Both toxins are suggested to contribute to the pathogenesis of type B-associated enterotoxaemia, which is characterized by diarrhea, neurological disorders and/or sudden death (Fernandez-Miyakawa et al., 2007). Different animal species but primarily sheep are susceptible to diseases caused by type B strains so far reported in Europe, South Africa and the Middle East (Li et al., 2013; Sayeed et al., 2010). Young lambs (1- 14 days of age) develop a condition known as “lamb dysentery” which is caused by type B infection and characterized by acute abdominal pain, distended abdomen and hemorrhagic diarrhea (Uzal et al., 2010). Epsilon toxin in type B strains is thought to induce brain lesions and neurological signs, which include opisthotonus, blindness and lack of coordination (Uzal et al., 2010). In addition to beta and epsilon toxin genes, type B isolates can possess additional plasmid-borne toxin genes e.g., genes encoding urease enzymes, lambda toxin and TpeL toxin (Li et al., 2013).

1.1.2.3 *C. perfringens* type C

C. perfringens type C can infect humans and animals. Type C strains primarily produce beta toxin (CPB) but also express other plasmid-encoded toxins e.g. enterotoxin, beta2 and TpeL but not the epsilon toxin (Rood et al., 2018). Diseases in animals include hemorrhagic necrotizing enteritis in lambs, piglets, calves and foals (Uzal and McClane, 2011). Newborn animals are typically most susceptible especially piglets (Nagahama et al., 2015; Niilo, 1988; Songer, 1996; Uzal et al., 2014). The clinical signs and pathology were reported to be similar in different animal species e.g. piglets, calves and lambs (Uzal and McClane, 2011). Adult sheep, however, develop a severe form of enterotoxaemia also known as “struck” because of the fulminant disease progression leading to the death of the animal (Sayeed et al., 2008; Uzal and McClane, 2011). *C. perfringens* type C disease in animals is mediated by the beta toxin, it was found that the *cpb* null mutant type C strain (CN3685) is avirulent in animal models. The intestinal virulence was restored upon *cpb* complementation (Garcia et al., 2012; Sayeed et al., 2008; Uzal et al., 2014; Uzal et al., 2009). CPB affects the epithelium of the jejunum and ileum (Sayeed et al., 2008; Uzal and McClane, 2011). The lesions are characterized by hemorrhagic necrosis that primarily affects the intestinal mucosa, then progresses deeply to involve all intestinal layers (Uzal and McClane, 2011). In humans, type C strains can cause enteritis necroticans known as “Pigbel” or “Darmbrand”, a life-threatening illness, which is characterized by hemorrhagic, inflammatory or ischemic necrosis of the jejunum associated with abdominal pain, vomiting and severe bloody diarrhea (Murrell and Walker, 1991; Nagahama et al., 2015). In enteritis necroticans strains, the beta toxin and the enterotoxin were found to act in synergism and contribute to the enteric lesions (Ma et al., 2014).

1.1.2.4 *C. perfringens* type D

The main factor in the *C. perfringens* type D diseases is the epsilon toxin, a highly potent clostridial toxin (Alves et al., 2014; Garcia et al., 2013). Diseases due to type D strains can be observed mainly in sheep and goats where the disease is characterized by sudden death or neurological and respiratory signs with infrequent intestinal lesions and diarrhea (Uzal et al., 2010). Diseases due to type D strains are among the most common clostridial diseases in sheep and goats worldwide and are sometimes referred to as “pulpy kidney” (Uzal and Songer, 2008; Uzal et al., 2016b; Uzal et al., 2010). Neurological microscopic damages are commonly observed in sheep and considered as pathognomonic for *C. perfringens* type D infection including perivascular edema, degeneration and necrosis of brain parenchyma (Uzal and Songer, 2008). Young goats can develop acute disease with sudden death similar to enterotoxaemia in sheep (Uzal, 2004; Uzal et al., 2016b). The subacute disease can, however, be observed more frequently in adult goats, and is characterized by hemorrhagic diarrhea, abdominal discomfort and intestinal damages with fibrinonecrotic (pseudomembranous)

inflammation (Uzal et al., 2004). Spontaneous type D enterotoxaemia in calves has also been described (Watson and Scholes, 2009) and cattle experimentally challenged with type D strains developed enterotoxaemia lesions similar to those in sheep (Filho et al., 2009).

1.1.2.5 *C. perfringens* type E

C. perfringens type E produces iota toxin, a clostridial binary toxin, which is encoded by two plasmid genes (Sakurai et al., 2009). Only limited information on the role of *C. perfringens* type E in animal disease is available. Type E strains are frequently isolated from cases of hemorrhagic enteritis and sudden death in neonatal calves and are infrequently detected in lambs with enterotoxaemia (Redondo et al., 2013; Songer and Miskimmins, 2004). Type E infections are generally characterized by diarrhea and intestinal hemorrhage (Songer, 2016). Rabbit enterotoxaemia is also linked to type E strains (Songer, 2016).

1.1.2.6 *C. perfringens* type F

C. perfringens type F was recently proposed to describe type A strains that produce enterotoxin (Rood et al., 2018). Type F strains are involved in cases of human food poisoning, one of the most commonly reported bacterial foodborne illnesses in the US with nearly 1 million cases per annum (Scallan et al., 2011). The disease is relatively mild and self-limiting with a short incubation period followed usually by recovery within 24 h (Shrestha et al., 2018). Fatalities can be observed in the elderly or debilitated individuals (Bos et al., 2005). This enterotoxin is also the main pathogenicity factor in cases of *C. perfringens* foodborne illness in humans (Freedman et al., 2016). Cases of non-foodborne gastrointestinal disease in humans are believed to involve also the enterotoxin (Lindström et al., 2011) e.g. antibiotic associated diarrhea. Enterotoxin positive strains were additionally found in cases of enteric infections in different animal species (Marks et al., 2002; Miyakawa et al., 2007). Of interest, the enterotoxin is not limited to type F but can also be detected in other toxin types (Freedman et al., 2016; Rood et al., 2018) (Table 1).

1.1.2.7 *C. perfringens* type G

C. perfringens type G includes strains that produce NetB toxin (Rood et al., 2018) and are found mainly in poultry necrotic enteritis disease (1.2.4.1). NetB is reported to act as the key toxin in the disease process, yet not all prevalence studies support this association (Keyburn et al., 2010b; Martin and Smyth, 2009; Park et al., 2015).

Table 2: Diseases caused by different toxin types of *C. perfringens*

Toxin type	Subtypes (Genotype)*	Diseases in human and animal	Reference (pathogenicity / prevalence study)
A	*cpa, pfoA*	• Gas gangrene (human and animal) • Bovine necro-hemorrhagic enteritis in calves	(Awad et al., 1995; Verherstraeten et al., 2013)
	*cpa, bec*** *(or cpile)*	• Human food poisoning (acute gastroenteritis)	(Yonogi et al., 2014)
	cpa, cpb2	• Piglet (necrotic) enteritis • Equine enteritis (gentamycin associated)	(Vilei et al., 2005)
	cpa, netE, netF and *netG*	• Canine hemorrhagic gastroenteritis • Foal necrotizing enterocolitis	(Mehdizadeh Gohari et al., 2015)
B	*cpa, cpb, etx*	• Necrotizing enteritis and enterotoxaemia in sheep (lamb dysentery), cattle, and horses	(Songer, 1996)
C	*cpa, cpb*	• Hemorrhagic necrotizing enteritis and enterotoxaemia in lambs, piglets, calves and foals. • Acute enterotoxaemia (struck) in adult sheep and other animal species.	(Sayeed et al., 2008)
	cpa, cpb, cpe	• Human enteritis necroticans (“Pigbel”, “Darmbrand”)	(Ma et al., 2012)
D	*cpa, etx*	• Enterotoxaemia of sheep and goats	(Garcia et al., 2013)
E	*cpa, itx*	• Enterotoxaemia in calves, rabbits and lambs • Possible enteritis in different animal species	(Redondo et al., 2013; Songer and Miskimmins, 2004)
F	*cpa, cpe*	• Human food poisoning • Antibiotic associated diarrhea • Sporadic non-foodborne illness	(Sarker et al., 1999)
G	*cpa, netB*	• Necrotic enteritis in poultry	(Keyburn et al., 2008)

*The chromosome encoded toxin genes such as collagenase, sialidases, and clostripain are usually detected in all *C. perfringens* strains. Beta 2, enterotoxin or TpeL have also been reported in different toxin types.
** Binary enterotoxin of *C. perfringens*

1.1.3 *C. perfringens* toxins

The following paragraphs describe the main features of the traditional typing toxins (alpha, beta, epsilon and iota) as well as perfringolysin O and enterotoxin as they provide insights into the pathogenicity mechanisms of *C. perfringens* diseases.

1.1.3.1 C*. perfringens* alpha toxin

C. perfringens alpha toxin (CPA) is a zinc-containing metalloenzyme with phospholipase C and sphingomylinase activities (Flores-Díaz et al., 2015). CPA hydrolyzes the membrane phospholipids, preferentially phosphatidylcholine (PC or lecithin) and sphingomyelin (SM) in cell membranes (Flores-Díaz et al., 2015). When present in high concentration, it elicits a massive PC and SM degradation that directly leads to membrane disruption and cell lysis (Popoff and Bouvet, 2009). However, lower CPA concentrations lead to formation of diacylglycerol and ceramide; both activate a signaling transduction pathway and lead to an activation of endogenous phospholipases triggering hemolysis (Popoff and Bouvet, 2009; Sakurai et al., 2004). Furthermore, alpha toxin induces the release of cytokines, notably IL-8, eliciting an acute inflammatory reaction through the recruitment of neutrophils (Popoff and Bouvet, 2009). In summary, CPA seems to exert its virulence via (I) inducing signal transduction pathways, (II) IL-8 activation, (III) hyperexpression of adhesion molecules, platelet aggregation and thrombus formation, and (IV) direct hemolytic and cytotoxic effects (Flores-Díaz et al., 2015).

The gene encoding for alpha toxin (*plc* or *cpa*) is carried on the chromosome of virtually all *C. perfringens* and so all types of *C. perfringens* have the potential to produce CPA (Rood, 1998; Rood et al., 2018). However, the produced amounts of CPA can vary between *C. perfringnes* strains, presumably due to the regulation system of *plc* gene expression (Abildgaard et al., 2009; Bullifent et al., 1996).

CPA is well known as a key virulence factor in the development of gas gangrene. Its role was proven based on murine infection models with CPA-inactivated *C. perfringens* mutant strains (Awad et al., 1995). In addition, mice immunized with a recombinant CPA-toxoid developed a solid protection against experimental gas gangrene (Williamson and Titball, 1993).

Enteric diseases due to *C. perfringens* type A is historically reported in association with alpha toxin. However, a number of newly identified toxins were reported in a significant association with various forms of intestinal *C. perfringens* type A infections in humans and animals e.g. enterotoxin in human food poisoning cases (now classified as type F strains) and NetB in avian necrotic enteritis (recently typed as type G strains) (Freedman et al., 2016; Keyburn et al., 2008). The importance of CPA in these intestinal diseases is discussed controversial. A study reported that the CPA role in avian necrotic enteritis could not be confirmed using an experimental chicken model (Keyburn et al., 2006). In contrast, a significant involvement of CPA in the pathogenesis of bovine necro-haemorrhagic enteritis was demonstrated in a calf intestinal loop model (Goossens et al., 2017).

1.1.3.2 Perfringolysin O

Perfringolysin O (PFO) or theta (θ) toxin is not a member of the *C. perfringens* typing toxins. However, it has frequently been reported to act in synergism with CPA to produce the characteristic lesions in gas gangrene (Awad et al., 2001; Verherstraeten et al., 2013). PFO belongs to the pore-forming cholesterol-dependent cytolysin (CDC) family (Verherstraeten et al., 2015). A mutational study proved the non crucial role of PFO in a gas gangrene infection model as the PFO-mutant was still able to cause murine myonecrosis (Awad et al., 1995).

The gene encoding for perfringolysin (*pfoA*) is chromosomally located and is carried by most *C. perfringens* strains. Yet, a distinct group of *C. perfringens* lacks this gene (Deguchi et al., 2009). This group forms a cluster based on multilocus sequence typing (MLST) and comprises type F strains that carry the enterotoxin gene on the chromosome (food poisoning strains) as well as those type C strains derived from human enteritis necroticans also known as Darmbrand strains (Ma et al., 2012; Xiao et al., 2012).

1.1.3.3 *C. perfringens* enterotoxin

C. perfringens enterotoxin (CPE) is a member of the aerolysin pore-forming toxin family, is produced by ~5% of the *C. perfringens* strains and is mostly associated with cases of foodborne illness in humans (Briggs et al., 2011; Sarker et al., 1999). CPE is synthetized during the sporulation and released with the lysis of the mother cell (Li et al., 2016). CPE exerts its action by the initial binding to receptors on the apical surface of the host cell resulting in an active pore which alters the membrane permeability and allows influx of small molecules (Freedman et al., 2016; Smedley et al., 2007). Cellular effects of CPE include epithelial cell damage and villous blunting (Freedman et al., 2016).

C. perfringens type F produces CPE, but the gene encoding CPE (*cpe*) is also reported in association with other types including types C, D and E (Shrestha et al., 2018). CPE amino acid composition was reported to be highly conserved (Freedman et al., 2016), but silent and variant *cpe* were reported in some type E strains (Billington et al., 1998; Miyamoto et al., 2006; Miyamoto et al., 2011). The *cpe* gene can be either located on the chromosome or on a large conjugative plasmid (Li et al., 2013; Shrestha et al., 2018). Of interest, *C. perfringens* strains which carry the chromosomal *cpe*, are predominantly associated with food poisoning (~70%), whereas strains that harbor the plasmid-encoded *cpe* gene are usually involved in non-foodborne gastrointestinal illness such as antibiotic-associated diarrhea and sporadic non-foodborne diarrhea (Robertson et al., 2014). The *cpe* gene is usually associated with insertion sequences located upstream (IS1469) and downstream (IS1470 or IS1151) of the gene associated with helping gene mobilization and transfer (Miyamoto et al., 2006). However, chromosomal *cpe* strains form a MLST cluster together with the type C human enteritis

necroticans strains (Xiao et al., 2012). They also lack the *pfoA* and produce a variant small acid soluble protein (Ssp4), conferring an exceptional resistance of the strains to harsh environmental conditions such as heat, osmotic pressure and pH (Li et al., 2013; Li and McClane, 2006a; Li and McClane, 2008). The spore resistance properties of chromosomal *cpe* strains is additionally attributed to the production of a smaller, more dehydrated spore cores (Novak et al., 2003; Orsburn et al., 2008). Furthermore, the chromosomal *cpe* isolates can grow much faster than other *C perfringens* with a ~10 min generation time at 43°C (Robertson et al., 2014).

1.1.3.4 *C. perfringens* beta toxin

C. perfringens beta toxin (CPB, β-toxin) is a pore-forming toxin, which is related to the hemolysin/leucocidin toxins produced by *Staphylococcus aureus* (β-pore-forming toxin family; β-PFT) (Hunter et al., 1993). CPB is described as dermonecrotic and lethal but non-hemolytic (Songer, 1996). CPB is highly sensitive to proteases such as trypsin (Theoret et al., 2015). The action of CPB is either locally within the intestine by targeting enterocytes and endothelia or systemically via absorption leading to enterotoxaemia (Uzal and McClane, 2011).

C. perfringens toxin types B and C can produce beta toxin. Diseases associated with type C strains involve enteritis and enterotoxaemia in both humans and livestock (Uzal and McClane, 2011). The CPB is (believed to be) an essential factor for the diseases caused by type C strains as evidenced by (I) the observation that CPB-toxin neutralization using monoclonal antibodies abolishes the lethality in mice challenged with type C culture supernatants (Fisher et al., 2006), (II) the purified CPB is highly lethal for mice (Fisher et al., 2006; Shatursky et al., 2000) and can induce necrotizing enteritis in a rabbit intestinal loop model (Sayeed et al., 2008; Vidal et al., 2008), and (III) a *cpb* mutant type C strain lost the ability to provoke the disease in several animal species models whereas knocking out the genes of CPA, PFO or both, did not abolish the virulence of the strains in a rabbit intestinal loop model (Sayeed et al., 2008). In type C strains that harbor the enterotoxin gene e.g. human enteritis necroticans strains, both toxins, CPE and CPB, were shown to act in synergism and contributed to the enteropathogenic lesions in rabbit small intestinal loops (Ma et al., 2014).

Purified CPB is thermolabile and highly sensitive to proteases e.g. trypsin and pepsin (Nagahama et al., 2015). The lack of trypsin activity is considered predisposing factor for the β-toxin induced pathology (Uzal et al., 2014). Infections due to type C are mostly observed when the trypsin activity in the host intestine is abolished or inhibited, e.g. in livestock neonates colostrum may inhibit the trypsin action, in malnourished or diabetics or in people who frequently consume food rich in trypsin inhibitors such as sweet potatoes (Uzal et al., 2014; Uzal and McClane, 2011). However, *C. perfringens* type B strains produce beta and epsilon

toxins and both were reported to be crucial in type B infections (Fernandez-Miyakawa et al., 2007). Interestingly, while CPB is highly sensitive to protease digestion, epsilon toxin requires protease activation to be completely functional (Li et al., 2013). This suggests that depending on the intestinal conditions the activity of either toxin can be selectively affecting the intestine (Fernandez-Miyakawa et al., 2007).

The CPB gene (*cpb*) is localized on large, most likely conjugative plasmids in type B and type C strains and is associated with insertion elements (Nagahama et al., 2015). In type B strains, the *cpb*-plasmid is distinct from the *etx*-plasmid (Li et al., 2013). *Cpb*-plasmids in most type B strains are ~90Kbp, but also ~65Kbp *cpb*- plasmids have been reported in some strains (Li et al., 2013). The size of *cpb* plasmids in type C strains may vary from ~65Kbp to ~110Kbp (Li et al., 2013). In both types (B and C), the *cpb*-plasmid may carry other toxin genes such as *tpeL* (Nagahama et al., 2015).

1.1.3.5 *C. perfringens* epsilon toxin

Epsilon toxin (ETX) is a lethal heat-labile toxin which is similar to the aerolysin family of pore-forming toxins (Cole et al., 2004). ETX is secreted as a prototoxin that can be activated by proteolytic enzymes such as trypsin (Miyata et al., 2001). The action mechanism of the activated ETX comprises (I) binding to a receptor on the host cell membrane, (II) oligomerization into a heptameric prepore complex using lipid rafts in the cell membrane, and (III) formation of an active pore via extending a β-hairpin loop into the lipid bilayer resulting in ion dysregulation, cellular damages and necrosis (Alves et al., 2014; Uzal et al., 2016b).

The gene encoding ETX (*etx*) in type B and D strains is reported to be located on a large plasmid which also carries a *tcp* conjugation locus (Li et al., 2013). Similar *etx* plasmids were observed in different type B strains which have a size of ~65Kbp and display the *cpb2* gene and an insertion sequence (IS1151) located adjacent to the *etx* gene (Miyamoto et al., 2008). In contrast, most *etx* plasmids of type D strains are variable in size (Sayeed et al., 2007). Larger plasmids (75 or 110Kb) are usually reported in type D strains that carry *cpe* or *cpb2* genes, while smaller plasmids (~48Kbp) are usually found in strains lacking *cpe* or *cpb2* (Li et al., 2013). Some type D strains were reported to carry a ~65Kbp plasmid likely similar to the *etx* plasmid of type B strains (Li et al., 2013; Sayeed et al., 2007).

ETX is believed to be the essential factor in *C. perfringens* type D diseases including fatal enterotoxaemia which is mainly observed in sheep and goats (Uzal et al., 2016b). Diseased animals can develop neurological and respiratory signs or may die suddenly (Uzal et al., 2010). ETX can affect the intestine causing mucosal damages, congestion and hemorrhages. Additionally, it can be also absorbed into the circulation and bind to internal organ tissue e.g.

brain, kidneys and lungs. Intercellular flow of the ETX to reach the circulation without intestinal damages seems also possible (Uzal et al., 2016b). ETX increases the vascular permeability in internal organs due to its effect on the endothelium of blood vessels (Popoff, 2011). In the brain, ETX disrupts the blood-brain barrier and leads to the development of perivascular edema but also binds to brain tissue thereby inducing the release of excitatory neurotransmitter and development of neurologic symptoms (Alves et al., 2014; Uzal et al., 2014). Focal symmetrical encephalomalacia has been described in the sub-acute and chronic forms of the disease (Alves et al., 2014).

1.1.3.6 *C. perfringens* iota toxin

Iota toxin (ITX) is a member of the clostridial binary toxin family which is produced by type E strains and was first described in 1943 by Bosworth (Bosworth, 1943; Rood et al., 2018). Type E strains are usually involved in cases of necrotizing enteritis and enterotoxaemia in calves and lambs (1.1.2) (Songer, 2016). Type E induced enterotoxaemia in rabbits was also reported (Songer, 2016; Songer and Miskimmins, 2004). ITX consists of two separate polypeptides (Ia and Ib) and both are required for toxin activity (Sakurai and Kobayashi, 1995; Sakurai et al., 2009). Following proteolytic activation, the toxin subunit Ib acts as a cell binding component that binds to a receptor on the plasma membrane and oligmerizes to form a heptamer (Sakurai et al., 2009). This subunit Ib heptamer interacts with the N-domain of Ia forming a complex that can then be internalized into the cell and will be transported to the early endosomes (Sakurai et al., 2009; Takehara et al., 2017). The Ia subunit represents the toxin enzymatic component, as ADP-ribosyl transferase, that translocates into the cytoplasm and acts on the cell cytoskeleton leading to collapse and cell death (Sakurai et al., 2009; Takehara et al., 2017).

The ITX binary toxin is encoded by two genes (*iap* and *ipb*) which are arranged as an operon and are separated by an intermediate region of 243 nucleotides (Li et al., 2013). The genes are carried by potentially conjugative plasmids that can additionally carry the enterotoxin gene (either silent or functional *cpe*) (Li et al., 2013). Two families of ITX plasmids have been described (Miyamoto et al., 2011). The pCPPB-1-like plasmids carry a functional atypical *cpe* gene as well as iota toxin genes and have a consistent size of ~65Kbp (Miyamoto et al., 2011). The other group of ITX plasmids resembles pCPF5603 (~ 97 or 135Kbp) and carry nonfunctional silent *cpe* sequences (Billington et al., 1998). The IS1151-like insertion sequence located adjacent to the *iap*-*ibp* operon as well as urease and lambda toxin genes were reported in association with the latter type of plasmids (Freedman et al., 2015).

1.2 Avian necrotic enteritis

Avian necrotic enteritis (NE) is an enteric disease of poultry that was first described by Parish in 1961 (Cooper et al., 2013; Parish, 1961). This disease is among the challenging problems of the poultry industry due to its economic impact which was estimated to sum up to a loss of $5-6 billion annually (Wade and Keyburn, 2015). Routine use of in-feed antimicrobial growth promoters (AGPs) in poultry production reduced the incidence of the disease in the past (Corpet, 2000). However, the risk of reemergence of the disease is reported because of voluntary and legal withdrawal of antimicrobials from animal feed (McDevitt et al., 2006; Shane, 2005; Smyth, 2016).

1.2.1 Clinical signs and pathology

Two forms of NE disease have been described: The acute clinical and the mild subclinical form (Cooper et al., 2013; Timbermont et al., 2011). Acute NE is typically of short duration and the birds die within 1-2 h after the onset of clinical signs. Clinical signs include diarrhea, depression, apathy and decreased feed consumption (Cooper and Songer, 2009; Cooper et al., 2013; Shojadoost et al., 2012). The pathological lesions are mainly restricted to the intestine i.e. a dilated intestine due to foul-smelling gases and dark greenish to dark brownish, bile stained liquid contents. The intestinal wall is usually thin. The mucosal surface bears diffuse necrotic patches and is covered by fibrinous diphtheritic pseudo-membranes (Cooper and Songer, 2009; Cooper et al., 2013). Due to the extensive intestinal mucosal damage, the bacteria reach the hepatic tissue through the portal vein and induce necrotic lesions (Timbermont et al., 2011). Hepatic lesions are prominent in the subclinical form (Kaldhusdal et al., 2016). Subclinical NE is the most prevalent form of the disease and causes significant economic loss to poultry producers because of the lower productivity of the birds, the cost of disease intervention and the high condemnation rates due to hepatic lesions (Kaldhusdal et al., 2016). Birds affected by the subclinical form usually have no observable clinical signs and so the diagnosis of these cases is largely based on the impaired productivity of the flock, poor condition of the birds and the presence of focal mucosal necrotic or ulcerative areas (Cooper et al., 2013; Johansson et al., 2010).

1.2.2 *C. perfringens* in avian necrotic enteritis disease

C. perfringens is the cause of NE in poultry. Also, it normally resides in the intestine of healthy birds (Antonissen et al., 2016). *C. perfringens* lacks many genes required for amino acid biosynthesis but it can release various degradative enzymes and nutrient transporter systems, thus it can efficiently utilize different nutrients available in the intestine (Shimizu et al., 2002; Timbermont et al., 2011). Thus, *C. perfringnes* can grow well and rapidly in the host intestine leading to tissue damages and destruction. Toxins produced by *C. perfringens* strains,

especially the necrotic entertitis toxin Beta-like (NetB toxin), are reported to induce the characteristic NE damages in poultry (Keyburn et al., 2008). However, the disease is believed to be multifactorial involving virulence elements e.g. toxins and bacteriocins, predisposing risk factors and host related factors (Moore, 2016). Additionally, several mechanisms were reported to contribute to NE disease including the degradation of the intestinal mucous layer, proteolytic activity of the strains and colonization of the epithelium by *C. perfringens* (Prescott et al., 2016).

1.2.3 Necrotic enteritis risk factors

Several predisposing factors can enhance the proliferation of *C. perfringens* and provide suitable conditions for NE development such as nutrients supply, alteration of the intestinal-nutritional dynamics, disruption of intestinal microbiota and impairment of the bird's immune system (Moran, 2014; Prescott et al., 2016). Consideration of predisposing factors is necessary for experimental NE infection models (Lee et al., 2011; Shojadoost et al., 2012) to enhance the *C. perfringens* growth and to facilitate the mechanisms by which *C. perfringens* causes NE (Moore, 2016).

NE predisposing factors include suboptimal diet, diseases and husbandry–related factors (Dahiya et al., 2006; Williams, 2005). Diets containing high levels of protein of animal sources such as fishmeal or diets rich in non-starch polysaccharides increase the intestinal viscosity and stimulate the overgrowth of intestinal *C. perfringens* (Stanley et al., 2014; Wu et al., 2014). In addition, potato-based diets were associated with elevated scores of intestinal and hepatic hemorrhages (Fernando et al., 2011; Palliyeguru et al., 2010).

Poultry infected with *Eimeria* species develops coccidiosis, a disease that causes intestinal lesions and can *per se* be fatal or predispose chickens to clostridial infections (Williams, 2005). Necrotic enteritis is commonly observed as a complication of coccidial pathogens (Bangoura et al., 2014; Lanckriet et al., 2010). Indeed, NE experimental models include a prior challenging of the bird with coccidia (Lee et al., 2011; Shojadoost et al., 2012). The damage induced by *Eimeria* spp. to the gut epithelium causes plasma leakage into the intestinal lumen that promotes *C. perfringens* replication (Moore, 2016). This damage also exposes the epithelium allowing colonization of *C. perfringens* (Timbermont et al., 2011; Uzal et al., 2015). Coccidial infection causes significant perturbations in gut microbiota (Stanley et al., 2014; Wu et al., 2014), exhausts the immune system of the bird and stimulates the production of intestinal mucus which in turn enhances the growth of mucolytic *C. perfringens* (Collier et al., 2008). Significant alterations in the gut microbiota were found in association with dietary fishmeal and *Eimeria* infection (Wu et al., 2014). Microbial imbalances of the bird's intestine may act as a predisposing factor for NE or as a consequence of the *C. perfringens* overgrowth (Antonissen

et al., 2016; Stanley et al., 2012; Stanley et al., 2014). Additionally, infections causing immune suppression e.g. infectious bursal disease (IBD) can also predispose chicken to NE (Moore, 2016). Indeed, IBD virus has been used instead of *Eimeria* oocysts to enable *C. perfringens* infection in experimental models (McReynolds et al., 2004). In contrast, circumstances that enhance the chicken's immune system e.g. availability of selenium-compounds decreases the risk for NE in birds (Lee et al., 2014a; Lee et al., 2014b; Xu et al., 2015a; Xu et al., 2015b).

Poultry may experience several adverse management stressors during rearing e.g. high environmental temperature, humidity, overstocking, bad ventilation, poor litter quality and change of feeding system i.e. from starter to grower. These stressors catalyze the secretion of corticosterone which consecutively inhibits several immune system functions and affect the ability of birds to resist infections caused by opportunistic bacteria (Virden and Kidd, 2009). Stocking density, prolonged heat and cold stress were shown to increase a bird's susceptibility to NE (Calefi et al., 2014; Lee et al., 2011; Tsiouris et al., 2015). Elevated stocking densities in poultry houses increase litter moisture and air ammonia content (Hermans and Morgan, 2007). Wet litter has been associated with intestinal necrotic lesions in broilers (Moore, 2016).

Recent studies have shown that host genetics also play a role in the susceptibility to NE (Oh and Lillehoj, 2016). Weight loss, gut lesions and NetB antibody levels were reported at higher values in Cobb chicken breed lines compared to Ross or Hubbard lines (Jang et al., 2013). Ross chickens show higher values of host innate immunity parameters such as higher β-defensin transcripts and IL-17F in chicken intestine compared with Cobb breed (Hong et al., 2012). Recently, Kim et al (2014) reported different susceptibilities to NE in two inbreed-chicken lines with identical MHC B2 genotype. In both lines, different immune pathways were reported with the upregulation of more than 15 immune-related genes in the more susceptible line (Kim et al., 2014; Oh and Lillehoj, 2016).

1.2.4 *C. perfringens* necrotic enteritis virulence factors

1.2.4.1 Toxins

Strains of *C. perfringens* produce a plethora of extracellular toxins and enzymes including the chromosomally encoded alpha toxin (CPA) (Kiu and Hall, 2018) (1.1.3.1). The role of CPA in NE pathogenicity is discussed controversially; its role could not be confirmed using an engineered *cpa*-negative strain as this strain was still able to cause NE in an experimental model despite the absence of CPA (Keyburn et al., 2006). However, CPA was found to be a protective immunogenic protein against NE strains and the levels of CPA-antibodies were detected at higher values in NE affected birds (Cooper and Songer, 2009; Cooper et al., 2009; Kulkarni et al., 2010). The CPA toxin was also detected in birds inoculated with a CPA-mutant strain and the amount of CPA correlated positively with the severity of necrotic damages

possibly due to resident intestinal *C. perfringnes* bacteria (Coursodon et al., 2010). Keyburn et al. 2018 (Keyburn et al., 2008) identified a pore-forming toxin named the Necrotic enteritis toxin B-like (NetB) in an Australian NE strain (EHE-NE18). The pathogenic role of NetB toxin in NE disease has been demonstrated by applying gene knockout technology (Keyburn et al., 2008; Li et al., 2013); in a chicken NE model a *netB* negative mutant strain was avirulent. The strain restored its ability to induce necrotic damages after mutant complementation with the wild-type *netB* gene (Keyburn et al., 2008). Also, challenging birds with a set of NetB-producing and non-producing strains from different sources (including one strain isolated from mammals) showed that only NetB producing strains were able to initiate the disease with different degrees of virulence regardless of the isolate origin (Keyburn et al., 2010b; Smyth and Martin, 2010). Very recently, a study showed that horizontal acquisition of a *netB*-plasmid could convert nonvirulent strains into virulent strains able to cause NE in a chicken model (Lacey et al., 2017). This horizontal transfer has been also demonstrated to happen naturally between *C. perfringens* strains that reside in the bird's intestine (Lacey et al., 2017). A recent toxin-typing scheme of *C. perfringens* classified the NetB positive strains as type G strains because of the clinical association of NetB to *C. perfringnes* avian necrotic enteritis disease (Rood et al., 2018). However, *netB* prevalence studies reported a percentage of *netB*-negative strains in NE cases and *netB*-positive strains in healthy birds (Chalmers et al., 2008a; Johansson et al., 2010; Keyburn et al., 2010b; Martin and Smyth, 2009). The frequent absence of the *netB* gene from NE isolates was explained as possible inadvertent selection of negative commensal strains during bacterial cultivation from clinical samples or due to loss of the *netB*-plasmid (Lepp et al., 2013). However, a study reported the stability of *netB* plasmids even after approximately 80 subcultures (Lepp et al., 2013). Studies also reported that *netB*-negative strains could induce NE lesions (Cooper and Songer, 2010; Li et al., 2017a; Li et al., 2017b). Moreover, NE isolates collected over 15 years ago in Albama showed the absence of the *netB* gene (Bailey et al., 2015). Interestingly, it was shown that not all *netB* positive strains are able to express and produce NetB toxin *in vitro* although sequence comparisons of *netB* genes from different strains revealed almost complete identity (Abildgaard et al., 2010; Lacey et al., 2016).

TpeL toxin has been detected in NE isolates (Chalmers et al., 2008a) in association with some *netB*-positive isolates from diseased birds (Bailey et al., 2015; Chalmers et al., 2008a; Keyburn et al., 2010b; Llanco et al., 2015). The gene encoding TpeL (*tpeL*) is plasmid encoded and has been reported in *C. perfringens* types A, B and C (Adams et al., 2014). In types B and C *tpeL* is located downstream the beta toxin gene (*cpb*) on ~65Kbp and 90Kbp plasmids (Gurjar et al., 2010; Sayeed et al., 2010). These plasmids carry transposases such as IS1151 and the *C. perfringens tcp* locus thus indicating a potential for a conjugative transfer for these plasmids (Gurjar et al., 2010). *Cpb* and *tpeL* were also reported on separate plasmids (Gurjar et al., 2010). The role of Tpel in NE has not been investigated yet. However, studies showed that

significant intestinal lesions are associated with strains carrying *netB* and *tpeL* genes, compared to lesions caused by only *netB* carrying strains (Coursodon et al., 2012).

Cpb2 encoding *C. perfringens* beta2 toxin was identified for the first time in 1997 (Gibert et al., 1997). Beta2 toxin is a pore-forming toxin which is frequently found in isolates from both healthy and diseased birds (Crespo et al., 2007; Tolooe et al., 2011a). Two variants of the *cpb2* gene have been described, a consensus variant and an atypical *cpb2* variant (Jost et al., 2005). The consensus *cpb2* variant is preferentially associated with porcine *C. perfringens* isolates and plays a role in the pathogenesis of porcine necrotic enteritis (Jost et al., 2006a). Atypical *cpb2* genes are mostly linked to *C. perfringens* isolates of non-porcine hosts. Their toxins displayed 62.3% protein identity to the consensus beta2-toxin (Jost et al., 2005). Studies showed that *cpb2* genes from non-porcine *C. perfringens* isolates are not always expressed *in vitro* (Jost et al., 2005; Popoff and Bouvet, 2013). The *cpb2* genes of *C. perfringens* strains isolated from avian species were classified as atypical forms (Bannam et al., 2011; Jost et al., 2005; Tolooe et al., 2011a). The incidence of *cpb2* gene in *C. perfringens* isolates from poultry is remarkably high with no bias in distribution between healthy and diseased birds suggesting no direct correlation with NE in birds (Cooper and Songer, 2010; Tolooe et al., 2011a). However, a recent report described a close association between the presence of beta2-positive *C. perfringens* isolates and the detection of focal duodenal necrosis in egg-laying chickens in the United States (França et al., 2015).

1.2.4.2 NE-associated plasmids

Plasmids are additional DNA molecules that are found in all types of bacteria and carry information of significant value for bacterial adaptation, virulence and evolution (Snyder et al., 2013). NE-derived *C. perfringens* strains harbor 2 to 5 independent low copy number plasmids of variable sizes from 45Kbp to 100Kbp (Bannam et al., 2011; Lepp et al., 2010; Parreira et al., 2012). This is in contrast to the *netB*-negative strains in healthy birds which carry fewer and smaller plasmids as shown by PFGE analyses (Parreira et al., 2012). In NE strains, the *netB* gene is localized on a large conjugative plasmid (~ 80 to 100Kbp) that is distinct from the plasmid carrying *cpb2* (Bannam et al., 2011). However, both plasmids are large and conjugative (Lepp et al., 2010; Parreira et al., 2012). In addition, a tetracycline resistance plasmid (47Kbp) has been frequently found in the poultry strains of *C. perfringens* (Bannam et al., 2011; Li et al., 2017d).

Sequence analysis of NE-derived *C. perfringens* plasmids has shown that these possess a conserved syntenous backbone of 39 genes (~41Kbp of the plasmid size) in addition to a variable region designated as a unique pathogenicity locus for each plasmid (Parreira et al., 2012). The conserved region includes a (plasmid) central control region (CCR) and a transfer

of clostridial plasmid (*tcp*) conjugation locus (Parreira et al., 2012). The CCR contains a ~6Kbp sequence and encodes genes responsible for plasmid replication (*rep* gene), regulation (*reg* gene) and partitioning (parMRC-like locus). The rep protein is highly conserved and specific for *C. perfringens* plasmids (Bannam et al., 2006). The compatibility mechanism of different plasmids within *C. perfringens* most likely included a *parMRC*-like partitioning locus (Adams et al., 2015). Different alleles of *parMRC* have been identified forming up to ten different partitioning families into which *C. perfringens* plasmids can be clustered (Adams et al., 2015). The plasmids of different partitioning families were found to be compatible (Adams et al., 2015; Parreira et al., 2012).

1.2.4.3 NE-associated pathogenicity loci

Avian *C. perfringens* NE strains carry three unique DNA segments i.e. pathogenicity loci: NELoc-1, NELoc-2 and NELoc-3 (Lepp et al., 2010). NELoc-1 (~ 42Kbp) is located on a large plasmid with 37 genes (Bannam et al., 2011; Lepp et al., 2010; Parreira et al., 2012) including the *netB* toxin gene along with other genes that may also contribute to NE (Zhou et al., 2017). A recent study described that the absence of NELoc-1 genes other than *netB* reduced the full virulence of an NE strain. The strain was still able to produce NetB and induced *in vitro* cytotoxicity (Zhou et al., 2017). Another study described the essentiality of some factors encoded by genes within NELoc-1 in an experimental NE model such as a diguanylate cyclase and a phosphodiesterase (cyclic-di-GMP) signaling system (Parreira et al., 2017). Moreover, NELoc-1 carries genes putatively encoding for an internalin-like protein and a ricin-type domain protein which may contribute to virulence (Cheung et al., 2010; Lepp et al., 2010). The NELoc-1 locus also possesses two putative small pore-forming toxin genes, leucocidin I and II sharing similarities with other known pore formers (Lepp et al., 2010). Genes homologous to chitinases, carbohydrate binding proteins and proteases have also been predicted within NELoc-1 (Lepp et al., 2010). NELoc-2 (11. 2Kbp) is located on the chromosome and consists of 11 contiguous genes encoding mainly membrane- and cell wall associated factors (Lepp et al., 2010). The NELoc-3 (5.6Kbp) is localized on another distinct plasmid that carries *cpb2* and consists of five genes including a putative gene for the NADP-dependent 7-alpha-hydroxysteroid dehydrogenase (7-HSDH)(Lepp et al., 2010). Interestingly, a later study reported a strong association between the plasmid NELoc-1 and the chromosomal NELoc-2 in a larger set of NE strains (Parreira et al., 2012). This is not the case for NELoc-3 which was not conserved among all analyzed NE strains (Parreira et al., 2012). Lepp et al., (2013) also reported that not all *cpb2* positive *C. perfringens* strains carry the NELoc-3 genes indicating that the association between the *cpb2* and the NELoc-3 is not always conserved (Lepp et al., 2013).

1.2.4.4 Additional virulence factors

Besides toxin production, virulence mechanisms of *C. perfringens* NE strains involve bacteriocin production (Barbara et al., 2008; Timbermont et al., 2014; Timbermont et al., 2009), proteolytic activity (Olkowski et al., 2006; Olkowski et al., 2008) and the capability to adhere to extracellular matrix molecules and to the host epithelium (Martin and Smyth, 2010; Timbermont et al., 2011).

Bacteriocins are substances commonly involved in intra-species growth inhibition resulting in a dominance of certain clones of bacteria (Barbara et al., 2008; Timbermont et al., 2009). Timbermont and colleagues (2014) identified a *C. perfringens* bacteriocin named perfrin in most *netB* positive strains (NE strains) (Timbermont et al., 2014). The antibacterial activities of these NE strains may provide a selective advantage in the chicken gut (Barbara et al., 2008; Timbermont et al., 2014; Timbermont et al., 2009). *C. perfringens* NE isolates have been frequently reported to be clonal within a single flock (Gholamiandekhordi et al., 2006). This is in contrast to *C. perfringens* isolates from healthy birds which were more divergent even within single birds (Chalmers et al., 2008b; Gholamiandekhordi et al., 2006). However, clonality of "disease strains" and divergence of "non-disease strains" is not a consistent finding in different studies (Chalmers et al., 2008b; Lyhs et al., 2013). PFGE analysis of a set of strains isolated from healthy and diseased turkeys showed high genetic diversity in the "disease strains" (Lyhs et al., 2013). Also, *C. perfringens* strains isolated from healthy birds on different farms were shown to be very similar (Chalmers et al., 2008a). In fact, bacteriocin production was not consistent in all NE strains and was also sometimes observed in "non-disease strains" (Timbermont et al., 2014).

Proteolytic factors are involved in early NE lesions (Olkowski et al., 2006; Olkowski et al., 2008). In the early stage of NE, the basal and lateral domain of the intestinal epithelium are mainly affected, but not the epithelial apical part (Olkowski et al., 2006; Olkowski et al., 2008; Timbermont et al., 2011). Studies showed that *C. perfringens* NE strains can secrete collagenolytic enzymes but can also activate the host's matrix metalloproteinases (MMPs). These proteolytic enzymes disrupt the extra-cellular matrix and cellular junctions causing significant changes in the basal and lateral domain of the intestinal epithelium and the lamina propria (Olkowski et al., 2006; Olkowski et al., 2008; Timbermont et al., 2011).

Adhesion to the extracellular matrix and to the host epithelium is another characteristic of the virulent NE strains (Martin and Smyth, 2010). Virulent strains are able to bind to different extracellular matrix molecules (ECMM) in the gut at higher levels than non-virulent strains suggesting a role of the adherence capability to ECMMs in enhancing pathogenicity (Martin and Smyth, 2010; Timbermont et al., 2011). This role has been elucidated using knockout gene

techniques and the resultant mutants showed an altered ability to bind to immobilized ECMMs *in vitro* (Timbermont et al., 2011). Additionally, the binding of C. *perfringens* to collagen has been recently described in the NE causing strains (Wade et al., 2015).

1.3 *C. perfringens* strain typing

Bacterial strain typing is essential for outbreak investigations, epidemiological surveillance and evaluation of control measures (Li et al., 2009b). Various methods have been employed for *C. perfringens* characterization based on phenotypic features e.g. capsular serotyping, bacteriocin typing, phage typing, plasmid profiling, multilocus enzyme electrophoresis, pathogenicity and antibiotic susceptibility profiles (Gross et al., 1989; Mahony et al., 1992; Pons et al., 1994; Schalch et al., 1998). Phenotypic methods, however, may not be able to provide enough discrimination between closely related strains. Instead, genotyping methods which are based on the strains genetic content provide a high typing resolution even between closely related strains (Li et al., 2009b). Genotyping methods can be categorized into three main groups: DNA banding pattern-, DNA sequencing-, and DNA hybridization-based methods (Li et al., 2009b). Genetic typing methods used for the characterization of *C. perfringens* include PCR-ribotyping (Schalch et al., 1998), Amplified Fragment Length Polymorphism (AFLP) (McLauchlin et al., 2000), Randomly Amplified Polymorphic DNA analyses (RAPD) (Baker et al., 2010), macro-restriction with Pulse Field Gel Electrophoresis (PFGE) (Lukinmaa et al., 2002), Arbitrarily Primed PCR analyses (AP-PCR) (Chukwu et al., 2017), Repetitive sequencing-based PCR analyses (REP-PCR) (Siragusa et al., 2006), classical and genome-wide Multi-locus Sequence Typing (MLST) (Jost et al., 2006b; Mehdizadeh Gohari et al., 2017), Multiple-locus Variable Number Tandem Repeat Analysis (MLVA) (Sawires and Songer, 2005), and DNA microarray hybridization-based techniques (Lahti et al., 2012).

1.3.1 Molecular typing of *C. perfringens* by means of multilocus sequence typing

Multilocus sequence typing (MLST) is a sequence-based gene-by-gene comparison for microbial typing that was first proposed in 1998 by Maiden et al. (1998) following the principles of phenotypic multilocus enzyme electrophoresis (Stanley and Wilson, 2003). The method relies on PCR amplification and sequence analysis of internal fragments of multiple housekeeping-genes that are spread across the genome (Maiden, 2006). MLST identifies slowly accumulating sequence variations in metabolic genes which are likely to be selection neutral (Maiden, 2006). By directly indexing nucleotide variation within metabolic genes, the method not only allows for high levels of discrimination between strains but also overcomes limitations introduced by previous typing methods such as PFGE and MLEE (Stanley and Wilson, 2003). MLST provides portable and reproducible data for easy comparison among different laboratories (Maiden, 2006; Maiden et al., 1998). For many species, MLST became

the "gold standard" for microbial genotyping (Larsen et al., 2012). Various MLST schemes have been developed, validated and widely accepted for diverse human and animal bacterial pathogens ensuring a uniform and reproducible bacterial isolate characterization (Larsen et al., 2012). MLST also allows for efficient evolutionary inference and global and long term epidemiological investigations (Maiden, 2006). Typically, traditional MLST schemes involve sequencing of 6 to 10 housekeeping gene loci. Isolates are characterized based on the allele types for which every unique sequence found at each locus is given a distinct allele number and each specific combination of alleles is referred to as a sequence type (ST) (Page et al., 2017). STs can be further grouped into clonal complexes (CCs) based on the number of shared alleles between different STs (Maiden, 2006; Page et al., 2017).

In the anaerobic pathogen *C. perfringens,* the concept of indexing nucleotide sequences of multiple housekeeping- and toxin genes was first presented by Rooney and colleagues (Rooney et al., 2006). The study aimed to explain variances in *C. perfringens* virulence based on the phylogenetic structure via analyzing polymorphism patterns across certain core regions (Rooney et al., 2006). The analyzed genomic portions in the study include the coding sequences of several housekeeping genes (*gyrA*: the gyrase subunitA gene, *pfoS*: a regulatory protein gene and *rplL*: the 50S ribosomal protein gene) and virulence related genes (*plc*: phopholipase gene and *cola*: collagenase gene) as well as fragments of non-coding intergenic flanking regions (*plc* 3', *pfoS* 5' and *rplL* 5' noncoding flanking region) (Rooney et al., 2006). The study investigated 247 *C. perfringens* type A strains sourced from human clinical cases, animals and retail food which were divided into five distinct lineages (Rooney et al., 2006). Lineage 1 was the most diverse (and mosaic) and comprised ~93% of the strains that were derived from different sources and diseases backgrounds. The other lineages (lineage 2 to 5) showed a certain degree of ecological specialization and evidence of an early stage of speciation (Rooney et al., 2006). A moderate to high level of recombination was reported for the analyzed genes (Rooney et al., 2006).

In 2006, a MLST scheme was proposed by Jost and colleagues (later referred to as Jost scheme) (Jost et al., 2006b) that is based on sequence data of one core virulence (*cpa*) gene and seven housekeeping genes: *ddlA* (the Dalanine-D-alanine ligase gene), *dut* (the deoxyuridine triphosphatase gene), *glpK* (the glycerol kinase gene), *gmk* (the deoxyguanylate kinase gene), *recA* (the recombinase gene), *sod* (the superoxide dismutase gene) and *tpiA* (the triose phosphate isomerase gene) (Jost et al., 2006b). The study analyzed sequence data of 132 *C. perfringens* isolates from different toxin types and ten different host species. A substantial degree of diversity among the isolates was observed. With the exception of most type E strains, clustering of *C. perfringens* based on their toxin type identity was not supported (Jost et al., 2006b). The study identified 80 sequence types (ST) with most of the STs (60 STs,

75%) represented by a single isolate. Three clonal complexes containing ~59% of the isolates were deduced based on double locus variants (DLV) i.e. 6 out of the 8 genes were identical (Jost et al., 2006b). A major finding by the application of the Jost scheme was an observed clonality among most of the porcine *C. perfringens* type A and type C strains that carry the consensus allele of *cpb2* as well as the *cna* gene encoding a putative collagen binding protein (Jost et al., 2006b). In addition, most of the type E strains from bovine sources also showed clonality (Jost et al., 2006b). This scheme has been subsequently used in various studies that investigated mostly isolates of veterinary sources for typing especially poultry isolates (Chalmers et al., 2008a; Hibberd et al., 2011; Nakano et al., 2017; Redondo et al., 2013). Redondo and colleagues employed the scheme to characterize two strains isolated from adult cows which suddenly died due to infection with type E *C. perfringens* in Argentina (Redondo et al., 2013). Interestingly, these Argentinian type E strains clustered together with the bovine type E strains described by Jost et al. (Jost et al., 2006b) which were isolated in North America adding further evidence for clonal relatedness of bovine type E strains (Redondo et al., 2013).

Deguchi et al. (Deguchi et al., 2009) presented a MLST system for the typing of *C. perfringens* which is based on 8 MLST loci including the *groEL* gene for chaperonin GroEL, *gyrB* gene for DNA gyrase subunit B, *nadA* gene for quinolinate synthetase, *pgk* gene for phosphoglycerate kinase, *sigK* gene for sporulation sigma factor SigK, *sodF* gene for Mn/Fe superoxide dismutase, *plc* gene for phospholipase and *colA* gene for collagenase A (Deguchi et al., 2009). A large number of the enterotoxigenic *C. perfringens strains* was characterized based on this scheme (Xiao et al., 2012). These strains were recently classified as type F strains in which the *cpe* gene can reside on a chromosome or a plasmid (Rood et al., 2018). Interestingly, studies showed that chromosomal *cpe C. perfringens* and the type C Darmbrand strains form a phylogenetically distinct group (Ma et al., 2012; Xiao et al., 2012). In contrast, *C. perfringens* strains that carry the plasmid encoded *cpe* gene are more variable (Deguchi et al., 2009).

1.3.2 Molecular typing of avian *C. perfringens* using Jost MLST scheme

To date, the Jost scheme was used to type 61 avian isolates obtained from Canadian poultry farms in Ontario (Chalmers et al., 2008a), 139 avian isolates derived from 8 different states in the United States (US) (Hibberd et al., 2011) and 20 avian strains from Brazil (Nakano et al., 2017). The latter study additionally involved also 20 *C. perfringens* derived from healthy human children *inter alia* strains that carry the *netB* gene (Nakano et al., 2017). In total, 220 avian isolates from three different countries were tested with this scheme. Of these, 84 strains were isolated from cases of avian necrotic enteritis (22 from the US, 54 from Canada and 20 from Brazil), 75 strains were isolated from healthy birds (55 from the US and 20 from Canada) and further 65 strains were derived from cases of poultry gangrene in the US (Chalmers et al., 2008a; Hibberd et al., 2011; Nakano et al., 2017). Using the Jost MLST scheme, Chalmers et

al. (Chalmers et al., 2008a) reported a clustering of most *C. perfringens* NE isolates that were derived from different NE outbreaks. Isolates from healthy birds were phylogenetically interspersed among the outbreak strains but interestingly a distinct and close clustering of non-outbreak isolates was obtained from 7 separate farms with a history of unrestricted use of antibiotics in animal feed (Chalmers et al., 2008a). Additionally, NE strains isolated from the same farm or from the same bird were usually observed to be very similar if not identical (Chalmers et al., 2008a). Hibberd et al. (Hibberd et al., 2011) reported two main MLST subtypes in close association with necrotic enteritis cases: ST-31 (10 strains) and CC-4 (STs 32, 33 and 39) which were found to be closely related to the ST-1 and ST-10 described by (Chalmers et al., 2008a). In addition, the NE isolates were found to be genetically distinct from poultry gangrene strains as both have no common clade designation (Hibberd et al., 2011). A report from Brazil used the Jost scheme to characterize 20 NE strains deficient in *netB* (Nakano et al., 2017). This study reported a distinct segregation of Brazilian strains from the strains so far described in North America (Chalmers et al., 2008a; Hibberd et al., 2011). Despite useful results, the utility of the Jost scheme for global comparisons is hindered by the unavailability of the scheme at the MLST databases e.g. MLST.net (http://www.mlst.net) or PubMLST (http://pubmlst.org) and probably also due to slight modifications of the protocol introduced during subsequent studies (Chalmers et al., 2008a; Hibberd et al., 2011; Nakano et al., 2017). The original Jost scheme is based on eight loci with a sum nucleotide length of 2,987bp (Jost et al., 2006b). Studies conducted by Chalmers et al. (Chalmers et al., 2008a) and Hibberd et al. (Hibberd et al., 2011) analyzed nine loci with a total nucleotide length of 4,398bp and eleven loci of a sum length of 5,779bp, respectively. Nakano et al. (Nakano et al., 2017), however, investigated a reduced length of 2,449bp for the eight original loci.

1.3.3 Molecular typing of *C. perfringens* using core genome MLST

The recent significant drops in DNA sequencing costs has extend the application of classical MLST to include many hundreds of genes with the so-called core genome MLST (cgMLST) and whole genome MLST (wgMLST) providing high resolution for optimal pathogen typing (Maiden et al., 2013). In *C. perfringens,* cgMLST system was firstly introduced by Mehdizadeh Gohari et al. to describe the clonal relatedness among strains involved in cases of foal necrotizing enteritis (FNE) and canine acute hemorrhagic diarrhea (CHD) (Mehdizadeh Gohari et al., 2017). These strains carry the gene of NetF, a toxin which is related to the pore-forming leukocidin/hemolysin toxins (Mehdizadeh Gohari et al., 2015). A total of 1,349 genes form the basis of the developed scheme which was used to type 47 *C. perfringens* strains. Results showed that 32 *netF*-strains were partitioned into two distinct cluster types (CT), CT-I (n=26) and CT-II (n=6) but no other CTs were observed for the remaining strains. The sequences from the study were released in the NCBI public repository but the developed scheme is

currently not available for subsequent typing. This actually underlines the need to standardize an accessible, genome-wide typing approach for a uniform and replicable characterization of *C. perfringens.*

1.4. The genome of *C. perfringens*

In 2002, the complete genome of *C. perfringens* strain 13, a type A soil isolate known to cause gas gangrene, was fully sequenced using the whole genome shotgun strategy coupled with PCR direct sequencing for gap closure (Shimizu et al., 2002). The strain's genome has a low G+C content (28.6%) and comprises a single circular chromosome of 3,031,430bp and ~50Kbp plasmid (pCP13). A total of 2,660 protein coding sequences, ten rRNA operons, 96 tRNA genes and few mobile genetic elements were predicted within the genome (Shimizu et al., 2002). Of significance, the genome of the *C. perfringens* strain 13 contains genes that encode enzymes for anaerobic fermentation pathways and lacks genes coding for tricarboxylic acid (TCA) cycle- or respiratory chain, as expected for an anaerobe, but also lacks genes for enzymes intended for amino acid biosynthesis (Shimizu et al., 2002). A total of 61 chromosomal genes was predicted to be involved in sporulation and germination based on similarities to genes from the spore-forming *Bacillus subtilis* (Kunst et al., 1997). Furthermore, most of the virulence genes did not appear to be located within discrete pathogenicity islands and the G+C content of these genes was not significantly different from that of other genes (Shimizu et al., 2002).

Myers et al. (2006) reported the full genome sequences of two additional *C. perfringens* strains: the species type strain ATCC 13124 and the enterotoxin-producing food poisoning strain SM101. In addition, a comparative genomic analysis of these two genomes together with the genome of strain 13 was performed (Myers et al., 2006). Interestingly, the study reported a high degree of genomic diversity between the three genomes with unusual clustering of variable genomic regions towards one chromosomal replichore (chromosomal halves between the origin and terminus of replication) (Blattner et al., 1997). However, the three genomes have 2,126 genes that were conserved among the strains with regard to the gene order (Myers et al., 2006). Unlike strain 13 and ATCC 13124, the SM101 strain was rich in insertion sequences (IS) and the distribution of these ISs was also biased to one replichore similar to the variable genomic regions distribution (Myers et al., 2006). A total of 300 variable regions of more than 1Kbp length were identified in the three genomes. Myers et al. (2006) reported a large region of 243Kbp (236 genes) which was found in strain ATCC 13124 but was completely absent in strain 13 and partially fragmented and dispersed in strain SM101. A later study reported that the genes located in this large genomic region were found in other *C. perfringens* strains which may indicate that the region was specifically lost in strain 13 rather than being acquired by ATCC 13124 (Hassan et al., 2015).

A comparative genomic analysis including twelve *C. perfringens* strains from different toxin-types (except the newly defined type G strain) was recently described (Hassan et al., 2015). A total of 1,715 genes was found to form the core genome of the twelve strains. Additionally, whole genome alignments and phylogenetic analyses did not support the distinct separation of the strains based on their toxin types (Hassan et al., 2015). Very recently, a comparative analysis comprising 56 *C. perfringens* sequences revealed that the species pangenome is highly versatile and still in its open state (Kiu et al., 2017). The pangenome comprises the entire gene families in all strains which was calculated to be 11,667 genes in the analyzed strains. However, the core genome was significantly small, constituting 12.6% of the species pangenome only (Kiu et al., 2017; Kiu and Hall, 2018).

The first draft genomes of poultry *C. perfringens* strains were described in 2010 including seven NE strains and one non-outbreak strain (Lepp et al., 2010). Comparative genomics of these strains led to the identification of three NE-associated pathogenic loci (NELoc-1 to 3) (1.2.4.3). Recently, a complete genome was reported for a *netB* positive NE strain (Del1 strain) which was very closely related to the Canadian NE strain CP4 sequenced in 2010 by Lepp et al. with similar genomic features. The Del1 genome is made up of a 3.5Mbp chromosome (3,361CDS) and four large plasmids (49.7Kbp – 82.5Kbp) including a *netB* carrying plasmid, a *cpb2*- plasmid and a plasmid with a tetracycline resistance gene (Li et al., 2017c). Additionally, the genome of Del1 carries several phage-encoded regions and six antimicrobial resistance genes (Li et al., 2017c).

Two independent genomic studies applied whole genome sequencing to *C. perfringens* strains from healthy and NE afflicted birds from different countries and showed that the NE strains are highly divergent based on the core genome SNPs (Lacey et al., 2018; Ronco et al., 2017). However, despite this high level of diversity, Lacey et al. (2018) found that the NE strains were grouped into different pathogenic clades based on their chromosomal gene content, thus demonstrating that pathogenicity did not correlate with the content of the core genome but with mobile genetic elements (Lacey et al., 2018). In addition, they identified specific chromosomal regions in the NE strains that are putatively involved in adhesion, capsule production, prophages and bacteriocins secretion.

1.5. Aim of this thesis

1. *In silico* investigation of the genomic variability, phylogenetic relatedness and virulence assessment of *C. perfringens* by means of comparative genome analysis employing publically available genomic data.
2. Isolation, characterization and genotyping of *C. perfringens* strains from healthy and diseased poultry collected from different farms and slaughterhouses in Egypt.
3. Development and application of a core genome-based multilocus sequence typing system for *C. perfringens*.

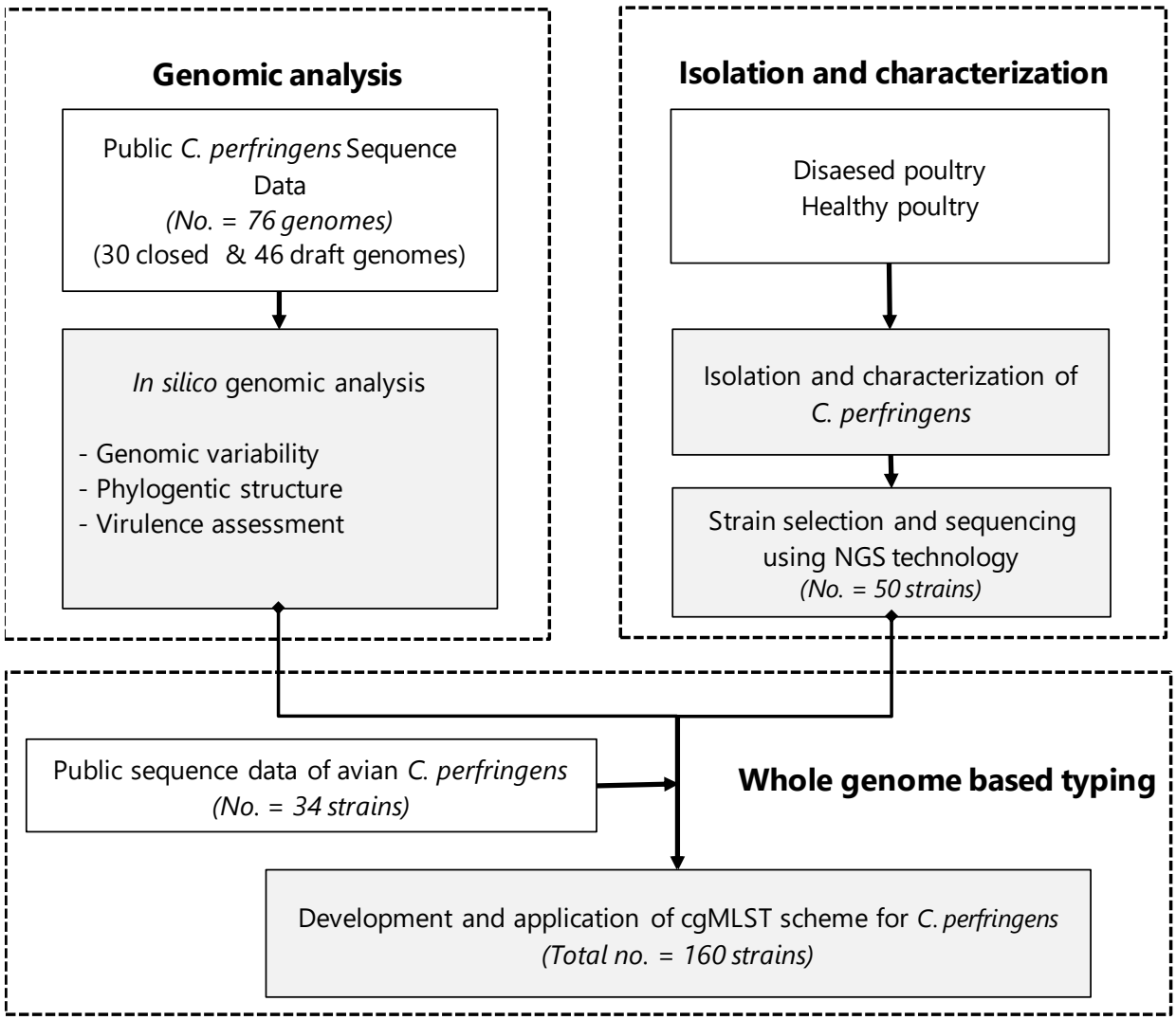


Figure 1: Aims of this thesis. Description of the main objectives of the thesis with the datasets involved in each study.

Chapter 2: Materials and methods

2.1. *In silico* investigation of the genomic variability, phylogenetic relatedness and virulence assessment of *C. perfringens* by means of comparative genome analysis employing publically available genomic data

2.1.1 Acquisition of publically accessible *C. perfringens* sequence data

To investigate the genomic diversity of *C. perfringens*, the raw Pacific Bioscience sequence data of a panel of 23 *C. perfringens* strains that were sequenced within the framework of the NCTC 3000 project (http://www.sanger.ac.uk/resources/downloads/bacteria/nctc/) were downloaded (June 2017). The strains were isolated between 1928 and 1975 from diverse sources including human strains (n = 7), foodborne strains (n = 7), animal strains (n = 2) and undefined (n = 7). These strains include 14 type F CPE-encoding human food poisoning strains (13 chromosomal *cpe* and one plasmid *cpe*) and one type C Darmbrand strain (NCTC 8081), a strain that was isolated during an outbreak of human enteritis necroticans in Germany after World War II (Zeissler et al., 1949).

In addition, a set of publicly available *C. perfringens* genome sequences (n = 53) at the NCBI (https://www.ncbi.nlm.nih.gov/genome/genomes/158) was downloaded including ten closed genomes and 43 draft assemblies. Out of these, 32 genomes were derived from cases of foal necrotizing enteritis (n = 16) and canine hemorrhagic enteritis (n = 16) (Mehdizadeh Gohari et al., 2016; Mehdizadeh Gohari et al., 2017).

Taken together, all downloaded data comprised a total of 76 genomes that represent to some extent epidemiologically unrelated strains from different ecological (human, animal, food and environment), geographical (America, Europe, Asia and Australia) and temporal (1920s to 2010s) niches (Table 3). Figure 2 shows the diversity of the used dataset and the main steps of the implemented genomic analysis workflow.

2.1.2 Assembly, annotation and genome comparison

Pacific Bioscience sequence raw data of 23 *C. perfringens* NCTC strains were *de novo* assembled using RS_HGAP_Assembly version 3 via SMRT Analysis system version 2.3.014 provided by Pacific Biosciences (Chin et al., 2013). For the strain NCTC 8081, canu (Koren et al., 2016), a fork of the Celera Assembler was used to perform the assembly instead of HGAP as a continuous merge of the plasmid to the chromosome was observed. The corrected preassembled reads from HGAP were used as input for canu which was run under default parameters but with a "correctedErrorRate" of 0.075. The circularization protocol was performed as recommended by PacBio (https://github.com/PacificBiosciences/Bioinformatics-

Training/wiki/Finishing-Bacterial-Genomes). Briefly, Gepard (Krumsiek et al., 2007) was used to define similar parts at the ends of each contig. Detected overlapping ends were merged and the genomes were circularized using Circlator (Hunt et al., 2015) or PERL script "check_circularity.pl" from SPRAI (SPRAI). Errors in the circularized region were iteratively refined with Quiver consensus algorithm (RS.Resequencing.1 module) resulting in contigs with at least 99.99% concordance to the reference contig.

Genome annotation was performed using Prokka version 1.12 (Seemann, 2014) and Rapid Annotation using the Subsystem Technology (RAST) server (Aziz et al., 2008). CRISPR sequences prediction was done using the CRISPR Recognition Tool plugin within the Geneious software (v 10.0.9). Genome comparison was carried out using progressiveMauve (Darling et al., 2010).

Insertion sequences (IS) were predicted using ISEscan version 1.5.4 (Xie and Tang, 2017) searching for complete and partial IS elements. Plotting the distribution of the IS elements across the *C. perfringens* chromosome genomes was done using R ggplot package (Wickham H., 2016).

Prophage prediction was performed using PHASTER (http://phaster.ca/) (Arndt et al., 2016), and the FASTA formatted genomes. PHASTER (http://phaster.ca/) implements a BLAST search against custom prophage databases and further assigns a completeness score for the predicted prophage regions based on the proportion of phage genes and the genetic structure of the identified region as described (Arndt et al., 2016; Knight et al., 2016). Predicted prophage regions with completeness scores less than 70, between 70 and 90 or greater than 90 were marked as incomplete, questionable or intact phage, respectively.

Genomic islands were predicted using the Islandviewer4 web service (http://www.pathogenomics.sfu.ca/islandviewer/) (Bertelli et al., 2017) and the GenBank formatted RAST annotated genomes. We used the R ggplot package (Wickham H., 2016) to plot the distribution of the IS elements across the *C. perfringens* chromosome genomes.

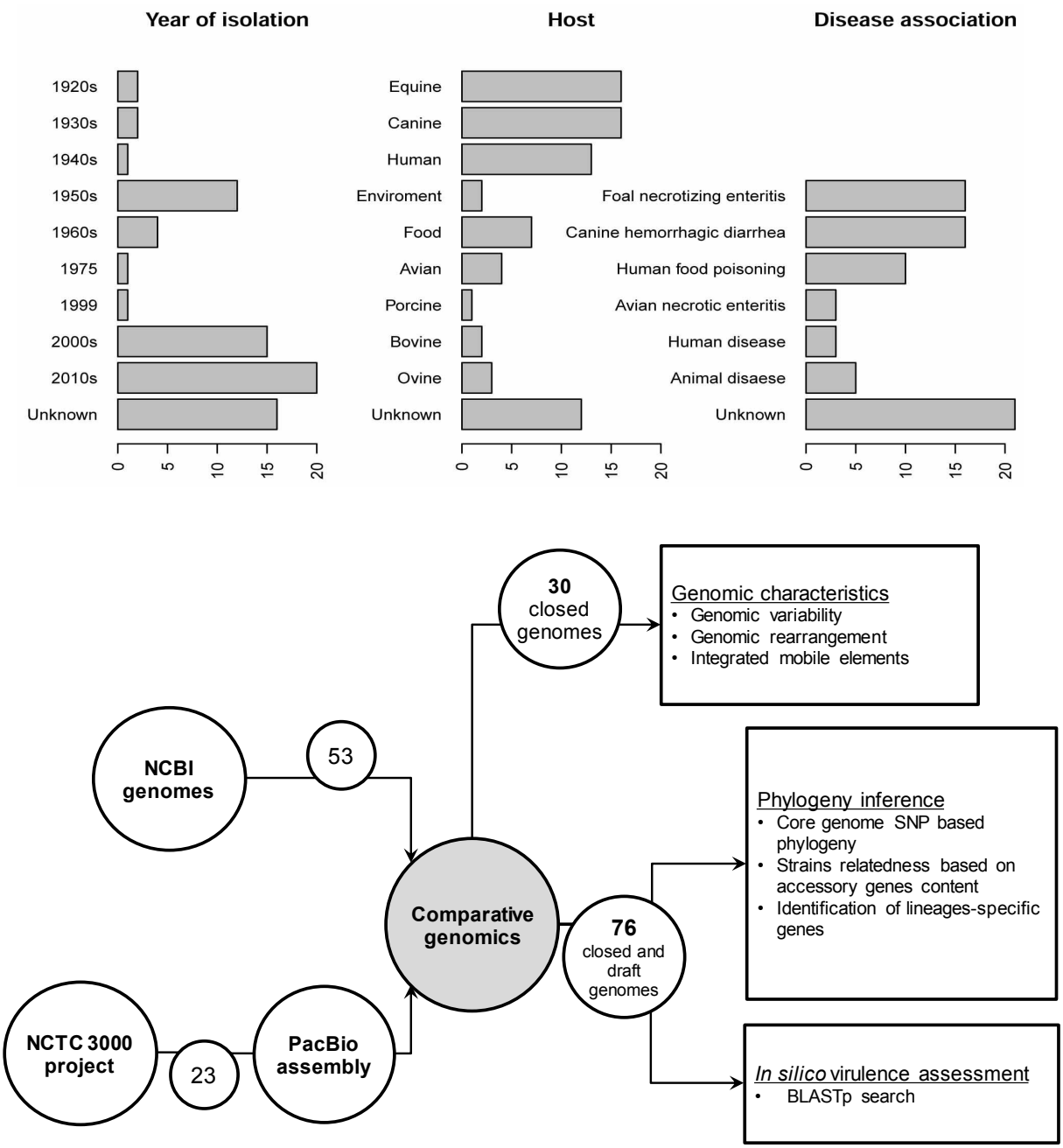


Figure 2: Schematic description of the involved data set and the main steps of the genomic analysis workflow (A) Origin of *Clostridium perfringens* isolates included in the study. Bar-plots represent the number of isolates in respect to year of isolation, host and disease association. **(B)** Steps of the genomic analysis included initially an investigation of the closed genomes (n = 30 genomes) describing genomic variability, mobile elements and genomic organization within these strains. Then, the full data set (n = 76 genomes) was used to investigate the strains phylogenetic structure as well as to predict *in silico* the virulence potential of the strains.

Table 3: Strains data of *C. perfringens* genomes used for *in silico* genomic analyses

No.	Strain name	Accession (run/bioproject)	Isolation source	Strain significance / disease	Region / country	Year of isolation	Toxin genes	Toxin-type	Study / source
Strains assembled in the study (originally sequenced within the NCTC 3000 project)									
1	NCTC 10239	ERR1681950	Food (rissoles)	Food poisoning?	UK	1958	*plc, cpe*	F	NCTC3000 project
2	NCTC 10240	ERR1681951	Bird (chicken)	-	UK	1959	*plc, cpe*	F	NCTC3000 project
3	NCTC 10578	ERR1787550	-	-	UK	1968	*plc*	A	NCTC3000 project
4	NCTC 10612	ERR1588634	Human stool	Food poisoning	UK	1966	*plc, cpe*	F	NCTC3000 project
5	NCTC 10613	ERR1787552	Food (minced beef)	Food poisoning	UK	1967	*plc, cpe*	F	NCTC3000 project
6	NCTC 10614	ERR1800584	Human stool	Food poisoning	UK	1968	*plc, cpe*	F	NCTC3000 project
7	NCTC 11144	ERR1954484	Food (beef)	Food poisoning	UK?	1975	*plc, cpe*	F	NCTC3000 project
8	NCTC 13170	ERR1377189	-	-	UK?	-	*plc*	A	NCTC3000 project
9	NCTC 2544	ERR1456745	Human gall bladder	Infected gall bladder	UK	1928	*plc*	A	NCTC3000 project
10	NCTC 2837	ERR1656459	-	-	Germany?	1929	*plc*	A	NCTC3000 project
11	NCTC 3182	ERR1599940	Sheep	Gangrene (struck?)	UK?	1930	*plc, cpb*	C	NCTC3000 project
12	NCTC 8081	ERR1656460	Human	Enteritis necroticans	Germany	1946 - 1948	*plc, cpb, cpe*	C	NCTC3000 project
13	NCTC 8238	ERR1656456	Food (salted beef)	-	UK?	1952	*plc, cpe*	F	NCTC3000 project
14	NCTC 8239	ERR1681948	-	Food poisoning?	UK?	1950s	*plc, cpe*	F	NCTC3000 project
15	NCTC 8246	ERR1656458	-	-	UK?	1950s	*plc*	A	NCTC3000 project
16	NCTC 8247	ERR1674568	Human stool	-	UK?	1950s	*plc, cpe*	F	NCTC3000 project
17	NCTC 8359	ERR1805687	-	-	UK?	1950s	*plc, cpe*	F	NCTC3000 project
18	NCTC 8503	ERR1407347	-	-	Australia	1930	*plc, etx*	D	NCTC3000 project
19	NCTC 8678	ERR1377187	Human stool	Food poisoning	UK	1950s	*plc*	A	NCTC3000 project
20	NCTC 8679	ERR1787549	Human stool	Food poisoning	UK	1950	*plc, cpe*	F	NCTC3000 project
21	NCTC 8797	ERR1377188	Salt beef	-	UK	1950	*plc, cpe*	F	NCTC3000 project

No.	Strain name	Accession (run/bioproject)	Isolation source	Strain significance / disease	Region / country	Year of isolation	Toxin genes	Toxin-type	Study / source
22	NCTC 8799	ERR1466824	Roast meat	Food poisoning	UK	1952	*plc, cpe*	F	NCTC3000 project
23	NCTC 9851	ERR1681949	Braised heart	-	UK	1950s	*plc, cpe*	F	NCTC3000 project
Genome assemblies available at the NCBI									
24	ATCC 13124	PRJNA304	-	Type strain	-	-	*plc*	A	Myers et al., 2006
25	Strain 13	PRJNA79	-		-	-	*plc*	A	Shimizu et al., 2002
26	SM101	PRJNA12521	-	Food poisoning	-	-	*plc, cpe*	F	Myers et al., 2006
27	FORC_003	PRJNA259939	Aquarium water	-	South Korea	2014	*plc*	A	NCBI
28	JP55	PRJNA276414	Horse / foal	Necrotizing enteritis	Canada	1999	*plc, cpe, netF*	F	Mehdizadeh Gohari et al., 2016
29	JP838	PRJNA276513	Dog	Hemorrhagic diarrhea	USA	2009	*plc, cpe, netF*	F	
30	FORC_025	PRJNA299533	Human stool	Foodborne pathogen	South Korea	2015	*plc*	A	NCBI
31	Del1	PRJNA363042	Chicken	Necrotic enteritis	USA	2009	*plc, netB*	G	Li et al., 2017
32	CP15	PRJNA363041	Bird intestine	Sick bird	USA	2009	*plc*	A	Li et al., 2017
33	F262	PRJNA51969	Bovine	Clostridial abomasitis	Canada		*plc*	A	Nowell et al., 2012
34	WAL-14572	PRJNA46397	-	-	-	-	*plc*	A	NCBI
35	CP4	PRJNA294154	Chicken	Necrotic enteritis	Canada	-	*plc, netB*	G	Lepp et al., 2010
36	MJR7757A	PRJNA272114	Human vagina		-	-	*plc*	A	NCBI
37	JGS1495	PRJNA20025	Pig	Diarrhea	-	-	*plc, cpb*	C	NCBI
38	ATCC 3626	PRJNA20027	Lamb intestine	-	-	-	*plc, cpb, etx*	B	NCBI
39	JGS1987	PRJNA20029	Cow	Enteritis	-	-	*plc, ibp, iap, cpe*	E	NCBI
40	F4969	PRJNA20031	-	Non foodborne disease	-	-	*plc, cpe*	F	NCBI
41	JGS1721	PRJNA28587	Sheep	Enteritis	-	-	*plc, etx*	D	NCBI
42	JJC	PRJNA215269	Sludge	-	Malaysia		*plc*	A	Wong et al., 2014
43	2789STDY5608889	PRJEB10915	Human stool	Healthy	UK	2013	*plc*	A	NCBI

No.	Strain name	Accession (run/bioproject)	Isolation source	Strain significance / disease	Region / country	Year of isolation	Toxin genes	Toxin-type	Study / source
44	JFP718	PRJNA327316	Dog	Hemorrhagic diarrhea	Canada	2011	*plc, cpe, netF*	F	Mehdizadeh Gohari et al., 2017
45	JFP727	PRJNA327317	Horse / foal	Necrotizing enteritis	Canada	2011	*plc, cpe, netF*	F	
46	JFP728	PRJNA327319	Horse / foal	Necrotizing enteritis	Canada	2011	*plc, cpe***	F	
47	JFP771	PRJNA327398	Dog	Hemorrhagic diarrhea	Canada	2011	*plc, cpe***	F	
48	JFP774	PRJNA327402	Dog	Hemorrhagic diarrhea	Canada	2011	*plc, cpe***	F	
49	JFP795	PRJNA327404	Dog	Hemorrhagic diarrhea	Canada	2012	*plc, cpe, netF*	F	
50	JFP796	PRJNA327405	Dog	Hemorrhagic diarrhea	Canada	2012	*plc, cpe, netF*	F	
51	JFP801	PRJNA327406	Horse / foal	Necrotizing enteritis	USA	2002	*plc, cpe, netF*	F	
52	JFP804	PRJNA327409	Horse / foal	Necrotizing enteritis	USA	2010	*plc, cpe, netF*	F	
53	JFP810	PRJNA327410	Dog	Hemorrhagic diarrhea	Canada	2012	*plc, cpe, netF*	F	
54	JFP826	PRJNA327411	Dog	Hemorrhagic diarrhea	USA	2012	*plc, cpe, netF*	F	
55	JFP828	PRJNA327440	Horse / foal	Necrotizing enteritis	USA	2011	*plc, cpe, netF*	F	
56	JFP829	PRJNA327441	Horse / foal	Necrotizing enteritis	USA	2010	*plc, cpe, netF*	F	
57	JFP833	PRJNA327442	Horse / foal	Necrotizing enteritis	USA	2000	*plc, netF*	F	
58	JFP834	PRJNA327443	Horse / foal	Necrotizing enteritis	USA	2002	*plc, cpe, netF*	F	
59	JFP836	PRJNA327444	Dog	Hemorrhagic diarrhea	USA	2008	*plc, cpe, netF*	F	
60	JFP914	PRJNA327445	Dog	Hemorrhagic diarrhea	Switzerland	2009	*plc, cpe, netF*	F	
61	JFP916	PRJNA327446	Dog	Hemorrhagic diarrhea	Switzerland	2009	*plc, cpe, netF*	F	

No.	Strain name	Accession (run/bioproject)	Isolation source	Strain significance / disease	Region / country	Year of isolation	Toxin genes	Toxin-type	Study / source
62	JFP921	PRJNA327447	Dog	Hemorrhagic diarrhea	USA	2007	*plc, cpe, netF*	F	Mehdizadeh Gohari et al., 2017
63	JFP922	PRJNA327448	Dog	Hemorrhagic diarrhea	USA	2006	*plc, cpe, netF*	F	
64	JFP923	PRJNA327449	Dog	Hemorrhagic diarrhea	USA	2006	*plc, cpe, netF*	F	
65	JFP941	PRJNA327450	Dog	Hemorrhagic diarrhea	Canada	2013	*plc, cpe, netF*	F	
66	JFP961	PRJNA327451	Dog	Hemorrhagic diarrhea	Canada	2014	*plc, cpe, netF*	F	
67	JFP978	PRJNA327452	Horse / foal	Necrotizing enteritis	USA	2011	*plc, cpe, netF*	F	
68	JFP980	PRJNA327453	Horse / foal	Necrotizing enteritis	USA	2006	*plc, cpe, netF*	F	
69	JFP981	PRJNA327455	Horse / foal	Necrotizing enteritis	USA	2004	*plc, cpe, netF*	F	
70	JFP982	PRJNA327456	Horse / foal	Necrotizing enteritis	USA	2001	*plc, cpe, netF*	F	
71	JFP983	PRJNA327457	Horse / foal	Necrotizing enteritis	USA	2004	*plc, cpe, netF*	F	
72	JFP986	PRJNA327458	Horse / foal	Necrotizing enteritis	USA	2011	*plc, cpe, netF*	F	
73	JFP992	PRJNA327459	Horse / foal	Necrotizing enteritis	USA	2008	*plc, cpe, netF*	F	
74	1207_CPER	PRJNA267549	-	-	-	-	*plc*	A	NCBI
75	str_0522A_28_397	PRJNA327106	Human gut	-	USA	2013-2014	*plc*	A	NCBI
76	CBA7123	PRJDB5093	Human gut	-	Korea	-	*plc*	A	Kim et al., 2017

? Metadata labeled with question marks indicate that ambiguities were found while extracting information from Genbank records and the NCTC website. ** The original study reported the strains to harbor the *netF* gene although the gene could not be detected during the *in silico* processing of the sequence data. UK, United Kingdom. USA, United States of America.

2.1.3 Core genome based phylogeny

Consensus clusters of core orthologous gene families were computed from the whole genome sequences of the 76 *C. perfringens* strains following the protocol previously described for *C. difficile* (Knight et al., 2016). Briefly, a stringent workflow was used to define core genes in which all strains should be represented (size = 100%) and each core gene cluster family should be simultaneously called by three different clustering algorithms (OrthoMCL, COGtriangles, and bidirectional best-hit (Contreras-Moreira and Vinuesa, 2013)) with thresholds (E-value = 1e-10, coverage = 0.9, identity = 0.9) and excluding paralogues. For that, the software GET_HOMOLOGUES (Contreras-Moreira and Vinuesa, 2013) was used to compute consensus clusters of orthologous gene families. The identified genes were each aligned using MAFFT v7.307 (L-INS-i method) (Katoh and Standley, 2013). An in-house bash script was then implemented to concatenate all the alignments of the identified core genes in to a single core genome alignment which was used to construct a maximum-likelihood (ML) phylogenetic tree using RAxML version 8.2.10 (Stamatakis, 2014). As parameters for RAxML, the general time reversible (GTR)-gamma model and 100 bootstrap replicates was set (Stamatakis, 2014). Additionally, the pairwise average nucleotide divergence (evolutionary distance) was calculated across the core genome using Mega (*P* distance) (Tamura et al., 2013). Clades were assigned with RAMI based on patristic distance (sum of branch length) using the best tree produced by RAxML (patristic distance threshold= 0.01) (Pommier et al., 2009; Stamatakis, 2014). The iTol web tool (https://itol.embl.de/) (Letunic and Bork, 2011) was used to visualize the phylogenetic tree and to overlay the associated metadata information. To display the evolutionary relationships of the strains in the form of networks, SplitsTree4 (Huson and Bryant, 2006) was used with the core genome Single Nucleotide Polymorphisms (SNPs) as input and the NeighbourNet method plus the Uncorrected *P* model of substitution as parameters (Bryant and Moulton, 2004).

2.1.4 Average nucleotide identity

The Average Nucleotide Identity (ANI) of the 76 *C. perfringens* strains was calculated using pyani (Pritchard et al., 2016). The ANI values were calculated based on the pairwise alignment using MUMmer (module ANIm) (Richter and Rossello-Mora, 2009).

2.1.5 k-mer based SNP analysis

A k-mer based SNP analysis (alignment-independent) was performed using kSNP 3.0 (Gardner et al., 2015). The optimal k-mer size (k-mer =21) was calculated using Kchooser (Gardner et al., 2015). From the SNP matrix file, the distance was calculated using kSNPdist

(https://sourceforge.net/projects/ksnp/) based on the equation $1 - D/N$ where D is the number of SNP differences and N is the number of SNP sites that are present in pairs of genomes.

2.1.6 *In silico* multilocus sequence typing

The following MLST genes *groEL*, *gyrB*, *nadA*, *pgk*, *sigK*, *sodF*, *plc* and *colA* were extracted from the WGS data of the 76 *C. perfringens* using an in-house script. Then *in silico* multilocus sequence typing was carried out based on a MLST system developed in 2009 (Deguchi et al., 2009). Additionally MLST data from 71 strains investigated in prior studies were included (Deguchi et al., 2009; Ma et al., 2012; Xiao et al., 2012).

2.1.7 Gene content analysis

Pangenome profiles for the 76 *C. perfringens* strains were determined using Roary software (Page et al., 2015) at 90% identity and paralogues clustering enabled selecting only genes differentially present or absent rather than gene allelic variants. Roary uses GNU Parallel (Tange, 2011) for running processes in parallel under Linux. Genes found in 2-95% of the genomes were defined as accessory genes.

2.1.8 Determination of strain relatedness based on the presence/absence pattern of accessory genes

To investigate the relatedness between the 76 strains based on the pattern of gene presence or absence, the pairwise gene content distances was inferred as Jaccard distance between each pair of genomes. The approach was previously described for *Klebsiella pneumoniae* (Holt et al., 2015). The Jaccard distance was calculated using *vegan* in R (Dixon, 2003) based on the calculation $J(A,B)=1-(A\cap B) / (A\cup B)$ where $(A\cap B)$ is the number of accessory genes present in both genomes and $(A\cup B)$ is the total number of accessory genes found in either genome. Fripan (http://drpowell.github.io/FriPan/) was used to create a multidimensional scaling plot of the pan genome using the Roary output file (gene_presence_absence.csv). A binary alignment of accessory genes present and absent in all strains was constructed and a ML phylogenetic tree was created using PhyML3.1 using the GTR model with 100 bootstraps (Guindon et al., 2010).

2.1.9 *In silico* prediction of virulence genes

Search for homologous genes related to virulence was conducted using BLASTp searches of non-redundant genes constituting the pangenome of the 76 strains against the core protein set of Virulence Factor Database (VFDB set A; http://www.mgc.ac.cn/VFs/) with thresholds (e-value <1e-20 and query alignment length >70%) (Chen et al., 2016; Pearson, 2013; Tekedar

et al., 2017). Functional annotation of the identified data set was performed using the Clusters of Orthologous Groups (COG) database (Tatusov et al., 2000) via the webmga web service (Wu et al., 2011). Gene Ontology (GO) terms were assigned using databases implemented in the Blast2GO software (Gotz et al., 2008). Additionally, BLASTn was used to search the 76 genomes for the putative iron systems that were previously described in *C. perfringens* strain 13 (Awad et al., 2016). These systems comprise two heme acquisition systems, one ferrous iron acquisition system (feoAB), three siderophore-mediated acquisition systems and one ferric citrate iron acquisition system (Awad et al., 2016).

2.2 Characterization and typing of *C. perfringens* isolates from heathy and diseased poultry in Egypt

2.2.1 Sample collection from diseased and heathy birds

Samples were obtained from poultry flocks reared in two governorates in Egypt (Sharkia and Dakahliyah) between September 2014 and December 2015. Private veterinary clinics were asked to contribute samples from birds suspected for clinical necrotic enteritis (NE) or subclinical NE. Clinical NE samples were collected from birds with necrotic patches or fibrino-necrotic membrane that covered the intestinal mucosa. Subclinical NE samples were collected from birds with yellowish necrotic or ulcerative spots on the intestinal mucosae or the liver, or from birds with abnormal intestinal fluid content. Twenty birds from ten flocks (7 layer and 3 broiler) were submitted as cases suspected for clinical NE. In addition, 34 birds from 17 flocks (15 broiler, 1 layer and 1 breeder) were suspected for subclinical NE (Table 4).

The flock sizes ranged from 3,000 to 25,000 and the age of the birds was between 10 days to 6 weeks in case of broiler farms and up to 64 weeks in case of layer farms. All birds were raised using conventional methods i.e. without restrictions of in-feed antibiotics. One to four birds per flock were taken and the clinical picture was documented. Two to four samples per bird were taken (duodenum, jejunum, caecum and liver) for further bacteriological investigation. All samples were stored at -20°C until processing. Additionally, eight different local slaughterhouses were visited and the caeca of 50 presumably healthy birds (broilers = 40, ducks = 10) were collected directly after evisceration (one sample per bird) (Table 4). All samples were stored at - 20°C until processing.

2.2.2 Isolation of *C. perfringens*

For the isolation of *C. perfringens*, material from specimens (duodenum, jejunum, caecum and liver) was thawed, then one loop was directly plated on Schaedler agar supplemented with 5% sheep blood (PB5034A, Oxoid, Germany) and Tryptose-Sulphite-Cycloserine (TSC) agar (CM0587B, Oxoid). The plates were incubated for 24-48 h at 37˚C under anaerobic conditions (anaerobic chamber MACS VA500, Don Whitley Scientific, UK). Suspected *C. perfringens* (colonies surrounded by dual hemolytic zone on Schaedler agar, or large grayish colonies on TSC agar) were selected and sub-cultured 3 times for microbiological purity. Lactose egg yolk agar plates (lactose 5 g [Merck, Germany], 0.4% bromocresol purple solution 5 ml [Aldrich, Germany], egg yolk emulsion 50 ml [Oxoid, Germany] per 100 ml) were also used to assess the lecithinase activity of some *C. perfringens* isolates.

Table 4: Total number of samples collected from healthy and diseased birds

Health status	No. of farms / slaughter houses	No. of birds	No. of samples
Samples from poultry farms			
Clinical NE	10 (7 layer, 3 broiler)	20	60
Subclinical NE	17 (15 broilers, 1 layer, 1 breeder)	34	63
Slaughterhouse samples			
Presumably healthy	8 local slaughterhouses	50 (40 broiler, 10 ducks)	50

2.2.3 DNA extraction

DNA was isolated using the DNeasy Blood and Tissue kit (Qiagen, Germany) following the manufacturers' recommendations. For DNA extraction, 2-3 colonies were taken from the dishes and re-suspended in 180µl enzymatic lysis buffer (TE buffer with 1.2 tritionX100 + 20µl of lysozyme (200mg/ml)). The suspension was then incubated at 37˚C for 1 h. Next, 25µl of proteinase K (20mg/ml) followed by 200µl of AL buffer were added, the sample vortexed and heated at 56°C for 1 h. The quality and quantity of extracted DNA was assessed using a NanoDrop apparatus (Thermo Fisher Scientific, USA).

2.2.4 Toxin-genotyping of *C. perfringens* by PCR

PCR was performed for the detection of different toxin genes of *C. perfringens*. For that, primers described by Baums et al. (2004) used to detect the genes of alpha-, beta-, epsilon-, iota- and enterotoxin (Table 5). To detect the two variants of beta2 toxin (*cpb2*)-genes (consensus and atypical alleles), a PCR analysis using primers described by van Asten et al. 2008 was carried out (Table 5). In addition, the primer set mentioned by Baums et al. (2004) were used to determine the consensus variant of the *cpb2* gene (Table 5). To detect the necrotic enteritis Beta-like toxin (*netB*) gene, primer sequences were used according to Keyburn et al. (2008) (Table 5). The PCR cycling program consisted of one cycle at 95ºC for 15 min (initial denaturation); 35 cycles of denaturation at 95ºC for 60 sec, annealing at 55ºC for 30 sec and extension at 72ºC for 60 sec; and one cycle of final extension at 72ºC for 3 min (Mastercycler, Eppendorf). For the primer set described by van Asten et al. (2008), the annealing temperature was set at 48ºC. PCR products were separated on a 1.5% agarose gel (1.5 g of agarose dissolved in 100 ml 1xTBE buffer) for a run time of 20 min at 200 V. DNA amplicons were visualized by ethidium bromide staining (250 µg/µl, Roth) under UV light and compared to the size of the positive controls.

Table 5: Primers used for toxin-genotyping

Amplified genes	Primer	Primer sequences (5' to 3')	Annealing temp (˚C)	Amplicon size (bp)	Reference
cpa	CPA5-L	AGTCTACGCTTGGGATGGAA	55	900	(Baums et al., 2004)
	CPA5-R	TTTCCTGGGTTGTCCATTTC			
cpb	CPB-L	TCCTTTCTTGAGGGAGGATAAA	55	611	
	CPB-R	TGAACCTCCTATTTTGTATCCCA			
etx	CPETX-L	TGGGAACTTCGATACAAGCA	55	396	
	CPETX-R	TTAACTCATCTCCCATAACTGCAC			
cpi	CPI-L	AAACGCATTAAAGCTCACACC	55	293	
	CPI-R	CTGCATAACCTGGAATGGCT			
cpe	CPE-L	GGGGAACCCTCAGTAGTTTCA	55	506	
	CPE-R	ACCAGCTGGATTTGAGTTTAATG			
netB	AKP78	GCTGGTGCTGGAATAAATGC	55	383	(Keyburn et al., 2008)
	AKP79	TCGCCATTGAGTAGTTTCCC			
cpb2-Both variants	CPBeta2totalF2	AAATATGATCCTAACCAAMAA	48	548	(van Asten et al., 2008)
	CPBeta2totalR	CCAAATACTYTAATYGATGC			
Consensus *cpb2*	CPB2-L	CAAGCAATTGGGGGAGTTTA	55	200	(Baums et al., 2004)
	CPB2-R	GCAGAATCAGGATTTTGACCA			

2.2.5 Sanger sequencing of specific *cpa* PCR products

The amplified alpha toxin gene (*cpa*) PCR product of two isolates from two different farms was purified using High Pure PCR Product Purification kit (Roche, Germany). The sequencing of the alpha-toxin gene was performed in both directions using primers listed for the c*pa*-gene (Baums et al., 2004) (Eurofins Genomics GmbH, Germany). Sequence reads of forward and reverse directions were manually verified and assembled to obtain the full amplicon length using Geneious software (v10.0.9). The consensus sequence was BLASTed for the best match in the GenBank nucleotide database under default settings.

2.2.6 Multilocus sequence typing of *C. perfringens* using Sanger sequencing

A pilot study was performed with the aim to identify the within host diversity of *C. perfringens* in cases suspected for NE using MLST. This pilot analysis included twelve isolates from three birds from three different farms (Table 6). These three birds were clinically suspicious for NE and isolates were collected from the intestinal tract (duodenum, jejunum and cecum) and liver (four isolates per bird). The twelve isolates were subjected to MLST based on the partial sequences of eleven housekeeping genes (*dut*, *ddl*, *sod*, *tpi*, *dnaK*, *glpK*, *gmk*, *gyrA*, *plc*, *recA* and *groEL*) (Hibberd et al., 2011). The PCR amplification was performed using oligonucleotide primer pairs and PCR conditions according to the literature (Hibberd et al., 2011) (Table 7).

The amplification products were purified using High Pure PCR Product Purification kit (Roche, Germany). Bidirectional Sanger sequencing was performed by Eurofins Genomics (Eurofins Genomics GmbH, Germany).

Table 6: Isolates subjected to multilocus sequence typing

Farm	**A**	**H**	**I**
No. birds	**1**	**1**	**1**
No. of isolates	4 (D/J/C/L)*	4 (D/J/C/L)	4 (D/J/C/L)
Host	Layer	Layer	Layer
Presumptive diagnosis	Clinical NE	Clinical NE	Clinical NE

*D: duodenum, J: jejunum, C: cecum, and L: liver

2.2.7 MLST data analysis

Allelic profiles and sequence types were assigned using MLSTest software (Tomasini et al., 2013). MLSTest was used to calculate typing efficiency, discriminatory power (Simpson's index (Hunter and Gaston, 1988; Simpson, 1949)), number of genotypes and polymorphisms for each locus. The ratio of non-synonymous (*dN*) to synonymous (*dS*) substitutions per nucleotide site following the method described by Nei and Gojobori (Nei and Gojobori, 1986) was evaluated using the START2 programme (Jolley et al., 2001).

For comparison, representative sequences from a prior study (Hibberd et al., 2011) describing poultry *C. perfringens* isolates were imported using GeneBank accession numbers (Hibberd et al., 2011). In addition, MLST genes were extracted from the draft genome of CP4, a Canadian NE field isolate (acc. No. LIYI01000001:LIYI01000098) and included. All sequences were aligned for each locus for uniform length as previously described (Hibberd et al., 2011).

Sequence types were analyzed by eBURST software (http://eburst.mlst.net) to identify groups of related genotypes and the founding genotype in each group (Feil et al., 2004). Different definitions of ST complexes based on single, double and triple-locus variants (SLV, DLV and TLV) were inferred to show relations among sequence types. MLST sequences were concatenated for each isolate and aligned using Clustal W alignment (Thompson et al., 1994) in MEGA software (Tamura et al., 2013). A phylogeny tree based on a total of 5,565 nucleotide positions (excluding 21 positions with gaps) was created using the neighbour-joining method in MEGA software with 1,000 bootstrap replications (Tamura et al., 2013).

Table 7: Primers used for multilocus sequence typing (Hibberd et al., 2011)

Amplified genes	Primer	Primer sequences (5' to 3')	annealing temp (˚C)	amplicon size (bp)
ddl	ddl-F	AGTGGTAATTCTGTAGTTCATGCCT	55	764
	ddl-R §	GGGAATAAACTATTCTTTGTCATACCAGG		
dnaK	dnaK-F §	TGCGTAGCTGTTATGGAAGGTGGA	58	754
	dnak-R	TTGGACCAGTTGCATCAGCAGT		
glpK	glpK F	TCCTAGAGAAGGATGGGTTGAGCA	57	619
	glpK-R	CCCAGCTATTCCAGCTATAGGAAC		
recA	recA-F	GCTTTAGGAATTGGTGGAGTACCA	57	734
	recA-R	CCATATGAGAACCAAGCTCCACTC		
gyrA	gyrA-F	CAGTTTATGATGCATTAGTTAGAATGGCAC	57	888
	gyrA-R §	CTTCTAATATGTGAGCTCTCTCTTTAGCTT		
groEL	groEL-F	AGCAAGAGAAATTGAACTTGAAGATGC	58	750
	groEL-R	CCGCCTACTTCATCAGATATAACAACGCC		
tpi	tpi-F	ACTCCAATAATCGCAGGAAACTGG	57	711
	tpi-R	TGCAACTAAGCTAGCTCCACCAAC		
plc	plc-F	AGCTTATTCTATACCTGACACAGGG	57	630
	plc-R	CCTGGGTTGTCCATTTCCCATTCT		
dut	dut-F	TTAAGTATTTTGATAACGCAAC	50	441
	dut-R	CTGTAGTACCAAATCCACCACG		
gmk	gmk-F	TAAGGGAACTATTTGTAAAGCC	50	475
	gmk-R	TACTGCATCTTCTACATTATCG		
sod	sod-F	GATGCTTTAGAGCCATCAATAG	50	475
	sod-R	AATAATAAGCATGTTCCCAAAC		

§ Primer sequences were obtained by direct personal communication with the author after a discrepancy was observed between the sequence data and the primer set provided in the original study (Hibberd et al., 2011) regarding three MLST loci (*dnaK, ddl* and *gyrA*).

2.3 Development and application of a core genome-based multilocus sequence typing system for *C. perfringens*

2.3.1 Bacterial strains and whole genome sequence data retrieval

2.3.1.1 *C. perfringens* WGS data retrieval for cgMLST

Three sets of publicly available *C. perfringens* sequence data totaling 110 WGS data were included in this study. Of these, the genomes of 76 strains were previously involved in a *in silico* genomic analysis to investigate the genomic variability and phylogenetic structure of *C. perfringens* (1.5, 2.1, Figure 1, Table 3). The genomes of the other 34 strains represent isolates from poultry recently deposited in GenBank (Table 8). Below is a description of the involved data sets:

Data set I comprises the raw MiSeq sequence data of 30 avian *C. perfringens* isolates previously sequenced and described by Ronco et al. (2017) (Table 8). These strains span a time period between 1997 and 2010 and include 17 isolates derived from both healthy (n = 4) and NE infected (n = 13) chickens in Denmark (from 1997 to 2002), in addition to 13 isolates obtained from both healthy (n = 4) and NE infected (n = 9) turkeys in Finland (from 1997 to 2010). The birds from both countries were raised under conventional methods in indoor flocks (Ronco et al., 2017).

Data set II involves a set of 23 NCTC *C. perfringens* strains sequenced within the NCTC 3000 project (2.1.1). Sections 2.1, 3.1 and 4.1 describe the *in silico* analyses of these strains with the aim to investigate the genomic variability as well as the phylogenetic structure of the strains (2.1, 3.1, 4.1). Details about these strains have been described previously (2.1.1, Table 3).

Data set III includes a set of assembled *C. perfringens* genomes (n = 57) including avian isolates (n = 7) and non-avian isolates (n = 50). Of these 57 strains, Table 3 summarises the details on the 53 strains (2.1.1) which were also included in the investigation on the genomic variability and the phylogenetic structure (2.1, 3.1, 4.1). The additional genomes of four strains were of poultry origin and recently released to NCBI (Table 7). Metadata of the downloaded *C. perfringens* genomes were extracted from GenBank records and respective publications (Keyburn et al., 2008; Li et al., 2017b).

2.3.1.2 Selection of *C. perfringens* strains for whole genome sequencing

A selection of 50 *C. perfringens* strains isolated from poultry (layer chicken, broiler chicken and ducks) in Egypt was selected for WGS (Table 9 and 4). These isolates comprise I) 23 isolates derived from cases suspected for necrotic enteritis disease from 14 different farms including the clinical disease (n = 17) and subclinical presentation (n = 6), II) 21 isolates collected from

clinically asymptomatic birds at slaughterhouses and III) 6 *C. perfringens* isolates from retail chicken meat parts. These isolates were obtained from samples collected from Sharkia and Dakahliya governorates in Egypt during the study period 2014 to 2015.

In summary, 160 WGS data (50 sequenced in the study [see above] and 110 publicly available WGS data [2.3.1.1]) were used to develop and test the applicability of a cgMLST scheme. Of these, 87 strains were poultry *C. perfringens* isolates representing the following geographic origins: Egypt = 50 strain, Denmark = 17 strains, Finland = 13 strains, USA = 3 strains, Canada = 1 strain, Australia = 1 strain, and Czech Republic = 2 strains. They were collected between 1997 and 2016. Table 10 summarises all strains included in the study.

Table 8: Information about the retrieved *C. perfringens* WGS data

Strain	Accession numbers	collection date	Isolation source	Host	Location	Health status
Retrieved raw Illumina MiSeq data of 30 poultry *C. perfringens* strains (Ronco et al., 2017)						
T18	SRR4434747	2009	Intestine	Turkey	Finland	Healthy
T22	SRR4434748	2009	Intestine	Turkey	Finland	Healthy
T14	SRR4434749	2006	Intestine	Turkey	Finland	NE
T16	SRR4434750	2008	Intestine	Turkey	Finland	NE
T6	SRR4434751	2005	Intestine	Turkey	Finland	NE
T11	SRR4434752	2006	Intestine	Turkey	Finland	NE
T1	SRR4434753	1998	Intestine	Turkey	Finland	NE
T5	SRR4434754	2005	Intestine	Turkey	Finland	NE
T34	SRR4434755	2009	Intestine	Turkey	Finland	Healthy
T43	SRR4434756	2009	Intestine	Turkey	Finland	Healthy
T46	SRR4434757	2010	Intestine	Turkey	Finland	NE
T53	SRR4434758	2010	Intestine	Turkey	Finland	NE
T84	SRR4434759	2011	Intestine	Turkey	Finland	NE
C7	SRR4448028	2001	Intestine	Chicken	Denmark	Healthy
C8	SRR4448029	2001	Intestine	Chicken	Denmark	Healthy
C1	SRR4448030	2001	Intestine	Chicken	Denmark	Healthy
C3	SRR4448031	2001	Intestine	Chicken	Denmark	Healthy
C33	SRR4448032	1998	Liver	Chicken	Denmark	NE
C32	SRR4457397	1998	Liver	Chicken	Denmark	NE
C37	SRR4457398	1999	Intestine	Chicken	Denmark	NE
C36	SRR4457399	1998	Liver	Chicken	Denmark	NE
C125	SRR4457400	2000	Intestine	Chicken	Denmark	NE
C124	SRR4457401	1997	Liver	Chicken	Denmark	NE
C27	SRR4457402	1997	Liver	Chicken	Denmark	NE
C26	SRR4457403	2000	Intestine	Chicken	Denmark	NE
C25	SRR4457404	2000	Intestine	Chicken	Denmark	NE

Strain	Accession numbers	collection date	Isolation source	Host	Location	Health status
C24	SRR4457405	2001	Intestine	Chicken	Denmark	NE
C31	SRR4457406	1998	Liver	Chicken	Denmark	NE
C48	SRR4457407	2002	Intestine	Chicken	Denmark	NE
C41	SRR4457408	2002	Intestine	Chicken	Denmark	NE
Retrieved genome data of poultry *C. perfringens* recently released in NCBI						
An68	NZ_NFHR00000000	2016	Caecum	Chicken	Czech Republic	Healthy
An185	NZ_NFKH00000000	2016	Caecum	Chicken		Healthy
LLY_N11	NZ_CP023410	2016	-	-	USA	Virulent
EHE-NE18	CP017106	2002	Intestine	Bird	Australia	Virulent

Table 9: Information about *C. perfringens* strains isolated and sequenced within this study

Strain	Flock No.	Bird No.	Culture source	Host	Place#	Health status (suspected)	*cpb2*	Toxin-type
Samples from diseased birds								
15S0027	A	29	Duodenum	Layer	Sh	NE**	+	A
15S0031	B	31	Jejunum	Broiler	Sh	SNE***	+	A
15S0055	C	41	Mid Intestine	Broiler	Dk	SNE	+	A
15S0056	C	42	Mid Intestine	Broiler	Dk	SNE	+	A
15S0059-1	D	45	Mid Intestine	Broiler	Dk	SNE	+	A
15S0060	E	46	Mid Intestine	Broiler	Dk	SNE	+	A
15S0069	F	51	Mid Intestine	Broiler	Sh	SNE	+	A
16S0002	G	1	Jejunum	Layer	Dk	NE	-	A
16S0003	G	1	Cecum	Layer	Dk	NE	-	A
16S0004	G	2	Duodenum	Layer	Dk	NE	-	A
16S0019	H	12	Duodenum	Layer	Sh	NE	+	A
16S0023	I	13	Duodenum	Layer	Sh	NE	+	A
16S0027	J	14	Duodenum	Layer	Sh	NE	+	A
16S0031	K	15	Duodenum	Layer	Sh	NE	+	A
16S0037	L	17	Duodenum	Layer	Dk	NE	-	A
16S0038	L	17	Jejunum	Layer	Dk	NE	-	A
16S0039	L	18	Duodenum	Layer	Dk	NE	-	A
16S0041	L	19	Duodenum	Layer	Dk	NE	-	A
16S0043	M	20	Duodenum	Broiler	Sh	NE	-	A
16S0051	M	22	Duodenum	Broiler	Sh	NE	+	A
16S0055	M	23	Duodenum	Broiler	Sh	NE	+	A
16S0057	N	24	Duodenum	Broiler	Dk	NE	+	A
16S0058	N	25	Duodenum	Broiler	Dk	NE	+	A
Samples from asymptomatic birds								
16S0071	O	H001	Caecum	Broiler	Dk	Healthy	+	A

Strain	Flock No.	Bird No.	Culture source	Host	Place#	Health status (suspected)	*cpb2*	Toxin-type
16S0072-1	O	H002	Caecum	Broiler	Dk	Healthy	-	A
16S0086-1	O	H016	Caecum	Broiler	Dk	Healthy	-	A
16S0088	O	H018	Caecum	Broiler	Dk	Healthy	+	A
16S0093-1	O	H023	Caecum	Broiler	Dk	Healthy	+	A
16S0095	O	H025	Caecum	Broiler	Dk	Healthy	-	A
16S0096	P	H026	Caecum	Duck	Dk	Healthy	-	A
16S0098-1	Q	H028	Caecum	Broiler	Dk	Healthy	+	A
16S0099-1	Q	H029	Caecum	Broiler	Dk	Healthy	+	A
16S0100-3	Q	H030	Caecum	Broiler	Dk	Healthy	-	A
16S0105-1	R	H035	Caecum	Duck	Dk	Healthy	+	A
16S0105-3	R	H035	Caecum	Duck	Dk	Healthy	-	A
16S0106-1	R	H036	Caecum	Duck	Dk	Healthy	-	A
16S0107	R	H037	Caecum	Duck	Dk	Healthy	+	A
16S0108-1	R	H038	Caecum	Duck	Dk	Healthy	+	A
16S0108-3	R	H038	Caecum	Duck	Dk	Healthy	-	A
16S0111-1	S	H041	Caecum	Broiler	Dk	Healthy	+	A
16S0112-1	S	H042	Caecum	Broiler	Dk	Healthy	-	A
16S0117-1	T	H047	Caecum	Duck	Dk	Healthy	-	A
16S0118-1	T	H048	Caecum	Duck	Dk	Healthy	+	A
16S0120-1	U	H049	Caecum	Duck	Dk	Healthy	+	A
Samples from poultry meat parts and edible internal organs								
16S0142	5	F024	Breast muscle	Broiler	Dk		+	A
16S0209	10	F091	Leg muscle	Broiler	Dk		-	A
16S0212	11	F094	Drum sticks	Broiler	Dk		+	A
16S0243e	13	F125	Chicken wings	Chicken	Dk		-	A
16S0139	4	F021	Liver	Chicken	Dk		+	A
16S0267-1	14	F149	Chicken wings	Chicken	Dk		+	A

** NE: clinical necrotic enteritis

***SNE: subclinical necrotic enteritis

#Dk, Dakahliyah; Sh, Sharkia

Table 10: Overview of all *C. perfringens* strains included in the core genome multilocus sequence typing scheme definition and validation

Host	Country	*netB* pos.	*netB* neg.	1997	1998	1999	2000	2001	2002	2005	2006	2008	2009	2010	2011	2015	2016	Total
Avian	**Isolates from avian NE**																	
	Egypt*	-	23													7	16	23
	Denmark	10	3	2	4	1	3	1	2									13
	Finland	2	7		1					2	2	1		2	1			9

Host	Country	*netB* pos.	*netB* neg.	1997	1998	1999	2000	2001	2002	2005	2006	2008	2009	2010	2011	2015	2016	Total
	USA	1	2**										2				1	3
	Canada	1						1										1
	Australia	1	-						1									1
	Total	15	33															50
	Isolates from healthy birds																	
	Egypt		21														21	21
	Finland		4										4					4
	Denmark		4					4										4
	Czech Republic		2														2	2
	Total		31															31
	Poultry meat isolates																	
	Egypt		6														6	6
	Total avian	15	70															87
	Non-avian isolates																	
Non-avian	Diverse origins		73															73
Total																		160

* Isolates were collected from cases suspected for NE disease as described earlier (see 2.2.1)

**One strain (LLY_N11) that was isolated from a healthy bird was found to be able to induce NE (Li et al., 2017a; Li et al., 2017b)

2.3.2 Extraction of genomic DNA and WGS of *C. perfringens* from poultry from Egypt

Strains were cultured on Schaedler Agar with Sheep Blood (supplemented with Haemin and Vitamin K, PB5034A Oxoid, Germany), incubated overnight at 37°C under anaerobiosis in an anaerobic chamber MACS VA500 (Don Whitley Scientific, UK). 1-5 fresh overnight colonies were then inoculated in 13 ml Selzer broth (Selzer et al., 1996) (tryptone - 30g, beef extract - 20g, glucose - 4g, L-cysteine hydrochloride - 1g in 1000 ml H2O, pH 7.2), incubated under anaerobiosis for 2.5 -3 hrs. 1.5 ml of the culture was then pelleted (13500 rpm for 10 min) and used for the extraction of genomic DNA. DNA was extracted using DNeasy Blood and Tissue Kit (Qiagen, Germany) according to the manufacturers' instructions for Gram-positive bacteria. For eluting nucleic acids, Buffer EB (10 mM Tris-Cl, pH 8.5, Qiagen, Germany) was used. The quality of DNA was assessed as follows: (I) the concentration and the purity (absorbance ratio 260/230 and 260/280) of the extracted DNA were measured using a NanoDrop spectrometer (Thermo Fisher Scientific, USA). (II) The integrity of the DNA was analyzed by agarose gel electrophoresis. (III) A Qubit 2.0 fluorometer (Life Technologies, Germany) was used to calculate the concentration of the double stranded DNA (dsDNA). The Qubit dsDNA BR Assay

Kit (Q32851, Life Technologies, Germany) was used for DNA quantification according to the manufacturers' instructions. Whole genome sequencing was performed by GATC Biotech (Konstanz, Germany) with paired-end libraries on an Illumina HiSeq platform (Illumina, USA) generating reads of 151 bp in length.

2.3.3 Genome *de novo* assembly and annotation

First, the quality of the paired-end Illumina sequence data was checked for each isolate using FastQC (Babraham Bioinformatics, Babraham Institute, Cambridge). Next, the data were processed with Sickle (v1.33) available at (https://github.com/najoshi/sickle) to trim low quality base calls using default parameters (Joshi and Fass, 2011). After trimming, a *de novo* genome assembly for all strains was performed using SPAdes version 3.9.1 (Bankevich et al., 2012) with enabled mismatch correction settings (careful option) using BayesHammer (Nikolenko et al., 2013). Given the versatile nature of *C. perfringens* in its life style and possibly in its genome, the use of a reference guided assembly or doing post assembly improvements that implement alignment of whole genome sequences to a reference genome was avoided as this may eventually omit parts of the non-reference sequences. Genome assembly evaluation as well as assembly statistics were generated using the QUality ASsessment Tool v4.3 (QUAST) (Gurevich et al., 2013). Generated contigs were filtered for a k-mer coverage threshold of 3x and a minimum contig length of 500 bp using an in-house script. The rapid genome annotation pipeline (Prokka v1.11) was used for annotation for all strains (Seemann, 2014).

2.3.4 Core genome multilocus sequence typing scheme definition

The SeqSphere+ cgMLST Target Definer (version 1.5 within the SeqSphere+ software version 4.1.9, Ridom GmbH, Münster, Germany) was used to define a fixed set of conserved genome-wide genes likely representing the species core genome (Junemann et al., 2013). The cgMLST Target Definer filters and extracts candidate genes from a reference strain (seed genome) querying the exact presence of these genes in a number of strains (query genomes) using BLAST algorithm (Altschul et al., 1997). Thus, the shared loci between the selected isolates (100% presence with identity >90% and coverage =100%) are considered as cgMLST targets.

2.3.4.1 Reference genome selection and filtration criteria

The complete sequence of the type strain ATCC 13124 (accession no. CP000246, size=3.2 Mb, CDS = 2,876) was used as reference (seed) genome to create the cgMLST scheme (Myers et al., 2006). The strain was reported to have no plasmid sequence. By default, SeqSphere+ applies the following filtering parameters to filter the cgMLST targets: (I) a "minimum length filter" to discard genes of length less than 50 bp, (II) a "start codon filter" and "stop codon filter" to discard genes that lack one start and /or one stop codon, (III) a

"homologous gene filter" for discarding genes that occur in multiple copies at a sequence identity of 90% and more than100 bp coverage, and (IV) a "gene overlap filter" that if 2 genes overlap by more than 4 bp length, the shorter gene will be discarded. Additionally, these reference genes were BLASTed against a set of 90 *C. perfringnes* plasmids that were either fully constructed from the complete genomes (2.1.2, additional data 1) or independently sequenced and deposited in the GenBank. Genes with a BLAST match to any of the plasmid genes at 90% identity were excluded.

2.3.4.2 Query genomes selection

To determine the core genes, the genes of the reference strain are BLASTed against a number of genomes called "query genomes". For the query genomes, a set of 38 *C. perfringens* genome data were selected that I) represented different phylogenetic clusters of *C. perfringens*, II) were publically available in the GenBank, and III) when possible, represented a closed genome. Query genomes included 26 closed genomes and 12 draft assemblies that belong to three major phylogroups, including phylogroup I (12 strains), phylogroup II (one strain) and phylogroup III (23 strains) in addition to 2 outgroup strains (3.1.2, additional data 2). The target genes extracted from the seed genome were compared to these query genomes using BLAST version 2.2.12 under default parameters (90% identity and 100% coverage). BLAST options included match reward 1, word size 11, gap open costs 5, mismatch penalty -1 and gap extension cost 2. Under default parameters, the SeqSphere+ (Ridom GmbH, Münster, Germany) applies "query genomes filters" to exclude genes found in more than 80% of the query genomes with missing or wrongly placed stop codons (default Stop Codon Percentage Filter). The SeqSphere+ software assigns alleles to each gene and generates an allelic profile for all strains.

2.3.4.3 Initial evaluation and refinement of the cgMLST targets

The output of the SeqSphere+ cgMLST Target Definer (Junemann et al., 2013) represents the core genes found in the 38 strains and were filtered for truncated, repetitive and overlapped genes. These genes were evaluated in order to keep only the complete genes that are most likely stable in the *C. perfringens* population. For this purpose, an initial data set of 80 genomes was used including the above-mentioned 38 genomes (data sets II and III described in 2.3.1.1, additional data 2). The genomes in this data set are publically available and highly divergent in terms of ecology, geography and isolation time. These genomes were processed using the SeqSphere+ built-in BLASTn searching each of the defined core genes in the assembled genomes with "required identity to reference sequence of 90%" and "required aligned to reference sequence with 99%". Only the complete genes were assigned to allele numbers, while genes with frameshifts or in frame stop codons, genes without a single start and stop

codon or genes with ambiguities were reported as failed targets. In-frame multiple insertions or deletions (indel) were allowed up to five codons per gene relative to the reference genes. Uncalled genes (due to real absence or failed allele assignment) in more than 5 % from the initial data set were excluded. This final set of the filtered core genes served as cgMLST targets. To detect recombination in the cgMLST loci, the Pairwise Homoplasy Index (PHI) (Bruen et al., 2006) was calculated for the final set of cgMLST loci using the PHIPack software (http://www.maths.otago.ac.nz/~dbryant/software/PhiPack.tar.gz).

2.3.5 Application of the core genome multilocus sequence typing scheme

According to the final set of the cgMLST targets, the poultry specific *C. perfringens* strains were investigated with regard to the clinical disease, geography and the NetB toxin gene carriage. For that, a set of 80 poultry *C. perfringens* strain data was used (Table 8 and 9). These strains comprise (I) 50 *C. perfringens* isolates from Egypt (2.3.1.2); (II) 30 *C. perfringens* from healthy and NE infected chickens from Denmark and Finland (data set I in 2.3.1.1). Allelic profiles were assigned to all strains (n = 160) according to the final cgMLST targets.

2.3.6 Whole genome based SNP detection

To detect SNP variations in the whole genome, the full data set of 160 *C. perfringens* strains was used and SNP detection using the RedDog pipeline v1beta.10.3 (Edwards et al., 2015) available at (https://github.com/katholt/reddog) was performed. The pipeline implements a mapping of Illumina reads to a reference genome followed by a series of steps to extract high quality SNPs. However, Illumina reads were not available for 61 genomes that were derived from GenBank. To include these genomes in the analyses, simulated reads (one million 100 bp reads) were generated from these 61 genomes using the wgsim command (https://github.com/lh3/wgsim) in the SAMTools package (Li et al., 2009a) as previously described (Holt et al., 2015). The genome of the strain ATCC 13124 (accession no. CP000246) (Myers et al., 2006) was used as reference genome. Paired-end reads as well as simulated reads were mapped to the reference sequence using Bowtie 2 version 2.3.0 with a sensitive-local algorithm and a maximum insert length of 2000 (set via -x option) (Langmead and Salzberg, 2012). In each isolate, SAMTools v1.4-14 (Li et al., 2009a) was used to call SNPs with Phred quality score ≥ 30 and read depth ≥5x. Ambiguous SNPs including heterozygous allele calls were discarded. In addition, repeat sequences and prophage regions were identified in the reference strain using nucmer alignment (Kurtz et al., 2004) and phaster (Arndt et al., 2016), respectively. SNPs located in these regions were excluded. The whole genome SNP sites that were present in the coding and non-coding sequences and were conserved in 99% of the genomes were selected. RAxML version 8.2.10 was used to generate a ML

phylogenetic tree using the general time reversible (GTR)-gamma model and 100 bootstrap replicates (Stamatakis, 2014).

2.3.7 Comparison between the cgMLST and the whole genome SNP typing

To visualize the topological concordance between the cgMLST neighbor joining tree (generated using SeqSphere+) and the SNP-based maximum likelihood tree (2.3.6), a tanglegram algorithm was applied (Scornavacca et al., 2011) using the software Dendroscope v.3.2.1027 (Huson and Scornavacca, 2012). The tanglegram method allows for a direct comparison between both trees via plotting them side by side and matching the corresponding taxa. Prior to generating tanglegrams, both trees were rooted with the strain CBA7123.

With the aim to generate a recombination filtered SNP tree, Gubbins v.2.2.1 was used to identify and mask putative recombination regions from the genome alignment (Croucher et al., 2015). As input for Gubbins, a pseudomolecule was constructed for each strain composed of the reference sequence (ATCC 13124) but updated with sample-specific SNPs using snpTable2GenomeAlignment.py available at (https://github.com/katholt/reddog).

To compare the discrimination levels and assess the congruence of the typing results using the cgMLST as well as the recombination filtered and unfiltered SNP typing methods, the Simpson's index of diversity (Hunter and Gaston, 1988; Simpson, 1949) and the adjusted Wallace index of concordance (Severiano et al., 2011) as calculated using the Comparing Partitions tool (Carrico et al., 2006) was applied. The Simpson's diversity index (Hunter and Gaston, 1988; Simpson, 1949) is a quantitative measure of diversity which denotes the probability that two individuals selected randomly from a sample will be classified into two different types. The value of the Simpson's diversity index ranges between zero and one, where one represents the maximum diversity in the sample. The adjusted Wallace's coefficient (Severiano et al., 2011) measures the probability that if two individuals classified together using method A, they will also share the same classification type by method B.

2.3.8 Comparison of the cgMLST typing to classical MLST typing methods

Classical MLST schemes have been previously described for *C. perfringens* (Deguchi et al., 2009; Hibberd et al., 2011; Jost et al., 2006b). To compare the resolution obtained by cgMLST to the resolution of classical MLSTs, the MLST genes described in the classical MLST studies (Deguchi et al., 2009; Hibberd et al., 2011; Jost et al., 2006b) were extracted from the whole genome data of the 160 strains using BLASTn (Altschul et al., 1997). Three classical MLST schemes previously described for *C. perfringens* were involved in this analysis. The MLST genes described by Deguchi et al., (2009) include *groEL*, *gyrB*, *nadA*, *pgk*, *sigK*, *sodF*, *plc* and *colA,* the genes described by Jost et al., (2006) include *cpa*, *ddlA*, *dut*, *glpK*, *gmk*, *recA*, *sod*

and *tpiA*, while the MLST genes described by Hibberd et al., (2011) include *dut*, *ddl*, *sod*, *tpi*, *dnaK*, *glpK*, *gmk*, *gyrA*, *plc*, *recA* and *groEL*. For each MLST locus, the extracted sequences from the 160 genomes were aligned using MAFFT (Katoh and Standley, 2013) then assigned allelic profiles and sequence types using the MLSTest software (Tomasini et al., 2013). MLSTest (Tomasini et al., 2013) was also used to calculate the typing efficiency, the number of genotypes and the polymorphisms for each locus. The Simpson's Index of Diversity (Hunter and Gaston, 1988; Simpson, 1949) and adjusted Wallace coefficients (Severiano et al., 2011) were calculated using the Comparing Partitions tool (Carrico et al., 2006) to assess the discriminatory power and to evaluate the congruence of the typing results between the cgMLST and the classical MLSTs methods. The topological concordance of the neighbor joining trees from the cgMLST and classical MLSTs was visualized using the tanglegram algorithm (Scornavacca et al., 2011) as implemented in the software Dendroscope v.3.2.1027 (Huson and Scornavacca, 2012). All trees were rooted with the strain CBA7123 before generating the tanglegrams.

2.3.9 Whole genome SNP and cgMLST population structure

Phylogeny dependent and independent approaches were used to define the population structure for the investigated 160 genomes of *C. perfringens*. As a phylogeny free approach, the BAPS (Bayesian Analysis of Population Structure) algorithm (Corander et al., 2008) was used, particularly its module hierBAPS (Cheng et al., 2013) to define the population structure based on the SNP variations in the whole genome. The software hierBAPS (Cheng et al., 2013) initially clusters the sample population and then again iteratively applies the BAPS algorithm on each of the resulting cluster thus evaluating for different hierarchal levels of clustering. In parallel, a phylogeny based analysis was inferred via construction of a phylogenetic network for the 160 strains using SplitsTree (Huson and Bryant, 2005) (NeighbourNet method (Bryant and Moulton, 2004)). As input for SplitsTree, two distance matrices were used, one was calculated using SeqSphere+ (Junemann et al., 2013) based on the cgMLST allelic profiles between the genomes and the other was calculated using MEGA7 (Kumar et al., 2016) based on the whole genomes SNP differences.

To support the results of the distinct separation of different lineages from the population structure analysis, the number of SNP variation between and within distinct lineages was calculated. For that, the distance matrices from the cgMLST and whole genome SNPs (see above) were converted into a column format using an in-house script. Then, the R package: ggplot2 (Wickham H., 2016) (function geom_density) was used to create a plot showing the distribution of the pairwise SNP differences between each pair of genomes. The values of the pairwise allelic differences and the corresponding values of the pairwise SNP variation between pairs of genomes were also plotted using the ggplot2 library in R (Wickham H., 2016).

Chapter 3: Results

3.1 *In silico* analysis of the genomic variability, phylogenetic relatedness and virulence genes of *C. perfringens*

3.1.1 *C. perfringens* genome overview, mobile genetic elements and genome alignment of 30 closed genomes

3.1.1.1 Genome overview

The *de novo* assembly of raw PacBio sequence data of the 23 NCTC *C. perfringens* genomes (2.1.1) yielded 20 circularized chromosomes with a mean final coverage for each strain ranging from 116 to 275x (2.1.2). In addition, a panel of 45 extrachromosomal elements, 32 of them circularized were identified (Table 11, 2.1.2).

Considering only closed genomes (20 assembled to closure in this study and 10 available at the NCBI, Table S1) revealed that the genome of *C. perfringens* is composed of a circular chromosome of variable size from 2.9 to 3.5Mbp (average = 3.12Mbp) with up to six (range 0 to 6) extrachromosomal elements (plasmids or episomal phages) (Table S1). The food poisoning (chromosomal *cpe*) strains (14 out of the 30 closed genomes) had a consistently smaller genome size (≤ 3Mbp) compared to other strains (Table S1). The *C. perfringens* chromosome contains on average 3,126 protein coding sequences (CDS) (Table S1). The number of CDS was found to be a function of the genome size i.e. larger genomes carry more CDS and vice versa (Table S1). The calculated coding density (the size of coding regions over the genome size) for the strains was ~83%. Members of the species *C. perfringens* belong to bacteria with low G+C content (~28%) and carry 10 rRNA operons and on average 93 tRNAs (Table S1, 2.1.2). The type strain (ATCC 13124) carries however, only eight rRNA operons (Myers et al., 2006). Plasmids of *C. perfringens* vary in size from 2.4Kbp to 404Kbp and some of them harbor the conjugation locus of the species (transfer of clostridial plasmids; *tcp*) facilitating plasmid spread within the species (Table S1). Plasmids contribute an average of 125 coding sequences (from 19 to 704 CDS) constituting up to 18% of the coding capacity of the *C. perfringens* genome (Table S1, 2.1.2).

Table 11: Results of *de novo* genome assembly for PacBio sequence data of 23 NCTC *C. perfringens* strains

						Extrachromosomal elements											
				Chromosome		A		B		C		D		E		F	
Strain	SRA Run	Toxin-type	Toxin genes	Status	Size (Mbp)	Status	Size (Kbp)	Status	Size (Kbp)	Status	Size (Kbp)	Status	Size (Kbp)	Status	Size (Kbp)	Status	Size (Kbp)
NCTC8678	ERR1377187	A	*cpa*	●	2.9	◐	38	◐	4.7								
NCTC8797	ERR1377188	F	*cpa, cpe*	●?	2.9	◐	25.3	◐	19								
NCTC13170	ERR1377189	A	*cpa*	●	3.29	-	-										
NCTC8503	ERR1407347	D	*cpa, etx*	◐	3.4 (6 ctgs*)	●§	54.6	●	19.3								
NCTC2544	ERR1456745	A	*cpa*	●	3.18	-	-										
NCTC8799	ERR1466824	F	*cpa, cpe*	●	2.9	◐	63.1	◐	10.2	◐	7.8						
NCTC10612	ERR1588634	F	*cp, cpe*	●	3.0	●	13.6	◐	8.8								
NCTC3182	ERR1599940	C	*cpa, cpb*	◐	3.5 (1ctg)	●	49.4	●§	59.1	●	37.2						
NCTC8238	ERR1656456	F	*cp, cpe*	●	2.9	●	56.6										
NCTC8246	ERR1656458	A	*cpa, cpb2*	●	3.3	●	54.3										
NCTC2837	ERR1656459	A	*cpa*	●	3.3	-	-										
NCTC8081	ERR1656460	C	*cpa, cpb, cpe*	●	3.1	●§	116.6	●	73.5	●§	67.9	●	58.1	●	53.7	◐	20.7
NCTC8247	ERR1674568	F	*cpa, cpe*	●	2.9	◐	57.02										
NCTC8239	ERR1681948	F	*cpa, cpe*	●	2.9	●	56.6										
NCTC9851	ERR1681949	F	*cpa, cpe*	●	2.9	●	56.6	●	12.3								
NCTC10239	ERR1681950	F	*cpa, cpe*	●	2.9	●	57.9	●	12.4	●	12.2						

						Extrachromosomal elements											
				Chromosome		A		B		C		D		E		F	
Strain	SRA Run	Toxin-type	Toxin genes	Status	Size (Mbp)	Status	Size (Kbp)	Status	Size (Kbp)	Status	Size (Kbp)	Status	Size (Kbp)	Status	Size (Kbp)	Status	Size (Kbp)
NCTC10240	ERR1681951	F	*cpa, cpe*	●	2.9	●	56.6	◐	9.7								
NCTC8679	ERR1787549	F	*cpa, cpe*	●	2.9	●	54.4	●	12.1								
NCTC10578	ERR1787550	A	*cpa*	●	3.3	-	-										
NCTC10613	ERR1787552	F	*cpa, cpe*	◐	2.94 (ctgs)	●	56.6	●	38.1	●	12.2	●	12.3	◐	8.9		
NCTC10614	ERR1800584	F	*cpa, cpe*	●	2.9	●	56.6	●	12.2								
NCTC8359	ERR1805687	F	*cpa, cpe*	●?	2.9	●	56.6	●	13.4	●	12.2						
NCTC11144	ERR1954484	F	*cpa, cpe*	●	3.2	●§	78	●	58.6	●	50.9						

*ctgs; contigs: a contiguous fragment of DNA

● Circularized finished genetic element

●? The genome is circularized but there is uncertainty about its finished status as deep valleys (assembly area with little support of PacBio reads) were noticed in the coverage plot of the genome of strain NCTC8797 that may indicate a misassembly issue (additional data 3). Also, the genome of strain NCTC8359 showed a large spike in the coverage plot which may indicate collapsed repeats (additional data 3)

◐ Draft genetic element

●§; indicates possible conjugative plasmids carrying known significant toxin genes (epsilon, beta, entero-toxin genes) and the species conjugation locus

3.1.1.2 CRISPR elements

Within the 30 closed genomes, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) elements were not detected in 14 strains (Table S1, 2.1.2). Fifteen strains carry a CRISPR-Cas system of either class I (similar to class I-B, described for *Clostridium klyveri)* or class II (similar to class II-C, reported in *Neisseria lactamica*) based on a recent CRISPR-Cas classification (Koonin et al., 2017) (Table S2, 2.1.2). Strain NCTC 10578 was predicted to harbor a CRISPR-Cas system of type I and type II as well as one additional CRISPR repeat flanked by a transposase gene (IS605 family) (Table S2, 2.1.2).

3.1.1.3 Mobile genetic elements

A search for mobile genetic elements (integrated phages, insertion sequences [IS], and genomic islands [GI]) was done within the 30 closed genomes (2.1.2, Table S3 and S4).

3.1.1.3.1 Prophages

Prophages were observed with an average number of two (range 1 to 6) and a total average size of 60.23Kbp (range 17.6 to 192Kbp) per chromosome, representing 1.9% (range 0.5 to 5.4%) of the chromosome size (Figure 3A and Table S4, 2.1.2). The G+C percentage of the predicted phages was 28.4% (range 26.1 to 31.8%). Prophage size varied between 7 and 58.4Kbp (average = 30Kbp). No direct correlation could be observed between CRISPR absence and prophages content per chromosome. The most commonly predicted phage was ΦCP51 (intact = 6, incomplete = 1, and questionable = 5) (Gervasi et al., 2013) (Table S4).

3.1.1.3.2 Insertion sequences

Prediction of IS elements was performed using ISEscan (2.1.2) (Xie and Tang, 2017) and revealed a highly variable occurrence of IS elements in the genomes with an average number of 48 IS elements (range 5 to 141) per chromosome and an average of 51 IS elements (range 5 to 187) per whole genome (chromosome and plasmids) (Figure 3A and Table S3). The mean total size of IS elements per chromosome was 63.2Kbp (range 8.8 to 187.8Kbp), representing 2.1% of the genome (range 0.26 to 6.1%). The major identified IS family was IS200/IS605 (n = 433) followed by IS30 (n = 316), IS6 (n = 303), IS 607 (n = 172), IS3 (n = 72), IS21 (n = 69), IS256 (n = 45) and others (Table S3). The number and size of ISs was significantly higher for the chromosomal *cpe* strains as well as the Darmbrand strain (Figure 3A). In these strains, ISs constitute an average of 3.7% (range 2.9 to 6.1%) of the genome with an average total size of 109.9Kbp (range 86.1 to 187.8Kbp) and an average total number of 84 (range 65 to 141).

3.1.1.3.3 Genomic islands

Genomic islands (GI) were predicted via submitting the GenBank formatted and RAST annotated genomes (2.1.2) to the IslandViewer4 web service (Table S3, 2.1.2). As for IS elements, a high variability was also observed for the GIs (Figure 3A) which were detected at an average number of 9.8 (range 1 to 23) and size of 4.2% (range 0.26 to 10.7%), corresponding to a mean total size of 130Kbp (range 8.2 to 330Kbp) per chromosome (Figure 3A and Table S3). Chromosomal *cpe* strains and the Darmbrand strain were highly enriched in GIs (average genome percentage = 6.7%, range 4.6 to 10.7%, average number = 14, range 11 to 23, average total size = 197.7Kbp, range 135.4 to 330Kbp) (Table S3).

In total, investigated mobile DNA elements (prophages, IS elements and GIs) constitute on average 8.29 ± 4.5% (approx. 253 ± 131Kbp) of the *C. perfringens* chromosome. The Darmbrand strain genome (NCTC 8081) was the one with the most mobile elements (18% of the genome, ~571Kbp) followed by the chromosomal *cpe* strains (9.25 to 15%, 271 to 453Kbp), and then the avian necrotic enteritis strain Del1 (9.25 %, 329Kbp). Although Del1 harbors few GIs (n = 9) and ISs (n = 11), the strain carries five intact copies of prophages (192Kbp) (Table S3 and S4).

The distribution patterns of both, IS and GI across the chromosome of the 30 closed *C. perfringens* genomes was investigated (see 2.1.2). As shown in Figure 3B (left side), the distribution pattern of ISs and GIs were biased toward one replichore, notably in the small sized genomes (chromosomal *cpe* strains). While the Darmbrand strain was also highly enriched in ISs and GIs, this distribution pattern of ISs and GIs toward one replichore was not observed (Figure 3B middle part).

3.1.1.4 Multiple genome alignment

Using a multiple genome alignment approach (see 2.1.2), 320 Locally Collinear Blocks (LCBs) of more than one Kbp size shared between the 30 strains were identified (additional data 4). Of these, 51 LCBs were present in all strains which on average constitute 87% of each genome sequence. The physical locations and relative order of genes in the chromosome of most genomes (27 out of 30) were conserved (Figure 4A). Limited inversions and shifts were observed and in most cases included integrated phages or genes flanked by insertion sequences. Large inversions were observed in three strains (NCTC 11144, NCTC 8081 and NCTC 8359) (Figure 4B). In strain NCTC 11144, the inversion was bordered by rRNA operons while the inversions and shifts in the strains NCTC 8081 and NCTC 8359 were bordered by IS elements (ISCpe7).

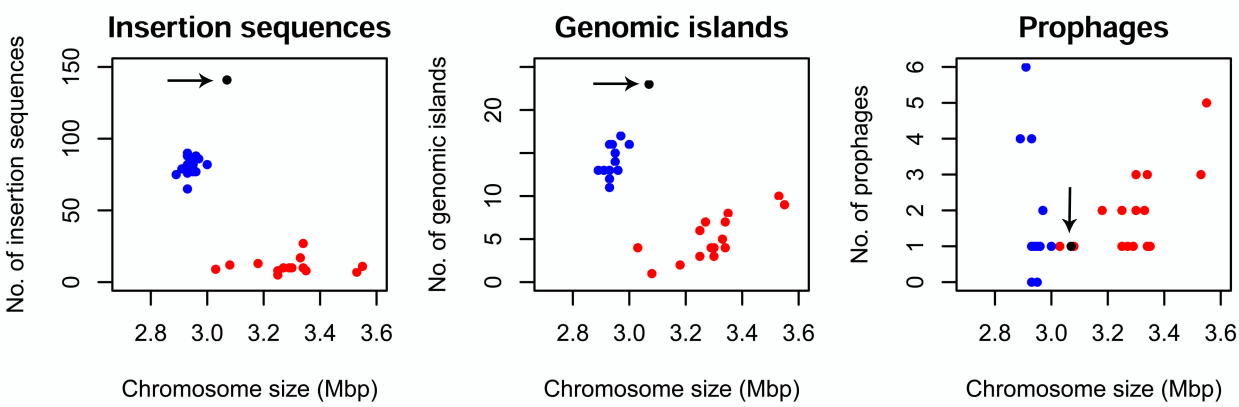


Figure 3A: Mobile genetic elements in the closed chromosome of 30 *C. perfringens* genomes. The predicted number of insertion sequences, genomic islands and prophages is plotted against the chromosome size. Human food poisoning strains carrying the enterotoxin gene on a chromosome (chromosomal *cpe* strains) are displayed in blue. Black color (marked with an arrow) represents the Darmbrand (human enteritis necroticans) strain. Other strains are colored red.

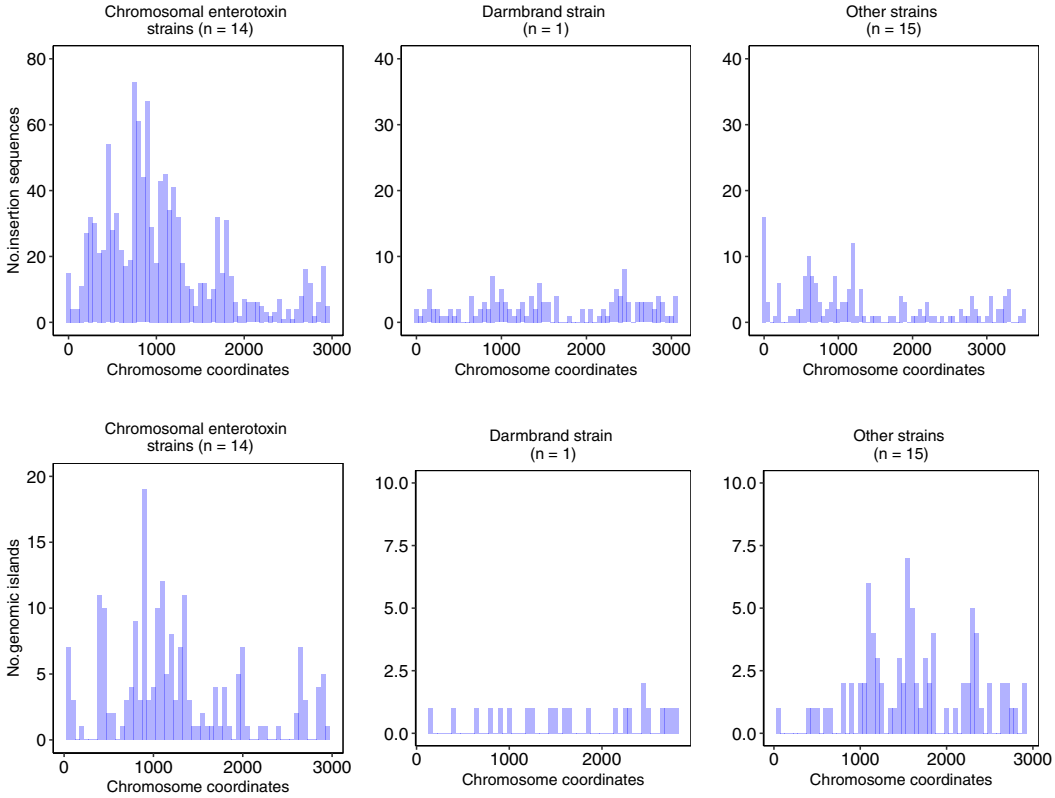


Figure 3B: The distribution of the ISs (first line) and GIs (second line) across the chromosome in the closed genomes of *C. perfringens*. An uneven distribution of IS and GI in the chromosomal *cpe* strains with an accumulation in the first chromosomal half was seen (left). The NCTC 8081 genome (middle) was also rich in the IS- and GI however, the uneven distribution of IS and GI toward a chromosomal half was not observed. Note the different scaling of y-axes. Position 1 represents the origin of replication as previously determined (Shimizu et al., 2002).

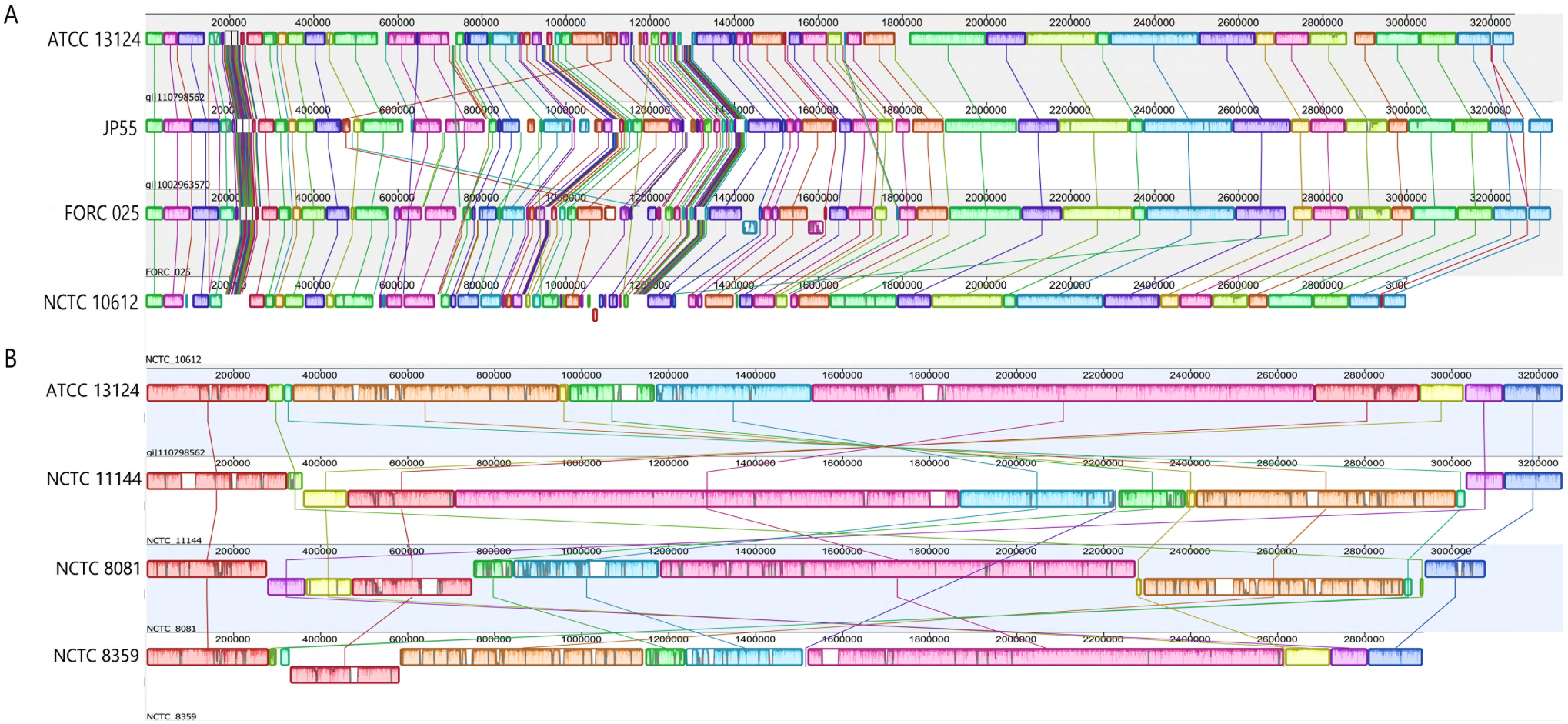

A
ATCC 13124
gi|110798562
JP55
gi|1002963570
FORC 025
FORC_025
NCTC 10612
B
NCTC_10612
ATCC 13124
gi|110798562
NCTC 11144
NCTC_11144
NCTC 8081
NCTC_8081
NCTC 8359
NCTC_8359

Figure 4 (page 57): Whole genome alignment of *C. perfringens* strains. (A) A representative detail of the multiple genome alignment depicting genomic conservation within the strains in respect to the physical location of Locally Collinear Blocks (LCBs) and relative order of the genes. Please see additional data 4 for a detailed alignment of the 30 genomes **(B)** Alignment of the *C. perfringens* genomes showing inversions and shifts (NCTC 11144, NCTC 8081 and NCTC 8359) relative to the reference strain (ATCC 13124). In strain NCTC 11144, a large inversion bordered by rRNA operons was observed while the inversions and shifts in the strains NCTC 8081 and NCTC 8359 were bordered by IS elements (ISCpe7). Position 1 in all strains corresponds to the origin of replication as determined in strain 13 (Shimizu et al., 2002).

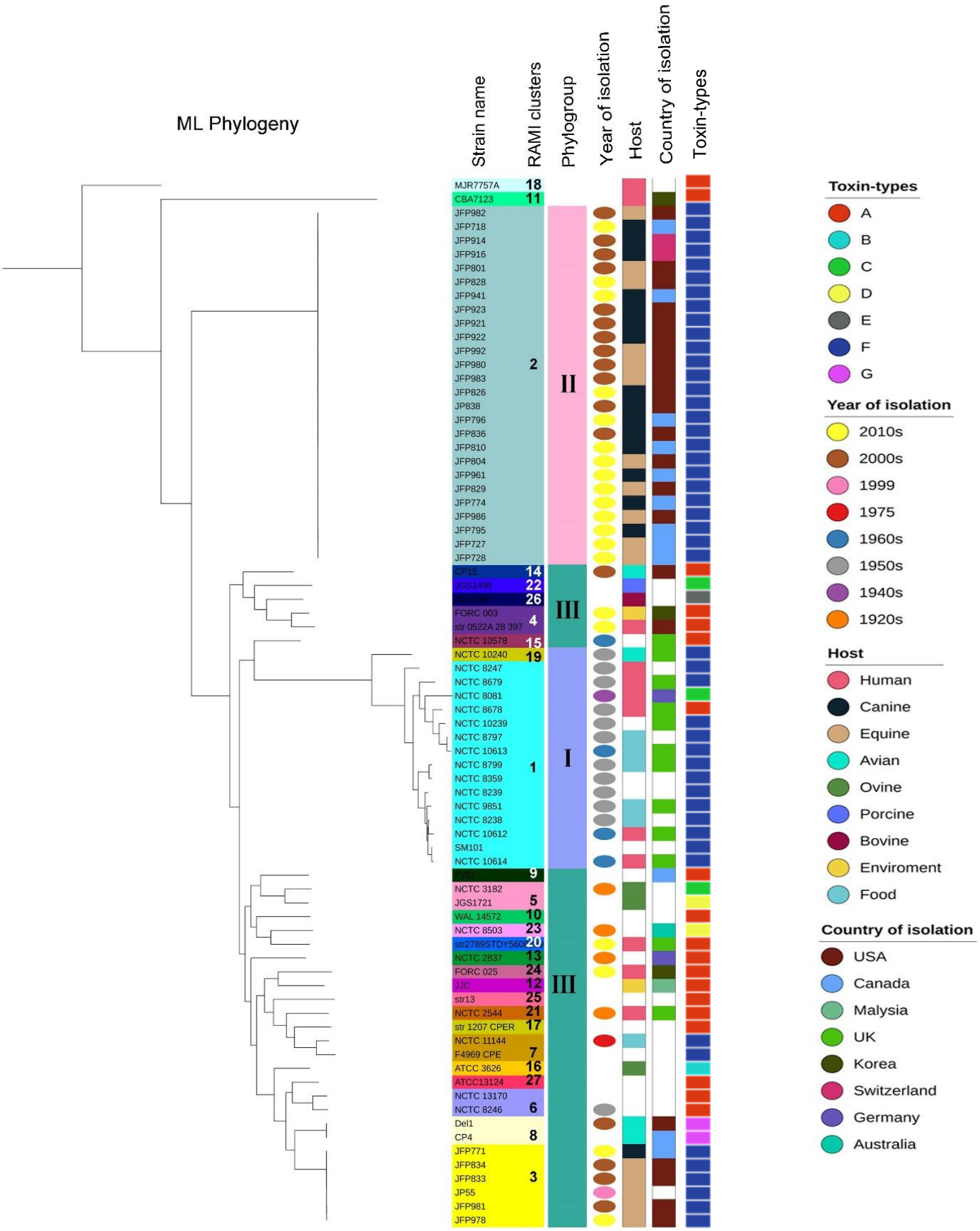
3.1.2 Phylogenetic relationships among *C. perfringens* genomes

The data set of 76 *C. perfringens* genomes (23 genomes sequenced in the NCTC 3000 project plus 53 genomes available at the NCBI; Table 3) was included to calculate the core genome (2.1.3, Figure 2B). For this purpose, the software GET_HOMOLOGUES was used with its implemented three algorithms: OrthoMCL, COGtriangles, and bidirectional best hit (2.1.3) which yielded 1,293, 1,310 and 1,246 orthologous gene families, respectively, shared by all isolates. 1,207 gene families (1,097,891bp; 37% of the average genome) were predicted simultaneously by the three algorithms (additional data 5) and were used to construct a maximum-likelihood (ML) tree (Figure 5, 2.1.3). Clusters were assigned based on patristic distance using RAMI tool (2.1.3) which identified 27 distinct clusters (Figure 5). Of these, clusters 1 - 3 include 48 strains; cluster 1 comprises 15 strains (mostly chromosomal *cpe* strains and the Darmbrand strain); cluster 2 comprises 16 strains (*netF* positive strains, including the complete genome JP838); cluster 3 comprise 6 strains (*netF* positive strains, including the complete genome JP55). The other clusters include five clusters of two strains each, and 19 clusters containing a single strain (Figure 5). The mean average nucleotide divergence within the clusters is 0.22% (calculated across the core genome within clusters ≥ 2 strains) and 1.4% between the clusters (additional data 6, Figure S1A, 2.1.3). This distinct clustering of the strains is further supported by the whole genome Average Nucleotide Identity (ANI) (calculated pairwise via identifying the matching regions across the whole genome) (2.1.4). The main three clusters showed a larger ANI (median ANI = 99.7%) compared to others (median ANI = 97%) (Figure S1B). Though these approaches are alignment based, the relatedness between strains was also inferred in an alignment independent manner (k-mer based analysis; 2.1.5). SNPs were identified and the number of SNP differences between all pairs of genomes were calculated revealing lower SNP distances between genomes of the same group (Figure S1C).

A set of 83,316 SNPs were additionally extracted from the resulting core genome and used to construct a phylogenetic network (Figure 6, 2.1.3, additional data 7). The analyzed 76 strains were grouped into three distinct phylogroups (I, II and III). The phylogroups match up with the clusters derived from the core genome ML tree including cluster 1 (mostly chromosomal *cpe* genomes) and cluster 2 (*netF* genomes related to JP838 strain) (Figure 5 and Figure 6). Phylogroup III collects most of the *C. perfringens* strains with a deep (separate) branching toward the root (Figure 6). Two strains (CBA7123 and MJR7757A) form separate branches and have the lowest ANI (median 95.7% – 96.2%, additional data 8). The analyzed 76 *C. perfringens* genomes can be partitioned into 27 distinct clusters (based on patristic distance, referred to later as cluster) and three phylogroups (based on network analysis, later referred to as phylogroup) (2.1.3, Figure 5 and Figure 6).

Next, *In silico* multilocus sequence typing (MLST) was performed using a MLST system developed in 2009 (Deguchi et al., 2009) (2.1.6). MLST genes were extracted from the WGS data of the 76 *C. perfringens* strains under study. The gene for *sigK* (phosphoglycerate kinase) was truncated in some genomes and therefore only a reduced length of the gene was analyzed, 388bp instead of 589bp analyzed in the original study (Deguchi et al., 2009). Additionally, classical MLST data from another 71 strains (additional data 9, 2.1.6) investigated and published in prior studies were included (Deguchi et al., 2009; Ma et al., 2012; Xiao et al., 2012). The final MLST tree (Figure S2) was based on 5.07Kbp sequence alignments (2.1.6). As indicated in Figure S2, the food poisoning strains and enteritis necroticans strains (including strains of phylogroup 1) clustered together based on the MLST phylogeny (Figure S2).

Figure 5 (page 60): Phylogenetic relationships of 76 *C. perfringens* genomes. A maximum likelihood phylogenetic tree (ML phylogeny) was computed using RAxML based on 1,207 core genes. Coloration around the strain names delineates the 27 different clusters as identified based on patristic distance using the RAMI tool (Pommier et al., 2009). The other columns indicate the phylogroups (3.1.2) and strain characteristics.


ML Phylogeny
Strain name
RAMI clusters
Phylogroup
Year of isolation
Host
Country of isolation
Toxin-types
MJR7757A 18
CBA7123 11
JFP982
JFP718
JFP914
JFP916
JFP801
JFP828
JFP941
JFP923
JFP921
JFP922
JFP992
JFP980 2
JFP983
JFP826
JP838
JFP796
JFP836
JFP810
JFP804
JFP961
JFP829
JFP774
JFP986
JFP795
JFP727
JFP728
II
CP15 14
JGS1495 22
26
FORC 003
str 0522A 28 397 4
NCTC 10578 15
III
NCTC 10240 19
NCTC 8247
NCTC 8679
NCTC 8081
NCTC 8678
NCTC 10239
NCTC 8797
NCTC 10613
NCTC 8799 1
NCTC 8359
NCTC 8239
NCTC 9851
NCTC 8238
NCTC 10612
SM101
NCTC 10614
I
9
NCTC 3182
JGS1721 5
WAL 14572 10
NCTC 8503 23
str2789STDY560 20
NCTC 2837 13
FORC 025 24
JJC 12
str13 25
NCTC 2544 21
str 1207 CPER 17
NCTC 11144
F4969 CPE 7
ATCC 3626 16
ATCC13124 27
NCTC 13170
NCTC 8246 6
III
Del1
CP4 8
JFP771
JFP834
JFP833 3
JP55
JFP981
JFP978
Toxin-types
A
B
C
D
E
F
G
Year of isolation
2010s
2000s
1999
1975
1960s
1950s
1940s
1920s
Host
Human
Canine
Equine
Avian
Ovine
Porcine
Bovine
Enviroment
Food
Country of isolation
USA
Canada
Malysia
UK
Korea
Switzerland
Germany
Australia

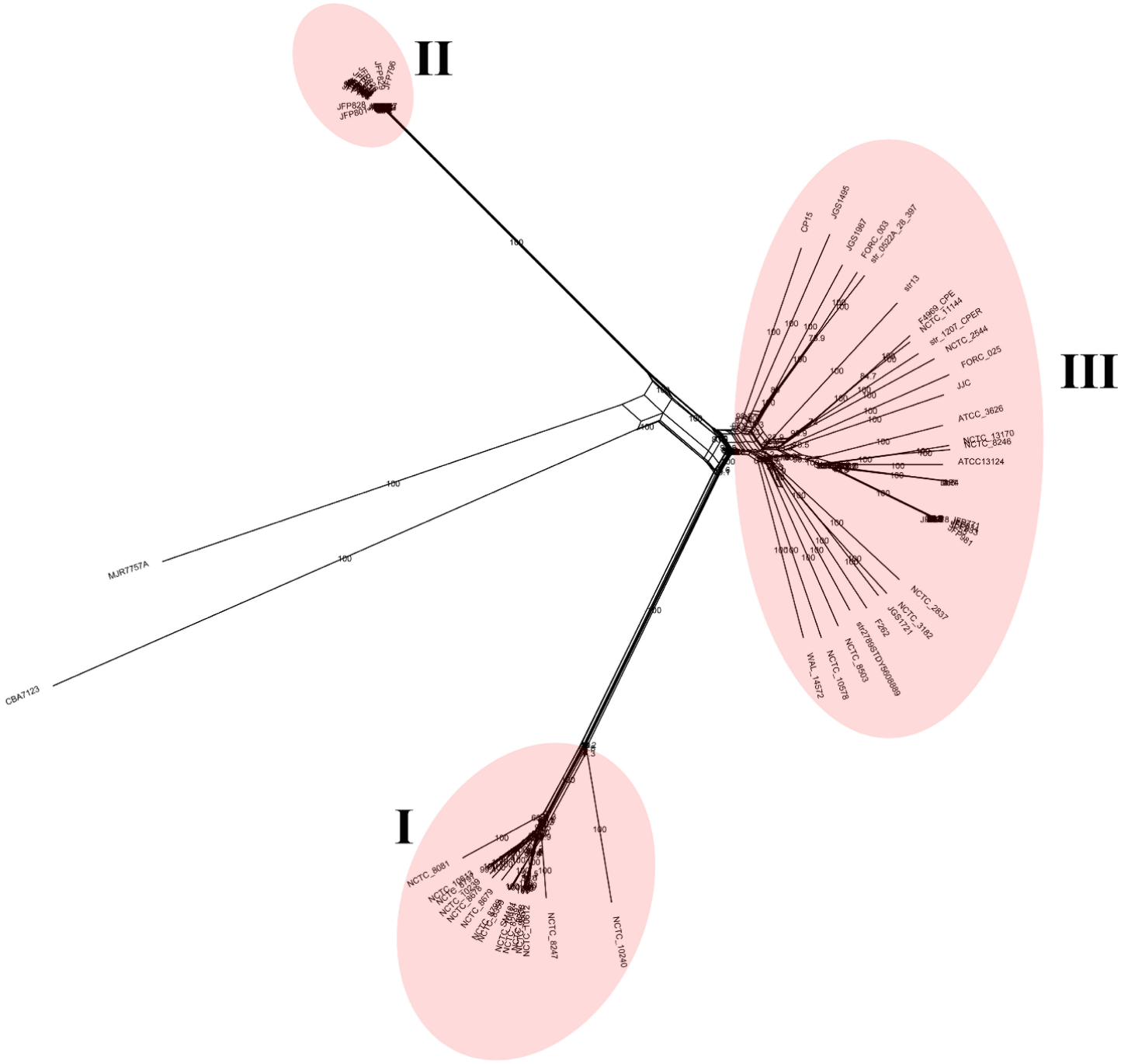


Figure 6: SplitsTree graph based on 83,316 unique SNP sites in the core genome of 76 *C. perfringens* strains. For a scalable image, please see additional data 7.

3.1.3 *C. perfringens* genome relatedness based on the accessory genes presence-absence pattern

A total of 10,098 gene clusters comprises the pangenome of the investigated 76 *C. perfringens* strains (additional data 11). These genes could be further divided into 2,057 core genes (genes present in more than 95% of the strains), 4,022 accessory genes (genes shared by less than 95% of the strains) and 4,019 strain specific genes (genes only present in one strain) (2.1.7).

Independent of SNP-based phylogenies, this study also aimed to infer the relatedness between the strains based on the pattern of gene presence-absence. For this purpose, the Jaccard distance of the accessory gene content (4,022 genes with a frequency 2-95%) was computed between each pair of genomes i.e. calculating the number of genes present in both genomes over the number of genes present in either genome (2.1.8). This allows a representation of the identity level of the accessory genome between the strains. The resulting matrix of the calculated pairwise gene content distances (2.1.8) was highly correlated to SNP-based matrices (Figure 7B and Figure S1). A ML tree of the accessory genome which was based on the binary gene presence-absence pattern was constructed (2.1.8). The resultant tree topology was in a good concordance with SNP-based trees (Figure 7A and Figure 5). Lastly, a multidimensional scaling plot was generated based on the variation in gene content between isolates (2.1.8) (Figure 7C). All strains were roughly grouped into three main categories, similar to what was inferred from the SNP-based network analysis (Figure 6).

Since the accessory gene content profiles were in good harmony with the core genome phylogeny, the next step aimed at identifying accessory genes that are distinctly associated to different phylogroups (additional data 12). At 90% specificity (genes present or absent in 90% of the target group and not present or not absent in 90% of the other strains), it was found that phylogroup I has 64 specific gene families not common in other strains but lack 304 chromosomal gene families (additional data 12). However, the phylogroup II genomes (*netF* positive strains related to JP838) carry additional 313 specific genes and lack 37 chromosomal genes (90% specificity) (additional data 12). Of note is that 49 genes were common in *netF* strains (splitted phylogenetically into phylogroup II and a cluster within phylogroup III) but are absent in 90% of the other strains (additional data 12). A list of these genes and their function is provided in additional data 12 as well as the abundance and distribution of the genes absent in the phylogroup I and phylogroup II relative to the chromosome of the type strain ATCC13124 (additional data 12).

A

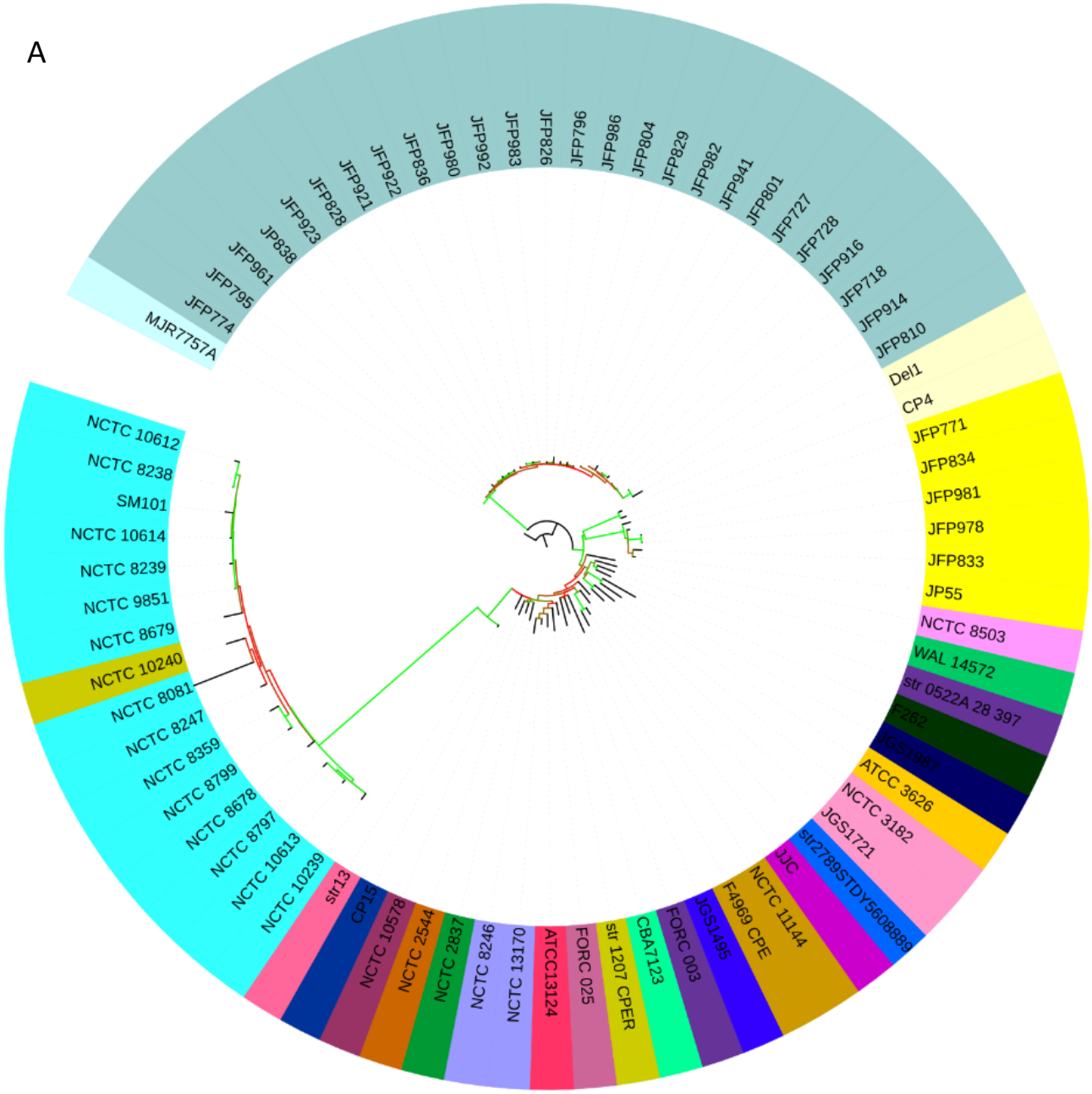


Figure 7A: A maximum likelihood phylogenetic tree of the accessory genome based on a binary pattern of gene presence/absence. Branch colors denote the values of bootstrap support from zero (red) to 100 (green).

B

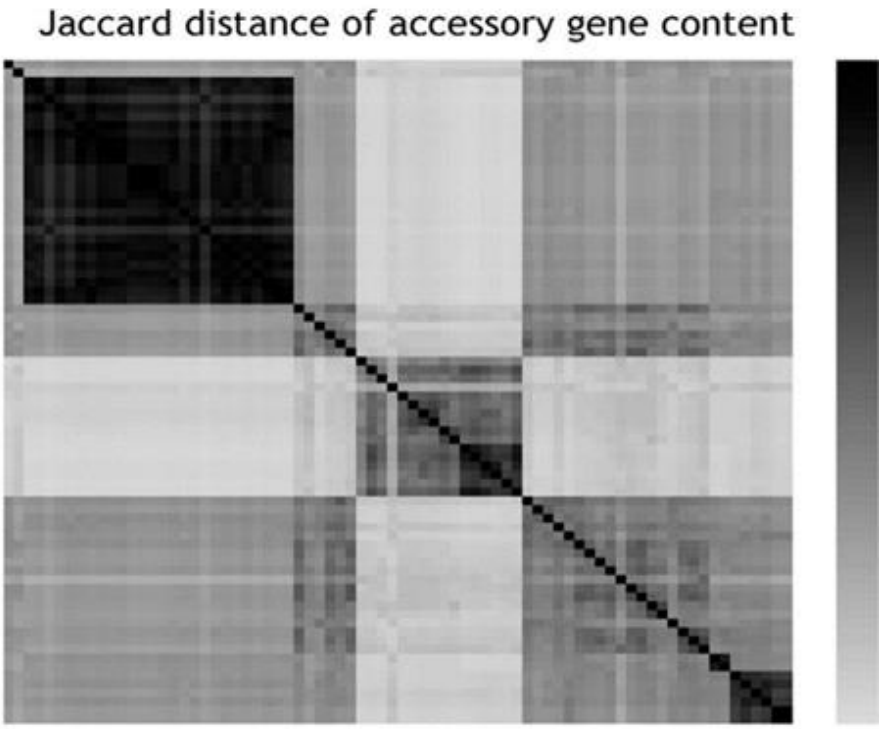


Figure 7B: Pairwise comparison of the accessory gene content calculated as Jaccard distance. Strains are ordered as in Figure 5

C

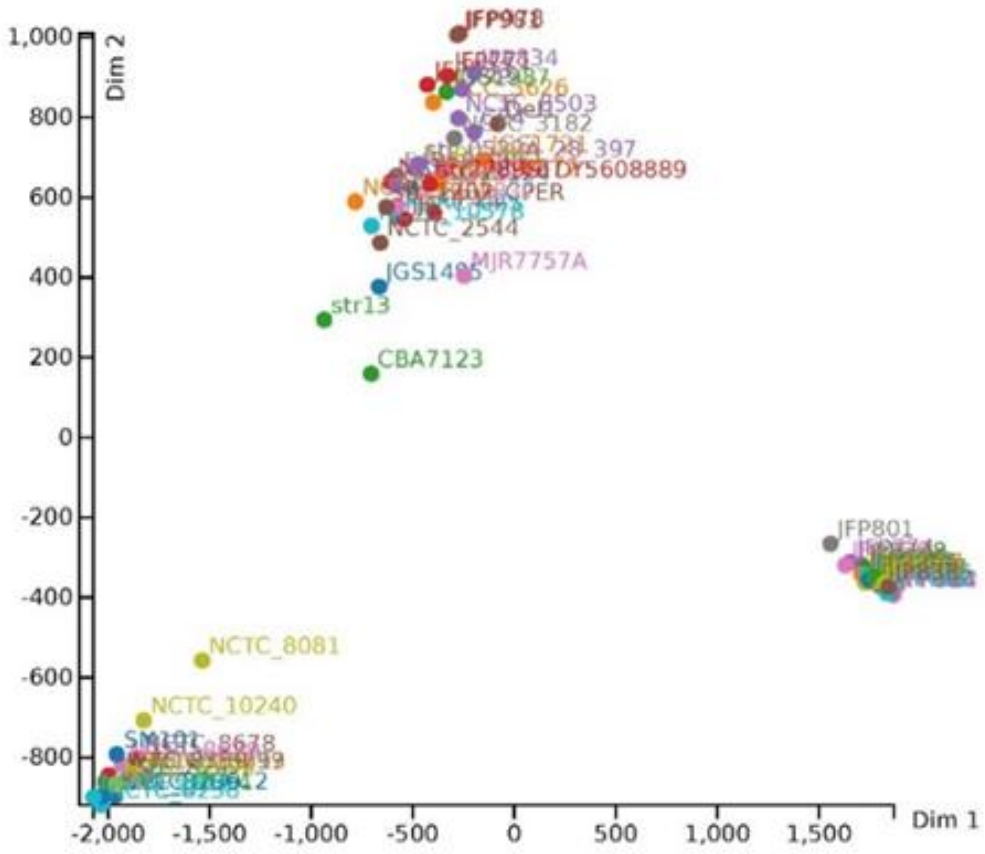


Figure 7C: Multidimensional scaling plot of the *C. perfringens* pangenome based on the gene presence/absence.

3.1.4 Identification of potential virulence factors and their distribution in the 76 *C. perfringens* genomes

Using the 76 *C. perfringens* genomes, a search for homologous genes related to virulence was performed by BLASTp searches of the pangenome against the core protein set of the Virulence Factor Database (VFDB set A). This dataset comprises non-redundant experimentally verified virulence factors in various pathogens (2.1.9) (Chen et al., 2016). Results show a set of 397 genes that share homology (of variable amino acid identity) with genes in VFDB. The distribution of the virulence factors homologs in the 76 *C. perfringens* strains is shown in additional data 13. The identified set belongs to core (123 genes), accessory (164 genes) and strain-specific genomes (110 genes) (additional data 13). The percent aminoacid identity of the genes compared to the VFDB reference genes varied between 20 to 50% for 339 genes and was higher than 50% for 58 genes. Interestingly, the distribution of the homolog genes among the strains showed a good concordance with the assigned (RAMI) phylogenetic cluster. For example, chromosomal *cpe* strains and the Darmbrand strain carry the lowest number of virulence homologs (149 – 161 genes). This is in contrast to clusters related to the *netF* positive strains in which cluster 2 (genomes related to JP838) carries 186 – 190 virulence gene homologs and cluster 3 (genomes related to JP55) carries 177 – 181 virulence gene homologs. Almost the same number of potential virulence genes were found in the necrotic enteritis strains, CP4 and Del1, which have 178 and 179 virulence gene homologs, respectively (additional data 13). The largest number of virulence gene homologs was detected in the type C strain JGS1987 (200 genes). The predicted virulence associated genes were scattered across the genome and the G+C content of the virulence related genes was not significantly different from those of other genes, an observation that has been already described in the first complete genome report of *C. perfringens* (Shimizu et al., 2002). Annotation of the identified gene set was performed (2.1.9) and showed that 40% of the genes belong to COG category “cell wall/membrane/envelope biogenesis”. This percentage was calculated based on the well characterized genes using the COG database i.e. after excluding genes that were not annotated and genes that belong to COG category S (function unknown) and R (general function prediction only) (2.1.9, additional data 13). Other categories including “defense mechanisms” and “inorganic ion transport and metabolism” involve 8.4% and 8% of the identified genes, respectively. Additionally, 94 genes revealed a variable degree of homology (23% - 84%) to known capsular genes. For example, homologs were identified for type 8 capsular genes (*cap8B* to *cap8G* and *cap8L* to *cap8P*) present in *Staphylococcus aureus*, also for the capsular genes *cpsC* to *cpsO* and capsular genes *cpsA*, *cpsC*, *cpsE*, *cpsG* and *cpsI* present in *Streptococcus* spp. and *Enterococcus* faecalis, respectively (Caimano et al., 1998;

O'Riordan and Lee, 2004). Only 10% of these capsular gene homologs were present in all strains (additional data 13).

Among toxin genes of *C. perfringens*, only the genes of alpha toxin (phospholipase C; *cpa*), clostripain (protease; *cloSI*) and kappa toxin (collagenase; *colA*) were present in all 76 strains. Therefore, they represent the core toxin genes (additional data 13). However, the presence of a degenerate *colA* gene with a premature stop codon or reduced length of the gene was frequently observed (*colA* FASTA files are provided in additional data 14). Deletion mutations and nonsense mutations were observed in the *colA* gene of strain CBA7123, nonsense (in-frame) mutation was observed in the Darmbrand strain (NCTC 8081) and frameshift mutations were detected in necrotic enteritis strains (Del1, CP4) and in all strains of cluster 3 (JP55-related).

The *pfoA* gene was additionally investigated in the analyzed strains. The *pfoA* gene was not detected in the chromosomal *cpe* strains and Darmbrand strain as well as in the strains JJC, NCTC10578 and MJR7757A. A copy of a *pfoA* like sequence was identified downstream of the typical PFO gene (JP838 locus tag: JFP838_00985) (*pfoA* FASTA files are provided in additional data 15). It was present in all phylogroup II strains as well as in the MJR7757A strain (2.1.9). The *pfo*-like region (length 1,578bp, positions in strain JP838: 222,029 -223,606) was predicted and annotated as perfringolysin O (Thiol-activated cytolysin) gene using Prokka and RAST. However, the NCBI annotation predicted the region as a pseudogene in strain JP838 (positions 222,212 - 223,606).

Further on, strain NCTC8081 Darmbrand was predicted to carry a sequence homologue (locus tag NCTC 8081_02938) to the pore-forming Leukocidin/Hemolysin family closely related to NetG and beta toxin gene (Figure 8) with a 65.5% and 50.5% protein identity, respectively. This sequence homologue was found only in the Darmbrand strain in which the sequence is carried on a large plasmid (size = 116Kbp) (Table 11) and is associated with upstream (IS1469, ISCpe2 and IS1470) and downstream (IS1470) mobile elements. The locus of conjugation was also identified together with the enterotoxin gene (*cpe*) on this large plasmid. This plasmid is distinct from another putatively conjugative plasmid of these strains which carries the beta toxin gene (size= 67.9Kbp) (Table 11)

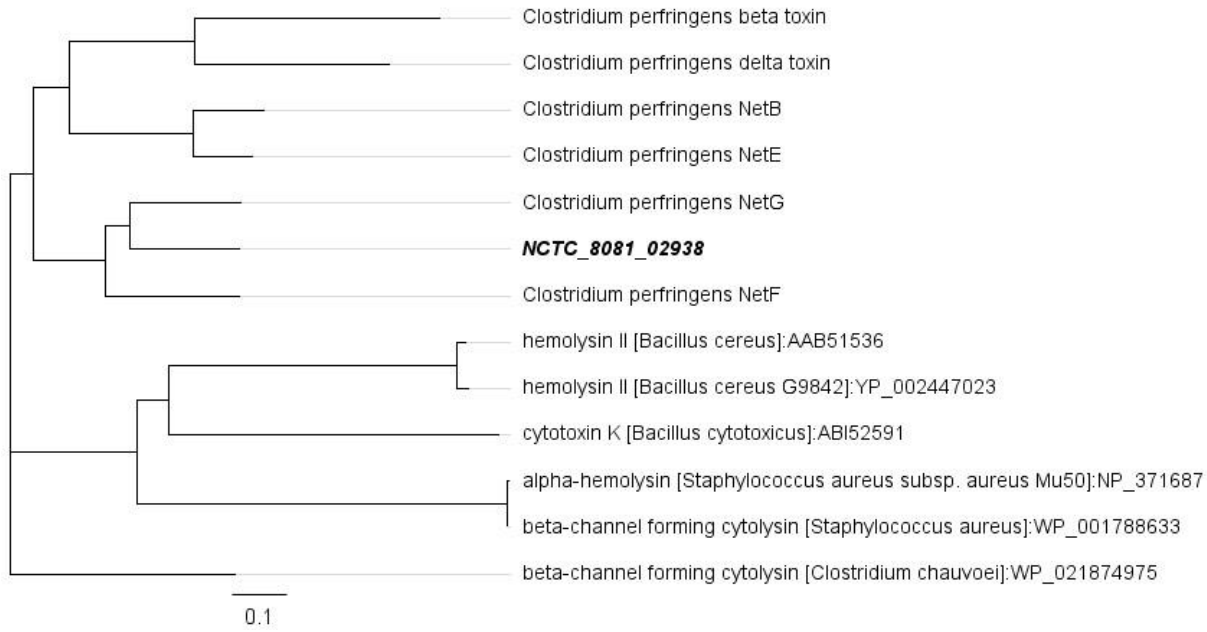


Figure 8: Phylogenetic tree based on the deduced aminoacid sequences of representative members of Leukocidin/Hemolysin superfamily and the homolog (locus tag 02938) of the Darmbrand NCTC 8081 strain showing the close relatedness of the identified homolog to the Leukocidin/Hemolysin superfamily.

Seven putative iron acquisition systems were previously identified in the genome of *C. perfringens* strain 13 (Awad et al., 2016). BLASTn search was performed to investigate the presence of these systems in the 76 genomes (2.1.9). The three siderophore based systems in strain 13 (coordinates 994,044 to 997,117, 1,242,218 to 1,244,913 and 1,423,030 to 1,420,122) were found at a frequency of 97%, 100% and 75% , respectively while the ferric citrate iron acquisition system was present in 85.5% of the investigated strains (additional data 16). One heme acquisition system encoded by the *Cht* locus was detected in all strains (additional data 16) with 98.8% identity. Other putative heme acquisition systems were found in 75% of the strains. The ferrous iron acquisition system (*feoAB*) was detected in all strains (pairwise identity 98.4%) while the additional two copies of *feoAB* identified in the type strain ATCC 13124 (*feoAB* 2: position 1,290,365 to 1,287,671 and *feoAB* 3: position 1,162,515 to 1,160,156) were detected at a frequency of 96% and 97%, respectively (additional data 16). In addition, there was mostly a general 100% gene linkage within each of these iron acquisition systems i.e. all genes were present or absent (additional data 16). In the genomes of chromosomal *cpe* strains as well as the Darmbrand strain (phylogroup I), two putative iron uptake systems (one heme - and one siderophore-iron uptake systems) could not be identified. In contrast, a putative additional iron system could be predicted in the *netF* positive strains that form a cluster with JP838 (phylogroup II; positions in strain JP838: 496,574 to 499,367).

3.2 Characterization and typing of *C. perfringens* isolates from heathy and diseased poultry in Egypt

3.2.1 Sampling and *C. perfringens* isolation

Private veterinary clinics in Sharkia and Dakahliyah governorates in Egypt contributed 20 birds from 10 flocks (7 layer flocks and 3 broiler flocks) as cases suspected for clinical necrotic enteritis (NE) in addition to 34 birds from 17 flocks (15 broiler, 1 layer and 1 breeder) as cases suspected for subclinical NE (2.2.1, Table 4). From these 27 farms, *C. perfringens* was isolated from samples from 14 farms (clinical cases = 9, subclinical cases = 5) (Table 12). These cases are reported as suspected for NE based on the gross lesions as well as the dominating presence of *C. perfringens* strains in bacteriological examination. However, comprehensive differential diagnosis of these enteritis cases was not done.

In cases suspected for clinical NE, the intestinal gross lesions were more prominent and severe in the jejunum of the diseased birds. However, in some cases the duodenum was also affected. Gross lesions consisted of dilated, hyperemic and thin walled or inflamed thick walled intestine distended with dark, brownish fluid exudate and gases (example in Figure 9B). Petechial hemorrhages on intestinal serosa were observed in two cases of clinical NE. In some cases, a green to yellow-brown fibrinous membrane loosely adhered to the entire gut mucosae (Figure 9D).

For cases suspected for subclinical NE, lesions were difficult to recognize. However, light yellowish necrotic spots were observed on the mucosal surface of the duodena (Figure 9A). Moreover, the hepatic tissue was slightly enlarged and the surface exhibited some yellowish necrotic spots.

In total, 51 *C. perfringens* strains were isolated from birds suspected of NE from 14 different farms in different intestinal segments and liver when available (Table 11, additional data 17). From the caeca of apparently healthy birds (n = 29), a total of 83 *C. perfringens* isolates (Table 12) were cultured (three isolates per plate if possible). Of interest is that ten *C. perfringens* strains recovered from five healthy ducks did not produce the typical dual hemolytic zone on blood agar plates. Instead, they yielded a divergent hemolytic pattern on sheep blood agar and no hemolysis on bovine blood agar (Figure 10).

Table 12: Samples from cases suspected for necrotic enteritis, subclinical necrotic enteritis and presumably healthy birds

Processor*	sampled birds per site	Sample origin**	production unit / Host	Place	Age (days)	Flock size	Health status	No. of isolates	*plc*	*cpe*	*netB*	*cpb2*#	Toxin-genotype
A	1	Dou/Jej/Cecum/Liver	Layer	Sharkia	170	4000	Clinical NE	4	+	-	-	+	A
B	1	Jejunum	Broiler	Sharkia	49	25000	SNE***	1	+	-	-	+	A
C	3	Mid-Intestine	Broiler	Dakahliyah	40	9000	SNE ***	4	+	-	-	+	A
D	2	Mid-Intestine	Broiler	Dakahliyah	28	20000	SNE***	2	+	-	-	+	A
E	1	Mid-Intestine	Broiler	Dakahliyah	29	3000	SNE ***	1	+	-	-	+	A
F	1	Mid-Intestine	Broiler	Sharkia	36	6000	SNE***	1	+	-	-	+	A
G	2	Dou/Jej/Cecum	Layer	Dakahliyah	120	18000	Clinical NE	3	+ §	-	-	-	A
H	1	Dou/Jej/Cecum/Liver	Layer	Sharkia	120	5000	Clinical NE	4	+	-	-	+	A
I	1	Dou/Jej/Cecum/Liver	Layer	Sharkia	120	5000	Clinical NE	4	+	-	-	+ (3)/-(1)	A
J	1	Dou/Jej/Cecum/Liver	Layer	Sharkia	120	5000	Clinical NE	4	+	-	-	+	A
K	1	Dou/Jej/Cecum/Liver	Layer	Sharkia	120	5000	Clinical NE	4	+	-	-	+	A
L	4	Dou/Jej	Layer	Dakahliyah	200	15000	Clinical NE	4	+ §	-	-	-	A
M	4	Dou/Jej/Cecum/Liver	Broiler	Sharkia	32	16000	Clinical NE	11	+	-	-	+(6)/-(5)	A
N	2	Dou/Jej/Cecum	Broiler	Dakahliyah	18	8000	Clinical NE	4	+	-	-	+(3)/-(1)	A
O	25	Caecum	Broiler	Dakahliyah	38	Unkown	Healthy	19	+	-	-	+(5)/-(14)	A
P	2	Caecum	Duck	Dakahliyah	90	Unkown	Healthy	1	+	-	-	-	A
Q	6	Caecum	Broiler	Dakahliyah	40	Unkown	Healthy	18	+	-	-	+(17)/-(1)	A
R	4	Caecum	Duck	Dakahliyah	Unkown	Unkown	Healthy	9	+	-	-	+(4)/-(5)	A
S	8	Caecum	Broiler	Dakahliyah	38	Unkown	Healthy	24	+	-	-	+(8)/-(16)	A
T	2	Caecum	Duck	Dakahliyah	Unkown	Unkown	Healthy	6	+	-	-	+(4)/-(2)	A
U	2	Caecum	Duck	Dakahliyah	Unkown	Unkown	Healthy	6	+	-	-	+(4)/-(2)	A

* Processor A to N refers to poultry flocks, processor O to U refers to slaughterhouses

Dou: Duodenum, Jej: Jejunum, * SNE, Subclinical necrotic enteritis

§ Type II intron is inserted in the alpha toxin gene in these isolates, # number in brackets for the *cpb2* positive and *cpb2* negative isolates

-: absent, +: present

Figure 9: Exemplary gross pathological findings of cases suspected for *C. perfringens* induced enteritis. (A) Duodenum of a broiler with suspected subclinical necrotic enteritis. Several yellowish (necrotic) spots are visible on the mucosa (arrows). **(B)** Intestine of a layer hen with severe necro-hemorrhagic enteritis of the jejunum. The lumen of the jejunum is distended and filled with hemorrhagic exudate that contains pieces of necrotic debris (brownish area). **(C)** Jejunum of a broiler chicken with suspected clinical necrotic enteritis. The jejunum is dilated and the mucosa is markedly thickened and irregular. **(D)** The intestine of a layer hen with suspected clinical necrotic enteritis showing a fibrinous content that is loosely attached to the intestinal mucosa.

3.2.2 Toxin genotyping of *C. perfringens* isolates by PCR

Toxin genotyping of *C. perfringens* strains was performed by PCR (2.2.4). All of the 51 isolates from NE-suspected (clinical and subclinical) cases and the 83 isolates from asymptomatic birds were genotyped as toxin-type A (Table 12). None of the isolates was found to harbor the enterotoxin gene or the *netB* gene. Seven isolates from two farms (G and L) suspected for clinical NE showed a larger PCR amplicon of the *cpa* gene (Table 12, Figure 11). Subsequent Sanger sequencing of the amplicon (2.2.5) showed an insertion of an 834 bp DNA segment, identical to a group II intron (acc. No. DQ787115.1) previously described (Ma et al., 2007). The

flanking region of the type II intron sequences showed 99 % identity to the *cpa*-gene of *C. perfringens* ATCC 13124 strain (acc. No. CP000246.1). The intron-carrying isolates were assessed for their hemolytic and phospholipase C activities by plating on blood and lactose egg yolk agar. Hemolytic and phospholipase C activities of the isolates were observed *in vitro* (data not shown).

All isolates were tested by PCR for the presence of the NetB toxin gene (2.2.4) (51 from NE-suspected cases and 83 from asymptomatic birds). All tested strains were *netB* negative. In contrast, the beta2 toxin gene (*cpb2*) was detected in 37 out of 51 NE isolates and 42 of the 83 isolates from healthy birds (Table 12). All *cpb2* genes were detected using primers that recognize the typical and atypical variants of *cpb2* (van Asten et al., 2008). All isolates tested negative when using primers by (Baums et al., 2004) which detect only the typical *cpb2* allele.

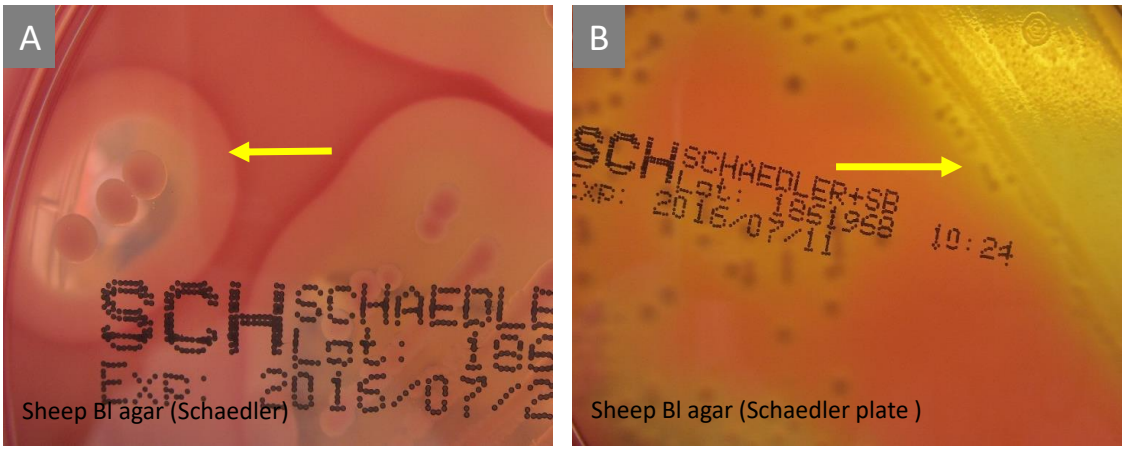


Figure 10: (A) *C. perfringens* with a typical dual hemolytic pattern (B) *C. perfringens* without the characteristic double hemolytic zone.

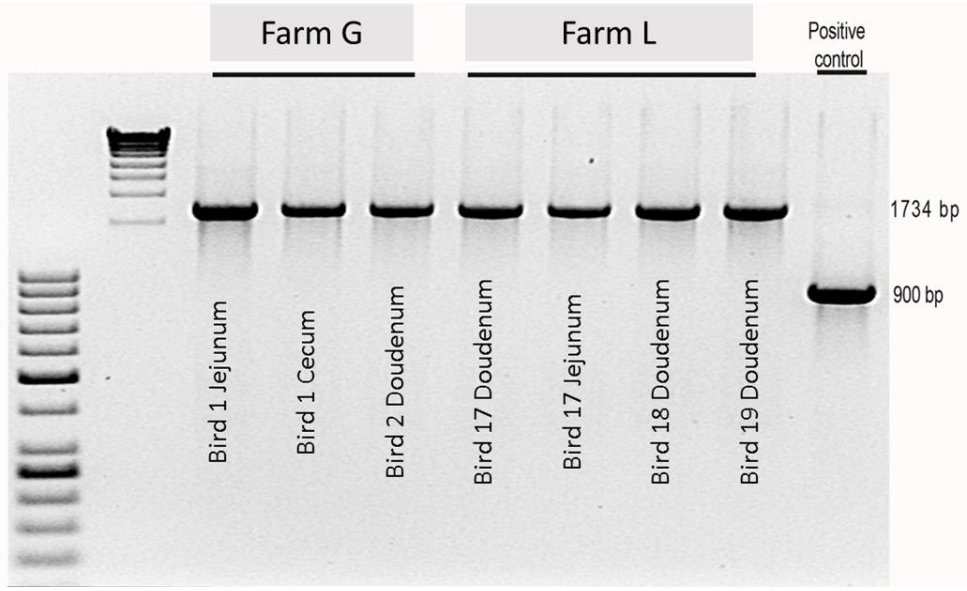


Figure 11: Agarose gel electrophoresis of the *plc* PCR amplicon of seven *C. perfringens* isolates from two different farms compared to the positive control.

3.2.3 Multilocus sequence typing of avian *C. perfringens* strains from suspected NE cases

MLST data from recent *C. perfringens* studies (Chalmers et al., 2008a; Hibberd et al., 2011; Jost et al., 2006b) are not available on MLST.net (http://www.mlst.net) or PubMLST (http://pubmlst.org). However, available accession numbers were used to retrieve the sequences from GenBank (Hibberd et al., 2011) (2.2.7). Sequencing the MLST loci from twelve isolates using Sanger technology was performed (2.2.6). MLSTest software was used to assign allele numbers and STs for the retrieved data with caution to match the already assigned ST-numbers defined in the published study (Hibberd et al., 2011) (2.2.7). To investigate the genetic diversity of poultry *C. perfringens* in clinical NE, MLST was applied to twelve isolates from three birds; four isolates were investigated per bird originating from duodenum, jejunum, cecum and liver. These birds were suspected for clinical NE and were selected from three different farms (2.2.6). Based on this analysis, the twelve strains were assigned new STs (Figure 12, additional data 18). In addition, two new alleles were identified for *glpK* and *groEL* genes. Table 13 shows the allelic diversity of the MLST loci for these twelve *C. perfringens* strains as calculated using START and MLSTest (2.2.7). Four isolates were investigated per each bird. In two cases, these isolates belong to a single ST*,* ST45 and ST46 for bird no. 1 and bird no. 2, respectively (Figure 13). However, the isolates from bird 3 belong to three different STs (ST46, ST47, and ST48), two of them (ST46 and ST48) form a clonal complex (CC1) based on Double-Locus Variants (DLV) i.e. nine out of the eleven MLST genes tested are identical (additional data 18). However, the investigated twelve strains do not group into one clonal complex (Figure 12 and Figure 13). ST46 was detected in farms 2 and 3.

Table 13: MLST loci diversity for the twelve *C. perfringens* strains investigated from Egypt

	ddl	***dnaK***	***dut***	***glpK***	***gmk***	***groEL***	***gyrA***	***plc***	***recA***	***sod***	***tpi***
Analyzed size	531	663	340	522	373	597	602	440	564	366	588
No of alleles	15	14	13	17	5	15	10	15	8	13	11
No of polymorphisms	30	37	21	20	4	28	18	26	27	28	14
Typing efficiency*	0.5	0.378	0.619	0.85	1.25	0.536	0.556	0.577	0.296	0.464	0.786
Discriminatory power**	0.592	0.892	0.851	0.839	0.614	0.88	0.688	0.856	0.514	0.772	0.539
*dN/dS****	0,0943	0,0586	0,817	0,0533	0,5211	0,0346	0,0694	0,0918	0	0	0,0157

*Typing efficiency (TE): defined as the number of different genotypes per polymorphic site

**Discriminatory Power (DP): defined as the ability of the typing system to distinguish between unrelated strains based on Simpson's diversity index (Hunter and Gaston, 1988) (2.2.7).

****dN/dS* ratio: The ratio of non- synonymous (dN) to synonymous (dS) substitutions per nucleotide site based on the method Nei and Gojobori (Nei and Gojobori, 1986)

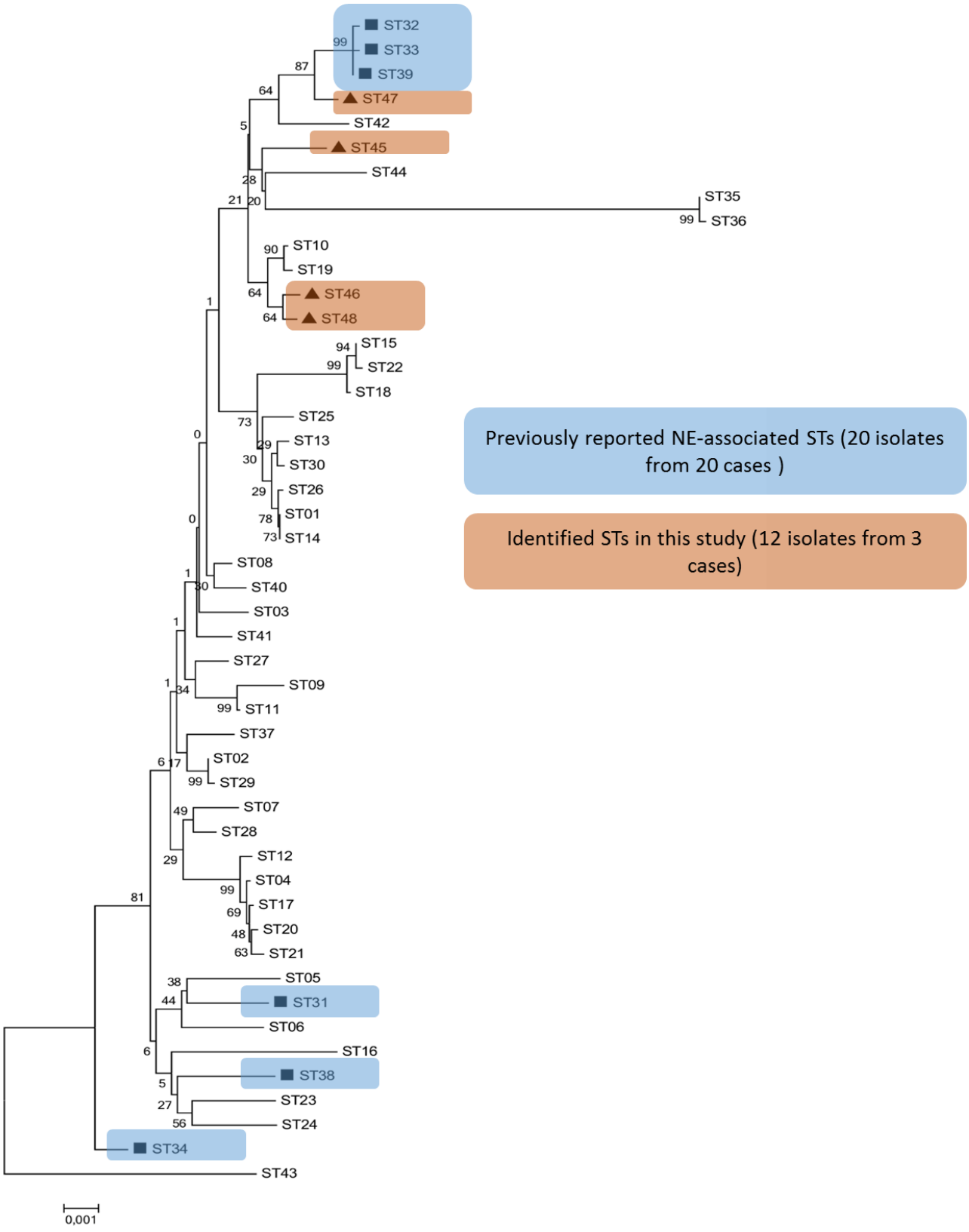


Figure 12: Neighbor-joining phylogeny of *C. perfringens* strains based on the concatenated MLST loci for sequence types (STs) for twelve strains from Egypt as well as 41 representative STs for the strains described by Hibberd et al. (2011).

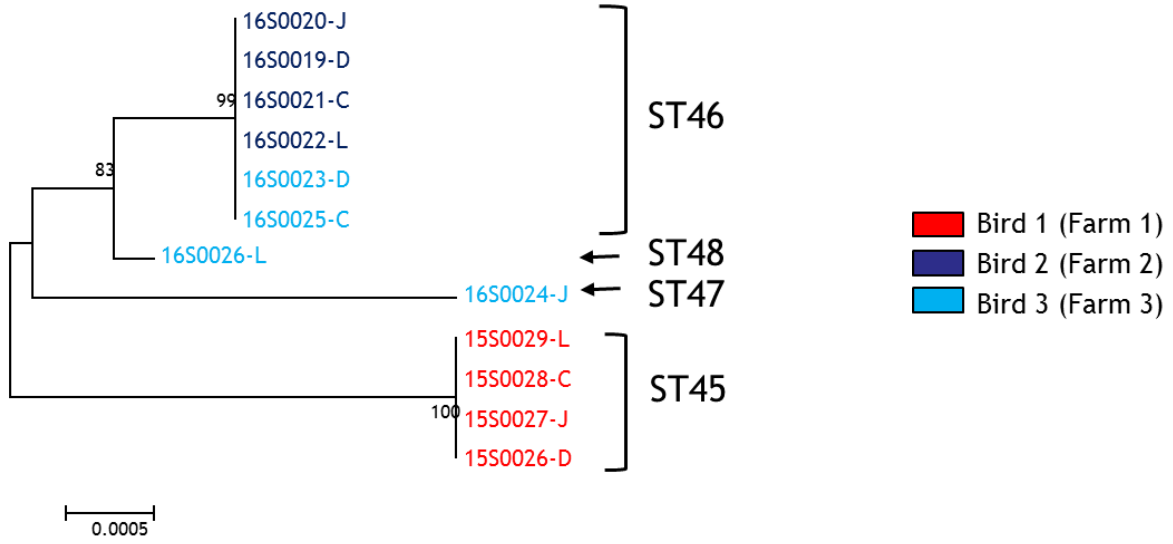


Figure 13: Neighbor-joining phylogeny based on the concatenated nucleotide sequences of the eleven MLST genes for twelve *C. perfringens* isolates from three bird.

*D: duodenum, J: jejunum, C: cecum, and L: liver

3.3 Development and application of a core genome-based multilocus sequence typing system for *C. perfringens*

3.3.1 Development of a cgMLST scheme

The SeqSphere+ software (Junemann et al., 2013) was used to develop a cgMLST scheme for *C. perfringens* (2.3.4). The complete genome sequence of the type strain ATCC 13124 was selected as a seed genome (2.3.4.1) which is annotated to have 2,876 CDS. Of these, 2,731 genes were retained and served as target genes after applying the basic filters of SeqSphere+ (Junemann et al., 2013) (filtering against overlapped genes [n = 116 genes], paralogues [n = 24 genes], missing or improperly positioned stop codons [n = four genes] and homology to a plasmid sequences [n = one gene]) (Figure 14). 1,510 out of the 2,731 retained genes were identified as strict core genes (100% presence) in 38 query genomes (2.3.4.2).

For the initial evaluation of these 1,510 genes, a set of 80 *C. perfringens* genomes (data sets II and III in 2.3.4.3, additional data 2) including strains used for target genes definition was screened using SeqSphere+ build-in BLASTn (Altschul et al., 1997; Junemann et al., 2013) for the detection of the genes and allele assignment (Figure 14). As previously described (Moura et al., 2016), alleles were only assigned to the complete genes without ambiguities i.e. alleles were not assigned to genes when they had frameshifts, in frame stop codons or non-GATC characters (2.3.4.3). Following these parameters, further 60 genes were discarded (Figure 14, additional data 19) as they were not called in more than 5 % of the genomes (missing genes or no allele assignment). The final set of the cgMLST targets comprises 1,450 genes that correspond to 1.35Mbp and represent 41.5% of the genome sequence of the reference strain ATCC 13124. The core target genes were distributed across the reference genome. However, this distribution was biased toward one chromosomal replichore (Figure 15A). The average G+C content of the cgMLST loci was 29.6 ± 3.1% (Figure 15B) and the locus lengths varied between 93 to 4,350bp (average = 932.9 ± 578) (Figure 15C and 15D). A complete list of the cgMLST loci can be found in Table S5 and additional data 19.

3.3.2 Application and performance of the cgMLST

The whole genome assemblies of an independent set of 80 *C. perfringens* isolates (data set I in 2.3.1.1, 2.3.1.2, and additional data 2) were additionally added and genotyped according to the final 1,450 cgMLST targets. These 80 genomes comprised 50 strains isolated from poultry samples from Egypt and sequenced in this study (2.3.1.2) as well as genomes of 30 strains from Finland and Denmark that were downloaded from the NCBI (data set I in 2.3.1.1). These 80 WGS data represent epidemiologically related and unrelated *C. perfringens* strains from poultry with clinical pictures suspected for necrotic enteritis as well as from healthy birds.

The 50 *C. perfringens* isolates from Egypt (2.3.1.2) were sequenced using Illumina HiSeq technology, with 151bp read length (2.3.2). The median read count per strain was 7.78 million reads. Additionally, the publically available MiSeq reads of 30 *C. perfringens* strains (read length 35 to 251bp) were downloaded from NCBI (data set I in 2.3.1.1). These HiSeq and MiSeq reads were used for *de novo* assembly using SPAdes assembler (Bankevich et al., 2012) after applying quality trimming procedures with Sickle (Joshi and Fass, 2011) (2.3.3). The average number of assembled contigs was 79.4 per strain and the median N50 value was 367Kbp (range11.8Kbp to 2.02Mbp) per strain (additional data 20). The total size of the assembled genomes ranged from 3.17 to 4.01Mbp. The G+C content ranged between 27.8 and 28.9% (median 28.1%). For the strains sequenced using HiSeq technology, an average coverage of 754.4 ± 241.1x was detected while for the downloaded MiSeq data, an average coverage of 102.7 ± 49.4x was observed (additional data 20). Two strains showed the lowest N50 value and the highest contig number: strain T43 (N50 value = 11.8Kbp, contig number = 498) and strain C3 (N50 value = 32.6 Kb, contig number = 203). However, for most of the strains, a high quality draft genome sequence (with low number of assembled contigs and high N50 value) could be generated (additional data 20). Genome annotation based on Prokka software (Seemann, 2014) showed that *C. perfringens* genomes encode on average 3,148 (range 2,781 to 3,792) protein coding sequences (2.3.3). The calculated size of the coding region was on average 83.3% of the whole genome size (additional data 20).

The screened genomes have on average 99.5% of the cgMLST targets. However, a sum of 552 target genes (additional data 21) were reported as nontypeable (genes without allele assignment) in at least one genome, ranging from 0 to 68 (median = 4) nontypeable gene per genome (additional data 21). Genes were reported nontypeable if the gene was absent (BLASTn cutoffs 90% identity and 99% coverage) or did not meet the quality thresholds e.g. due to frameshifts, internal stop codons or ambiguous nucleotides (2.3.4.3).

A total of 290 loci were absent in the analyzed genomes averaging of 2.7 ± 6.4 (range 0 to 53; median = 1) gene per genome. Three loci (CPF_1492, CPF_1860 and CPF_1958) were found to be absent in more than 10 genomes. The highest number of absent genes was observed in the genomes of the strains T43, T34, JGS1721, C7 and C3 that lack 53, 42, 28, 28 and 19 target genes, respectively (additional data 21). The evaluated genome data of these strains are characterized by high contig numbers and low N50 values (additional data 20).

Additionally, errors due to frameshifts, internal stop codons or ambiguous nucleotides resulted in failure of allele assignment to 341 target genes in at least one genome with an average of 4.3 ± 5.2 (range 0 to 37, median = 3) genes per strain (additional data 21). Of these, seven loci (CPF_2459, CPF_1195, CPF_1689, CPF_0908, CPF_2216, CPF_1448 and CPF_2763)

were found to fail in more than 10 genomes. It is noteworthy to mention that failed targets due to frameshifts and internal stop codons were reported at high frequency in the closed genome of strain 13 in which alleles were not assigned to 30 targets, 27 were due to indel-associated frameshifts and three loci gene fails were due to substitution that led to an internal stop codon. In total, indel mutations (either true or artefacts) were observed in a sum of 310 genes ranging from 1 to 13 nucleotide differences per gene relative to the cgMLST targets (additional data 19). Of these, 107 genes carry in frame indels that result in length variations equal 1 to 4 codons (one codon = 3bp) differences relative to the reference cgMLST genes (additional data 19).

The average number of alleles reported for each locus including the entire data set was 37.2 ± 13.8 alleles (range 1 to 74) (Figure 15C and Table S5). The cgMLST system assigned nearly each isolate as a distinct sequence type (the number of distinct allelic profiles was 158 for the 160 genomes). The Simpson's index of discrimination for the developed cgMLST that was 1.0 (95% confidence interval [0.999-1.0]) (2.3.5). Of interest, homologous recombination was found to affect nearly half of the loci as determined by PHI statistic test (759 out of 1,450 loci with *P* value less than 0.05) (2.3.4.3, additional data 19).

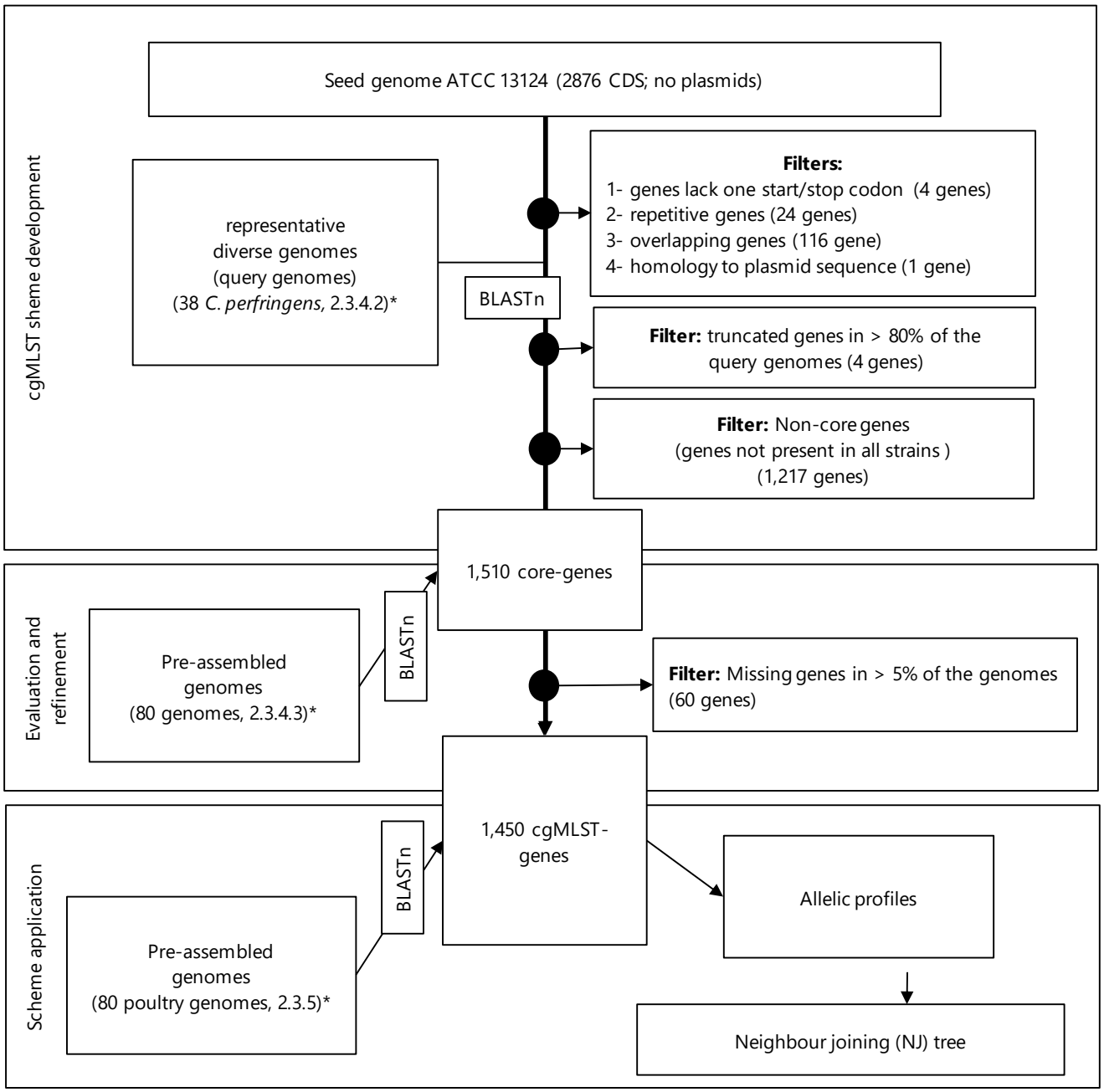


Figure 14: Scheme of the workflow to define and evaluate the core genome MLST targets.

*for details about different data sets, please see additional data 2.

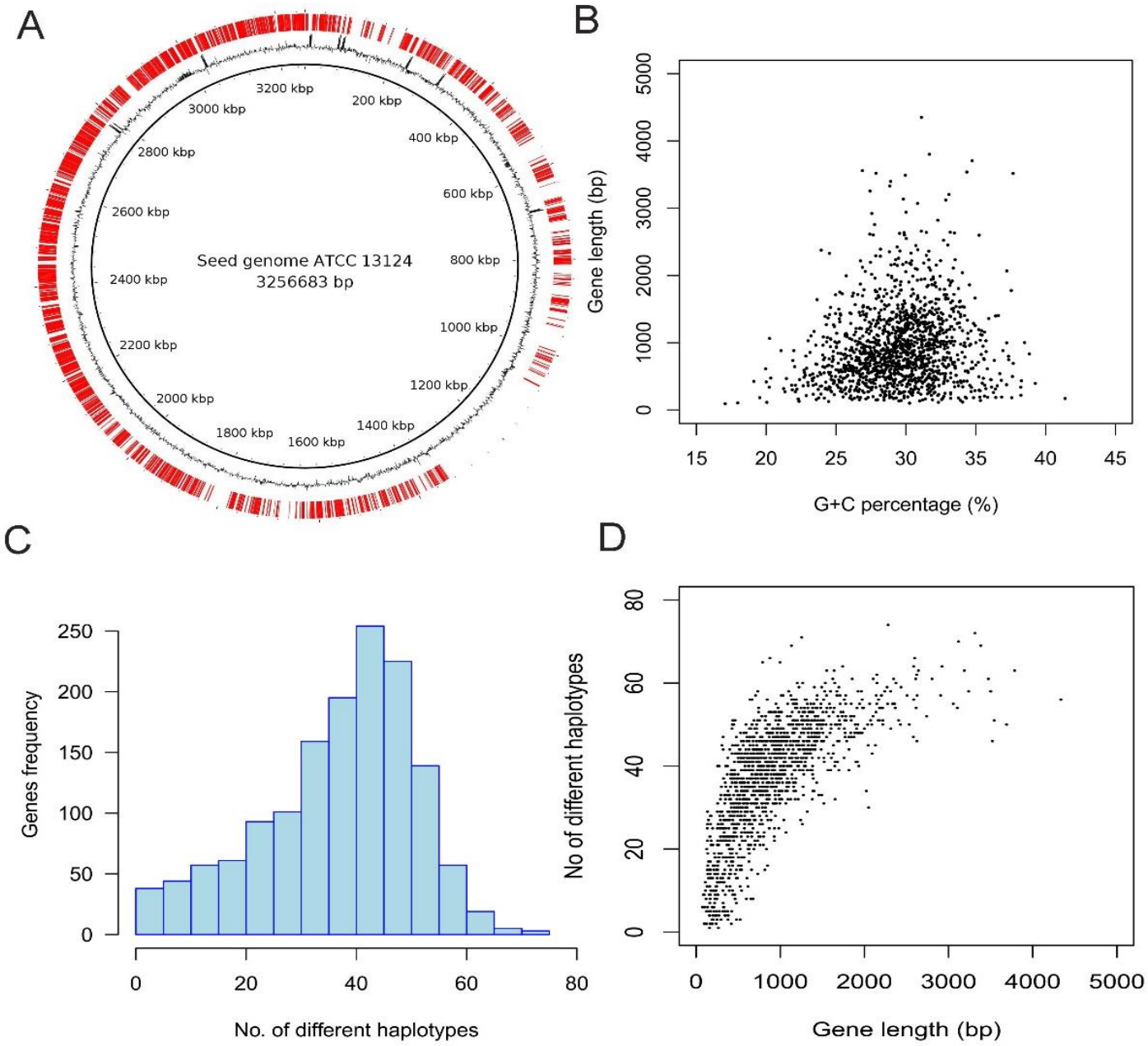


Figure 15: **Characteristics of the 1,450 core genome MLST genes. (A)** The distribution of the cgMLST targets (red, outer circle) across the *C. perfringens* ATCC 13124 reference genome. Inner circles represent the G+C content and genomic positions, respectively. **(B)** Plot showing the lengths and G+C content of the cgMLST targets. **(C)** Plot showing the frequency of different allelic variants observed for each of the cgMLST targets. **(D)** Plot showing direct association between the length of cgMLST target genes and the number of variants detected for each gene. For a scalable image, please see additional data 22.

3.3.3 Definition of cgMLST Cluster Types (CT)

Similar to clonal complexes for classical MLST, cgMLST profiles can be further grouped into Cluster Types (CT) defining a group of very similar cgMLST profiles differing from one another by up to a certain number of alleles defined as CT threshold. In the data set analyzed herein, a CT threshold of 60 allelic differences was defined as it showed a good concordance with a previous study that grouped the *netF*-positive strains into two distinct CTs based on the cgMLST (Mehdizadeh Gohari et al., 2017). These strains share the same epidemiological background as they were isolated from cases of foal necrotizing enteritis and canine hemorrhagic diarrhea (Mehdizadeh Gohari et al., 2017). Additionally, the allelic distances between each pair of genomes was calculated (2.3.9). The pairwise allelic distances was less than 60 for more than half of the genomes (84 strains) accounting for ~3% of the overall pairwise allelic distances (~392 comparisons out of 12,720) (Figure 16).

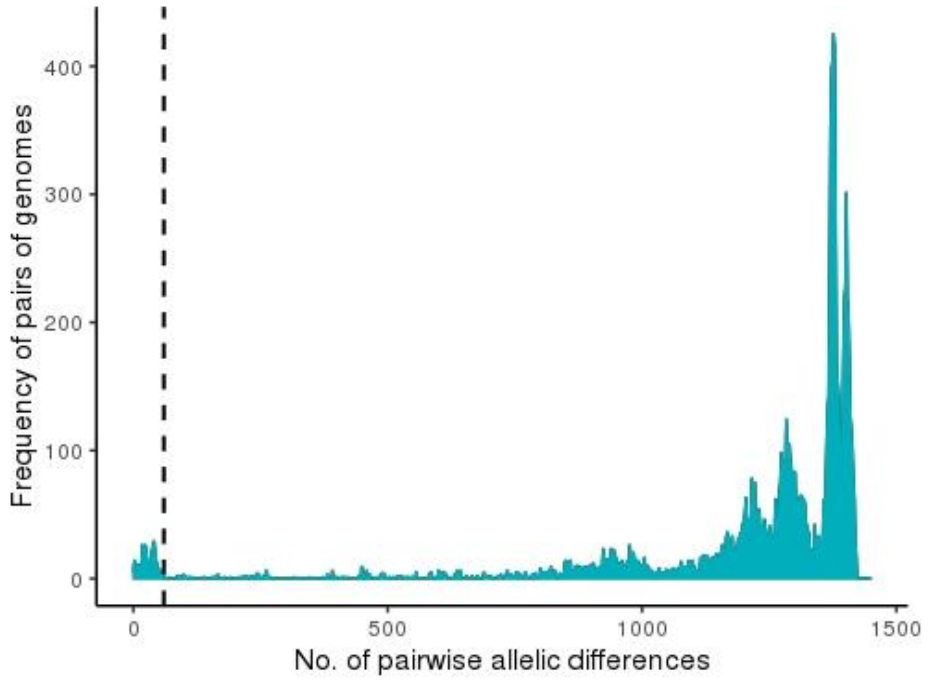


Figure 16: Distribution of the pairwise allelic differences between the 160 *C. perfringens* isolates. The number of allelic differences between each pair of genomes was calculated using SeqSphere+ totaling 12,720 non-redundant pairwise comparisons. The distribution of these pairwise allelic distances are represented in a histogram. The values on the x-axis denotes the number of allelic differences between isolate pairs while the y-axis denotes how many times these pairwise allelic distances are present in the dataset. The dashed line refers to the cutoff value of 60 allelic mismatches to define cluster types.

Using a CT threshold of 60, 84 strains were allocated to 21 CTs while 76 strains were reported as singletons (Figure 17). As previously described (Mehdizadeh Gohari et al., 2017), 31 of 32 *netF* positive strains clustered into two CTs: CT01 (n = 26 strains) and CT02 (n = 5 strains). The remaining 19 CTs (CT03 to CT21) included only poultry strains of *C. perfringens* (Figure 18A) that were isolated from cases suspected for NE as well as from healthy birds (Figure 18B). The clustered poultry strains of *C. perfringens* comprised 53 (out of 87) strains while 34 strains of poultry origin did not cluster (singletons).

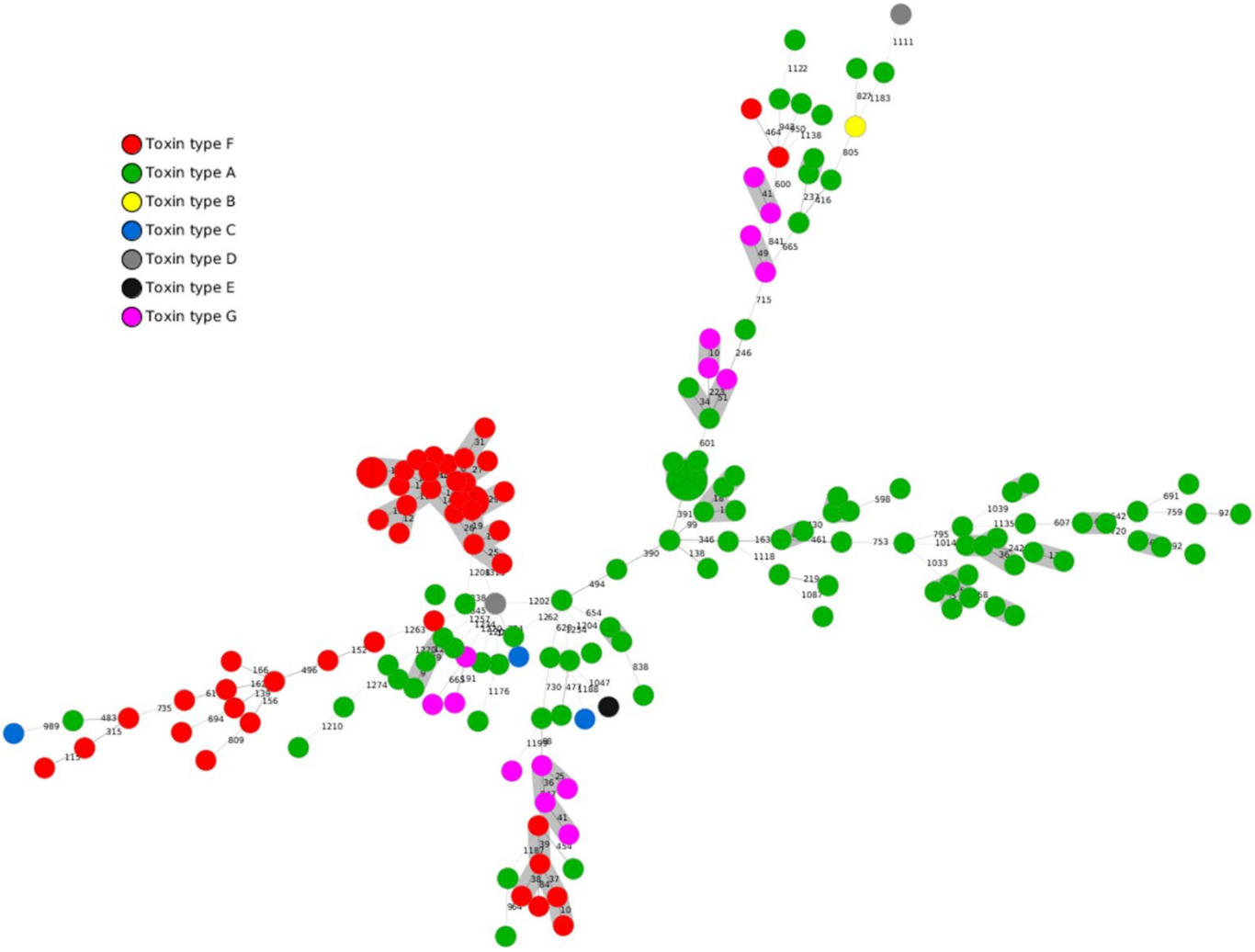


Figure 17: Minimum spanning tree (MST) based on the 1,450 core genome MLST (cgMLST) genes of 160 *C. perfringens* isolates. Circles represent sequence types identified for the isolates based on the core genome. Cluster types (CT) were identified based on less than 60 allelic mismatches (grey shading). The numbers on the connecting lines illustrate the numbers of target genes with different alleles.

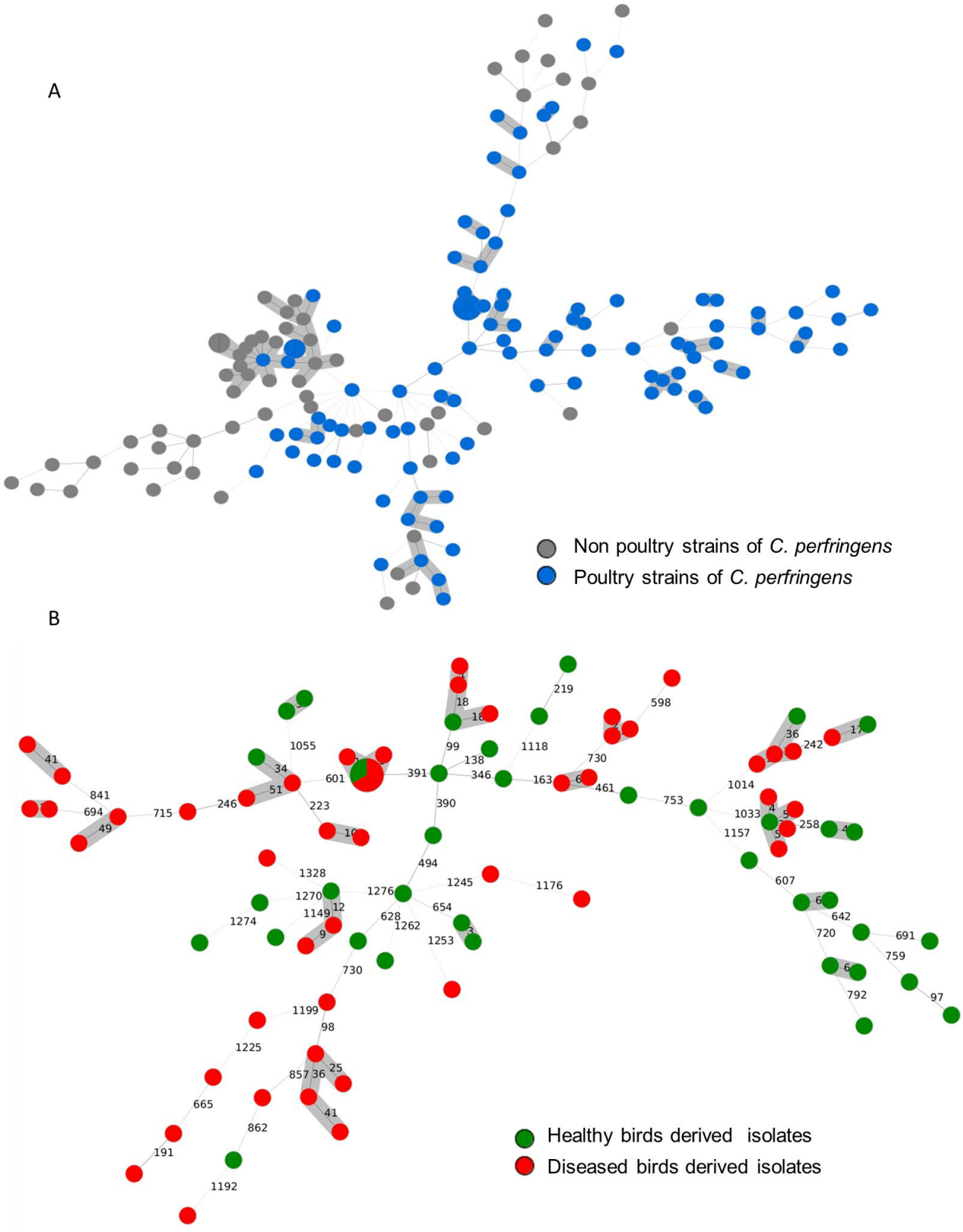


Figure 18: Minimum spanning tree (MST) based on the 1,450 core genome MLST (cgMLST) genes for 160 *C. perfringens* strains highlighting the strains of poultry origin (A). Minimum spanning tree (MST) for the 87 *C. perfringens* poultry strains indicating strains deriving from diseased birds and from healthy birds (B).

3.3.4 Genotyping of poultry *C. perfringens* based on the cgMLST scheme

The cgMLST scheme was applied next to analyze the poultry isolates of *C. perfringens*. In total, the data set analyzed herein includes 87 strains of poultry origin. These strains comprise 50 *C. perfringens* isolates from Egypt (2.3.1.2, Table 9), 30 avian *C. perfringens* isolates from healthy and NE infected chickens in Denmark and Finland (2.3.1.1, Table 8) and seven *C. perfringens* genomes available at the NCBI representing the following geography (USA = 3, Canada = 1, Australia = 1, and Czech Republic = 2) (Table 10). 53 of the 87 poultry strains clustered into 19 cluster types based on less than 60 allelic mismatches (Figure 18B and Figure 19). The remaining 34 poultry strains did not cluster (singletons) (Figure 18B).

3.3.4.1 cgMLST of *C. perfringens* poultry strains and country of isolation

Sequence data of the 87 poultry isolates investigated in the data set originated from four different continents (Egypt = 50, Finland = 13, Denmark = 17, USA = 3, Canada = 1, Australia = 1, and Czech Republic = 2) (additional data 2). Hence, cgMLST was applied to explore the relatedness of strains compared to their country of isolation (2.3.5). As mentioned above, 53 of the 87 poultry strains of *C. perfringens* were clustered into 19 CTs (Figure 18B, Figure 19). Of interest is that more than 70% (36 out of 50) of the Egyptian strains were included in 13 different CTs (Figure 19 and Table 13). Based on the 60 allelic difference threshold, none of the Egyptian strains grouped with strains from the other countries. Further, CTs comprised strains that derived from one country e.g. CT09 (T46, T43 and T84) and CT10 (T16 and T18) involved strains from Finland, while CT11 (C124 and C33) and CT21 (C31 and C26) comprise strains from Denmark. Moreover, hybrid CTs were also detected e.g. CT06 and CT12. CT06 encompasses four strains: CP4 (isolate from Canada), C24 and C36 (isolates from Denmark) and Del1 isolated in the United States while CT13 contains two strains: C48 from Denmark and EHE_NE18 from Australia (Figure 19).

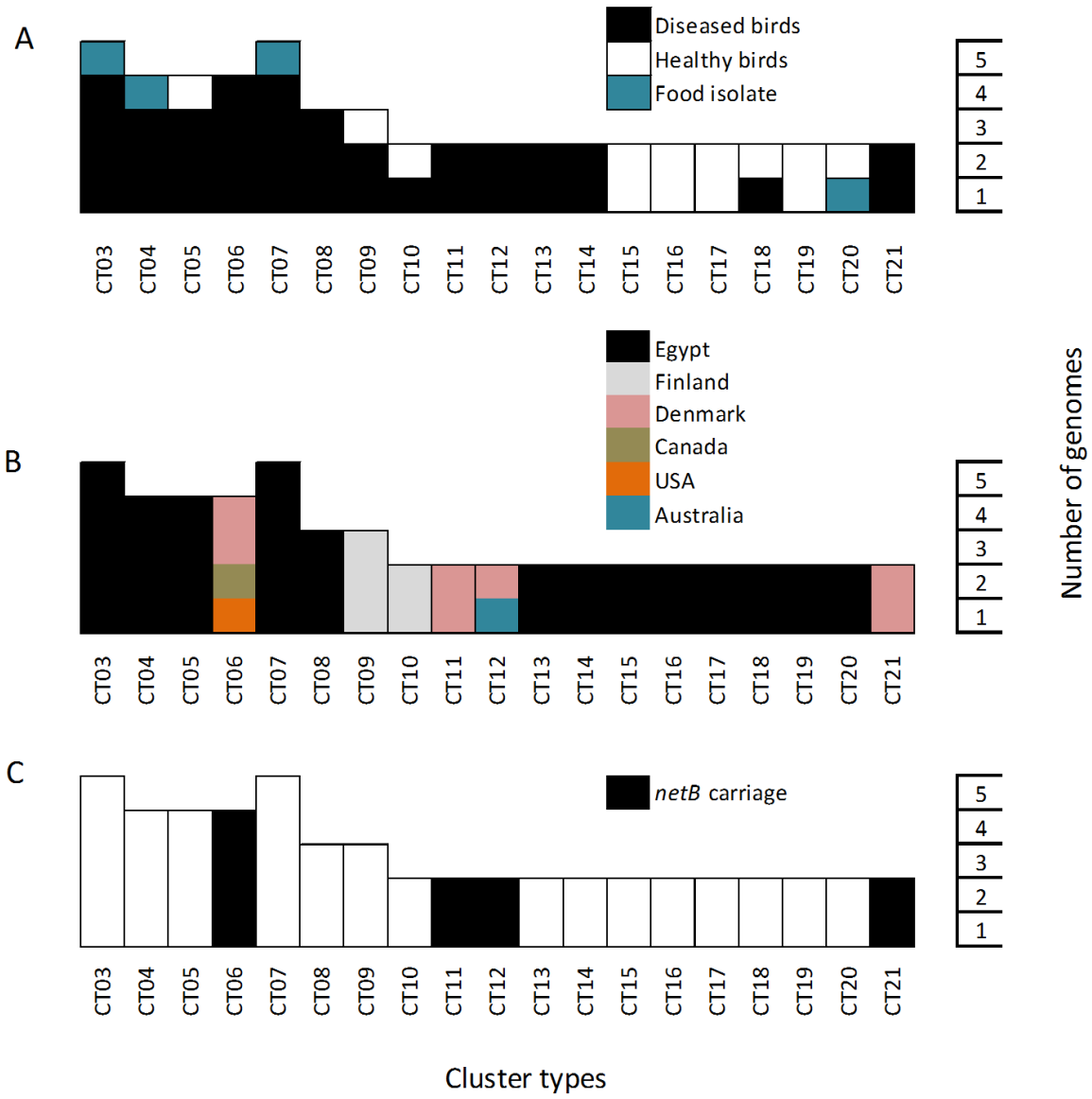


Figure 19: Association of the identified Cluster Types (CT) of *C. perfringens* strains regarding clinical disease (A), geography (B) and Necrotic Enteritis Toxin Beta-like gene (*netB*) carriage (C). CTs shown in the figure encompasses 19 CTs that include 53 (out of 87) poultry strains of *C. perfringens*

3.3.4.2 cgMLST of *C. perfringens* poultry strains and *netB* gene carriage

Although none of the Egyptian strains carries the *netB* gene, the dataset analyzed here included 15 *netB*-positive NE strains (Denmark =10, Finland = 2, Canada=1, Australia =1 and the US=1). The developed cgMLST was additionally employed to possibly identify clades associated with *netB* positive strains (2.3.5). Based on a neighbor allelic profiles distance of less than 60, 10 out of the 15 strains were clustered into four different CTs: CT11 (C124 and C33), CT6 (CP4, Del1, C36 and C24), CT12 (C48 and EHE-NE18), and CT21 (C31 and C26) (Figure 20B). A fifth clade including four strains (T11, C41, C125 and C37) could be identified based on the phylogenetic relatedness between the strains (Figure 20A).

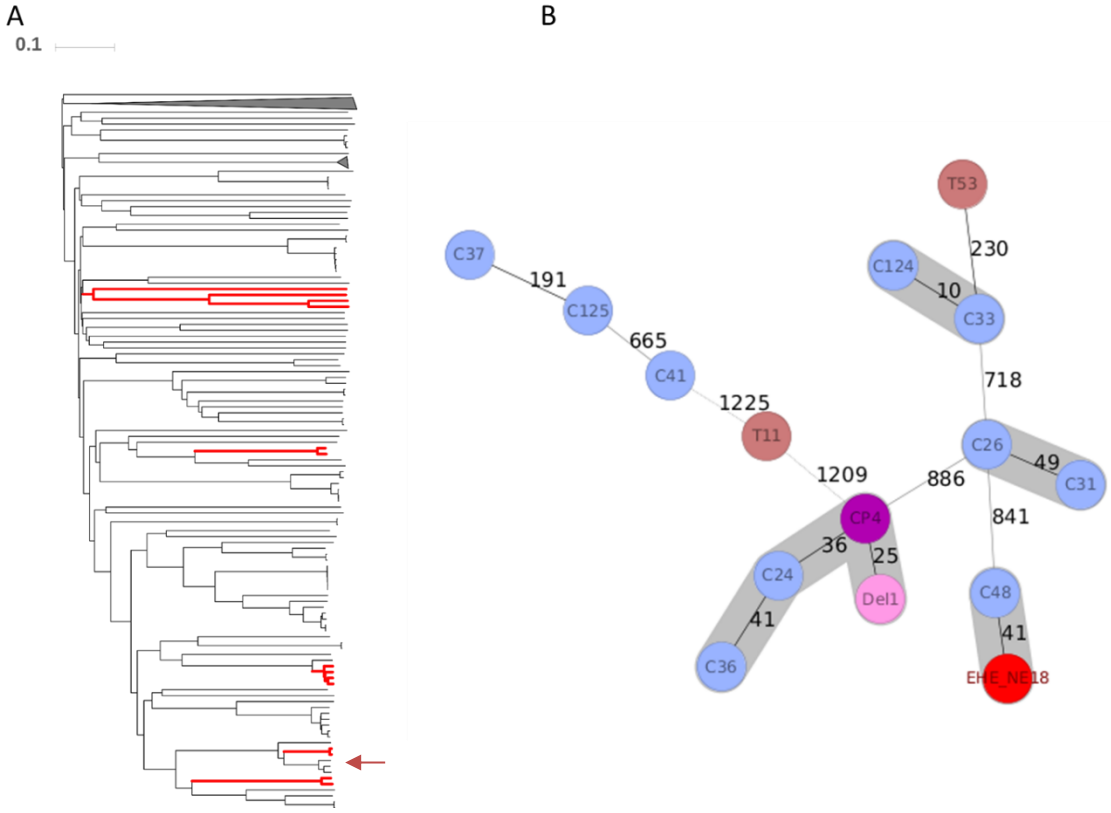


Figure 20: Diversity of *netB*-carrying *C. perfringens* poultry strains. (A) Neighbor-joining phylogeny of poultry *C. perfringens* strains (total = 87) based on the cgMLST allelic profiles show the five clades (marked in red) that comprise 14 *netB* positive strains. Red arrow refers to strain T53 which was not clustered. **(B)** A minimum spanning tree (MST) generated for the 15 *netB* positive strains show the neighbor allelic distance between the genomes. Gray shadow indicates that the strains belong to the same cluster type based on a cutoff of less than 60 different alleles. Circles were colored according to the strain origin as follow: Denmark: blue, Canada: violet, Australia: red, Finland: brown, USA: pink.

3.3.4.3 cgMLST of *C. perfringens* poultry strains and clinical disease

Based on less than 60 allelic mismatches, 53 of the 87 poultry strains were clustered into 19 cluster types. 14 CTs of the poultry strains included strains that were isolated from diseased birds (Figure 18B and 19A). However, 7 CTs of these also contained strains from healthy birds and food isolates (Figure 18B and 19A).

The cgMLST scheme was applied to analyze the 50 poultry *C. perfringens* isolates from Egypt (Table 9 and 14). These strains comprised 23 isolates from suspected NE cases from 14 different farms, 21 isolates from slaughterhouses from healthy birds and six *C. perfringens* isolates from retail chicken meat (Table 9 and 14). Based on the cgMLST results, isolates originating from the same farm with NE grouped together and were assigned to the same CT. Strains from farm C (15S0055 and 15S0056), strains from farm G (16S0002, 16S0003 and 16S0004), strains from farm L (16S0037, 16S0038, 16S0039 and 16S0041) and strains from farm N (16S0057 and 16S0058) were clustered and assigned to CT08, CT05, CT03 and CT07, respectively (Table 14). However, two CTs were observed for the strains in farm M, CT07 (16S0051 and 16S0055) and CT18 (16S0043). Moreover, despite the strong clustering of the suspected NE isolates in each farm, strains from different farms were sometimes clustered together. For example, CT07 included strains from farms M and N, CT08 included three strains from farms C and D, CT13 included strains from farms A and B, CT14 included two strains from farm H and I and CT04 included three strains from farms J, K and E (Table 14). Within clusters that include mainly suspected NE isolates, also isolates from healthy birds or from poultry meat samples were observed. For example, CT03, CT04 and CT07 included meat isolates and suspected NE isolates (Table 14). In addition, CT05 and CT18 comprise isolates from suspected NE cases as well as isolates from healthy birds.

In comparison, the isolates from healthy birds were more diverse even within a single site. Some isolates formed few clusters with a maximum of two isolates per cluster as observed for CT15 (16S0105_1 and 16S0108_1), CT16 (16S0105_3 and 16S0106_1), CT19 (16S0086_1 and 16S0072_1) and CT17 (16S0096 and 16S0108_3) (Table 14). Except for CT17 where isolates were derived from two different sites, isolates within each of the aforementioned CTs are from a single site.

After increasing the resolution of the CTs by defining the groups based on less than 10 allelic differences, it was observed that the suspected NE isolates did not cluster with any of the non-disease isolates except for one strain (16S0212 in CT07). Based on less than 10 allelic differences, seven clusters that included suspected NE isolates (CT03-5, 7, 8, 13, 14) and four clusters that included non-clinical isolates (CT15, CT16, CT17 and CT20) were found (Table 14).

Table 14: Cluster types (CT) identified for the *C. perfringens* strains isolated from suspected NE, healthy birds and poultry retail meat in Egypt

Processor*	Isolates	Cluster Types (CT) identified based on core-genome MLST typing																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	singletons
Strains isolated from cases suspected for avian necrotic enteritis (clinical and subclinical disease)																							
A	15S0027													■									
B#	15S0031													■									
C#	15S0055								■														
	15S0056								■														
D#	15S0059_1								■														
E#	15S0060				■																		
F#	15S0069																						■
G	16S0002					■																	
	16S0003					■																	
	16S0004					■																	
H	16S0019														■								
I	16S0023														■								
J	16S0027				■																		
K	16S0031				■																		
L	16S0037			■																			
	16S0038			■																			
	16S0039			■																			
	16S0041			■																			
M	16S0043																		■				
	16S0051							■															
	16S0055							■															
N	16S0057							■															
	16S0058							■															
Strains obtained from caecum from healthy birds in slaughterhouses																							
O	16S0071																						■
	16S0072_1																			■			
	16S0086_1																			■			
	16S0088																						■
	16S0093_1																				■		
	16S0095					■																	
P	16S0096																	■					
Q	16S0098_1																						■
	16S0099_1																						■
	16S0100_3																						■
R	16S0105_1															■							
	16S0105_3																■						
	16S0106_1																■						
	16S0107																						■
	16S0108_1															■							
	16S0108_3																	■					
S	16S0111_1																		■				
	16S0112_1																						■
T	16S0117_1																						■
	16S0118_1																						■
U	16S0120_1																						■
Strains obtained from retail meat parts																							
4	16S0139				■																		
5	16S0142																				■		
10	16S0209			■																			
11	16S0212							■															
13	16S0243																						■
14	16S0267_1																						■

* Processors A to N refer to poultry flocks, processors O to U refer to slaughterhouses, processor 4 refers to house rearing facitity, processors 5, 10, 11 and 14 refer to slaughterhouses, processor 13 refers to retail outlet

Cases of subclinical necrotic enteritis

Colored cells (blue and black cells) show strains that belong to the same CT based on a neighbor distance of less than 60 allelic differences used to assign clusters. Black cells show the strains which are related within the cluster type based on less than 10 allelic differences.

3.3.5 Whole genome SNPs and cgMLST population structure

A set of 201,095 whole genome SNP positions were identified in the coding and non-coding regions among the 160 *C. perfringens* strains after a reference mapping and the discarding of ambiguous SNPs (2.3.9). Of the retained 220,374 unique SNPs with distinct alleles, 125,186 (56.8%) were classified as synonymous SNPs resulting in no change in the protein sequence while 67,852 (30.78%) coded for non-synonymous changes that alter the protein sequences. The non-coding SNP group comprised 27,336 (12.4%) SNPs.

Based on the whole genome SNPs, a phylogeny free population genetics approach was implemented using hierBAPS to define the population structure for the 160 *C. perfringens* strains (2.3.9). The software hierBAPS divided 159 of the 160 *C. perfringens* strains into four main lineages at the first level of clustering and 16 clusters at the second partitioning level (additional data 23). Strain (CBA7123) was an outgroup consistent with our previous analysis (3.1.2). The number of strains per lineage ranged from 7 to 109 (Figure 21 and 22, additional data 23). These results were in congruence with the phylogenetic network (3.1.2) that was created based on the cgMLST allelic profiles (Figure 21A) as well as the phylogenetic network based on the whole genome SNPs (Figure 21B). Both methods divided the 159 isolates into four major lineages. However, as shown in Figure 21A and 21B, the inference of the four different lineages from the phylogenetic network that was constructed based on the SNP distances between the genomes was clearer when compared to the phylogenetic network that was created based on the cgMLST allelic distances between the genomes.

Interestingly, calculating the total number of SNPs between each pair of genomes (2.3.9) revealed a bimodal distribution of these pairwise SNP distances Figure 21D. This distribution was consistent with the distinct separation of the strains into four major clades as it was observed that the minimum number of SNPs between a pair of genomes from different clades was 40,571 while the maximum number of SNPs between the pair of genomes within one clade was 34,872 (Figure 21D, additional data 24). Moreover, there was a direct correlation between the number of SNPs and cgMLST allelic differences that were calculated between pairs of genomes for certain range (Figure 21C). However, this direct correlation was not observed for larger values of SNP differences between the isolates (>20,000 SNPs), in which the corresponding cgMLST profile differences were approaching a plateau (Figure 21C).

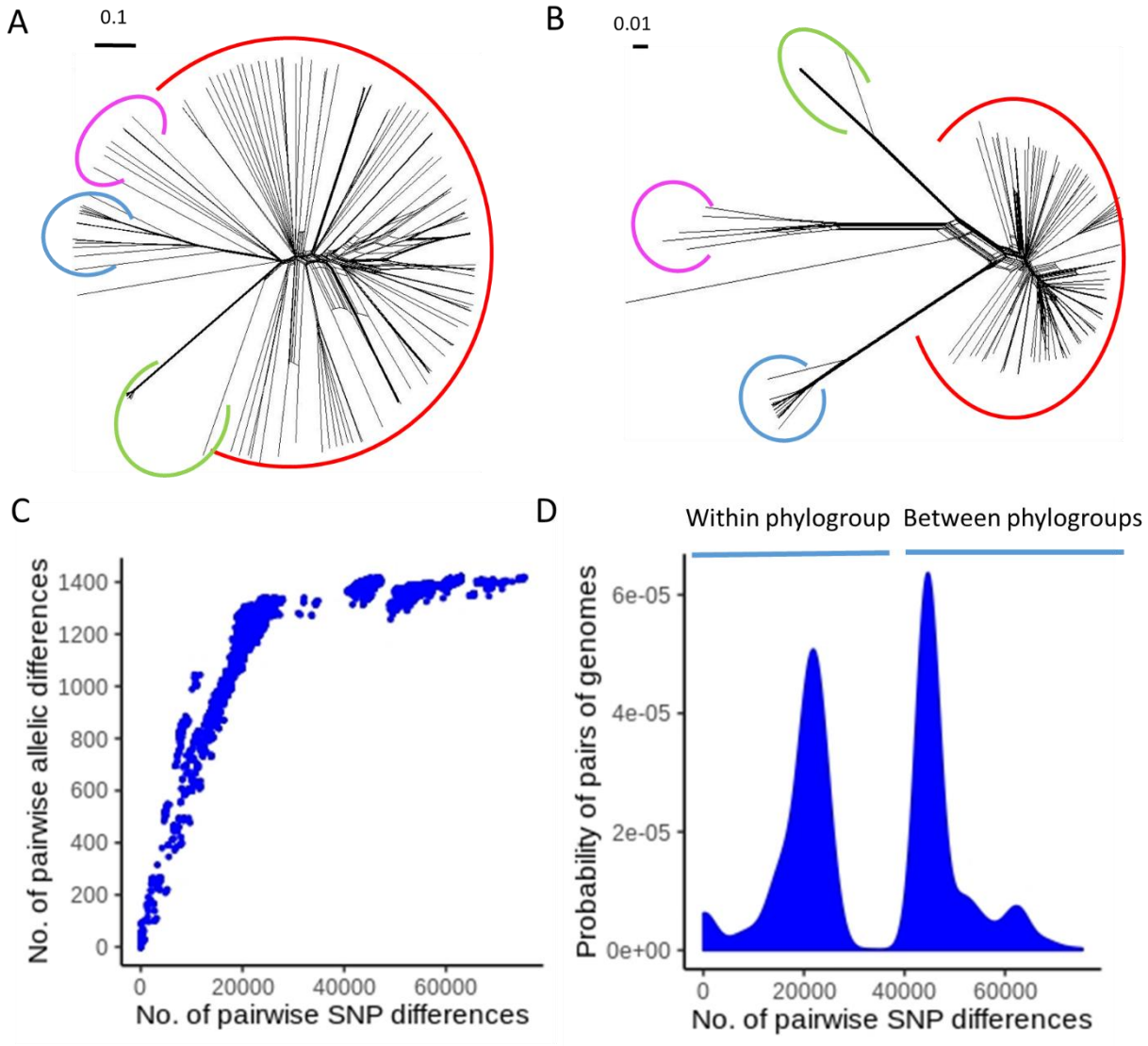


Figure 21: Comparison of core genome MLST phylogenetic network and the whole genome SNP phylogenetic network. Phylogenetic networks are shown based on the cgMLST allelic profiles **(A)** and the whole genomes SNPs **(B)**. Circles around branches were colored based on the clades identified from hierBAPS (phylogroup I: blue; phylogroup II: green; phylogroup III: red; and phylogroup IV: violet). **(C)** Correlation between the numbers of pairwise SNP distances to the pairwise cgMLST allelic differences. **(D)** Density distribution of the pairwise SNP differences between the strains. Note the bimodal distribution of the pairwise SNP differences.

To provide a consistent nomenclature for these clades (Figure 21B), the same numbers as described based on the network analysis were used (3.1.2). Phylogroup I comprises 16 strains that mainly include the chromosomal *cpe* strains and the Darmbrand strain (3.1.2 and additional data 23). Phylogroup II includes 27 strains, 26 were isolated from enteritis cases in

foal and dogs and were highly similar based on the cgMLST typing (CT 01) (3.1.2). One strain (T1) of poultry origin was additionally identified in this group and was highly divergent from the other 26 strains by an average of 6,356 SNPs (additional data 24). Phylogroup III is the largest heterogeneous phylogroup that involved ~68% of the investigated *C. perfringens* genomes. Strains within this phylogroup were derived from diverse hosts and were associated with different diseases (additional data 23). Phylogroup IV was the group that was newly defined in this investigation. SNP differences between the isolates in this phylogroup (IV) were very high (additional data 24) and consisted of seven isolates, all from poultry except strain (MJR7757A) which is isolated from human. All the analyzed poultry isolates of *C. perfringens* (except strain T1) belonged to the phylogroups III and IV (additional data 23). In addition, the Egyptian poultry strains were observed only as a part of phylogroup III (additional data 23).

3.3.6 Comparison between the core genome MLST and the SNP-based phylogenies

By using the tanglegram algorithm (Scornavacca et al., 2011) the degree of topological congruence between the cgMLST phylogenetic tree and the SNPs-based phylogenetic tree could be visualized (2.3.7). As shown in Figure 22, a very good concordance between the two phylogenies could be observed where most connecting lines between the corresponding taxa of both phylogenetic trees were in parallel except for very few crossing lines. The 160 *C. perfringens* strains were divided into 158 unique types by the cgMLST while the SNP-based typing methods resulted in 156 (recombination-filtered) (2.3.7) and 160 (recombination-unfiltered) (2.3.6) different types for the strains (additional data 25). This resulted in a Simpson's index of diversity (Hunter and Gaston, 1988; Simpson, 1949) of 1.000; 95% confidence interval (CI): 0.999-1.000 for the cgMLST typing and a Simpson's index of diversity of 0.999; 95% confidence interval (CI): 0.998-1.000 for the recombination-filtered SNP-based typing (Table 15). The recombination-unfiltered SNP-based typing assigned each strain as independent type (Simpson's index of diversity =1). These results indicate that both methods provided a maximum resolution for these strains based on the Simpson's index of diversity (Hunter and Gaston, 1988; Simpson, 1949).

3.3.7 Comparison of the core genome MLST and the classical MLSTs

Sequence types (ST) were defined for the 160 genomes using classical MLSTs based on schemes developed by Jost et al. (2006b), Hibberd et al. (2011) and Deguchi et al. (2009) (2.3.8). For the Jost scheme, eight loci with a total length of 2,987bp were analyzed (2.3.8) while for the scheme described by Hibberd et al., (2011) eleven loci with a total length of 5,779bp were analyzed (2.3.8). For the Deguchi scheme, reduced lengths of three loci (*plc*, *colA* and *sigK*) were investigated because it was found that the *sigK* gene was truncated in 16 genomes, the *colA* gene had a frameshift in five genomes and the *plc* gene in eight genomes

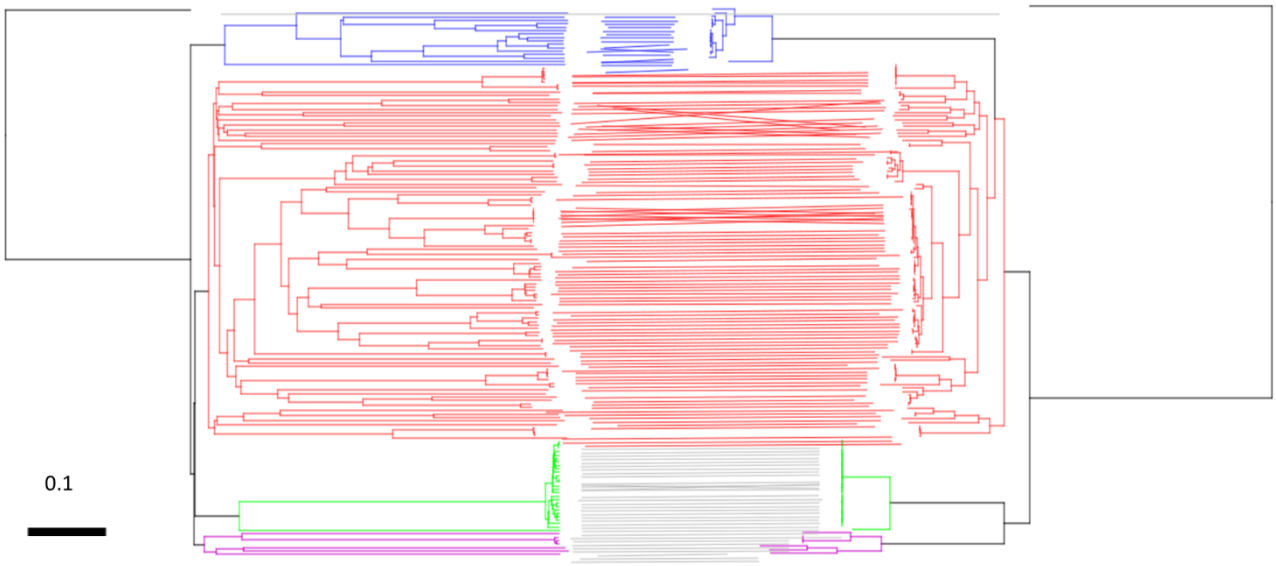


Figure 22: Topological concordance between the cgMLST neighbor-joining tree (left) and the SNP-based maximum likelihood tree (right) is represented as tanglegram. Tree branches were colored based on the clades identified from hierBAPS: clade1: blue, clade2: green, clade3: red and clade 4: violet.

harbored an intron resulting in a truncated *plc* at a contig break. In sum, a total nucleotide length of 4,874bp (instead of the originally used 5,274bp) for the eight loci was analyzed based on the Deguchi scheme (2.3.8).

MLST schemes described by Jost et al. (2006b), Hibberd et al. (2011) and Deguchi et al. (2009) resulted in 77 STs (Jost) (Simpson's diversity index = 0.962, 95% confidence interval CI = [0.945-0.979]), 88 STs (Hibberd) (Simpson's diversity index = 0.967, 95% confidence interval CI = [0.949-0.984]) and 91 STs (Deguchi) (Simpson's diversity index = 0.972, 95% confidence interval CI = [0.956-0.987]), respectively (Table 15). As expected, sequence types inferred from the classical MLST could be further divided into different cgMLST cluster types indicating the higher discrimination of the cgMLST compared to classical methods. The degree of agreements between the different typing methods were described as Adjusted Wallace's coefficient (Table 15). In addition, the tanglegram algorithm was used also (Scornavacca et al., 2011) to visualize the topological congruence between the neighbor-joining phylogenetic tree from the cgMLST and the classical MLSTs (2.3.8, Figure S3). Classical MLSTs were less concordant with the whole genome based typing methods despite the high discrimination level of all methods (based on the Simpson's diversity indices).

Table 15: Analysis of genotyping results as determined by the discrimination power (based on the Simpson's Index of Diversity) and typing concordance (calculated by using the adjusted Wallace's coefficient)

Typing method	No. geno-types	Simpson's index of diversity (95% confidence interval)	Adjusted Wallace index of concordance (95% confidence interval)						
			Classical MLST			Core genome-based MLST		Whole genome-based SNP typing	
			Jost scheme	Deguchi scheme	Hibberd scheme	Core genome ST	Cluster types CT-60	No filter for Recombination	Filtered for Recombination
Classical MLST									
Jost scheme	77	0.962 (0.945-0.979)		0.733 (0.599-0.866)	0.336 (0.190-0.481)	0.006 (0.000-0.031)	0.803 (0.768-0.838)	0.000 (0.000-0.000)	0.020 (0.000-0.044)
Deguchi scheme	91	0.972 (0.956-0.987)	0.986 (0.972-1.000)		0.344 (0.181-0.507)	0.008 (0.000-0.038)	0.926 (0.903-0.949)	0.000 (0.000-0.000)	0.027 (0.000-0.055)
Hibberd scheme	88	0.967 (0.949-0.984)	0.385 (0.201-0.569)	0.293 (0.125-0.462)		0.000 (0.000-0.025)	0.365 (0.180-0.550)	0.000 (0.000-0.000)	0.000 (0.000-0.025)
Core genome-based MLST									
Core genome ST*	158	1.000 (0.999-1.000)	1.000 (1.000-1.000)	1.000 (1.000-1.000)	0.000 (0.000-0.563)		1.000 (1.000-1.000)	0.000 (0.000-0.000)	1.000 (1.000-1.000)
Cluster types (CT-60)**	97	0.969 (0.951-0.987)	0.989 (0.981-0.998)	0.848 (0.690-1.000)	0.392 (0.216-0.568)	0.007 (0.000-0.034)		0.000 (0.000-0.000)	0.025 (0.000-0.049)
Whole genome-based SNP typing									
No filter for Recombination	160	1.000 (NaN-NaN)	1.000 (1.000-1.000)	1.000 (1.000-1.000)	1.000 (1.000-1.000)	1.000 (1.000-1.000)	1.000 (1.000-1.000)		1.000 (1.000-1.000)
Filtered for Recombination	156	0.999 (0.998-1.000)	1.000 (1.000-1.000)	1.000 (1.000-1.000)	0.000 (0.000-0.262)	0.300 (0.000-0.737)	1.000 (1.000-1.000)	0.000 (0.000-0.000)	

*ST, sequence types, **CT-60, Cluster types as identified using the cgMLST based on 60 allelic mismatches.

Chapter 4: Discussion

4.1 *In silico* analysis of the genomic variability, phylogenetic relatedness and virulence genes of *C. perfringens*

4.1.1 Genomic overview and chromosomal (re)arrangement in *C. perfringens*

C. perfringens has a sympatric lifestyle which may have some impacts on its genomic features and influence the genomic variability within this species. One aim of this study was to computationally investigate the *C. perfringens* genomic diversity through analyzing the publically available WGS data of *C. perfringens* strains (3.1). Therefore, the Pacific Bioscience raw sequence data of 23 NCTC strains originally sequenced by the NCTC 3000 project were collected (2.1.1) and assembled (2.1.2). Among these sequence data were the genome sequence of 13 food poisoning strains with chromosomal *cpe* (type F) and one type C Darmbrand strain (Table 11). Additionally, these data were combined with 53 NCBI publically available genomes, 32 of them were *netF* positive strains derived from cases of foal necrotizing enteritis (n = 16) and canine hemorrhagic diarrhea (n = 16) (Mehdizadeh Gohari et al., 2016; Mehdizadeh Gohari et al., 2017) (2.1.1). This procedure may cause a potential bias in the analyzed data set as the *netF* positive strains are overrepresented. However, the inclusion of these strains was important to identify the phylogenetic structure of *C. perfringens* based on all available genomic data. In total, the genomes of 76 *C. perfringens* strains were investigated (3.1). From the information available, these 76 strains are temporally and geographically diverse spanning a time period from 1920s to 2010s and originate from different ecosystems (humans, animals, food and environment) of various continents (America, Europe, Asia and Australia) (2.1.1, Figure 2 and Table 3). The analyzed 76 strains were confirmed as members of the species *C. perfringens* as the ANI values were above 95% between all genomes (3.1.2, Figure S1B) (Richter and Rossello-Mora, 2009).

Among the 76 analyzed genomes, 30 were in their closed state of assembly (3.1.1.4). A substantial degree of diversity was observed among the genome content of these 30 strains regarding episome content (up to six episomes per strain), chromosome size (2.9 to 3.5Mbp) and the presence of mobile genetic elements (Table S1-S4). However, despite this variability, there was a general genomic conservation among the strains in respect to the physical location and relative order of genes in each chromosome. Few inversions mostly confined to integrated phages or genes flanked by IS elements were detected (3.1.1.4, Figure 4A and additional data 4). However, strains NCTC 11144 (a food poisoning isolate), NCTC 8081 (Darmbrand type C strain) and NCTC 8359 presented much less conservation in their genome organization with reversals and shifts that were observed along the genome segments. In these strains the

distribution of the PacBio reads across the chromosome was uniform without discontinuities in the mapping pattern excluding a possible misassembly in these strains. However, one coverage spike was observed in the strain NCTC 8359 probably due to a repeat collapse (additional data 3). Strain NCTC 11144 showed an inversion of a large genomic region around the terminus of replication (3.1.1.4, Figure 4B). This large inversion was bordered by rRNA operons. Large chromosomal inversions were already reported in various bacterial species to occur symmetrically around the replication origin and terminus (Eisen et al., 2000; Iguchi et al., 2006; Raeside et al., 2014) as it was previously suggested that most recombination events occur in relation to the replication fork (recombination hot spots) (Mackiewicz et al., 2001; Tillier and Collins, 2000). However, strains NCTC 8081 and NCTC 8359 do not follow the same pattern of rearrangement as strain NCTC 11144 (Figure 4B). Several blocks of genes in strain NCTC 8081 were translocated and inverted whereas one block in strain NCTC 8359 was translocated (Figure 4B). The inversions and shifts in these two strains were bordered by identical copies of the IS element ISCpe7. Chromosomal rearrangements in association with IS elements as found here have been also previously reported e.g. for *Bordetella* spp. (Parkhill et al., 2003). These results - when taken together - imply a considerable conserved genomic synteny (physical location and relative order of homologous blocks) between the genomes of most *C. perfringens* which was generally observed in 90% of the investigated strains (27 out of 30). Few genomes (3 out of 30) were less conserved with inversions and shifts associated with IS elements in two genomes (3.1.1.4, Figure 4B). For unknown reasons, the frequency of such events seems to be very low in *C. perfringens* compared to other bacteria e.g. pathogenic *Yersinia* (*Y.*) spp. (Darling et al., 2008). The influence of chromosome rearrangements on the bacterial phenotype has been previously described to affect the cell viability of *Escherichia (E.) coli* (Darmon and Leach, 2014; Esnault et al., 2007). However, it is unclear how these inversions could influence the phenotypic characteristics in *C. perfringens*.

4.1.2 Impact of mobile genetic elements on *C. perfringens* genome size and variability

Mobile genetic elements (MGE) affect bacterial genome structure and function. These effects include gene inactivation (Simonet et al., 1996) or activation, altering the gene order and deletions of large DNA segments that may result in a reduction of genome size (Darling et al., 2008; Liang et al., 2010; Siguier et al., 2014). A search of the closed genomes for the presence of MGE was performed (3.1.1.3) to detect specifically integrated phages (3.1.1.3.1), insertion sequences (IS) (3.1.1.3.2) and genomic islands (GI) (3.1.1.3.3) as well as CRISPR elements (3.1.1.2) which confer protection against bacteriophage invaders in bacteria (Koonin et al., 2017). Prophages were detected at a variable range with no direct correlation of their frequency to the absence or presence of CRISPR elements (3.1.1.3.1, Table S1). CRISPR elements were found in 16 strains but were absent in the chromosome of 14 strains (Table S1, 2.1.2,

3.1.1.2). IS elements and GIs were identified with highly variable occurrence between strains (3.1.1.3.2, 3.1.1.3.3, Table S3). Strikingly, a significant accumulation of ISs and GIs in certain genomes was found. These comprise the Darmbrand strain and chromosomal *cpe* strains which are also characterized by a smaller genome size (2.9 - 3Mbp) compared to other *C. perfringens* strains (> 3Mbp) (3.1.1, Table S1). In these strains, ISs and GIs were predicted on average of 3.7% and 6.7% of the genome size, respectively in contrast to other strains (average genome size of 0.49% and 1.87% for IS and GI, respectively; 3.1.1.3.2, 3.1.1.3.3, Figure 3A). According to the phylogenetic analysis (3.1.2), these genomes are closely related forming a single phylogenetic group (referred to as phylogroup I) (3.1.2, Figure 5). Genome reduction linked to large expansions of IS elements has been reported in different bacteria in association with bacterial specialization to certain niches e.g. *Y. pestis*, *Bordetella pertussis, Burkholderia mallei*, *Salmonella enterica* and *Shigella* (Moran and Plague, 2004; Parkhill et al., 2003; Preston et al., 2004; Song et al., 2010). Of interest, phylogroup I strains cause mainly human foodborne illnesses (Ma et al., 2012; Shrestha et al., 2018). It has been proposed that chromosomal *cpe C. perfringens* strains have a different epidemiology and are adapted to a different environment than other *C. perfringens* strains (Lindström et al., 2011). It is therefore highly plausible to assume that IS elements are driving the evolution in these strains towards a certain niche streamlining these genomes by discarding genes which are no longer necessary and thereby favour replication efficiency. It has to be mentioned that two strains (JGS1721 and JGS1495) which are not members of phylogroup I were found also to be enriched with IS elements (additional data 26) indicating that the feature of IS expansion is not strictly limited to this phylogroup.

A biased genomic variability of *C. perfringens* was previously described after a genomic comparison of strain 13 (soil isolate), SM101 (food poisoning isolate) and ATCC 13124 (type strain) (Myers et al., 2006). One of the genomes analyzed (SM101) was enriched for IS elements which were unevenly distributed and biased to a more variable replichore (Myers et al., 2006). With the advantage of having 30 chromosomes in their closed state, this study aimed to investigate the variability within these genomes to portray the distribution pattern of GIs and ISs across chromosomes (3.1.1.4, Figure 3B). The alignment of the genomes showed that variable regions are present across the whole chromosome. However, their distribution was to some extent shifted toward one replichore with the exception of the three strains with different chromosomal rearrangements (described in 4.1.1). Locally collinear blocks (LCBs) within one replichore (left side in Figure 4A) were shorter with several rearrangements, breakpoints and abundances of regions that lack detectable homology compared to the other replichore (right side in Figure 4A). Plotting the distribution of IS elements and genomic islands (3.1.1.3.3) across the chromosome revealed asymmetrical distribution of GIs and ISs in chromosomal *cpe* strains toward the less stable replichore (Figure 3B). The Darmbrand strain

was also highly enriched for ISs and GIs. However, because of DNA rearrangements in the strain a distribution bias was not observed (Figure 3B). The concentration of IS elements and GIs towards one replichore may indicate that non-essential gene families are primarily located in this part of the genome or that a certain replication mechanism favours this bias. The different stability of the replichores may accelerate or restrain the evolution in certain gene families according to their location.

4.1.3 Distinct *C. perfringens* clustering based on core genome SNPs and accessory gene content

To infer the phylogenetic relationships of *C. perfringens* (3.1.2), the genomes of 76 *C. perfringens* strains were analyzed including 30 closed genomes plus 46 draft genomes as these analyses are not significantly affected by the assembly status of the genomes (Figure 1 and Figure 2). Core genome SNP based ML phylogeny and split network analysis grouped the 76 *C. perfringens* strains into 27 distinct clusters (Figure 5) and 3 phylogroups (Figure 6). Cluster 1 comprises 15 strains mainly involved in cases of human food poisoning plus the NCTC 8081 strain (Darmbrand strain). These strains carry the enterotoxin gene (*cpe*) on a chromosome except strain NCTC 8081 in which the gene is located on a large plasmid (116Kbp) and NCTC 8678 in which the gene is absent. Additionally, Cluster 1 genomes carry at least one extrachromosomal element and none of these except NCTC 8081 harbor the *C. perfringens tcp* locus of conjugation. Strain NCTC 10240 clustered separately based on patristic distance (cluster 19) (3.1.2, Figure 5). However, this strain also carries the enterotoxin gene on a chromosome and is very closely related to cluster 1 on the phylogenetic tree (3.1.2, Figure 5) with a mean nucleotide divergence of 0.85%. Both cluster 1 and the strain NCTC 10240 belong to a single phylogroup (phylogroup I) based on network analysis (3.1.2, Figure 6). Interestingly, MLST analyses based on 8 housekeeping genes (Deguchi et al., 2009) grouped these phylogroup I strains together with a larger number of food poisoning strains that carry chromosomal *cpe* and enteritis necroticans (Darmbrand) isolates which have not been whole genome sequenced so far (3.1.2, Figure S2). In summary, these results corroborate the genetic relatedness between chromosomal *cpe* strains (and Darmbrand strains) suggesting a common evolutionary history as previously hypothesized (Lindström et al., 2011; Ma et al., 2012).

Cluster 2 and 3 contain 26 and 6 strains, respectively. These strains were involved in cases of foal necrotizing enteritis and canine hemorrhagic diarrhea (3.1.2, Figure 5 and 6). They carry the gene encoding NetF toxin and the plasmid encoded *cpe* gene (Mehdizadeh Gohari et al., 2017; Mehdizadeh Gohari et al., 2015). The complete genomes of these groups (strains JF55 and JF838) show a larger size (3.3 - 3.5Mbp) and many plasmids (n = 5). Some of these

plasmids are possibly conjugative carrying the *tcp* conjugation locus (Mehdizadeh Gohari et al., 2015). Cluster 2 formed a single phylogroup (phylogroup II) while cluster 3 together with further 21 clusters collectively formed phylogroup III (Figure 5 and 6).

Grouping the isolates based on their accessory gene content and correlating this pattern to the core genome phylogeny led to the identification of genes which are distinctly associated with different phylogroups and thus may contribute to the characteristic phenotype of some strains e.g. disease outcome or spore resistance (3.1.3, additional data 12). In phylogroup I, 90% of the strains lack 304 chromosomal genes which are present in 90% of the phylogroup II and III strains (3.1.3). In parallel, 90% of the strains of phylogroup I carry 64 additional gene families which are absent in 90% of other strains (3.1.3, additional data 12). It was striking to find that the pattern of gene loss (304 chromosomal genes) in phylogroup I was more prominent than gene gain (64 genes) which correlates well with the characteristic smaller genome size. The loss of chromosomal genes in phylogroup I was in sharp contrast to phylogroup II genomes (*netF* positive strains related to JP838) where additional 313 genes and simultaneous absence of 37 chromosomal genes were found in 90% of the phylogroup II genomes but not in phylogroup I and III (3.1.3, additional data 12). This indicates that the gain and loss of genetic elements within the species *C. perfringens* is not homogenous in the different phylogroups. It seems that phylogroup II is directed to gain new genetic material while phylogroup I is directed toward gene loss. To date, such a divergent pattern of evolution within a single species has not been described in other bacteria.

A common epidemiological background (disease and host) was previously reported for the *netF* positive genomes (Mehdizadeh Gohari et al., 2017; Mehdizadeh Gohari et al., 2015). However, phylogenetic analysis split the *netF* strains (n = 32) into two lineages: phylogroup II and a cluster within phylogroup III as previously described (Mehdizadeh Gohari et al., 2017). Interestingly, both lineages have 49 genes in common which are absent in 90% of the other strains (3.1.3, additional data 12).

4.1.4 *In silico* analysis of potential virulence factors

A set of 397 genes that share homology (of variable amino acid identity) with experimentally verified virulence factors in various pathogens was identified (3.1.4, additional data 13). Thus, these genes may contribute to the virulence of *C. perfringens*. 40% of the identified genes belong to COG category “cell wall/membrane/envelope biogenesis” followed by the categories “defense mechanisms” and “inorganic ion transport and metabolism”. In concordance with these results, the majority of the identified virulence gene candidates (94 genes = 24%) were found to have a variable degree of homology (23% - 84%) to known capsular genes in some Gram positive and Gram negative bacteria (3.1.4). Many strains of *C. perfringens* are known

to produce an extracellular capsule. Also, a serotyping system based on the bacterium's capsular variation was developed (Hughes et al., 1976). Interestingly, only 10% of the capsular gene-homologs are present in all strains (core capsular genes) while 90% are part of the variable genome (3.1.4, additional data 13). In line with this observation is a previous report showing that most of the capsular genes belong to the variable genomic regions in *C. perfringens* (Myers et al., 2006). Many bacteria produce an extracellular capsule which renders the bacterium resistant to phagocytosis (Nanra et al., 2013). However, the capsule of *C. perfringens* strain 13 does not inhibit phagocytosis (O'Brien and Melville, 2003).

C. perfringens is known for its ability to produce a wide variety of extracellular toxins and enzymes (at least 20) to retrieve essential nutrients for its growth (Kiu and Hall, 2018). Recently, Kiu et al., (2017) reported the presence of three toxin genes; phospholipase (*cpa*), clostripain (*closI*) and collagenase (*colA*) genes in all *C. perfringens* strains (n = 56), thus representing core toxin genes (Kiu et al., 2017). This finding was also corroborated by this study (3.1.4, additional data 13). However, *colA* was frequently (additional data 14) subjected to disruptive events such as frameshifts or large deletions indicating that the gene may not be translated or may produce a truncated protein (3.1.4).

Some of the *C. perfringens* toxins are pore forming e.g. beta toxin, delta toxin, NetB, NetE, NetF and NetG (Li et al., 2013; Mehdizadeh Gohari et al., 2015). These toxins are related to the pore-forming Leukocidin/Hemolysin family which also includes some members of *Staphylococcus (S.) aureus* toxins e.g. the α-toxin (Berube and Bubeck Wardenburg, 2013). In this study, another so far unknown homolog to this family was identified in the Darmbrand NCTC 8081 strain (3.1.4, Figure 15). The gene homolog in this strain was flanked by mobile elements, and located on a large plasmid which also carries the enterotoxin gene (*cpe*) and the transfer of clostridial plasmids (*tcp*) locus suggesting that this plasmid could be a conjugative plasmid (Adams et al., 2015). Further investigations are needed to decipher the biological characteristics and possible contribution to virulence of this new homolog.

Perfringolysin (PFO) of *C. perfringens* is also a pore-forming toxin related to the cholesterol-dependent cytolysins (thiol-activated cytolysin) (Verherstraeten et al., 2015). Interestingly, an additional copy of a PFO-like sequence was identified downstream the typical PFO gene with 85.7% nucleotide identity and 81.7% protein identity (3.1.4, additional data 15). This PFO-like sequence was predominantly present in the genomes of phylogroup II (*netF* positive strains related to JP838) and was predicted as a coding region by RAST and Prokka (length 1,578bp, positions 222,029 - 223,606) (3.1.4). However, the NCBI annotation predicted a disrupted gene (pseudogene) in JP838 (positions 222,212 - 223,606) (3.1.4). All strains in phylogroup I (chromosomal *cpe* strains and the Darmbrand strain) lack the *pfoA* gene, as well as the strains

JJC, NCTC10578 and MJR7757A (3.1.4). The latter strain (MJR7757A) however, carries only the divergent form of *pfoA*. Phenotypically, PFO plays a role in the virulence of *C. perfringens* e.g. in gas gangrene infection (1.1.2.1.1, 1.1.3.2). Whether the divergent form of *pfoA* could be expressed and contribute to the disease phenotype remains to be elucidated.

Prior analysis of the genome of *C. perfringens* strain 13 identified seven putative iron-acquisition systems: two heme acquisition systems, one ferrous iron acquisition system (*feoAB*), three siderophore-mediated acquisition systems and one ferric citrate iron acquisition system (Awad et al., 2016). Strains ATCC 13124 and SM101 were also previously reported to have three and two copies of *feoAB* operon, respectively (Myers et al., 2006). Two of these systems (ferrous iron acquisition system encoded by *feoAB* operon and a heme-uptake system encoded by *C. perfringens* heme transport "*Cht*" locus) were experimentally proven to be essential for the virulence of *C. perfringens* in gas gangrene models (Awad et al., 2016; Choo et al., 2016). Consistent with their importance, it was shown that both loci (*Cht* and *feoAB*) are maintained by all investigated *C. perfringens* genomes (100% presence, pairwise identity 98.4% and 98.8% respectively) (3.1.4, additional data 16). Further putative iron acquisition systems were observed in more than 75% of the strains (3.1.4, additional data 16). Interestingly, phylogroup I strains lack one heme- and one siderophore-iron uptake system, while phylogroup II strains (*netF*-positive JP838-related) carry an additional putative iron uptake operon (3.1.4, additional data 16). Taken together, the preservation of a variety of iron uptake systems in *C. perfringens* could enable the bacterium to survive iron shortage conditions in the environment and to retrieve iron sequestered by host proteins during infections.

Summary

The publically available genome sequence data of 76 *C. perfringens* strains from diverse ecological, geographical and temporal niches were analyzed. Data analyses included 30 complete genomes which were composed of a circular chromosome (2.9 to 3.5Mbp) and up to six extrachromosomal elements. A substantial degree of genomic variability was detected in respect to episome content, chromosome size and mobile elements. Insertion sequences were identified and revealed abundance of their occurrence in certain genomes. Comparative alignment of complete genomes displayed a considerable degree of conservation in the order of genes in each chromosome except for three of 30 genomes.

A phylogenetically related group of strains (phylogroup I) was identified consisting of human food poisoning strains that carry chromosomal *cpe* and a Darmbrand (enteritis necroticans) strain. This phylogroup is characterized by a high abundance of IS elements and genomic islands. The distribution of ISs and GIs in chromosomal *cpe* strains was shifted toward one

chromosomal replichore (biased distribution). This distribution pattern was not observed in the Darmbrand strain probably because of the inversions of DNA segments. Additionally, a marked loss of chromosomal genes was detected in this phylogroup which may explain the characteristic relative small genome size compared to other *C. perfringens* strains. Phylogroup I strains also lack one putative heme- and one putative siderophore-iron uptake system as well as the *pfoA* gene. These genomic features of phylogroup I (abundance of IS elements and genome reduction) provide indications that these strains adapt to a certain habitat causing human foodborne illness. It could be assumed that the absence of certain virulence genes (iron uptake systems and PFO) indicate also adaptation to a less competitive environment (food) for replication.

Chromosomal *cpe* strains as well as Darmbrand strains can form spores with high heat resistance properties (Ma et al., 2012; Shrestha et al., 2018). These strains produce a variant of small acid-soluble protein 4 (Ssp4) that tightly binds to the spore DNA conferring high spore resistance phenotype (Li et al., 2013). Chromosomal *cpe* strains are considered an important food contaminant (Lindström et al., 2011). These strains can survive cooking and replicate very fast while bacterial competitors are absent in the food matrix (Lindström et al., 2011; Shrestha et al., 2018). After ingestion, they can also survive the acidity of the stomach and passage to the intestine where they undergo sporulation and CPE production (Shrestha et al., 2018). These strains release the CPE toxin and the spore in the intestine upon the lysis of the mother cell, thereby inducing lesions in the intestine, diarrhea and abdominal cramps (Shrestha et al., 2018). The disease is relatively mild and self-limiting with few fatality cases that were reported in elderly or debilitated individuals (Freedman et al., 2016). Darmbrand strains are also involved in human foodborne illnesses but cause necrotizing enteritis referred to as “type C enteritis necroticans” or “Darmbrand”. However, the disease is life threatening in humans with severe abdominal pain and bloody diarrhea (Ma et al., 2012).

In contrast to phylogroup I, phylogroup II is directed to gain new genetic material. They also carry an additional putative iron uptake operon beside the previously described seven putative iron systems (Awad et al., 2016). Moreover, an additional copy of a *pfoA*-like gene was found in phylogroup II along with the typical *pfoA*. It is noteworthy to mention that phylogroup II strains can cause enteric lesions in different hosts (equine and canine).

In summary, phylogroup I (chromosomal *cpe* and Darmbrand strains) is subjected to certain evolutionary mechanisms and displays characteristics that indicate speciation of these strains.

4.2 Characterization and typing of *C. perfringens* isolates from heathy and diseased poultry in Egypt

4.2.1 Isolation and toxin-genotyping of *C. perfringens* strains

C. perfringens strains were isolated from cases suspected for NE in Egyptian poultry flocks and from apparently healthy poultry at local slaughterhouses (3.2.1). 51 isolates from suspected NE-cases and 83 isolates from healthy birds were characterized by PCR and were classified toxin-genotype A possessing only the *cpa* gene among the major "typing" toxins-genes (Table 12, 3.2.2). This is in concordance with previous studies that reported the isolation of type A strains of *C. perfringens* from enteritis cases in poultry from different countries (Brady et al., 2010; Tolooe et al., 2011b). Recently, former type A strains carrying the *netB* gene were renamed as type G (Rood et al., 2018). NetB toxin has been associated with the virulence of *C. perfringens* strains which induce NE disease in poultry (Keyburn et al., 2008; Keyburn et al., 2010b). Since its discovery in 2008 (Keyburn et al., 2008), *netB* prevalence studies report *netB*-negative strains in NE cases and also *netB*-positive strains in healthy birds (Abildgaard et al., 2010; Bailey et al., 2015; Gad et al., 2011; Keyburn et al., 2010a; Martin and Smyth, 2009; Park et al., 2015) (Table 16). The *netB* gene was not detected in suspected NE strains nor in strains from asymptomatic birds in Egypt similar to studies that reported zero to very low prevalence of *netB* in NE isolates (Table 16). This discrepancy in the *netB* occurrence is far from well understood. Several explanations have been proposed such as: I) Unintentional selection of *netB*-negative commensal strains during initial culturing (Lepp et al., 2013): II) Loss of the *netB*-carrying plasmid during subcultivation (Lepp et al., 2013): III) Misidentification of the cases as *C. perfringens* induced NE, especially in cases of mixed infection (Uzal et al., 2016a): IV) Sampling error (Smyth, 2016), or that V) NetB is not essential in all cases of NE and hitherto undefined factors are responsible for the NE pathogenesis in *netB*-deficient *C. perfringens* strains (Keyburn et al., 2008).

The beta2 toxin gene was detected with a high prevalence in both samples from diseased and non-diseased birds (Table 12, 3.2.2) with no direct correlation to a specific health status which is in agreement with previous reports (Tolooe et al., 2011a). In this study, only the atypical variant of the *cpb2* gene was detected in all *cpb2* positive isolates (37 isolates from suspected NE cases and 42 isolates from healthy birds) (3.2.2, Table 12). Prior investigations reported a variable occurrence of *cpb2* in NE cases (Gad et al., 2011; Hibberd et al., 2011; Park et al., 2015; Tolooe et al., 2011b). However, a direct comparison of these prevalence studies is difficult because of different primer sets used. Some primer sets can only detect the consensus variant of the *cpb2* gene e.g. the primers described by Baums et al. (2004). Consensus *cpb2* was reported to be linked to *C. perfringens* strains of porcine origin while the atypical *cpb2* gene was linked to strains from non-porcine hosts (Jost et al., 2005).

Table 16: Reported prevalence of *netB*-positive *C. perfringens* isolates in poultry

Country	Host	*C. perfringens* from NE		*C. perfringens* from healthy birds		References
		Total no. of isolates	*netB* positive (Prevalence)	Total no. of isolates	*netB* positive (Prevalence)	
Prevalence of *netB* more than 50%						
USA	Broiler	20	17 (85%)	54	10 (18.5%)	(Hibberd et al., 2011)
USA	Chicken	12	7 (58.3%)	80	7 (8.7%)	(Martin and Smyth, 2009)
Canada	Broiler	41	39 (95.1%)	20	7 (35%)	(Chalmers et al., 2008)
Canada	Chicken	6	6 (100%)	5	4 (80%)	(Brady et al., 2010)
Netherland	Layer	160*	148 (92.5%)	NR	-	(Alaart et al., 2012)
Sweden	Broiler	34	31 (91%)	NS	25%	(Johansson et al., 2010)
Denmark	Broiler	25	13 (52%)	23	14 (61%)	(Abildgaard et al., 2010)
Iran	Broiler	36	19 (52.7%)	43§	0	(Tolooe et al., 2011b)
Italy	Chicken	30	16 (53%)	22	4 (17.4%)	(Drigo et al., 2009)
Australia	Chicken	18	14 (77.7%)	NR	-	(Keyburn et al., 2008)
Australia	Chicken	24	19 (79.1%)	55	2 (3.6%)	(Keyburn et al., 2010b)
Belgium	Chicken	4	3 (75%)			
Canada	Chicken	9	4 (45%)			
Denmark	Chicken	7	5 (71.4%)			
Prevalence of *netB* less than 50%						
Korea	Chicken	17	8 (47.1%)	50	2 (4%)	(Park et al., 2015)
Germany	Turkey	NR	-	267	53 (20%)	(Gad et al., 2011)
Zero to very low prevalence of *netB*						
USA	Broiler	49	2 (4%)	NR	-	(Bailey et al., 2015)
Brazil	Chicken	22	0	NR	-	(Llanco et al., 2015)
India	Chicken	1	0	NR	-	(Thomas et al., 2014)
Italy	Turkey	42	0	NR	-	(Giovanardi et al., 2016)
Finland	Turkey	174	14 (8%)	38	0	(Lyhs et al., 2013)

* These 160 strains were isolated from 19 birds with mild NE disease
§ Ten isolates were from layer and four from broiler breeder
NS: not specified, NR: not reported

An insertion of an 834bp DNA segment was observed in the PCR amplicon of the alpha toxin gene (3.2.2, Figure 11) in seven isolates of *C. perfringens* which were isolated from two suspected NE farms. The inserted DNA segment was identical to a group II intron already described (Ma et al., 2007). The hemolytic and phospholipase C activities of these intron-carrying isolates were assessed by plating on blood agar and lactose egg yolk agar, respectively (3.2.2). As previously reported (Ma et al., 2007), the insertion did not inhibit these activities *in vitro*. The detection of this intron in *C. perfringens* isolates from Japan (Ma et al., 2007), Italy (Drigo et al., 2010), Denmark (Abildgaard et al., 2009) and now from Egypt (this

study) reflects that this is a feature of *C. perfringens* strains worldwide. So far the insertion has been detected in chicken isolates only. The function and biological meaning of this group II intron in *C. perfringens* strains is not clear to date. Future studies may provide further insights into the evolution of *C. perfringens* strains carrying this group II intron.

Ten isolates from five healthy ducks did not produce the typical dual hemolytic pattern on blood agar plates while the identity of these *C. perfringens* isolates has been confirmed by PCR (*cpa* gene) (3.2.1). It can be assumed that production or functionality of hemolysins is affected in these strains, notably alpha toxin (responsible for the outer zone of green hemolysis) and perfringolysin (responsible for the inner beta hemolysis) (Hatheway, 1990). Although the detection of non-hemolytic *C. perfringens* strains has been reported previously (Bailey et al., 2015; Schoepe et al., 2001), the pathogenic potential of these strains has not been described in detail. However, the potential use of the non-hemolytic alpha toxin variant as vaccine candidate has been investigated (Schoepe et al., 2001). Interestingly, the hemolysins alpha toxin and perfringolysin are required for the formation of *C. perfringens* biofilms (Vidal et al., 2015). Biofilms have recently been suggested to contribute to NE pathogenesis (Prescott et al., 2016). Intestinal colonization of *C. perfringens* involves the formation of adherent biofilms that provide a matrix for survival and persistence of *C. perfringens* in the intestinal lumen (Charlebois et al., 2016). Therefore, a reduced virulence of these strains seems possible.

4.2.2 Within-host *C. perfringens* diversity in diseased birds

A pilot investigation based on MLST was performed to determine the genetic relatedness of *C. perfringens* in individual diseased birds (3.2.3). For this purpose, MLST was carried out on twelve isolates from three birds, four isolates (from different locations) were investigated for each bird. The isolates from two birds (bird no. 1 and no. 2) belonged to a single MLST sequence type each (ST45 and ST46) (3.2.3, Figure 13). This is in concordance with various reports that described the limited diversity of strains in NE affected birds where a single genetic type is generally detected per flock (Lacey et al., 2016). In contrast, healthy birds and birds with mild NE were reported to harbor several different PEGE subtypes in different tissues (Johansson et al., 2010; Nauerby et al., 2003). The four isolates investigated per bird were obtained from samples collected along the intestinal tract and the liver and were kept separate during culture (3.2.3, Figure 13). The presence of a single MLST type (MLST ST) may indicate that a certain strain of *C. perfringens* dominates in the diseased bird and is most likely suppressing other strains. Indeed, recent reports by Timbermount et al. (2009, 2014) described the ability of NE strains to produce bacteriocins which enable virulent strains to outcompete other *C. perfringens* strains. While this intra-species growth-inhibition trait was found to be more prevalent in NE strains than non-outbreak strains, Timbermont et al., (2009) also reported that ~17% of NE strains were unable to inhibit the growth of other strains *in vitro*

(Timbermont et al., 2009). The *C. perfringens* isolates investigated in bird no. 3 belonged to three different MLST STs (ST46, ST47 and ST48) (3.2.3). This may indicate the possibility of a multiple-strain infection or the inability of these strains to produce bacteriocins. Interestingly, one sequence type (ST46) was found to be present in farms 2 and 3 (3.2.3). Tracing back the history of these two farms revealed ownership to the same producer. Therefore, it can be assumed that the birds had the same origin and were handled by the same staff.

The application of a MLST system (Hibberd et al., 2011) (3.2.3) provided also a good opportunity to investigate the relatedness between the *C. perfringens* isolates from this study and strains described in prior investigations (Chalmers et al., 2008a; Hibberd et al., 2011). In prior studies, distinct MLST genotypes were reported in association with NE in the poultry population in North America (Chalmers et al., 2008; Hibberd et al., 2011). In this study, the investigated twelve strains from Egypt (see 3.2.3) were assigned to new STs that do not belong to the previously described "NE-associated" genotypes (ST31, CC-4). Table 17 is summarizing the STs reported in association with NE.

Table 17: MLST sequence types for the *C. perfringens* strains that were isolated from cases suspected for necrotic enteritis disease in the USA, Canada (Hibberd et al., 2011) and Egypt

ST*	Location	No. of isolates	Host	Beta2	NetB	Lineage	
						SLV**	DLV***
ST31	Minnesota, USA	4	Broiler	+	+	-	-
	Pennsylvania, USA	6	Broiler	+	+ / -		
ST32	Minnesota, USA	3	Broiler	+	+	CC4#	CC4
ST33	Pennsylvania, USA	1	Broiler	+	-	CC4	CC4
ST34	Georgia, USA	1	Broiler	+	-	-	-
ST38	Pennsylvania, USA	1	Broiler	+	+	-	-
ST39	Minnesota, USA	3	Broiler	+	+	CC4	CC4
	Georgia, USA	1	Broiler	+	+		
	Ontario, Canada	1	Broiler	+	+		
ST45	Sharkia, Egypt	4	Layer	+	-	-	-
ST46	Sharkia, Egypt	6	Layer	+	-	-	CC1
ST47	Sharkia, Egypt	1	Layer	-	-	-	-
ST48	Sharkia, Egypt	1	Layer	+	-	-	CC1

*ST, sequence type description based on the study of (Hibberd et al., 2011). Note that ST45, ST46, ST47 and ST48 were newly described for the *C. perfringens* strains isolated in this study
**SLV, Single locus variant;
***DLV, double locus variant;
#CC, clonal complex

Summary

A total of 54 birds from 27 farms suspected for NE disease as well as 50 healthy birds (10 ducks, 40 chickens) from eight slaughterhouses were sampled. *C. perfringens* was isolated from birds suspected for NE in 14 different farms (n = 51 isolates) as well as from apparently healthy poultry at slaughterhouses (n = 83 isolates). The *C. perfringens* isolates from suspected NE cases in Egypt were of toxin type A and *netB* negative, despite the fact that NetB toxin was reported to play an important role in NE pathogenesis. The beta2 toxin gene was detected in both diseased and non-diseased birds. Ten isolates from five healthy ducks that did not produce the typical dual hemolysis on blood agar were identified. In addition, seven isolates from two cases suspected for NE were found to have an insertion of an 834bp DNA segment identical to group II introns within the amplicon of the alpha toxin gene. Such strains have been previously reported in Japan, Italy and Denmark. Interestingly, the insertion has been detected so far within chicken isolates only. A pilot investigation based on MLST was performed to determine the genetic relatedness of *C. perfringens* isolates in individual diseased birds in Egypt. The isolates of two birds (bird no. 1 and no. 2) belong to a single MLST sequence type each (ST45 and ST46) which is in accordance with various reports that described limited diversity of strains in NE affected birds. However, the isolates investigated in bird no. 3 belonged to three different STs which may indicate a possibility of multiple-strain infection or the inability of the strains to produce bacteriocins. Comparing MLST data for the twelve strains investigated in this study and MLST data described in prior investigations, the twelve strains were assigned new STs and do not belong to the previously described “NE-associated” genotypes (ST31, CC-4).

4.3 A Development and application of a core genome-based multilocus sequence typing system for *C. perfringens*

4.3.1 Development of a core genome MLST scheme for *C. perfringens*

Whole genome sequencing (WGS) can provide a complete overview for an organism's genetic information but also represents a powerful molecular epidemiological tool for pathogen subtyping and outbreak investigations (Quainoo et al., 2017). In this study, a cgMLST scheme for *C. perfringens* typing was developed based on the allelic profiling of the entire core genes (2.3.4, 3.3.1). To provide a stable set of cgMLST genes, a strict core genome of 1510 genes was initially calculated in 38 *C. perfringens* genomes (2.3.4.2) that represent the different phylogenetic groups of *C. perfringens*. To evaluate this gene set and add more robustness to the scheme, cgMLST typing was performed on an initial data set of 80 genomes that were highly divergent in terms of ecology, geography and isolation time (2.3.4.3, additional data 2). As a result, further 60 genes that were not assigned to allele numbers (nontypeable) in more than 5% of the investigated genomes were discarded (3.3.1, Figure 14). Thus, the final set of the refined cgMLST scheme comprises 1,450 typeable genes in a collection of 160 genomes (99.5%, range 95.3% to 100% per genome) (3.3.1, Figure 14). A recent study by Mehdizadeh Gohari et al. (2017) described a cgMLST scheme for *C. perfringens* typing with a primary focus on the NetF positive strains. This scheme, which included 1,349 genes was applied to a total of 47 *C. perfringens* genomes and has not been made publicly available yet. In comparison, the cgMLST scheme described in this study includes more genes (1,450 genes) and was evaluated (3.3.1) and applied (3.3.2) on a larger number of genomes (Figure 14). The scheme will be made publically available and should facilitate interlaboratory comparison of MLST typing results.

4.3.2 Comparison of cgMLST to whole genome SNP typing and classical MLSTs

Typing results based on the cgMLST scheme were compared to the results of classical MLSTs methods (3.3.7) and whole genome SNP typing (3.3.6) (Table 15). The developed cgMLST scheme has a greater discriminatory power than the classical MLSTs methods (3.3.7). The discriminatory power was indeed comparable to the results obtained by whole genome SNP typing (3.3.6). However, in comparison to the SNP typing methods, cgMLST data are easily portable and accessible using online databases such as (http://www.cgmlst.org/) and PubMLST (https://pubmlst.org/). Therefore, the cgMLST system enables the establishment of a consistent nomenclature between different laboratories (Maiden et al., 2013) which is an essential prerequirement for global comparison and rapid outbreak communications (Jolley and Maiden, 2010). Standardization may represent a difficulty with whole genome SNP typing methods due to different algorithms generally used to detect SNPs (Pearce et al., 2018;

Pightling et al., 2014). SNP calling is usually performed via mapping raw reads to a reference genome, alignment of pre-assembled genomes or *k-mer*-based SNP detection (Gardner et al., 2015; Olson et al., 2015). Due to differences between alleles and SNP-calling methods, it was proposed that a combined analysis of WGS-based MLST and SNP typing could provide the maximum discrimination for accurate outbreak investigations (Chen et al., 2017).

Additionally, the cgMLST typing also considers gene variation due to insertion or deletion (indel). The inclusion of indel variations in the SNP typing methods is variable and frequently ignored (Lüth et al., 2018; Pearce et al., 2018; Quainoo et al., 2017). 310 (21%) out of 1,450 of the cgMLST genes were affected by indel mutations (3.3.2, additional data 19). Interestingly, the translational frame of 7% (107 genes) of the target genes was not disturbed, resulting in 1 to 4 amino acid-length variation in the resulting protein relative to the reference genes (3.3.2, additional data 19). These results may indicate that indel mutations might be common in the *C. perfringens* genome and by using the cgMLST typing method these indel variations are also considered.

4.3.3 Influence of data quality on the cgMLST

Accuracy in detecting genetic variation in a population sample is fundamental to our understanding of bacterial diversity and evolution (Endrullat et al., 2016). However, true variants may be difficult to distinguish from sequencing, assembly and alignment errors (Olson et al., 2015). The cgMLST method as implemented in the SeqSphere+ software is an assembly based system (Junemann et al., 2013). Therefore, sequencing errors and assembly artefacts could negatively affect the detection of the cgMLST target genes (Olson et al., 2015). Indeed, the more fragmented genomes in the data set exhibited a significant higher number of undetected cgMLST genes (3.3.2). This is most likely attributed to the poor sequencing and/or the assembly status because the cgMLST algorithm will only detect the gene and call the allele if the gene is fully constructed during assembly (Junemann et al., 2013). Moreover, detecting genetic variants directly from the assembly does not involve the mapping of the FASTQ reads to identify spurious variants and genome contaminations (Jolley and Maiden, 2010). These facts delineate the importance of quality assessment of sequencing and assembly data that are used as input for MLST genotyping. Although no experiment was done in this study to define the optimal quality threshold parameters for *C. perfringens* typing, a prior investigation on *Listeria monocytogenes* reported the reproducibility of cgMLST allelic calls using different assemblers when the depth of the sequencing was more than 40x (Moura et al., 2016). In the data set analyzed here, the quality of the assembly (no. of contigs and N50 value) was low if the coverage was less than 30x (3.3.2) e.g. strain T43 (coverage = 19x, N50 value = 11.8Kbp, contigs n = 498) and strain C3 (coverage = 29x, N50 value = 32.6Kbp, contigs n = 203). These

two strains, T43 and C3 lack 53 and 19 cgMLST genes, respectively. However, a high number of missing genes was also observed for few genomes that had a higher coverage e.g. T34 (coverage = 64x, N50 = 245, contigs n = 94) and C7 (coverage = 78x, N50= 359.8, contigs n = 32) with 42 and 28 missing genes, respectively. Therefore, coverage may not be the only parameter that influences the detection of the cgMLST targets in the assembled genomes but additional unknown factors may also influence the results of cgMLST. The number of missing genes should be provided as additional information in cgMLST results.

4.3.4 cgMLST Cluster Type (CT) threshold

Different CT thresholds were proposed for different bacterial species in the recent literature. Studies reported less than ten differing alleles as a threshold for a CT distance between epidemiologically linked isolates e.g., for *Enterococcus faecalis* (CT threshold = 7 (Neumann et al., 2019)), *Listeria monocytogens* (CT threshold = 7 (Moura et al., 2016) or 10 (Ruppitsch et al., 2015)), *Brucella melitensis* (CT threshold = 6 (Janowicz et al., 2018)), *Clostridium difficile* (CT threshold = 6 (Bletz et al., 2018)), *Mycoplasma gallisepticum* (CT threshold = 10 (Ghanem et al., 2018)) and *Mycobacterium tuberculosis* (CT threshold = 5 (Kohl et al., 2018)). In this study, some of the investigated isolates were epidemiologically related by geography and isolation time comprising 50 isolates from diseased and non-diseased poultry in Egypt. For these strains, it was shown that the 10 allelic differences provided an optimal resolution as the isolates from diseased birds clustered with only one of the non-disease isolates (3.3.4.3). A recent cgMLST scheme described for *Enterococcus faecium* (de Been et al., 2015) involved different CT thresholds for epidemiological investigations. They reported a threshold of more than 40 allelic mismatches between unrelated isolates, less than 20 allelic mismatches between closely related isolates (most likely from one outbreak) and between 21 and 40 allelic mismatches for isolates that are possibly related (de Been et al., 2015). Additionally, Moura et al. (2016) identified 150 allelic differences between strains of *Listeria monocytogens* that belong to the same sublineages (Moura et al., 2016). In the study presented here a cluster type (CT) threshold of 60 allelic differences was defined between the genomes (3.3.3, Figure 16) as the frequency of less than 60 pairwise allelic differences was significant (3.3.3, Figure 16). Setting a CT threshold at 60 enabled the clustering of more than half of the investigated genomes (n = 84) into CTs (n = 21) (Figure 17 and 18). Results showed also a good agreement with a previous cgMLST study of *C. perfringens* as described in (3.3.3).

4.3.5 Application of the cgMLST system on poultry strains of *C. perfringens*

The developed cgMLST scheme was applied to analyze a set of 87 genomes of poultry *C. perfringens* with regard to country of isolation, NetB toxin gene carriage and clinical disease (3.3.4). Based on a threshold of 60 allelic differences, 53 of the 87 poultry strains were

clustered into 19 cluster types (Figure 18B and 19). Interestingly, 17 CTs thereof involved only strains that were derived from a single country (n = 46) (3.3.4.1, Figure 19B) and none of the clustered Egyptian strains (n = 37) nor the clustered strains from Finland (n = 5) were found to be grouped with strains from other countries (Figure 19B). However, two CTs (CT06 and CT12) included strains from different countries (Denmark, USA, Canada and Australia) (3.3.4.1, Figure 19B). These strains also carry the *netB* gene (3.3.4.2, Figure 19C), three of them (CP4, Del1 and EHE_NE18) were proven virulent based on animal experiments (Keyburn et al., 2008; Lepp et al., 2010; Li et al., 2017a). In total, the cgMLST grouped 10 of the 15 investigated *netB* positive strains into four different CTs (3.3.4.2, Figure 19C). Four further *netB* positive strains (T11, C41, C125 and C37) were found to be closely related based on the phylogenetic analysis of the cgMLST allelic profiles (3.3.4.2, Figure 20). Taken together, the resolution obtained by the cgMLST enabled the identification of CTs that comprised isolates derived from a single country (3.3.5.2, Figure 19). However, the presence of certain CTs in different countries is intriguing and could be attributed to the poultry commercial system worldwide which is likely maintained by few companies i.e. few sources only (Lacey et al., 2018).

Because of the availability of the detailed epidemiological data of the 50 poultry *C. perfringens* isolates from Egypt, these strains were investigated separately (Table 9 and 14). The Egyptian strains include 23 isolates from suspected NE cases of 14 different farms, 21 isolates from healthy birds at slaughterhouses and 6 *C. perfringens* isolates from retail chicken meat parts (Table 9 and 14). It was found that suspected NE isolates from a single farm belong to the same CT. However, isolates of the same CT could be detected in different farms (3.3.4.3, Table 14 and Figure 19A). Few CTs could be identified for the suspected NE isolates (Table 14 and Figure 19A). This is in contrast to isolates from healthy birds and meat samples. They have a high variety of CTs and most isolates represent singletons. Additionally, suspected NE isolates clustered together with isolates from healthy birds or meat samples. These results are in agreement with a recent genomic study with focus on NE strains (Lacey et al., 2018). These authors reported similar genomic features for NE strains from different geographic regions at various times. This study also showed that the pathogenic clades of NE isolates could not be identified based on the core genome SNP method (Lacey et al., 2018). In summary, these results show that suspected NE strains within a single farm are usually similar and unrelated isolates derived from different suspected NE cases, healthy birds or meat samples could sometimes cluster together.

4.3.5 Whole genome SNPs and cgMLST population structure

This study further aimed to compare whole genome SNP analysis and cgMLST to define the population structure of the *C. perfringens* genomes investigated. For this purpose, phylogeny

dependent and independent approaches were implemented (2.3.9) and applied to 160 *C. perfringens* genomes (2.3.1). The software hierBAPS was used as a phylogeny independent method to define the population structure based on the SNPs of the whole genome (2.3.9, 3.3.5) and grouped 159 of 160 strains into four main lineages (phylogroups) (additional data 23). These results were compared to two SplitsTree NeighbourNet networks which is a phylogeny dependent analysis (2.3.9, 3.3.5). A phylogenetic network based on the allelic distances of the cgMLST profiles (Figure 21A) as well as a network based on the whole genome SNP variations were constructed (2.3.9, 3.3.5, Figure 21B). The four main phylogroups identified based on hierBAPS were superimposed on the phylogenetic network from cgMLST (Figure 21A) and whole genome SNPs (Figure 21B), respectively. The phylogenetic network based on the whole genome SNP differences (Figure 21B) showed the delineation of the four different phylogroups more accurately when compared to the phylogenetic network that was constructed based on the cgMLST allelic profiles (Figure 21A). This is possibly due to the cgMLST collapsing the SNP variations into alleles (de Been et al., 2015; Ghanem et al., 2018; Lüth et al., 2018). In cgMLST, sequence differences are calculated in the strain coding regions only and the variations in each of the genes are collapsed and described as a single allelic variant regardless of how many SNPs are within the gene (Lüth et al., 2018; Maiden et al., 2013). This is in contrast to the SNP typing methods which take multiple variations in a single gene into account and consider the coding and non-coding regions of the genomes when applicable (Lüth et al., 2018; Pearce et al., 2018; Quainoo et al., 2017). This “variation collapse” in the cgMLST reduces the discriminatory power of the cgMLST compared to the whole genome based SNP typing methods. However, for the data set investigated herein (2.3.1) both cgMLST and whole genome SNP typing methods provided maximum discrimination based on the Simpson’s diversity index (Hunter and Gaston, 1988) (3.3.7). On the other hand, when the mismatches between the genomes are very high, the differences between the cgMLST allelic profiles were not found to reflect the actual SNP differences between the genomes (Figure 21C). This was observed when the SNP variation between the genomes exceeded 20,000 SNPs (Figure 21C). Roughly, above this SNP value the corresponding cgMLST allelic distances were no longer directly correlated to the SNP distance between genomes (Figure 21C). Based on this study, 40,571 SNPs represent the minimum SNP difference between the different phylogroups (Figure 21D). Taken together, these results may show a better applicability of the whole genome SNP typing in comparison to the cgMLST to delineate the four major phylogroups of *C. perfringens*.

Summary

This study describes a cgMLST scheme for *C. perfringens* typing which is based on the allelic profiling of 1,450 core genes. The cgMLST scheme was applied to a data set of 160 genomes

and an average of 99.5% (range 95.3% to 100%) of the 1,450 cgMLST targets was found typeable (assigned allele number) per each genome. Additionally, the developed cgMLST scheme has a greater discriminatory power than the classical MLSTs methods. In addition, a whole genome based SNP typing was performed. The discriminatory power between the cgMLST and the whole genome SNP typing was comparable.

The developed cgMLST scheme was applied using a cluster type (CT) threshold of 60 allelic differences to analyze 87 poultry strains genomes of *C. perfringens*. Based on cgMLST results, most CTs comprised isolates derived from a single country only. However, few CTs harbour strains which were isolated in different countries. This could be due to the poultry commercial system which is maintained by few companies i.e. few sources worldwide (Lacey et al., 2018). Interestingly, these strains carry also the NetB toxin gene, the proposed main virulence factor for necrotic enteritis in poultry (Keyburn et al., 2010a).

Further 50 isolates which were epidemiologically related and isolated from diseased and non-diseased poultry in Egypt were investigated. Limited diversity was detected among the suspected necrotic enteritis (NE) isolates compared to the isolates that were derived from healthy birds and meat samples supporting the hypothesis that distinct isolates cause NE (Lacey et al., 2018). Isolates from diseased birds were found to group with isolates from healthy birds or meat samples highlighting the wide distribution of potentially virulent strains and the multifactorial character of the disease. However, *netB* positive isolates were not found, thus raising questions about the assumed exclusive role of NetB in the pathogenesis of NE in poultry.

In this study, the 160 investigated genomes were divided into four main phylogroups by hierBAPS to define the population structure. Whole genome SNP typing showed a superior applicability to delineate these phylogroups. A minimum SNP difference of ~ 40,000 SNPs was observed between the phylogroups.

In conclusion, a cgMLST scheme was developed for *C. perfringens*. The cgMLST is useful and applicable for epidemiological investigations while the whole genome SNP typing may better delineate the affiliation of single isolates to the main phylogroups of *C. perfringens*.

Summary

In silico* genome analysis and molecular typing of *Clostridium perfringens

Main objectives of the thesis:

1. *In silico* investigation of the genomic variability, phylogenetic relatedness and virulence assessment of *C. perfringens* by means of comparative genome analysis employing publically available genomic data.
2. Isolation, characterization and genotyping of *C. perfringens* strains from healthy and diseased poultry collected from different farms and slaughterhouses in Egypt.
3. Development and application of a core genome-based multilocus sequence typing system for *C. perfringens.*

1. *In silico* analysis of the genomic variability, phylogenetic relatedness and virulence genes of *C. perfringens*

The Gram-positive anaerobic spore forming bacterium *C. perfringens* is able to produce a large number of toxins by which it can cause various defined diseases in different hosts. With the aim to investigate its genomic diversity the publically available genome sequence data of 76 *C. perfringens* strains from diverse ecological, geographical and temporal niches were analyzed. Data analysis included 30 complete genomes which were composed of a circular chromosome (2.9 to 3.5Mbp) and up to six extrachromosomal elements. A substantial degree of genomic variability was detected in respect to episome content, chromosome size and mobile elements. Insertion sequences were identified and revealed abundance of their occurrence in certain genomes. Comparative alignment of complete genomes displayed a considerable degree of conservation in the order of genes in each chromosome except for three (out of the 30) genomes. The analyzed 76 *C. perfringens* strains were divided into three different phylogroups (I - III). Phylogroup I consisted of human food poisoning strains with chromosomal *cpe* as well as a Darmbrand (enteritis necroticans) strain. This phylogroup is characterized by a significant enrichment in mobile elements, relative small genome size and marked loss of chromosomal genes. Phylogroup I strains lack also two putative iron uptake systems as well as the *pfoA* gene. The genomic features of this phylogroup I (abundance of IS elements and genome reduction) provide indications that these strains adapt to a certain habitat causing human foodborne illnesses. Also, the absence of certain virulence genes (iron uptake systems and PFO) indicates the strains' adaptation to less competitive environment (food) for replication. The loss of chromosomal genes in phylogroup I was in contrast to phylogroup II, in which the genome size indicates the addition of new genetic material. Phylogroup II strains carry also an additional putative iron uptake operon and an additional

copy of the *pfoA* gene. Strains of this phylogroup were frequently reported in different animal hosts (equine and canine) in which they can cause enteric lesions. In sum, this study provides new insights into genomic variability and phylogenetic structure of *C. perfringens*. Phylogroup I (chromosomal *cpe* and Darmbrand strains) appears to be exposed to certain evolutionary mechanisms and displays characteristics that indicate speciation of these strains

2. Characterization and typing of C. perfringens isolates from healthy and diseased poultry in Egypt

To investigate the diversity of poultry strains of *C. perfringens* isolated from Egypt, 54 birds from 27 farms suspected for necrotic enteritis (NE) as well as 50 healthy birds (10 ducks, 40 chickens) from eight slaughterhouses were sampled. *C. perfringens* was isolated from birds suspected of NE in 14 different farms (n = 51 isolates) as well as from apparently healthy poultry at slaughterhouses (n = 83 isolates). The *C. perfringens* isolates from suspected NE cases in Egypt were of toxin type A and *netB* negative, despite fact that NetB toxin was reported to play an important role in NE pathogenesis. The beta2 toxin gene was detected in both diseased and non-diseased birds. Ten isolates from five healthy ducks that did not produce the typical dual hemolysis on blood agar were identified. In addition, seven isolates from two cases suspected for NE were detected with an insertion of 834bp DNA segment within the amplicon of alpha toxin gene. The inserted DNA segment was identical to group II intron. *C. perfringens* strains that carry the group II intron were reported previously in Japan, Italy and Denmark. Interestingly, the insertion was detected so far within chicken isolates only.

Additionally, a pilot investigation based on classical MLST was performed to determine the genetic relatedness of *C. perfringens* isolated in individual diseased birds. The investigated isolates of two birds (bird no. 1 and no. 2) belong to a single MLST sequence type each (ST45 and ST46) in concordance with various reports that described limited diversity of strains in NE affected birds. However, the isolates from bird no. 3 belonged to three different STs. Comparing MLST data for the 12 strains of this study and MLST data described in prior investigations revealed that the 12 strains were assigned to new STs and do not belong to the previously described “NE-associated” genotypes (ST31, CC-4).

3. Development and application of a core genome-based multilocus sequence typing system for *C. perfringens*

Whole genome sequencing can provide a complete overview on organism genetic information but also represents a powerful molecular epidemiological tool for pathogen subtyping and outbreak investigations. In this study, a cgMLST scheme of 1,450 genes was developed for *C. perfringens* typing. The developed scheme was applied on a set of 160 genomes. An average

of 99.5% of the cgMLST targets was found typeable per each genome. This scheme has a greater discriminatory power than the classical MLSTs methods. In addition, a whole genome based SNP typing was performed. The discriminatory power between the cgMLST and the whole genome SNP typing was comparable. The developed cgMLST scheme was applied using a cluster type (CT) threshold of 60 allelic differences to analyze 87 genomes of poultry strains of *C. perfringens*. Based on cgMLST results, most CTs comprised isolates derived from a single country only. However, few CTs harbor strains which were isolated in different countries. This could be due to the poultry commercial system which is likely maintained by few companies i.e. few sources worldwide. Compared to isolates from healthy birds and meat samples, a limited diversity was found in the suspected necrotic enteritis (NE) isolates from Egypt, supporting the hypothesis that distinct isolates cause NE. Isolates from diseased birds were found to group with isolates from healthy birds or meat samples highlighting the wide distribution of potentially virulent strains and the multifactorial character of the disease.

Additionally, the 160 genomes investigated in this study were divided into four main phylogroups by hierBAPS. Whole genome SNP typing showed a superior applicability to delineate these phylogroups. A minimum SNP difference of ~ 40,000 SNPs was observed between the phylogroups.

In sum, a useful cgMLST scheme *for C. perfringens* was developed that is applicable for broad and standardized epidemiological investigations. On the other hand, whole genome SNP typing can map the affiliation of individual isolates to the main phylogroups of *C. perfringens* in more detail.

Zusammenfassung

In silico Genomanalyse und molekulare Typisierung von *Clostridium perfringens*

Hauptziele der Arbeit:

1. *In silico*-Untersuchung der genomischen Variabilität, phylogenetischen Verwandtschaft und Virulenzeinschätzung von *C. perfringens* mittels vergleichender Genomanalyse unter Verwendung öffentlich zugänglicher genomischer Daten.
2. Isolierung, Charakterisierung und Genotypisierung von *C. perfringens* Stämmen von gesundem und erkranktem Geflügel, die in verschiedenen landwirtschaftlichen Betrieben und Schlachthöfen in Ägypten gesammelt wurden.
3. Entwicklung und Anwendung eines Core-genom basierten Multilocus-Sequenz-Typisierungssystems für *C. perfringens*.

1. In silico-Untersuchung der genomischen Variabilität, phylogenetischen Verwandtschaft und Virulenzgene von C. *perfringens*

Das Gram-positive, anaerobe, sporenbildende Bakterium *C. perfringens* ist in der Lage, eine große Anzahl von Toxinen zu produzieren, wodurch es bei verschiedenen Wirten unterschiedliche definierte Krankheiten verursachen kann. Mit dem Ziel die genomische Vielfalt von *C. perfringens* zu untersuchen, wurden öffentlich verfügbaren Genomsequenzdaten von 76 Stämmen verschiedener ökologischer, geographischer und zeitlicher Herkunft analysiert. Die Datenanalyse umfasste 30 vollständige Genome, die aus einem ringförmigen Chromosom (2,9 bis 3,5 Mbp) und bis zu sechs extrachromosomalen Elementen bestanden. Ein erhebliches Maß an genomischer Variabilität wurde in Bezug auf Episom, Chromosomengröße und mobile Elemente festgestellt. Insertionssequenzen wurden identifiziert und die Häufung ihres Auftretens in bestimmten Genomen aufgedeckt. Das vergleichende Alignement vollständiger Genome zeigte einen erheblichen Grad an Konservierung in der Anordnung der Gene in jedem Chromosom, mit Ausnahme von drei (von den 30) Genomen. Die analysierten 76 *C. perfringens* Stämme wurden in drei verschiedene Phylogruppen (I - III) unterteilt. Phylogruppe I bestand aus lebensmittelvergiftenden Stämmen mit chromosomalem *cpe* sowie einem Darmbrand (Enteritis necroticans) Stamm. Diese Phylogruppe zeichnet sich durch eine signifikante Anreicherung mobiler Elemente, eine relativ kleine Genomgröße und einen deutlichen Verlust chromosomaler Gene aus. Den Phylogruppe I-Stämmen fehlen zwei mutmaßliche Eisen-Aufnahmesysteme sowie das *pfoA* Gen. Die genomischen Merkmale dieser Phylogruppe I (Fülle von IS-Elementen und Genomreduktion) liefern Hinweise darauf, dass sich diese Stämme an einen bestimmten Lebensraum anpassen, in welchem sie lebensmittelbedingte Krankheiten beim Menschen verursachen. Auch das

Fehlen bestimmter Virulenzgene (Eisenaufnahmesysteme und PFO) deutet auf die Anpassung der Stämme an ein weniger kompetitives Umfeld (Lebensmittel) für die Replikation hin. Der Verlust chromosomaler Gene in der Phylogruppe I stand im Gegensatz zur Phylogruppe II, bei der die Genomgröße auf die Hinzugewinnung neuen Erbmaterials hinweist. Phylogruppe II-Stämme enthalten ein zusätzliches mutmaßliches Operon für Eisenaufnahme und eine zusätzliche Kopie des *pfoA* Gens. Stämme dieser Phylogruppe wurden häufig in verschiedenen Wirten (Pferd und Hund) beobachtet, bei denen sie enterische Läsionen verursachen können. In Summe liefert diese Studie neue Erkenntnisse über die genomische Variabilität und phylogenetische Struktur von *C. perfringens*. Phylogruppe I (chromosomale cpe- und Darmbrand-Stämme) scheint bestimmten evolutionären Mechanismen ausgesetzt und weist Merkmale auf, die auf eine Artbildung dieser Stämme hinweisen.

2. Charakterisierung und Genotypisierung von *C. perfringens* Stämmen aus gesundem und erkranktem Geflügel in Ägypten

Um die Vielfalt der *C. perfringens* Geflügelstämme in Ägypten zu untersuchen, wurden 54 Vögel aus 27 Betrieben mit Verdacht auf nekrotische Enteritis (NE), sowie 50 gesunde Vögel (10 Enten, 40 Hühner) aus acht Schlachthöfen untersucht. *C. perfringens* wurde aus NE-verdächtigen Vögeln von 14 verschiedenen Betrieben isoliert (n = 51 Isolate), sowie aus augenscheinlich gesundem Geflügel in Schlachthöfen (n = 83 Isolate). Die *C. perfringens* Isolate aus NE-Verdachtsfällen in Ägypten gehörten zum Toxin-Typ A und waren *netB* negativ, obwohl NetB-Toxin eine wichtige Rolle bei der NE-Pathogenese spielen soll. Das beta2-Toxin-Gen wurde sowohl bei kranken als auch bei gesunden Vögeln nachgewiesen. Zehn Isolate von fünf gesunden Enten, die auf Blutagar nicht die typische Doppelzonen-Hämolyse zeigten, wurden identifiziert. Darüaber hinaus wurde bei sieben Isolaten aus zwei NE-Verdachtsfällen die Insertion eines DNA-Segments von 834 bp im Amplikon des Alpha-Toxin-Gens nachgewiesen. Das inserierte DNA-Segment war identisch mit einem Gruppe II Intron. *C. perfringens* Stämme, die ein Intron der Gruppe II tragen, wurden zuvor in Japan, Italien und Dänemark gefunden. Interessanterweise wurde die Insertion bisher nur bei Hühnerisolaten nachgewiesen.

Zusätzlich wurde eine Pilotuntersuchung basierend auf klassischer MLST durchgeführt, um die genetische Verwandtschaft von *C. perfringens* bei einzelnen erkrankten Vögeln zu bestimmen. Die untersuchten Isolate von zwei Vögeln (Vogel Nr. 1 und Nr. 2) gehören jeweils zu einem einzigen MLST-Sequenztyp (ST45 und ST46) in Übereinstimmung mit verschiedenen Berichten, die eine begrenzte Vielfalt von Stämmen bei von NE betroffenen Vögeln beschrieben. Allerdings gehörten die Isolate von Vogel Nr. 3 zu drei verschiedenen STs. Der Vergleich der MLST-Daten für die 12 Stämme dieser Studie mit den in früheren

Untersuchungen beschriebenen MLST-Daten ergab, dass die 12 Stämme neuen STs zugeordnet wurden und nicht zu den zuvor beschriebenen "NE-assoziierten" Genotypen (ST31, CC-4) gehören.

3. Entwicklung und Anwendung eines Core-genom basierten Multilocus-Sequenz-Typisierungssystems für *C. perfringens*

Die Ganzgenomsequenzierung kann einen vollständigen Überblick über die genetische Information eines Organismus geben, stellt aber auch ein leistungsfähiges molekularepidemiologisches Werkzeug für die Subtypisierung von Krankheitserregern und die Untersuchung von Ausbrüchen dar. In dieser Studie wurde ein cgMLST-Schema mit 1.450 Genen für die Typisierung von *C. perfringens* entwickelt. Das entwickelte Schema wurde auf ein Set von 160 Genomen angewendet. Durchschnittlich 99,5% der cgMLST-Ziele wurden pro Genom als typisierbar gewertet. Das Schema hat eine größere diskriminatroische Power als die klassischen MLST-Methoden. Darüber hinaus wurde eine Gesamt-Genom basierte SNP-Typisierung durchgeführt. Die Trennschärfe zwischen der cgMLST und der Gesamt-Genom SNP-Typisierung war vergleichbar. Das entwickelte cgMLST-Schema wurde mit einem Clustertyp (CT)-Schwellenwert von 60 allelischen Differenzen angewendet, um 87 Genome von Geflügelstämmen von *C. perfringens* zu analysieren. Basierend auf den cgMLST-Ergebnissen enthielten die meisten Clustertypen (CTs) Isolate, die nur aus einem einzigen Land stammen. Andererseits beherbergen nur wenige CTs Stämme, die in verschiedenen Ländern isoliert wurden. Dies könnte auf das Geflügelhandelssystem zurückzuführen sein, das wahrscheinlich von wenigen Unternehmen unterhalten wird, d.h. von wenigen Quellen weltweit. Im Vergleich zu Isolaten von gesunden Vögeln und aus Fleischproben, wurde bei den vermeintlichen nekrotische Enteritis (NE) Isolaten aus Ägypten eine begrenzte Vielfalt festgestellt, was die Hypothese stützt, dass bestimmte Isolate NE verursachen. Isolate von kranken Vögeln wurden mit Isolaten von gesunden Vögeln oder Fleischproben gruppiert, dies unterstreicht den multifaktoriellen Charakter der Krankheit und die weite Verbreitung potenziell virulenter Stämme.

Zusätzlich konnten die 160 in dieser Studie untersuchten Genome mittels hierBAPS in vier Hauptphylogruppen eingeteilt werden. Die Gesamt-Genom SNP-Typisierung zeigte eine überlegene Eignung zur Abgrenzung dieser Phylogruppen. Zwischen den Phylogruppen wurde eine minimale SNP-Differenz von ~ 40.000 SNPs beobachtet.

In Summe wurde ein gut anwendbares cgMLST-Schema für *C. perfringens* entwickelt, das für breite und standardisierte epidemiologische Untersuchungen verwendbar ist. Andererseits kann die Gesamt-Genom SNP-Typisierung die Zugehörigkeit einzelner Isolate zu den Hauptphylogruppen von *C. perfringens* detaillierter abbilden.

References

Abildgaard, L., Engberg, R.M., Pedersen, K., Schramm, A., Hojberg, O., 2009. Sequence variation in the alpha-toxin encoding plc gene of Clostridium perfringens strains isolated from diseased and healthy chickens. Veterinary Microbiology 136, 293-299.

Abildgaard, L., Sondergaard, T.E., Engberg, R.M., Schramm, A., Højberg, O., 2010. In vitro production of necrotic enteritis toxin B, NetB, by netB-positive and netB-negative Clostridium perfringens originating from healthy and diseased broiler chickens. Veterinary Microbiology 144, 231-235.

Adams, V., Li, J., Wisniewski, J.A., Uzal, F.A., Moore, R.J., McClane, B.A., Rood, J.I., 2014. Virulence Plasmids of Spore-Forming Bacteria. Microbiology Spectrum 2.

Adams, V., Watts, T.D., Bulach, D.M., Lyras, D., Rood, J.I., 2015. Plasmid partitioning systems of conjugative plasmids from Clostridium perfringens. Plasmid.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389-3402.

Alves, G.G., Machado de Avila, R.A., Chavez-Olortegui, C.D., Lobato, F.C., 2014. Clostridium perfringens epsilon toxin: the third most potent bacterial toxin known. Anaerobe 30, 102-107.

Antonissen, G., Eeckhaut, V., Van Driessche, K., Onrust, L., Haesebrouck, F., Ducatelle, R., Moore, R.J., Van Immerseel, F., 2016. Microbial shifts associated with necrotic enteritis. Avian Pathology 45, 308-312.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic acids research 44, W16-21.

Awad, M.M., Bryant, A.E., Stevens, D.L., Rood, J.I., 1995. 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. Molecular microbiology 15, 191-202.

Awad, M.M., Cheung, J.K., Tan, J.E., McEwan, A.G., Lyras, D., Rood, J.I., 2016. Functional analysis of an feoB mutant in Clostridium perfringens strain 13. Anaerobe 41, 10-17.

Awad, M.M., Ellemor, D.M., Boyd, R.L., Emmins, J.J., Rood, J.I., 2001. Synergistic Effects of Alpha-Toxin and Perfringolysin O in Clostridium perfringens-Mediated Gas Gangrene. Infection and Immunity 69, 7904-7910.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9.

Bailey, M.A., Macklin, K.S., Krehling, J.T., 2015. Low Prevalence of netB and tpeL in Historical Clostridium perfringens Isolates from Broiler Farms in Alabama. Avian diseases 59, 46-51.

Baker, A.A., Davis, E., Rehberger, T., Rosener, D., 2010. Prevalence and Diversity of Toxigenic Clostridium perfringens and Clostridium difficile among Swine Herds in the Midwest. Applied and Environmental Microbiology 76, 2961-2967.

Bangoura, B., Alnassan, A.A., Lendner, M., Shehata, A.A., Kruger, M., Daugschies, A., 2014. Efficacy of an anticoccidial live vaccine in prevention of necrotic enteritis in chickens. Experimental parasitology 145, 125-134.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19, 455-477.

Bannam, T.L., Teng, W.L., Bulach, D., Lyras, D., Rood, J.I., 2006. Functional identification of conjugation and replication regions of the tetracycline resistance plasmid pCW3 from Clostridium perfringens. J Bacteriol 188, 4942-4951.

Bannam, T.L., Yan, X.X., Harrison, P.F., Seemann, T., Keyburn, A.L., Stubenrauch, C., Weeramantri, L.H., Cheung, J.K., McClane, B.A., Boyce, J.D., Moore, R.J., Rood, J.I., 2011. Necrotic enteritis-derived Clostridium perfringens strain with three closely related independently conjugative toxin and antibiotic resistance plasmids. mBio 2.

Barbara, A.J., Trinh, H.T., Glock, R.D., Glenn Songer, J., 2008. Necrotic enteritis-producing strains of Clostridium perfringens displace non-necrotic enteritis strains from the gut of chicks. Vet Microbiol 126, 377-382.

Baums, C.G., Schotte, U., Amtsberg, G., Goethe, R., 2004. Diagnostic multiplex PCR for toxin genotyping of Clostridium perfringens isolates. Vet Microbiol 100, 11-16.

Bertelli, C., Laird, M.R., Williams, K.P., Lau, B.Y., Hoad, G., Winsor, G.L., Brinkman, F.S.L., 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic acids research 45, W30-W35.

Berube, B.J., Bubeck Wardenburg, J., 2013. Staphylococcus aureus α-toxin: nearly a century of intrigue. Toxins 5, 1140-1166.

Billington, S.J., Wieckowski, E.U., Sarker, M.R., Bueschel, D., Songer, J.G., McClane, B.A., 1998. Clostridium perfringens Type E Animal Enteritis Isolates with Highly Conserved, Silent Enterotoxin Gene Sequences. Infection and Immunity 66, 4531-4536.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of Escherichia coli K-12. Science (New York, N.Y.) 277, 1453-1462.

Bletz, S., Janezic, S., Harmsen, D., Rupnik, M., Mellmann, A., 2018. Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Genome-Wide Typing of Clostridium difficile. J Clin Microbiol 56.

Bos, J., Smithee, L., McClane, B., Distefano, R.F., Uzal, F., Songer, J.G., Mallonee, S., Crutcher, J.M., 2005. Fatal Necrotizing Colitis Following a Foodborne Outbreak of Enterotoxigenic Clostridium perfringens Type A Infection. Clinical Infectious Diseases 40, e78-e83.

Bosworth, T.J., 1943. On a New Type of Toxin Produced by Clostridium Welchii. Journal of Comparative Pathology and Therapeutics 53, 245-255.

Brady, J., Hernandez-Doria, J.D., Bennett, C., Guenter, W., House, J.D., Rodriguez-Lecompte, J.C., 2010. Toxinotyping of necrotic enteritis-producing and commensal isolates of Clostridium perfringens from chickens fed organic diets. Avian pathology : journal of the W.V.P.A 39, 475-481.

Briggs, D.C., Naylor, C.E., Smedley, J.G., Lukoyanova, N., Robertson, S., Moss, D.S., McClane, B.A., Basak, A.K., 2011. Structure of the Food-Poisoning Clostridium perfringens Enterotoxin Reveals Similarity to the Aerolysin-Like Pore-Forming Toxins. Journal of Molecular Biology 413, 138-149.

Bruen, T.C., Philippe, H., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172, 2665-2681.

Bryant, A.E., Aldape, M.J., Stevens, D.L. 2015. Chapter 49 - Clostridium perfringens and Other Life-Threatening Clostridial Soft Tissue Infections A2 - Tang, Yi-Wei, In: Sussman, M., Liu, D., Poxton, I., Schwartzman, J. (Eds.) Molecular Medical Microbiology (Second Edition). Academic Press, Boston, 899-907.

Bryant, D., Moulton, V., 2004. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Molecular Biology and Evolution 21, 255-265.

Bullifent, H.L., Moir, A., Awad, M.M., Scott, P.T., Rood, J.I., Titball, R.W., 1996. The Level of Expression of α-toxin by Different Strains ofClostridium perfringensis Dependent on Differences in Promoter Structure and Genetic Background. Anaerobe 2, 365-371.

Caimano, M.J., Hardy, G.G., Yother, J., 1998. Capsule genetics in Streptococcus pneumoniae and a possible role for transposition in the generation of the type 3 locus. Microbial drug resistance (Larchmont, N.Y.) 4, 11-23.

Calefi, A.S., Honda, B.T., Costola-de-Souza, C., de Siqueira, A., Namazu, L.B., Quinteiro-Filho, W.M., Fonseca, J.G., Aloia, T.P., Piantino-Ferreira, A.J., Palermo-Neto, J., 2014. Effects of long-term heat stress in an experimental model of avian necrotic enteritis. Poult Sci 93, 1344-1353.

Carrico, J.A., Silva-Costa, C., Melo-Cristino, J., Pinto, F.R., de Lencastre, H., Almeida, J.S., Ramirez, M., 2006. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant Streptococcus pyogenes. J Clin Microbiol 44, 2524-2532.

Chalmers, G., Bruce, H.L., Hunter, D.B., Parreira, V.R., Kulkarni, R.R., Jiang, Y.-F., Prescott, J.F., Boerlin, P., 2008a. Multilocus Sequence Typing Analysis of Clostridium perfringens Isolates from Necrotic Enteritis Outbreaks in Broiler Chicken Populations. J. Clin. Microbiol. 46, 3957-3964.

Chalmers, G., Martin, S.W., Hunter, D.B., Prescott, J.F., Weber, L.J., Boerlin, P., 2008b. Genetic diversity of Clostridium perfringens isolated from healthy broiler chickens at a commercial farm. Vet Microbiol 127, 116-127.

Charlebois, A., Jacques, M., Archambault, M., 2016. Comparative transcriptomic analysis of Clostridium perfringens biofilms and planktonic cells. Avian Pathology 45, 593-601.

Chen, L., Zheng, D., Liu, B., Yang, J., Jin, Q., 2016. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. Nucleic acids research 44, D694-D697.

Chen, Y., Luo, Y., Carleton, H., Timme, R., Melka, D., Muruvanda, T., Wang, C., Kastanis, G., Katz, L.S., Turner, L., Fritzinger, A., Moore, T., Stones, R., Blankenship, J., Salter, M., Parish, M., Hammack, T.S., Evans, P.S., Tarr, C.L., Allard, M.W., Strain, E.A., Brown, E.W., 2017. Whole genome and core genome multilocus sequence typing and single nucleotide polymorphism analyses of Listeria monocytogenes associated with an outbreak linked to cheese, United States, 2013. Appl Environ Microbiol.

Cheng, L., Connor, T.R., Siren, J., Aanensen, D.M., Corander, J., 2013. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol Biol Evol 30, 1224-1228.

Cheung, J.K., Keyburn, A.L., Carter, G.P., Lanckriet, A.L., Van Immerseel, F., Moore, R.J., Rood, J.I., 2010. The VirSR two-component signal transduction system regulates NetB toxin production in Clostridium perfringens. Infect Immun 78, 3064-3072.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Meth 10, 563-569.

Choo, J.M., Cheung, J.K., Wisniewski, J.A., Steer, D.L., Bulach, D.M., Hiscox, T.J., Chakravorty, A., Smith, A.I., Gell, D.A., Rood, J.I., Awad, M.M., 2016. The NEAT Domain-Containing Proteins of Clostridium perfringens Bind Heme. PLoS One 11, e0162981.

Chukwu, E.E., Nwaokorie, F.O., Coker, A.O., Avila-Campos, M.J., Ogunsola, F.T., 2017. Genetic variation among Clostridium perfringens isolated from food and faecal specimens in Lagos. Microb Pathog 111, 232-237.

Cole, A.R., Gibert, M., Popoff, M., Moss, D.S., Titball, R.W., Basak, A.K., 2004. Clostridium perfringens ε-toxin shows structural similarity to the pore-forming toxin aerolysin. Nature Structural &Amp; Molecular Biology 11, 797.

Collier, C.T., Hofacre, C.L., Payne, A.M., Anderson, D.B., Kaiser, P., Mackie, R.I., Gaskins, H.R., 2008. Coccidia-induced mucogenesis promotes the onset of necrotic enteritis by supporting Clostridium perfringens growth. Veterinary immunology and immunopathology 122, 104-115.

Contreras-Moreira, B., Vinuesa, P., 2013. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. Applied and Environmental Microbiology 79, 7696-7701.

Cooper, K.K., Songer, J.G., 2009. Necrotic enteritis in chickens: A paradigm of enteric infection by Clostridium perfringens type A. Anaerobe 15, 55-60.

Cooper, K.K., Songer, J.G., 2010. Virulence of Clostridium perfringens in an experimental model of poultry necrotic enteritis. Vet Microbiol 142, 323-328.

Cooper, K.K., Songer, J.G., Uzal, F.A., 2013. Diagnosing clostridial enteric disease in poultry. Journal of Veterinary Diagnostic Investigation 25, 314-327.

Cooper, K.K., Trinh, H.T., Songer, J.G., 2009. Immunization with recombinant alpha toxin partially protects broiler chicks against experimental challenge with Clostridium perfringens. Veterinary Microbiology 133, 92-97.

Corander, J., Marttinen, P., Sirén, J., Tang, J.J.B.B., 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. 9, 539.

Corpet, D.E., 2000. Mechanism of antimicrobial growth promoters used in animal feed. Revue de Médecine Vétérinaire 151, 99-104.

Coursodon, C.F., Glock, R.D., Moore, K.L., Cooper, K.K., Songer, J.G., 2012. TpeL-producing strains of Clostridium perfringens type A are highly virulent for broiler chicks. Anaerobe 18, 117-121.

Coursodon, C.F., Trinh, H.T., Mallozzi, M., Vedantam, G., Glock, R.D., Songer, J.G., 2010. Clostridium perfringens alpha toxin is produced in the intestines of broiler chicks inoculated with an alpha toxin mutant. Anaerobe 16, 614-617.

Crespo, R., Fisher, D.J., Shivaprasad, H.L., Fernandez-Miyakawa, M.E., Uzal, F.A., 2007. Toxinotypes of Clostridium perfringens isolated from sick and healthy avian species. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 19, 329-333.

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., Harris, S.R., 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research 43, e15-e15.

Dahiya, J.P., Wilkie, D.C., Van Kessel, A.G., Drew, M.D., 2006. Potential strategies for controlling necrotic enteritis in broiler chickens in post-antibiotic era. Animal Feed Science and Technology 129, 60-88.

Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLOS ONE 5, e11147.

Darling, A.E., Miklós, I., Ragan, M.A., 2008. Dynamics of Genome Rearrangement in Bacterial Populations. PLOS Genetics 4, e1000128.

Darmon, E., Leach, D.R., 2014. Bacterial genome instability. Microbiology and molecular biology reviews : MMBR 78, 1-39.

de Been, M., Pinholt, M., Top, J., Bletz, S., Mellmann, A., van Schaik, W., Brouwer, E., Rogers, M., Kraat, Y., Bonten, M., Corander, J., Westh, H., Harmsen, D., Willems, R.J., 2015. Core Genome Multilocus Sequence Typing Scheme for High- Resolution Typing of Enterococcus faecium. J Clin Microbiol 53, 3788-3797.

Deguchi, A., Miyamoto, K., Kuwahara, T., Miki, Y., Kaneko, I., Li, J., McClane, B.A., Akimoto, S., 2009. Genetic characterization of type A enterotoxigenic Clostridium perfringens strains. PLoS One 4, e5598.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science 14, 927-930.

Drigo, I., Agnoletti, F., Bacchin, C., Bettini, F., Cocchi, M., Ferro, T., Marcon, B., Bano, L., 2010. Toxin genotyping of Clostridium perfringens field strains isolated from healthy and diseased chickens. 2010 7, 4.

Edwards, D.J., Pope, B.J., Holt, K.E., 2015. RedDog: comparative analysis pipeline for large numbers of bacterial isolates using high-throughput sequences. In prep. (working title only).

Eisen, J.A., Heidelberg, J.F., White, O., Salzberg, S.L., 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biology 1, research0011.0011-0011.0019.

el Idrissi, A.H., Ward, G.E., 1992. Evaluation of enzyme-linked immunosorbent assay for diagnosis of Clostridium perfringens enterotoxemias. Vet Microbiol 31, 389-396.

Endrullat, C., Glökler, J., Franke, P., Frohme, M., 2016. Standardization and quality management in next-generation sequencing. Applied & Translational Genomics 10, 2-9.

Esnault, E., Valens, M., Espéli, O., Boccard, F., 2007. Chromosome Structuring Limits Genome Plasticity in Escherichia coli. PLOS Genetics 3, e226.

Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., Spratt, B.G., 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186, 1518-1530.

Fernandez-Miyakawa, M.E., Fisher, D.J., Poon, R., Sayeed, S., Adams, V., Rood, J.I., McClane, B.A., Uzal, F.A., 2007. Both Epsilon-Toxin and Beta-Toxin Are Important for the Lethal Properties of Clostridium perfringens Type B Isolates in the Mouse Intravenous Injection Model. Infection and Immunity 75, 1443-1452.

Fernando, P.S., Rose, S.P., Mackenzie, A.M., Silva, S.S., 2011. Effect of diets containing potato protein or soya bean meal on the incidence of spontaneously-occurring subclinical necrotic enteritis and the physiological response in broiler chickens. British poultry science 52, 106-114.

Filho, E.J., Carvalho, A.U., Assis, R.A., Lobato, F.F., Rachid, M.A., Carvalho, A.A., Ferreira, P.M., Nascimento, R.A., Fernandes, A.A., Vidal, J.E., Uzal, F.A., 2009. Clinicopathologic features of experimental Clostridium perfringens type D enterotoxemia in cattle. Veterinary pathology 46, 1213-1220.

Fisher, D.J., Fernandez-Miyakawa, M.E., Sayeed, S., Poon, R., Adams, V., Rood, J.I., Uzal, F.A., McClane, B.A., 2006. Dissecting the contributions of Clostridium perfringens type C toxins to lethality in the mouse intravenous injection model. Infect Immun 74, 5200-5210.

Flores-Díaz, M., Monturiol-Gross, L., Alape-Girón, A. 2015. 22 - Membrane-damaging and cytotoxic sphingomyelinases and phospholipases A2 - Alouf, Joseph, In: Ladant, D., Popoff, M.R. (Eds.) The Comprehensive Sourcebook of Bacterial Protein Toxins (Fourth Edition). Academic Press, Boston, 627-676.

Fohler, S., Klein, G., Hoedemaker, M., Scheu, T., Seyboldt, C., Campe, A., Jensen, K.C., Abdulmawjood, A., 2016. Diversity of Clostridium perfringens toxin-genotypes from dairy farms. BMC microbiology 16, 199.

França, M., Barrios, M.A., Stabler, L., Zavala, G., Shivaprasad, H.L., Lee, M.D., Villegas, A.M., Uzal, F.A., 2015. Association of Beta2-Positive Clostridium perfringens Type A With Focal Duodenal Necrosis in Egg-Laying Chickens in the United States. Avian diseases 60, 43-49.

Freedman, J.C., Shrestha, A., McClane, B.A., 2016. Clostridium perfringens Enterotoxin: Action, Genetics, and Translational Applications. Toxins (Basel) 8.

Freedman, J.C., Theoret, J.R., Wisniewski, J.A., Uzal, F.A., Rood, J.I., McClane, B.A., 2015. Clostridium perfringens type A–E toxin plasmids. Research in Microbiology 166, 264-279.

Gad, W., Hauck, R., Krüger, M., Hafez, H.M., 2011. Prevalence of Clostridium perfringens in commercial turkey and layer flocks. European Poultry Science (EPS) 75, 6.

Garcia, J.P., Adams, V., Beingesser, J., Hughes, M.L., Poon, R., Lyras, D., Hill, A., McClane, B.A., Rood, J.I., Uzal, F.A., 2013. Epsilon toxin is essential for the virulence of Clostridium perfringens type D infection in sheep, goats, and mice. Infect Immun 81, 2405-2414.

Garcia, J.P., Beingesser, J., Fisher, D.J., Sayeed, S., McClane, B.A., Posthaus, H., Uzal, F.A., 2012. The effect of Clostridium perfringens type C strain CN3685 and its isogenic beta toxin null mutant in goats. Vet Microbiol 157, 412-419.

Gardner, S.N., Slezak, T., Hall, B.G., 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31, 2877-2878.

Gervasi, T., Curto, R.L., Narbad, A., Mayer, M.J., 2013. Complete genome sequence of PhiCP51, a temperate bacteriophage of Clostridium perfringens. Archives of virology 158, 2015-2017.

Ghanem, M., Wang, L., Zhang, Y., Edwards, S., Lu, A., Ley, D., El-Gazzar, M., 2018. Core Genome Multilocus Sequence Typing: a Standardized Approach for Molecular Typing of Mycoplasma gallisepticum. 56, e01145-01117.

Gholamiandekhordi, A.R., Ducatelle, R., Heyndrickx, M., Haesebrouck, F., Van Immerseel, F., 2006. Molecular and phenotypical characterization of Clostridium perfringens isolates from poultry flocks with different disease status. Veterinary Microbiology 113, 143-152.

Gibert, M., Jolivet-Renaud, C., Popoff, M.R., 1997. Beta2 toxin, a novel toxin produced by Clostridium perfringens. Gene 203, 65-73.

Goossens, E., Valgaeren, B.R., Pardon, B., Haesebrouck, F., Ducatelle, R., Deprez, P.R., Van Immerseel, F., 2017. Rethinking the role of alpha toxin in Clostridium perfringens-associated enteric diseases: a review on bovine necro-haemorrhagic enteritis. Veterinary research 48, 9.

Goossens, E., Verherstraeten, S., Valgaeren, B.R., Pardon, B., Timbermont, L., Schauvliege, S., Rodrigo-Mocholí, D., Haesebrouck, F., Ducatelle, R., Deprez, P.R., Van Immerseel, F., 2016. The C-terminal domain of Clostridium perfringens alpha toxin as a vaccine candidate against bovine necrohemorrhagic enteritis. Veterinary research 47, 52-52.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research 36, 3420-3435.

Gross, T.P., Kamara, L.B., Hatheway, C.L., Powers, P., Libonati, J.P., Harmon, S.M., Israel, E., 1989. Clostridium perfringens food poisoning: use of serotyping in an outbreak setting. J. Clin. Microbiol. 27, 660-663.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307-321.

Gupta, R.S., Gao, B., 2009. Phylogenomic analyses of clostridia and identification of novel protein signatures that are specific to the genus Clostridiumsensu stricto (cluster I). International Journal of Systematic and Evolutionary Microbiology 59, 285-294.

Gurjar, A., Li, J., McClane, B.A., 2010. Characterization of Toxin Plasmids in Clostridium perfringens Type C Isolates. Infection and Immunity 78, 4860-4869.

Hassan, K.A., Elbourne, L.D., Tetu, S.G., Melville, S.B., Rood, J.I., Paulsen, I.T., 2015. Genomic analyses of Clostridium perfringens isolates from five toxinotypes. Res Microbiol 166, 255-263.

Hatheway, C.L., 1990. Toxigenic clostridia. Clinical Microbiology Reviews 3, 66-98.

Hermans, P.G., Morgan, K.L., 2007. Prevalence and associated risk factors of necrotic enteritis on broiler farms in the United Kingdom; a cross-sectional survey. Avian pathology : journal of the W.V.P.A 36, 43-51.

Hibberd, M.C., Neumann, A.P., Rehberger, T.G., Siragusa, G.R., 2011. Multilocus Sequence Typing Subtypes of Poultry Clostridium perfringens Isolates Demonstrate Disease Niche Partitioning. J. Clin. Microbiol. 49, 1556-1567.

Holt, K.E., Wertheim, H., Zadoks, R.N., Baker, S., Whitehouse, C.A., Dance, D., Jenney, A., Connor, T.R., Hsu, L.Y., Severin, J., Brisse, S., Cao, H., Wilksch, J., Gorrie, C., Schultz, M.B., Edwards, D.J., Nguyen, K.V., Nguyen, T.V., Dao, T.T., Mensink, M., Minh, V.L., Nhu, N.T., Schultsz, C., Kuntaman, K., Newton, P.N., Moore, C.E., Strugnell, R.A., Thomson, N.R., 2015. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proceedings of the National Academy of Sciences of the United States of America 112, E3574-3581.

Hong, Y.H., Song, W., Lee, S.H., Lillehoj, H.S., 2012. Differential gene expression profiles of beta-defensins in the crop, intestine, and spleen using a necrotic enteritis model in 2 commercial broiler chicken lines. Poult Sci 91, 1081-1088.

Hughes, J.A., Turnbull, P.C., Stringer, M.F., 1976. A serotyping system for Clostridium welchii (C. perfringens) type A, and studies on the type-specific antigens. Journal of medical microbiology 9, 475-485.

Hunt, M., Silva, N.D., Otto, T.D., Parkhill, J., Keane, J.A., Harris, S.R., 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biology 16, 294.

Hunter, P.R., Gaston, M.A., 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26, 2465-2466.

Hunter, S.E., Brown, J.E., Oyston, P.C., Sakurai, J., Titball, R.W., 1993. 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.

Huson, D.H., Bryant, D., 2005. Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution 23, 254-267.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution 23, 254-267.

Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. Systematic Biology 61, 1061-1067.

Iguchi, A., Iyoda, S., Terajima, J., Watanabe, H., Osawa, R., 2006. Spontaneous recombination between homologous prophage regions causes large-scale inversions within the Escherichia coli O157:H7 chromosome. Gene 372, 199-207.

Jang, S.I., Lillehoj, H.S., Lee, S.H., Lee, K.W., Lillehoj, E.P., Hong, Y.H., An, D.J., Jeoung, D.H., Chun, J.E., 2013. Relative disease susceptibility and clostridial toxin antibody responses in three commercial broiler lines coinfected with Clostridium perfringens and Eimeria maxima using an experimental model of necrotic enteritis. Avian diseases 57, 684-687.

Janowicz, A., De Massis, F., Ancora, M., Camma, C., Patavino, C., Battisti, A., Prior, K., Harmsen, D., Scholz, H., Zilli, K., Sacchini, L., Di Giannatale, E., Garofolo, G., 2018. Core Genome Multilocus Sequence Typing and Single Nucleotide Polymorphism Analysis in the Epidemiology of Brucella melitensis Infections. J Clin Microbiol 56.

Johansson, A., Aspán, A., Kaldhusdal, M., Engström, B.E., 2010. Genetic diversity and prevalence of netB in Clostridium perfringens isolated from a broiler flock affected by mild necrotic enteritis. Veterinary Microbiology 144, 87-92.

Jolley, K.A., Feil, E.J., Chan, M.-S., Maiden, M.C.J., 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17, 1230-1231.

Jolley, K.A., Maiden, M.C., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11, 595.

Joshi, N., Fass, J., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

Jost, B.H., Billington, S.J., Trinh, H.T., Bueschel, D.M., Songer, J.G., 2005. Atypical cpb2 genes, encoding beta2-toxin in Clostridium perfringens isolates of nonporcine origin. Infection and Immunity 73, 652-656.

Jost, B.H., Billington, S.J., Trinh, H.T., Songer, J.G., 2006a. Association of genes encoding beta2 toxin and a collagen binding protein in Clostridium perfringens isolates of porcine origin. Veterinary Microbiology 115, 173-182.

Jost, B.H., Trinh, H.T., Songer, J.G., 2006b. Clonal relationships among Clostridium perfringens of porcine origin as determined by multilocus sequence typing. Vet Microbiol 116, 158-165.

Junemann, S., Sedlazeck, F.J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., Mellmann, A., Goesmann, A., von Haeseler, A., Stoye, J., Harmsen, D., 2013. Updating benchtop sequencing performance comparison. Nature biotechnology 31, 294-296.

Kaldhusdal, M., Benestad, S.L., Løvland, A., 2016. Epidemiologic aspects of necrotic enteritis in broiler chickens – disease occurrence and production performance. Avian Pathology 45, 271-274.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30, 772-780.

Keyburn, A.L., Bannam, T.L., Moore, R.J., Rood, J.I., 2010a. NetB, a Pore-Forming Toxin from Necrotic Enteritis Strains of Clostridium perfringens. Toxins 2, 1913-1927.

Keyburn, A.L., Boyce, J.D., Vaz, P., Bannam, T.L., Ford, M.E., Parker, D., Di Rubbo, A., Rood, J.I., Moore, R.J., 2008. NetB, a new toxin that is associated with avian necrotic enteritis caused by Clostridium perfringens. PLoS Pathog 4.

Keyburn, A.L., Sheedy, S.A., Ford, M.E., Williamson, M.M., Awad, M.M., Rood, J.I., Moore, R.J., 2006. Alpha-Toxin of Clostridium perfringens Is Not an Essential Virulence Factor in Necrotic Enteritis in Chickens. Infection and Immunity 74, 6496-6500.

Keyburn, A.L., Yan, X.X., Bannam, T.L., Van Immerseel, F., Rood, J.I., Moore, R.J., 2010b. Association between avian necrotic enteritis and Clostridium perfringens strains expressing NetB toxin. Veterinary research 41, 21.

Kim, D.K., Lillehoj, H.S., Jang, S.I., Lee, S.H., Hong, Y.H., Cheng, H.H., 2014. Transcriptional Profiles of Host-Pathogen Responses to Necrotic Enteritis and Differential Regulation of Immune Genes in Two Inbreed Chicken Lines Showing Disparate Disease Susceptibility. PLoS ONE 9, e114960.

Kiu, R., Caim, S., Alexander, S., Pachori, P., 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. Frontiers in microbiology 8, 2485.

Kiu, R., Hall, L.J., 2018. An update on the human and animal enteric pathogen Clostridium perfringens. Emerging Microbes & Infections 7, 141.

Knight, D.R., Squire, M.M., Collins, D.A., 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. Frontiers in microbiology 7, 2138.

Kohl, T.A., Harmsen, D., Rothgänger, J., Walker, T., Diel, R., Niemann, S., 2018. Harmonized Genome Wide Typing of Tubercle Bacilli Using a Web-Based Gene-By-Gene Nomenclature System. EBioMedicine 34, 131-138.

Koonin, E.V., Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. Current opinion in microbiology 37, 67-78.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Phillippy, A.M., 2016. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. bioRxiv.

Krumsiek, J., Arnold, R., Rattei, T., 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23, 1026-1028.

Kulkarni, R.R., Parreira, V.R., Jiang, Y.F., Prescott, J.F., 2010. A live oral recombinant Salmonella enterica serovar typhimurium vaccine expressing Clostridium perfringens antigens confers protection against necrotic enteritis in broiler chickens. Clinical and vaccine immunology : CVI 17, 205-214.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33, 1870-1874.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Cordani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denziot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Togoni, A., Tosato, V., Uchiyama, S., Vandebol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H., Danchin, A., 1997. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature 390, 249-256.

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

Labbé, R.G. 2003. CLOSTRIDIUM | Occurrence of Clostridium perfringens A2 - Caballero, Benjamin, In: Encyclopedia of Food Sciences and Nutrition (Second Edition). Academic Press, Oxford, 1398-1401.

Lacey, J.A., Allnutt, T.R., Vezina, B., Van, T.T.H., Stent, T., Han, X., Rood, J.I., Wade, B., Keyburn, A.L., Seemann, T., Chen, H., Haring, V., Johanesen, P.A., Lyras, D., Moore, R.J., 2018. Whole genome analysis reveals the diversity and evolutionary relationships between necrotic enteritis-causing strains of Clostridium perfringens. BMC Genomics 19, 379.

Lacey, J.A., Johanesen, P.A., Lyras, D., Moore, R.J., 2016. Genomic diversity of necrotic enteritis-associated strains of Clostridium perfringens: a review. Avian Pathology 45, 302-307.

Lacey, J.A., Keyburn, A.L., Ford, M.E., Portela, R.W., Johanesen, P.A., Lyras, D., Moore, R.J., 2017. Conjugation-mediated horizontal gene transfer of Clostridium perfringens plasmids in the chicken gastrointestinal tract results in the formation of new virulent strains. Appl Environ Microbiol.

Lahti, P., Lindstrom, M., Somervuo, P., Heikinheimo, A., Korkeala, H., 2012. Comparative genomic hybridization analysis shows different epidemiology of chromosomal and plasmid-borne cpe-carrying Clostridium perfringens type A. PLoS One 7, e46162.

Lanckriet, A., Timbermont, L., De Gussem, M., Marien, M., Vancraeynest, D., Haesebrouck, F., Ducatelle, R., Van Immerseel, F., 2010. The effect of commonly used anticoccidials and antibiotics in a subclinical necrotic enteritis model. Avian pathology : journal of the W.V.P.A 39, 63-68.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357-359.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Pontén, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. 50, 1355-1361.

Lee, K.W., Lillehoj, H.S., Jeong, W., Jeoung, H.Y., An, D.J., 2011. Avian necrotic enteritis: Experimental models, host immunity, pathogenesis, risk factors, and vaccine development. Poultry Science 90, 1381-1390.

Lee, S.H., Lillehoj, H.S., Jang, S.I., Jeong, M., Kim, D.K., Xu, S., Lee, S.K., Kim, J.B., Park, H.J., Kim, H.R., Bravo, D.M., 2014a. Immune and anti-oxidant effects of in ovo selenium proteinate on post-hatch experimental avian necrotic enteritis. Vet Parasitol 206, 115-122.

Lee, S.H., Lillehoj, H.S., Jang, S.I., Jeong, M.S., Xu, S.Z., Kim, J.B., Park, H.J., Kim, H.R., Lillehoj, E.P., Bravo, D.M., 2014b. Effects of in ovo injection with selenium on immune and antioxidant responses during experimental necrotic enteritis in broiler chickens. Poult Sci 93, 1113-1121.

Lepp, D., Gong, J., Songer, J.G., Boerlin, P., Parreira, V.R., Prescott, J.F., 2013. Identification of Accessory Genome Regions in Poultry Clostridium perfringens Isolates Carrying the netB Plasmid. Journal of Bacteriology 195, 1152-1166.

Lepp, D., Roxas, B., Parreira, V.R., Marri, P.R., Rosey, E.L., Gong, J., Songer, J.G., Vedantam, G., Prescott, J.F., 2010. Identification of Novel Pathogenicity Loci in <italic>Clostridium perfringens</italic> Strains That Cause Avian Necrotic Enteritis. PLoS ONE 5, e10795.

Letunic, I., Bork, P., 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Research 39, W475-W478.

Li, C., Lillehoj, H.S., Gadde, U.D., Ritter, D., Oh, S., 2017a. Characterization of Clostridium perfringens Strains Isolated from Healthy and Necrotic Enteritis-Afflicted Broiler Chickens. Avian diseases 61, 178-185.

Li, C., Yan, X., Lillehoj, H.S., 2017b. Complete Genome Sequence of Clostridium perfringens LLY_N11, a Necrotic Enteritis-Inducing Strain Isolated from a Healthy Chicken Intestine. Genome announcements 5.

Li, C., Yan, X., Lillehoj, H.S., 2017c. Complete genome sequences of Clostridium perfringens Del1 strain isolated from chickens affected by necrotic enteritis. Gut Pathog 9, 69.

Li, C., Yan, X., Lillehoj, H.S.J.G.P., 2017d. Complete genome sequences of Clostridium perfringens Del1 strain isolated from chickens affected by necrotic enteritis. 9, 69.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009a. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

Li, J., Adams, V., Bannam, T.L., Miyamoto, K., Garcia, J.P., Uzal, F.A., Rood, J.I., McClane, B.A., 2013. Toxin plasmids of Clostridium perfringens. Microbiology and molecular biology reviews : MMBR 77, 208-233.

Li, J., McClane, B.A., 2006a. Comparative effects of osmotic, sodium nitrite-induced, and pH-induced stress on growth and survival of Clostridium perfringens type A isolates carrying chromosomal or plasmid-borne enterotoxin genes. Appl Environ Microbiol 72, 7620-7625.

Li, J., McClane, B.A., 2006b. Further Comparison of Temperature Effects on Growth and Survival of Clostridium perfringens Type A Isolates Carrying a Chromosomal or Plasmid-Borne Enterotoxin Gene. Applied and Environmental Microbiology 72, 4561-4568.

Li, J., McClane, B.A., 2008. A Novel Small Acid Soluble Protein Variant Is Important for Spore Resistance of Most Clostridium perfringens Food Poisoning Isolates. PLOS Pathogens 4, e1000056.

Li, J., Paredes-Sabja, D., Sarker, M.R., McClane, B.A., 2016. Clostridium perfringens Sporulation and Sporulation-Associated Toxin Production. Microbiology spectrum 4, 10.1128/microbiolspec.TBS-0022-2015.

Li, W., Raoult, D., Fournier, P.E., 2009b. Bacterial strain typing in the genomic era. FEMS microbiology reviews 33, 892-916.

Liang, Y., Hou, X., Wang, Y., Cui, Z., Zhang, Z., Zhu, X., Xia, L., Shen, X., Cai, H., Wang, J., Xu, D., Zhang, E., Zhang, H., Wei, J., He, J., Song, Z., Yu, X.-j., Yu, D., Hai, R., 2010. Genome Rearrangements of Completely Sequenced Strains of Yersinia pestis. J. Clin. Microbiol. 48, 1619-1623.

Lindström, M., Heikinheimo, A., Lahti, P., Korkeala, H., 2011. Novel insights into the epidemiology of Clostridium perfringens type A food poisoning. Food microbiology 28, 192-198.

Liu, H., McCord, K.D., Howarth, J., Popham, D.L., Jensen, R.V., Melville, S.B., 2014. Hypermotility in Clostridium perfringens Strain SM101 Is Due to Spontaneous Mutations in Genes Linked to Cell Division. Journal of Bacteriology 196, 2405-2412.

Llanco, L.A., Nakano, V., Avila-Campos, M.J., 2015. Sialidase production and genetic diversity in Clostridium perfringens type A isolated from chicken with necrotic enteritis in Brazil. Current microbiology 70, 330-337.

Lucey, B.P., Hutchins, G.M., 2004. William H. Welch, MD, and the discovery of Bacillus welchii. Archives of pathology & laboratory medicine 128, 1193-1195.

Lukinmaa, S., Takkunen, E., Siitonen, A., 2002. Molecular Epidemiology of Clostridium perfringens Related to Food-Borne Outbreaks of Disease in Finland from 1984 to 1999. Applied and Environmental Microbiology 68, 3744-3749.

Lüth, S., Kleta, S., Al Dahouk, S.J.T.i.F.S., Technology, 2018. Whole genome sequencing as a typing tool for foodborne pathogens like Listeria monocytogenes–the way towards global harmonisation and data exchange. 73, 67-75.

Lyhs, U., Perko-Makela, P., Kallio, H., Brockmann, A., Heinikainen, S., Tuuri, H., Pedersen, K., 2013. Characterization of Clostridium perfringens isolates from healthy turkeys and from turkeys with necrotic enteritis. Poult Sci 92, 1750-1757.

Ma, M., Gurjar, A., Theoret, J.R., Garcia, J.P., Beingesser, J., Freedman, J.C., Fisher, D.J., McClane, B.A., Uzal, F.A., 2014. Synergistic effects of Clostridium perfringens enterotoxin and beta toxin in rabbit small intestinal loops. Infect Immun 82, 2958-2970.

Ma, M., Li, J., McClane, B.A., 2012. Genotypic and Phenotypic Characterization of Clostridium perfringens Isolates from Darmbrand Cases in Post-World War II Germany. Infection and Immunity 80, 4354-4363.

Ma, M., Ohtani, K., Shimizu, T., Misawa, N., 2007. Detection of a Group II Intron without an Open Reading Frame in the Alpha-Toxin Gene of Clostridium perfringens Isolated from a Broiler Chicken. Journal of Bacteriology 189, 1633-1640.

Mackiewicz, P., Mackiewicz, D., Kowalczuk, M., Cebrat, S., 2001. Flip-flop around the origin and terminus of replication in prokaryotic genomes. Genome Biology 2, interactions1004.1001-interactions1004.1004.

Mahony, D.E., Ahmed, R., Jackson, S.G., 1992. Multiple typing techniques applied to a Clostridium perfringens food poisoning outbreak. J Appl Bacteriol 72, 309-314.

Maiden, M.C., 2006. Multilocus sequence typing of bacteria. Annu Rev Microbiol 60, 561-588.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proceedings of the National Academy of Sciences of the United States of America 95, 3140-3145.

Maiden, M.C., Jansen van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., McCarthy, N.D., 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nature reviews. Microbiology 11, 728-736.

Marks, S.L., Kather, E.J., Kass, P.H., Melli, A.C., 2002. Genotypic and phenotypic characterization of Clostridium perfringens and Clostridium difficile in diarrheic and healthy dogs. Journal of veterinary internal medicine 16, 533-540.

Martin, T.G., Smyth, J.A., 2009. Prevalence of netB among some clinical isolates of Clostridium perfringens from animals in the United States. Veterinary Microbiology 136, 202-205.

Martin, T.G., Smyth, J.A., 2010. The ability of disease and non-disease producing strains of Clostridium perfringens from chickens to adhere to extracellular matrix molecules and Caco-2 cells. Anaerobe 16, 533-539.

McDevitt, R.M., Brooker, J.D., Acamovic, T., Sparks, N.H.C., 2006. Necrotic enteritis; a continuing challenge for the poultry industry. Worlds Poult. Sci. J. 62, 221-247.

McLauchlin, J., Ripabelli, G., Brett, M.M., Threlfall, E.J., 2000. Amplified fragment length polymorphism (AFLP) analysis of Clostridium perfringens for epidemiological typing. Int J Food Microbiol 56, 21-28.

McReynolds, J.L., Byrd, J.A., Anderson, R.C., Moore, R.W., Edrington, T.S., Genovese, K.J., Poole, T.L., Kubena, L.F., Nisbet, D.J., 2004. Evaluation of immunosuppressants and dietary mechanisms in an experimental disease model for necrotic enteritis. Poultry Science 83, 1948-1952.

Mehdizadeh Gohari, I., Kropinski, A.M., Weese, S.J., Parreira, V.R., Whitehead, A.E., Boerlin, P., Prescott, J.F., 2016. Plasmid Characterization and Chromosome Analysis of Two netF+ Clostridium perfringens Isolates Associated with Foal and Canine Necrotizing Enteritis. PLoS ONE 11, e0148344.

Mehdizadeh Gohari, I., Kropinski, A.M., Weese, S.J., Whitehead, A.E., Parreira, V.R., Boerlin, P., Prescott, J.F., 2017. NetF-producing Clostridium perfringens: Clonality and plasmid pathogenicity loci analysis. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 49, 32-38.

Mehdizadeh Gohari, I., Parreira, V.R., Nowell, V.J., Nicholson, V.M., Oliphant, K., 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.

Miyakawa, M.E., Saputo, J., Leger, J.S., Puschner, B., Fisher, D.J., McClane, B.A., Uzal, F.A., 2007. Necrotizing enterocolitis and death in a goat kid associated with enterotoxin (CPE)-producing Clostridium perfringens type A. The Canadian veterinary journal = La revue veterinaire canadienne 48, 1266-1269.

Miyamoto, K., Fisher, D.J., Li, J., Sayeed, S., Akimoto, S., McClane, B.A., 2006. Complete Sequencing and Diversity Analysis of the Enterotoxin-Encoding Plasmids in Clostridium perfringens Type A Non-Food-Borne Human Gastrointestinal Disease Isolates. Journal of Bacteriology 188, 1585-1598.

Miyamoto, K., Li, J., Sayeed, S., Akimoto, S., McClane, B.A., 2008. Sequencing and Diversity Analyses Reveal Extensive Similarities between Some Epsilon-Toxin-Encoding Plasmids and the pCPF5603 Clostridium perfringens Enterotoxin Plasmid. Journal of Bacteriology 190, 7178-7188.

Miyamoto, K., Yumine, N., Mimura, K., Nagahama, M., Li, J., McClane, B.A., Akimoto, S., 2011. Identification of Novel Clostridium perfringens Type E Strains That Carry an Iota Toxin Plasmid with a Functional Enterotoxin Gene. PLoS ONE 6, e20376.

Miyata, S., Matsushita, O., Minami, J., Katayama, S., Shimamoto, S., Okabe, A., 2001. Cleavage of a C-terminal peptide is essential for heptamerization of Clostridium perfringens epsilon-toxin in the synaptosomal membrane. The Journal of biological chemistry 276, 13778-13783.

Moore, R.J., 2016. Necrotic enteritis predisposing factors in broiler chickens. Avian Pathology 45, 275-281.

Moran, E.T., Jr., 2014. Intestinal events and nutritional dynamics predispose Clostridium perfringens virulence in broilers. Poult Sci 93, 3028-3036.

Moran, N.A., Plague, G.R., 2004. Genomic changes following host restriction in bacteria. Current opinion in genetics & development 14, 627-633.

Moura, A., Criscuolo, A., Pouseele, H., Maury, M.M., Leclercq, A., Tarr, C., Bjorkman, J.T., Dallman, T., Reimer, A., Enouf, V., Larsonneur, E., Carleton, H., Bracq-Dieye, H., Katz, L.S., Jones, L., Touchon, M., Tourdjman, M., Walker, M., Stroika, S., Cantinelli, T., Chenal-Francisque, V., Kucerova, Z., Rocha, E.P., Nadon, C., Grant, K., Nielsen, E.M., Pot, B., Gerner-Smidt, P., Lecuit, M., Brisse, S., 2016. Whole genome-based population biology and epidemiological surveillance of Listeria monocytogenes. Nature microbiology 2, 16185.

Murrell, T.G.C., Walker, P.D., 1991. The pigbel story of Papua New Guinea. Transactions of the Royal Society of Tropical Medicine and Hygiene 85, 119-122.

Myers, G.S., Rasko, D.A., Cheung, J.K., Ravel, J., Seshadri, R., DeBoy, R.T., Ren, Q., Varga, J., Awad, M.M., Brinkac, L.M., Daugherty, S.C., Haft, D.H., Dodson, R.J., Madupu, R., Nelson, W.C., Rosovitz, M.J., Sullivan, S.A., Khouri, H., Dimitrov, G.I., Watkins, K.L., Mulligan, S., Benton, J., Radune, D., Fisher, D.J., Atkins, H.S., Hiscox, T., Jost, B.H., Billington, S.J., Songer, J.G., McClane, B.A., Titball, R.W., Rood, J.I., Melville, S.B., Paulsen, I.T., 2006. Skewed genomic variability in strains of the toxigenic bacterial pathogen, Clostridium perfringens. Genome research 16, 1031-1040.

Nagahama, M., Kobayashi, K., Ochi, S., Sakurai, J., 1991. Enzyme-linked immunosorbent assay for rapid detection of toxins from Clostridium perfringens. FEMS microbiology letters 68, 41-44.

Nagahama, M., Ochi, S., Oda, M., Miyamoto, K., Takehara, M., Kobayashi, K., 2015. Recent Insights into Clostridium perfringens Beta-Toxin. Toxins 7, 396-406.

Nakano, V., Ignacio, A., Llanco, L., Bueris, V., Sircili, M.P., Avila-Campos, M.J., 2017. Multilocus sequence typing analyses of Clostridium perfringens type A strains harboring tpeL and netB genes. Anaerobe 44, 99-105.

Nanra, J.S., Buitrago, S.M., Crawford, S., Ng, J., Fink, P.S., Hawkins, J., Scully, I.L., McNeil, L.K., Aste-Amézaga, J.M., Cooper, D., Jansen, K.U., Anderson, A.S., 2013. Capsular polysaccharides are an important immune evasion mechanism for Staphylococcus aureus. Human Vaccines & Immunotherapeutics 9, 480-487.

Nauerby, B., Pedersen, K., Madsen, M., 2003. Analysis by pulsed-field gel electrophoresis of the genetic diversity among Clostridium perfringens isolates from chickens. Vet Microbiol 94, 257-266.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular biology and evolution 3, 418-426.

Neumann, B., Prior, K., Bender, J.K., Harmsen, D., Klare, I., Fuchs, S., Bethe, A., Zühlke, D., Göhler, A., Schwarz, S., Schaffer, K., Riedel, K., Wieler, L.H., Werner, G., 2019. A Core Genome Multilocus Sequence Typing Scheme for Enterococcus faecalis. 57, e01686-01618.

Niilo, L., 1988. Clostridium perfringens Type C Enterotoxemia. The Canadian Veterinary Journal 29, 658-664.

Nikolenko, S.I., Korobeynikov, A.I., Alekseyev, M.A., 2013. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics 14, S7.

Novak, J.S., Juneja, V.K., McClane, B.A., 2003. An ultrastructural comparison of spores from various strains of Clostridium perfringens and correlations with heat resistance parameters. Int. J. Food Microbiol. 86, 239-247.

O'Brien, D.K., Melville, S.B., 2003. Multiple effects on Clostridium perfringens binding, uptake and trafficking to lysosomes by inhibitors of macrophage phagocytosis receptors. Microbiology 149, 1377-1386.

O'Riordan, K., Lee, J.C., 2004. Staphylococcus aureus capsular polysaccharides. Clinical microbiology reviews 17, 218-234.

Oh, S.T., Lillehoj, H.S., 2016. The role of host genetic factors and host immunity in necrotic enteritis Avian Pathology 45, 313-316.

Olkowski, A.A., Wojnarowicz, C., Chirino-Trejo, M., Drew, M.D., 2006. Responses of broiler chickens orally challenged with Clostridium perfringens isolated from field cases of necrotic enteritis. Research in veterinary science 81, 99-108.

Olkowski, A.A., Wojnarowicz, C., Chirino-Trejo, M., Laarveld, B., Sawicki, G., 2008. Sub-clinical necrotic enteritis in broiler chickens: Novel etiological consideration based on ultra-structural and molecular changes in the intestinal tissue. Research in veterinary science 85, 543-553.

Olson, N.D., Lund, S.P., Colman, R.E., Foster, J.T., Sahl, J.W., Schupp, J.M., Keim, P., Morrow, J.B., Salit, M.L., Zook, J.M., 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Frontiers in Genetics 6, 235.

Orsburn, B., Melville, S.B., Popham, D.L., 2008. Factors Contributing to Heat Resistance of Clostridium perfringens Endospores. Applied and Environmental Microbiology 74, 3328-3335.

Page, A.J., Alikhan, N.-F., Carleton, H.A., Seemann, T., Keane, J.A., Katz, L.S., 2017. Comparison of classical multi-locus sequence typing software for next-generation sequencing data. Microbial genomics 3, e000124-e000124.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691-3693.

Palliyeguru, M.W., Rose, S.P., Mackenzie, A.M., 2010. Effect of dietary protein concentrates on the incidence of subclinical necrotic enteritis and growth performance of broiler chickens. Poult Sci 89, 34-43.

Parish, W.E., 1961. Necrotic enteritis in the fowl (Gallus gallus domesticus). I. Histopathology of the disease and isolation of a strain of Clostridium welchii. Journal of comparative pathology 71, 377-393.

Park, J.Y., Kim, S., Oh, J.Y., Kim, H.R., Jang, I., Lee, H.S., Kwon, Y.K., 2015. Characterization of Clostridium perfringens isolates obtained from 2010 to 2012 from chickens with necrotic enteritis in Korea. Poult Sci 94, 1158-1164.

Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D., Mungall, K.L., Cerdeno-Tarraga, A.M., Temple, L., James, K., Harris, B., Quail, M.A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N., Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P., Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H., Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H., O'Neil, S., Ormond, D., Price, C., Rabbinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrell, B.G., Maskell, D.J., 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nature genetics 35, 32-40.

Parreira, V.R., Costa, M., Eikmeyer, F., Blom, J., Prescott, J.F., 2012. Sequence of Two Plasmids from Clostridium perfringens Chicken Necrotic Enteritis Isolates and Comparison with C. perfringens Conjugative Plasmids. PLoS ONE 7, e49753.

Parreira, V.R., Ojha, S., Lepp, D., Mehdizadeh Gohari, I., Zhou, H., Susta, L., Gong, J., Prescott, J.F., 2017. Necrotic enteritis locus 1 diguanylate cyclase and phosphodiesterase (cyclic-di-GMP) gene mutation attenuates virulence in an avian necrotic enteritis isolate of Clostridium perfringens. Veterinary Microbiology 208, 69-73.

Pearce, M.E., Alikhan, N.F., Dallman, T.J., Zhou, Z., Grant, K., Maiden, M.C.J., 2018. Comparative analysis of core genome MLST and SNP typing within a European Salmonella serovar Enteritidis outbreak. Int J Food Microbiol 274, 1-11.

Pearson, W.R., 2013. An introduction to sequence similarity ("homology") searching. Current protocols in bioinformatics, 3.1. 1-3.1. 8.

Pightling, A.W., Petronella, N., Pagotto, F., 2014. Choice of Reference Sequence and Assembler for Alignment of Listeria monocytogenes Short-Read Sequence Data Greatly Influences Rates of Error in SNP Analyses. PLOS ONE 9, e104579.

Pommier, T., Canbäck, B., Lundberg, P., Hagström, Å., Tunlid, A., 2009. RAMI: a tool for identification and characterization of phylogenetic clusters in microbial communities. Bioinformatics 25, 736-742.

Pons, J.L., Combe, M.L., Leluan, G., 1994. Multilocus enzyme typing of human and animal strains of Clostridium perfringens. FEMS microbiology letters 121, 25-30.

Popoff, M.R., 2011. Epsilon toxin: a fascinating pore-forming toxin. The FEBS journal 278, 4602-4615.

Popoff, M.R., 2014. Clostridial pore-forming toxins: powerful virulence factors. Anaerobe 30, 220-238.

Popoff, M.R., Bouvet, P., 2009. Clostridial toxins. Future microbiology 4, 1021-1064.

Popoff, M.R., Bouvet, P., 2013. Genetic characteristics of toxigenic Clostridia and toxin gene evolution. Toxicon 75, 63-89.

Prescott, J.F., Parreira, V.R., Mehdizadeh Gohari, I., Lepp, D., Gong, J., 2016. The pathogenesis of necrotic enteritis in chickens: what we know and what we need to know: a review. Avian pathology : journal of the W.V.P.A 45, 288-294.

Preston, A., Parkhill, J., Maskell, D.J., 2004. The Bordetellae: lessons from genomics. Nature Reviews Microbiology 2, 379.

Pritchard, L., Glover, R.H., Humphris, S., Elphinstone, J.G., Toth, I.K., 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods 8, 12-24.

Quainoo, S., Coolen, J.P.M., van Hijum, S., Huynen, M.A., Melchers, W.J.G., van Schaik, W., Wertheim, H.F.L., 2017. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. Clin Microbiol Rev 30, 1015-1063.

Raeside, C., Gaffé, J., Deatherage, D.E., Tenaillon, O., Briska, A.M., Ptashkin, R.N., Cruveiller, S., Médigue, C., Lenski, R.E., Barrick, J.E., Schneider, D., 2014. Large Chromosomal Rearrangements during a Long-Term Evolution Experiment with Escherichia coli. mBio 5, e01377-01314.

Rainey, F., Hollen, B., Small, A. 2009. Genus I. Clostridium Prazmowski 1880, In: Paul De Vos, G.M.G., Dorothy Jones, Noel R. Krieg, Wolfgang Ludwig, Fred A. Rainey, Karl-Heinz Schleifer and William B. Whitman (Ed.) Bergey's Manual® of Systematic Bacteriology. Springer, 738-827.

Redondo, L.M., Farber, M., Venzano, A., Jost, B.H., Parma, Y.R., Fernandez-Miyakawa, M.E., 2013. Sudden death syndrome in adult cows associated with Clostridium perfringens type E. Anaerobe 20, 1-4.

Richter, M., Rossello-Mora, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences of the United States of America 106, 19126-19131.

Robertson, S., Li, J., McClane, B.A. 2014. Bacteria: Clostridium perfringens A2 - Motarjemi, Yasmine, In: Encyclopedia of Food Safety. Academic Press, Waltham, 395-402.

Ronco, T., Stegger, M., Ng, K.L., Lilje, B., Lyhs, U., Andersen, P.S., Pedersen, K., 2017. Genome analysis of Clostridium perfringens isolates from healthy and necrotic enteritis infected chickens and turkeys. BMC Research Notes 10, 270.

Rood, J.I., 1998. Virulence genes of Clostridium perfringens. Annu Rev Microbiol 52.

Rood, J.I., Adams, V., Lacey, J., Lyras, D., McClane, B.A., Melville, S.B., Moore, R.J., Popoff, M.R., Sarker, M.R., Songer, J.G., Uzal, F.A., Van Immerseel, F., 2018. Expansion of the Clostridium perfringens toxin-based typing scheme. Anaerobe.

Rooney, A.P., Swezey, J.L., Friedman, R., Hecht, D.W., Maddox, C.W., 2006. Analysis of core housekeeping and virulence genes reveals cryptic lineages of Clostridium perfringens that are associated with distinct disease presentations. Genetics 172, 2081-2092.

Ruppitsch, W., Pietzka, A., Prior, K., Bletz, S., Fernandez, H.L., Allerberger, F., Harmsen, D., Mellmann, A., 2015. Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of Listeria monocytogenes. J Clin Microbiol 53, 2869-2876.

Sakurai, J., Kobayashi, K., 1995. Lethal and dermonecrotic activities of Clostridium perfringens Iota toxin: biological activities induced by cooperation of two nonlinked components. Microbiol Immunol 39, 249-253.

Sakurai, J., Nagahama, M., Oda, M., 2004. Clostridium perfringens alpha-toxin: characterization and mode of action. J Biochem 136.

Sakurai, J., Nagahama, M., Oda, M., Tsuge, H., Kobayashi, K., 2009. Clostridium perfringens Iota-Toxin: Structure and Function. Toxins 1, 208-228.

Sarker, M.R., Carman, R.J., McClane, B.A., 1999. Inactivation of the gene (cpe) encoding Clostridium perfringens enterotoxin eliminates the ability of two cpe-positive C. perfringens type A human gastrointestinal disease isolates to affect rabbit ileal loops. Molecular microbiology 33, 946-958.

Sawires, Y.S., Songer, J.G., 2005. Multiple-locus variable-number tandem repeat analysis for strain typing of Clostridium perfringens. Anaerobe 11, 262-272.

Sayeed, S., Li, J., McClane, B.A., 2007. Virulence plasmid diversity in Clostridium perfringens type D isolates. Infection and Immunity 75, 2391-2398.

Sayeed, S., Li, J., McClane, B.A., 2010. Characterization of virulence plasmid diversity among Clostridium perfringens type B isolates. Infect Immun 78, 495-504.

Sayeed, S., Uzal, F.A., Fisher, D.J., Saputo, J., Vidal, J.E., Chen, Y., Gupta, P., Rood, J.I., McClane, B.A., 2008. Beta toxin is essential for the intestinal virulence of Clostridium perfringens type C disease isolate CN3685 in a rabbit ileal loop model. Molecular microbiology 67, 15-30.

Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.-A., Roy, S.L., Jones, J.L., Griffin, P.M., 2011. Foodborne Illness Acquired in the United States—Major Pathogens. Emerging infectious diseases 17, 7-15.

Schalch, B., Eisgruber, H., Schau, H., Wiedmann, M., Stolle, A., 1998. Strain differentiation of Clostridium perfringens by bacteriocin typing, plasmid profiling and ribotyping. Zoonoses and Public Health 45, 595-602.

Schoepe, H., Pache, C., Neubauer, A., Potschka, H., Schlapp, T., Wieler, L.H., Baljer, G., 2001. Naturally occurring Clostridium perfringens nontoxic alpha-toxin variant as a potential vaccine candidate against alpha-toxin-associated diseases. Infect Immun 69, 7194-7196.

Scornavacca, C., Zickmann, F., Huson, D.H., 2011. Tanglegrams for rooted phylogenetic trees and networks. Bioinformatics 27, i248-256.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068-2069.

Severiano, A., Pinto, F.R., Ramirez, M., Carriço, J.A., 2011. Adjusted Wallace coefficient as a measure of congruence between typing methods. J. Clin. Microbiol. 49, 3997-4000.

Shane, S.M., 2005. Re-emergence of necrotic enteritis in the broiler industry. International Journal of Poultry Science 4, 604-611.

Shatursky, O., Bayles, R., Rogers, M., Jost, B.H., Songer, J.G., Tweten, R.K., 2000. Clostridium perfringens Beta-Toxin Forms Potential-Dependent, Cation-Selective Channels in Lipid Bilayers. Infection and Immunity 68, 5546-5551.

Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S., Hayashi, H., 2002. Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater. Proceedings of the National Academy of Sciences of the United States of America 99, 996-1001.

Shojadoost, B., Vince, A.R., Prescott, J.F., 2012. The successful experimental induction of necrotic enteritis in chickens by Clostridium perfringens: a critical review. Veterinary research 43, 12.

Shrestha, A., Uzal, F.A., McClane, B.A., 2018. Enterotoxic Clostridia: Clostridium perfringens Enteric Diseases. Microbiol Spectr 6.

Siguier, P., Gourbeyre, E., Chandler, M., 2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS microbiology reviews 38, 865-891.

Simonet, M., Riot, B., Fortineau, N., Berche, P., 1996. Invasin production by Yersinia pestis is abolished by insertion of an IS200-like element within the inv gene. Infection and Immunity 64, 375-379.

Simpson, E.H., 1949. Measurement of Diversity. Nature 163, 688.

Siragusa, G.R., Danyluk, M.D., Hiett, K.L., Wise, M.G., Craven, S.E., 2006. Molecular Subtyping of Poultry-Associated Type A Clostridium perfringens Isolates by Repetitive-Element PCR. J. Clin. Microbiol. 44, 1065-1073.

Smedley, J.G., 3rd, Uzal, F.A., McClane, B.A., 2007. Identification of a prepore large-complex stage in the mechanism of action of Clostridium perfringens enterotoxin. Infect Immun 75, 2381-2390.

Smyth, J.A., 2016. Pathology and diagnosis of necrotic enteritis: is it clear-cut? Avian pathology : journal of the W.V.P.A 45, 282-287.

Smyth, J.A., Martin, T.G., 2010. Disease producing capability of netB positive isolates of C. perfringens recovered from normal chickens and a cow, and netB positive and negative isolates from chickens with necrotic enteritis. Veterinary Microbiology 146, 76-84.

Snyder, L., Peters, J.E., Henkin, T.M., Champness, W., 2013. Molecular Genetics of Bacteria, 4th Edition. American Society of Microbiology.

Song, H., Hwang, J., Yi, H., Ulrich, R.L., Yu, Y., Nierman, W.C., Kim, H.S., 2010. The Early Stage of Bacterial Genome-Reductive Evolution in the Host. PLOS Pathogens 6, e1000922.

Songer, J.G., 1996. Clostridial enteric diseases of domestic animals. Clinical Microbiology Reviews 9, 216-234.

Songer, J.G. 2016. Infections by Clostridium perfringens Type E, In: Clostridial Diseases of Animals (eds F. A. Uzal, J. G. Songer, J. F. Prescott and M. R. Popoff) doi:10.1002/9781118728291.ch14.

Songer, J.G., Miskimmins, D.W., 2004. Clostridium perfringens type E enteritis in calves: two cases and a brief review of the literature. Anaerobe 10, 239-242.

SPRAI. SPRAI: Single pass read accuracy improver. http://zombie.cb.k.u-tokyo.ac.jp/sprai/index.html. .

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Stanley, D., Keyburn, A.L., Denman, S.E., Moore, R.J., 2012. Changes in the caecal microflora of chickens following Clostridium perfringens challenge to induce necrotic enteritis. Vet Microbiol 159, 155-162.

Stanley, D., Wu, S.-B., Rodgers, N., Swick, R.A., Moore, R.J., 2014. Differential Responses of Cecal Microbiota to Fishmeal, <italic>Eimeria</italic> and <italic>Clostridium perfringens</italic> in a Necrotic Enteritis Challenge Model in Chickens. PLoS ONE 9, e104739.

Stanley, T., Wilson, I.G.J.M.B., 2003. Multilocus enzyme electrophoresis. 24, 203-220.

Stevens, D.L., Aldape, M.J., Bryant, A.E., 2012. Life-threatening clostridial infections. Anaerobe 18, 254-259.

Takehara, M., Takagishi, T., Seike, S., Oda, M., Sakaguchi, Y., Hisatsune, J., Ochi, S., Kobayashi, K., Nagahama, M., 2017. Cellular Entry of Clostridium perfringens Iota-Toxin and Clostridium botulinum C2 Toxin. Toxins 9, 247.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30, 2725-2729.

Tange, O., 2011. GNU Parallel - The Command-Line Power Tool, ;login: The USENIX Magazine, February 2011:42-47.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research 28, 33-36.

Tekedar, H.C., Karsi, A., Reddy, J.S., Nho, S.W., Kalindamar, S., Lawrence, M.L., 2017. Comparative Genomics and Transcriptional Analysis of Flavobacterium columnare Strain ATCC 49512. Frontiers in Microbiology 8, 588.

Theoret, J.R., Uzal, F.A., McClane, B.A., 2015. Identification and characterization of Clostridium perfringens beta toxin variants with differing trypsin sensitivity and in vitro cytotoxicity activity. Infect Immun 83, 1477-1486.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research 22, 4673-4680.

Tillier, E.R., Collins, R.A., 2000. Genome rearrangement by replication-directed translocation. Nature genetics 26, 195-197.

Timbermont, L., De Smet, L., Van Nieuwerburgh, F., Parreira, V.R., Van Driessche, G., Haesebrouck, F., Ducatelle, R., Prescott, J., Deforce, D., Devreese, B., Van Immerseel, F., 2014. Perfrin, a novel bacteriocin associated with netB positive Clostridium perfringens strains from broilers with necrotic enteritis. Veterinary research 45, 40-40.

Timbermont, L., Haesebrouck, F., Ducatelle, R., Van Immerseel, F., 2011. Necrotic enteritis in broilers: an updated review on the pathogenesis. Avian pathology : journal of the W.V.P.A 40, 341-347.

Timbermont, L., Lanckriet, A., Pasmans, F., Haesebrouck, F., Ducatelle, R., Van Immerseel, F., 2009. Intra-species growth-inhibition by Clostridium perfringens is a possible virulence trait in necrotic enteritis in broilers. Vet Microbiol 137, 388-391.

Tolooe, A., Shojadoost, B., Peighambari, S.M., 2011a. Molecular detection and characterization of cpb2 gene in Clostridium perfringens isolates from healthy and diseased chickens. Journal of Venomous Animals and Toxins Including Tropical Diseases 17, 59-65.

Tolooe, A., Shojadoost, B., Peighambari, S.M., Tamaddon, Y., 2011b. Prevalence of netB Gene among Clostridium perfringens Isolates Obtained from Healthy and Diseased Chickens. J. Anim. Vet. Adv. 10, 106-110.

Tomasini, N., Lauthier, J.J., Llewellyn, M.S., Diosque, P., 2013. MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 20, 188-196.

Tsiouris, V., Georgopoulou, I., Batzios, C., Pappaioannou, N., Ducatelle, R., Fortomaris, P., 2015. High stocking density as a predisposing factor for necrotic enteritis in broiler chicks. Avian pathology : journal of the W.V.P.A 44, 59-66.

Uzal, F.A., 2004. Diagnosis of Clostridium perfringens intestinal infections in sheep and goats. Anaerobe 10, 135-143.

Uzal, F.A., Freedman, J.C., Shrestha, A., Theoret, J.R., Garcia, J., Awad, M.M., Adams, V., Moore, R.J., Rood, J.I., McClane, B.A., 2014. Towards an understanding of the role of Clostridium perfringens toxins in human and animal disease. Future microbiology 9, 361-377.

Uzal, F.A., Kelly, W.R., Morris, W.E., Bermudez, J., Baison, M., 2004. The pathology of peracute experimental Clostridium perfringens type D enterotoxemia in sheep. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 16, 403-411.

Uzal, F.A., McClane, B.A., 2011. Recent progress in understanding the pathogenesis of Clostridium perfringens type C infections. Vet Microbiol 153, 37-43.

Uzal, F.A., McClane, B.A., Cheung, J.K., Theoret, J., Garcia, J.P., Moore, R.J., Rood, J.I., 2015. Animal models to study the pathogenesis of human and animal Clostridium perfringens infections. Vet Microbiol.

Uzal, F.A., Saputo, J., Sayeed, S., Vidal, J.E., Fisher, D.J., Poon, R., Adams, V., Fernandez-Miyakawa, M.E., Rood, J.I., McClane, B.A., 2009. Development and Application of New Mouse Models To Study the Pathogenesis of Clostridium perfringens Type C Enterotoxemias. Infection and Immunity 77, 5291-5299.

Uzal, F.A., Senties-Cue, C.G., Rimoldi, G., Shivaprasad, H.L., 2016a. Non-Clostridium perfringens infectious agents producing necrotic enteritis-like lesions in poultry. Avian pathology : journal of the W.V.P.A 45, 326-333.

Uzal, F.A., Songer, J.G., 2008. Diagnosis of Clostridium perfringens intestinal infections in sheep and goats. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 20, 253-265.

Uzal, F.A., Songer, J.G., Prescott, J.F., Popoff, M.R., Giannitti, F., Finnie, J.W., García, J.P. 2016b. Diseases Produced by Clostridium perfringens Type D. In Clostridial Diseases of Animals (eds F. A. Uzal, J. G. Songer, J. F. Prescott and M. R. Popoff). , In: Clostridial Diseases of Animals.

Uzal, F.A., Songer, J.G., Prescott, J.F., Popoff, M.R., Silva, R.O., Uzal, F.A., Oliveira, C.A., Lobato, F.C. 2016c. Gas Gangrene (Malignant Edema), In: Clostridial Diseases of Animals (eds F. A. Uzal, J. G. Songer, J. F. Prescott and M. R. Popoff).

Uzal, F.A., Vidal, J.E., McClane, B.A., Gurjar, A.A., 2010. Toxins Involved in Mammalian Veterinary Diseases. The open toxinology journal 2, 24-42.

van Asten, A.J., Allaart, J.G., Meeles, A.D., Gloudemans, P.W., Houwers, D.J., Grone, A., 2008. A new PCR followed by MboI digestion for the detection of all variants of the Clostridium perfringens cpb2 gene. Vet Microbiol 127, 412-416.

Varga, J.J., Nguyen, V., O'Brien, D.K., Rodgers, K., Walker, R.A., Melville, S.B., 2006. Type IV pili-dependent gliding motility in the Gram-positive pathogen Clostridium perfringens and other Clostridia. Molecular microbiology 62, 680-694.

Verherstraeten, S., Goossens, E., Valgaeren, B., Pardon, B., Timbermont, L., Haesebrouck, F., Ducatelle, R., Deprez, P., Wade, K., Tweten, R., Van Immerseel, F., 2015. Perfringolysin O: The Underrated Clostridium perfringens Toxin? Toxins 7, 1702.

Verherstraeten, S., Goossens, E., Valgaeren, B., Pardon, B., Timbermont, L., Vermeulen, K., Schauvliege, S., Haesebrouck, F., Ducatelle, R., Deprez, P., Van Immerseel, F., 2013. The synergistic necrohemorrhagic action of Clostridium perfringens perfringolysin and alpha toxin in the bovine intestine and against bovine endothelial cells. Veterinary research 44, 45.

Vidal, J.E., McClane, B.A., Saputo, J., Parker, J., Uzal, F.A., 2008. Effects of Clostridium perfringens Beta-Toxin on the Rabbit Small Intestine and Colon. Infection and Immunity 76, 4396-4404.

Vidal, J.E., Shak, J.R., Canizalez-Roman, A., 2015. The CpAL Quorum Sensing System Regulates Production of Hemolysins CPA and PFO To Build Clostridium perfringens Biofilms. Infection and Immunity 83, 2430-2442.

Vilei, E.M., Schlatter, Y., Perreten, V., Straub, R., Popoff, M.R., Gibert, M., Grone, A., Frey, J., 2005. Antibiotic-induced expression of a cryptic cpb2 gene in equine beta2-toxigenic Clostridium perfringens. Molecular microbiology 57, 1570-1581.

Virden, W.S., Kidd, M.T., 2009. Physiological stress in broilers: Ramifications on nutrient digestibility and responses. The Journal of Applied Poultry Research 18, 338-347.

Wade, B., Keyburn, A., 2015. The true cost of necrotic enteritis. World Poultry, 31, 16–17.

Wade, B., Keyburn, A.L., Seemann, T., Rood, J.I., Moore, R.J., 2015. Binding of Clostridium perfringens to collagen correlates with the ability to cause necrotic enteritis in chickens. Vet Microbiol 180, 299-303.

Watson, P.J., Scholes, S.F., 2009. Clostridium perfringens type D epsilon intoxication in one-day-old calves. Vet Rec 164, 816-817.

Weinstein , L., Barza , M.A., 1973. Gas Gangrene. New England Journal of Medicine 289, 1129-1131.

Welch, W.H., Nuttall, G.H.F., 1892. A gas producing bacillus (Bacillus aerogenes capsulatus, Nov. Spec.) capable of rapid development in the blood vessels after death. Bull. Johns Hopkins Bull. 3, 81-91.

Wickham H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Williams, R.B., 2005. Intercurrent coccidiosis and necrotic enteritis of chickens: rational, integrated disease management by maintenance of gut integrity. Avian pathology : journal of the W.V.P.A 34, 159-180.

Williamson, E.D., Titball, R.W., 1993. A genetically engineered vaccine against the alpha-toxin of Clostridium perfringens protects mice against experimental gas gangrene. Vaccine 11, 1253-1258.

Wu, S.-B., Stanley, D., Rodgers, N., Swick, R.A., Moore, R.J., 2014. Two necrotic enteritis predisposing factors, dietary fishmeal and Eimeria infection, induce large changes in the caecal microbiota of broiler chickens. Veterinary Microbiology 169, 188-197.

Wu, S., Zhu, Z., Fu, L., Niu, B., Li, W., 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics 12, 444.

Xiao, Y., Wagendorp, A., Moezelaar, R., Abee, T., Wells-Bennik, M.H., 2012. A wide variety of Clostridium perfringens type A food-borne isolates that carry a chromosomal cpe gene belong to one multilocus sequence typing cluster. Appl Environ Microbiol 78, 7060-7068.

Xie, Z., Tang, H., 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics 33, 3340-3347.

Xu, S., Lee, S.-H., Lillehoj, H.S., Hong, Y.H., Bravo, D., 2015a. Effects of dietary selenium on host response to necrotic enteritis in young broilers. Research in veterinary science 98, 66-73.

Xu, S.Z., Lee, S.H., Lillehoj, H.S., Bravo, D., 2015b. Dietary sodium selenite affects host intestinal and systemic immune response and disease susceptibility to necrotic enteritis in commercial broilers. British poultry science 56, 103-112.

Yonogi, S., Matsuda, S., Kawai, T., Yoda, T., Harada, T., Kumeda, Y., Gotoh, K., Hiyoshi, H., Nakamura, S., Kodama, T., Iida, T., 2014. BEC, a novel enterotoxin of Clostridium perfringens found in human clinical isolates from acute gastroenteritis outbreaks. Infect Immun 82, 2390-2399.

Yoo, H.S., Lee, S.U., Park, K.Y., Park, Y.H., 1997. Molecular typing and epidemiological survey of prevalence of Clostridium perfringens types by multiplex PCR. J. Clin. Microbiol. 35, 228-232.

Zeissler, J., Rassfeld-Sternberg, L., Oakley, C.L., Dieckmann, C., Hain, E., 1949. Enteritis Necroticans due to Clostridium Welchii Type F. British Medical Journal 1, 267-271.

Zhou, H., Lepp, D., Pei, Y., Liu, M., Yin, X., Ma, R., Prescott, J.F., Gong, J., 2017. Influence of pCP1NetB ancillary genes on the virulence of Clostridium perfringens poultry necrotic enteritis strain CP1. Gut Pathogens 9, 6.

Appendix

Table S1: Annotation summary for the 30 closed genomes of *C. perfringens* strains based on Prokka annotation

	Chromosome statistics								Extrachromosomal elements			Genome (chromosome and plasmids)
Strain	**Size (bp)**	**GC%**	**# rRNA**	**#rRNA operon**	**# CDS**	**# tRNA**	**Coding density**	**CRISPR (repeat No.)**	**# Episome***	**Episome sizes variation (Kb)**	**Episomes CDS**	**#CDS total**
ATCC_13124	3256683	28,38	24	8	2879	93	83,9	None	0			2879
CBA7123	3088370	28,54	30	10	2739	94	82,8	Class I (76)	1	46,6	55	2794
CP15	3354441	28,58	30	10	2977	94	83,2	Class I (35)	4	2,4 - 14,7	21	2998
Del1	3559163	28,32	29	10	3297	93	83,1	None	4	49,5 - 82,5	292	3589
FORC_003	3338532	28,43	30	10	2929	96	83,4	Class II (23)	1	56,5	65	2994
FORC_025	3343822	28,49	30	10	2986	95	83,6	None	0			2986
JP55	3347300	28,38	30	10	2992	95	83,2	None	5	14,06 - 72,5	241	3233
JP838	3530414	28,38	30	10	3150	93	83,7	None#	5	14,6 - 404,5	704	3854
SM101	2897393	28,25	30	10	2580	94	82,2	Class I (61)	3	12,2 - 38,09	78	2658
Str_13	3031430	28,57	30	10	2670	97	83,5	None	1	54,3	59	2729
NCTC_10239	2957462	28,31	30	10	2644	92	82,4	None	3	12,2 - 57,9	82	2726
NCTC_10240	2932430	28,32	30	10	2563	93	82,6	Class I (43)	2 (1)	9,7 - 56,6	64	2627
NCTC_10578	3301103	28,36	30	10	2849	94	83,1	Class I (108), II (33)	0			2849
NCTC_10612	3000433	28,2	30	10	2673	88	82,4	Class I (59)	2 (1)	8,8 - 13,6	19	2692

NCTC_10614	2938187	28,27	30	10	2608	94	82,2	Class I (61)	2 (1)	12,2 - 56,6	67	2675
NCTC_11144	3254757	28,44	30	10	2848	92	83,5	None	3	50,9 - 78	210	3058
NCTC_13170	3291135	28,46	30	10	2880	95	83,5	Class II (17)	0			2880
NCTC_2544	3184230	28,47	30	10	2754	93	83,1	None	0			2754
NCTC_2837	3273517	28,39	30	10	2868	95	83,2	None	0			2868
NCTC_8081	3079007	28,22	30	10	2765	93	82,4	None	6 (1)	20,7 - 116,6	428	3193
NCTC_8238	2967022	28,27	30	10	2647	94	82,4	Class I (61)	1	56,6	58	2705
NCTC_8239	2930120	28,28	30	10	2589	94	82,4	Class I (35)	1	56,6	58	2647
NCTC_8246	3300925	28,38	30	10	2887	94	83,5	None	1	54,3	58	2945
NCTC_8247	2959541	28,25	30	10	2650	94	82,3	Class I (34)	1 (1)	57,02	58	2708
NCTC_8359	2935354	28,29	30	10	2631	94	82,6	Class I (53)	3	12,2 - 56,6	79	2710
NCTC_8678	2937420	28,27	30	10	2621	92	82,3	Class I (60)	2 (2)	4,7 - 38	47	2668
NCTC_8679	2976781	28,28	30	10	2681	92	82,3	None	2	12,1 - 54,4	64	2745
NCTC_8797	2948187	28,31	30	10	2633	92	82,5	None	2 (2)	19 - 25,3	51	2684
NCTC_8799	2966670	28,3	30	10	2649	94	82,4	Class I (67)	3 (3)	7,8 - 63,1	78	2727
NCTC_9851	2918345	28,28	30	10	2586	92	82,3	Class I (34)	2	12,3 - 56,6	70	2656
Average	3.126.672,5	28,4	29,8		2.774,2	93,5	82,9		1,8		125,3	2.874,4
Standard deviation	199.972,0	0,1	1,1		182,5	1,6	0,5		1,6		156,4	283,3

*episome: plasmid or phages, the number in brackets refers to the incomplete genetic elements
§accession number of the strains; sequence data are provided in Table 3 (see 2.1.1)

CRISPR with 10 repeat units; repeat region was predicted on the large plasmid of strain JP838

Table S2: Identified CRISPR-*Cas* system in the chromosome of closed genomes of C*. perfringens*

CRISPR repeat site 1 (class I)	CRISPR repeat	GTTTTATATTAACTATGAGGAATGTAAAT
	Avg. repeat Length	29
	Avg. spacer length	36
	Associated *Cas* genes	cas4-cas1-cas2-cas6-cas8a1-cas7-cas5-cas3 (12 strains)
		cas6-cas8a1-cas7-cas5-cas3 (CP15, NCTC 8799)
CRISPR repeat site 2 (Class II)	CRISPR repeat	GTTATAGTTCCTAGTAAATTCTCGATATGCTATAAT
	Avg. repeat Length	36
	Avg. spacer length	29
	Associated *Cas* genes	cas2-cas1-cas9 (FORC_003 & NCTC 13170)
		cas1-cas9 (NCTC 10578)
CRISPR repeat site 3	CRISPR repeat	GTTGAACTTTAACAAAGGTTGTATTTAAAG
	Avg. repeat Length	30
	Avg. spacer length	35
	Flanking genes	transposase, IS605 family (NCTC 10578)

Table S3: Insertion sequences and genomic islands identified in the chromosome of the closed genomes of *C. perfringens*

	Insertion sequences														Genomic islands					
	Insertion sequences families																			
Strain	IS607	IS6	IS30	IS3	IS21	IS1182	IS256	new	IS200/IS605	IS110	IS1595	Total no.	Total length	%	No GIs	Total size of the GIs	%	Min GI size per genome	Max GI size per genome	Average GI size per genome
complete *C. perfringens* genomes available at the NCBI																				
ATCC_13124	3				2							5	9021	0,28	3	26110	0,80	4369	11037	
CBA7123	1	2	5	1	2				1			12	16902	0,55	1	8240	0,27			
CP15	1			3	2				2			8	8862	0,26	8	134388	4,01	8230	36563	16798,5
Del1	3				3				1	4		11	26875	0,76	9	110256	3,10	5106	28270	12250,6
FORC_003	4		2	1	3				7			17	20852	0,62	5	41532	1,24	2605	13319	8306,4
FORC_025	4			2	3				1			10	13997	0,42	7	97382	2,91	5512	30016	13911,7
JP55	7			2	3			1	12	2		27	32051	0,96	4	35565	1,06	6177	11893	8891,2
JP838	1				2		1	1	1	1		7	12260	0,35	10	75657	2,14	1762	10967	7565,7
SM101	7	14	17	3	2	5	3	2	22			75	99683	3,44	13	172636	5,96	5933	39668	13279,6
Str_13	4				2				3			9	13567	0,45	4	85985	2,84	10664	32822	21496,25
Genomes assembled to closure within the study																				
NCTC_10239	8	17	16	1	2	5			28			77	95772	3,24	14	225119	7,61	3800	39688	16079,9
NCTC_10240	15	2	10	3	2	1	1	2	29			65	86195	2,94	11	140647	4,80	5505	26598	12786,09
NCTC_10578	3				2				5			10	15805	0,48	3	42927	1,30	10703	21190	14309
NCTC_10612	7	12	18	4	3	6	7	2	23			82	107879	3,6	16	181094	6,04	5933	39668	11318,3
NCTC_10614	8	14	18	3	2	3	4		24			76	98591	3,36	13	176761	6,02	6303	39668	13597
NCTC_11144	4			1	2				1			8	13919	0,43	6	67563	2,08	7019	18761	11260,5

NCTC_13170	4			1	3				2			10	15672	0,48	4	30435	0,92	4348	11037	7608,75
NCTC_2544	4		2		2				5			13	17467	0,55	2	21708	0,68	10671	11037	
NCTC_2837	3	2			2				2		1	10	12982	0,4	7	96375	2,94	3217	27509	13767,8
NCTC_8081	9	68	29	6	2	2		2	20	3		141	187878	6,1	23	330095	10,72	4625	42539	14351,9
NCTC_8238	7	13	17	4	3	3	3	2	25			77	98959	3,34	13	181235	6,11	6303	39668	13941,15
NCTC_8239	7	14	21	3	2	4	3	2	26			82	105625	3,6	13	166137	5,67	6303	39668	12779,7
NCTC_8246	5			1	2				2			10	16225	0,49	4	61649	1,87	8230	27914	15412,25
NCTC_8247	6	15	20	4	2	5		3	27			82	106504	3,6	15	135481	4,58	3490	13911	9032,06
NCTC_8359	8	21	22	4	2	4	3	1	23			88	115511	3,94	12	180911	6,16	3294	39688	15075,9
NCTC_8678	8	25	20	1	3	4	2		27			90	109675	3,73	16	236853	8,06	6256	39688	14803,3
NCTC_8679	8	19	26	2	2	4	4		21			86	111075	3,73	17	266660	8,96	5094	40485	15685,8
NCTC_8797	7	18	19	2	3	5			31			85	108199	3,67	16	211971	7,19	3663	39688	13248,18
NCTC_8799	8	19	19	4	2	6	5	1	24			88	116111	3,91	13	202758	6,83	3800	39688	15596,7
NCTC_9851	7	14	18	4	2	4	3	2	25			79	102053	3,5	13	157659	5,40	4815	39668	12127,6
Total	171	289	299	60	69	61	39	21	420	10	1									
Minimum												5,0	8.862,0	0,3	1,0	8.240,0	0,3	1.762,0	10.967,0	7.565,7
Maximum												141,0	187.878,0	6,1	23,0	330.095,0	10,7	10.703,0	42.539,0	21.496,3
Median												46,0	59.123,0	2,0	10,5	134.934,5	4,3	5.505,0	32.822,0	13.597,0
Average												48,0	63.205,6	2,1	9,8	130.059,6	4,3	5.645,9	29.390,2	13.158,6
Standard deviation												39,4	50.356,7	1,7	5,5	81.555,5	2,8	2.316,1	11.760,4	3.098,6

Table S4: Prophage identified within the chromosome of closed genomes of *C. perfringens*

	Integrated prophages																												
	Intact phages												Incomplete phages										Questionable						
Strain	ΦCP51	Φ3626	phiSM101	phiCT19406C	Lactob_LLKu	phiCD6356	Staphy_80	Entero_phiFL4A	Strept_phiARI0746	Lactob_Ld25A	PhiS63	Bacill_vB_BhaS_171	ΦCP51	Φ3626	Bacill_AR9	Clostr_c_st	phiCTC2A	PhiS63	Entero_phi92	Lactob_phiAT3	Prochl_P_SSM2	Staphy_SPbeta_like	ΦCP51	Strept_phiARI0031	Prochl_P_SSM2	Lactob_PLE2	**Total number of predicted phages**	**Total predicted phages size**	**%**
ATCC_13124	1	1																									2	99977	3,07
CBA7123		1																									1	48492	1,57
CP15																							1				1	44372	1,32
Del1	2	1	1	1																							5	192104	5,40
FORC_003	1					1																					2	68247	2,04
FORC_025					1		1																1				3	125316	3,75
JP55	1																										1	55036	1,64
JP838				1							1		1				1							1			5	116418	3,30
SM101															1	1					1	1					4	60143	2,08
Str_13														1													1	17682	0,58
NCTC_10239									1																		1	42625	1,44
NCTC_10240								1																			1	44436	1,52
NCTC_10578															1				1		1						3	25351	0,77
NCTC_10612																							1				1	50063	1,67

NCTC_10614															1	1			1		1						4	74217	2,53
NCTC_11144										1																	1	38137	1,17
NCTC_13170	1																										1	55321	1,68
NCTC_2544															1						1						2	18820	0,59
NCTC_2837								1																			1	53042	1,62
NCTC_8081			1																								1	54007	1,75
NCTC_8238																							1				1	41465	1,40
NCTC_8239															1	1					1	1					4	60139	2,05
NCTC_8246																		1									1	20774	0,63
NCTC_8247		1																									1	46748	1,58
NCTC_8359											1																1	58454	1,99
NCTC_8678																									1		1	43833	1,49
NCTC_8679					1																					1	2	75568	2,54
NCTC_8797																							1				1	44969	1,53
NCTC_8799								1																			1	52480	1,77
NCTC_9851															1	1			1	1	1	1					6	78706	2,70
Minimum																											1,0	17.682,0	0,6
Maximum																											6,0	192.104,0	5,4
Median																											1,0	52.761,0	1,7
Average																											2,0	60.231,4	1,9
Standard deviation																											1,5	35.369,6	1,0

Table S5: List of cgMLST loci and different alleles found for each locus

Target	#alleles	Target	#alleles	Target	#alleles	Target	#alleles	Target	#alleles	Target	#alleles	Target	#alleles	Target	#alleles	Target	#alleles
CPF_0002	34	CPF_0310	54	CPF_0749	47	CPF_1447	39	CPF_1801	41	CPF_2048	46	CPF_2269	11	CPF_2486	46	CPF_2762	36
CPF_0003	13	CPF_0311	29	CPF_0750	54	CPF_1448	47	CPF_1802	47	CPF_2049	30	CPF_2270	31	CPF_2487	40	CPF_2763	45
CPF_0004	24	CPF_0315	42	CPF_0752	49	CPF_1449	27	CPF_1803	48	CPF_2050	40	CPF_2271	24	CPF_2496	43	CPF_2764	17
CPF_0005	8	CPF_0317	63	CPF_0764	53	CPF_1450	51	CPF_1804	36	CPF_2051	18	CPF_2272	33	CPF_2497	42	CPF_2765	43
CPF_0006	41	CPF_0318	52	CPF_0777	43	CPF_1460	52	CPF_1806	21	CPF_2052	55	CPF_2274	34	CPF_2498	43	CPF_2766	25
CPF_0010	43	CPF_0319	36	CPF_0778	35	CPF_1461	44	CPF_1807	49	CPF_2062	40	CPF_2275	37	CPF_2499	53	CPF_2767	31
CPF_0011	52	CPF_0322	54	CPF_0790	34	CPF_1462	47	CPF_1808	39	CPF_2063	33	CPF_2276	50	CPF_2500	32	CPF_2768	41
CPF_0012	47	CPF_0323	53	CPF_0793	39	CPF_1471	50	CPF_1809	27	CPF_2064	32	CPF_2277	42	CPF_2502	51	CPF_2774	43
CPF_0013	19	CPF_0324	49	CPF_0794	38	CPF_1482	55	CPF_1811	50	CPF_2065	44	CPF_2278	28	CPF_2503	55	CPF_2775	49
CPF_0014	45	CPF_0325	43	CPF_0840	63	CPF_1484	37	CPF_1812	8	CPF_2066	27	CPF_2279	10	CPF_2504	28	CPF_2776	50
CPF_0021	23	CPF_0337	61	CPF_0845	57	CPF_1488	26	CPF_1813	26	CPF_2068	41	CPF_2280	1	CPF_2508	36	CPF_2777	50
CPF_0022	41	CPF_0338	17	CPF_0847	41	CPF_1492	48	CPF_1817	49	CPF_2069	46	CPF_2281	22	CPF_2512	56	CPF_2779	45
CPF_0023	13	CPF_0339	45	CPF_0850	49	CPF_1496	51	CPF_1827	46	CPF_2070	17	CPF_2282	43	CPF_2536	43	CPF_2781	49
CPF_0027	49	CPF_0340	50	CPF_0851	48	CPF_1499	54	CPF_1829	33	CPF_2071	37	CPF_2283	34	CPF_2537	59	CPF_2782	9
CPF_0028	48	CPF_0341	47	CPF_0854	50	CPF_1500	50	CPF_1831	42	CPF_2072	39	CPF_2284	46	CPF_2539	37	CPF_2783	37
CPF_0029	51	CPF_0342	37	CPF_0855	52	CPF_1501	39	CPF_1832	35	CPF_2073	52	CPF_2285	31	CPF_2540	29	CPF_2784	38
CPF_0030	48	CPF_0343	32	CPF_0857	53	CPF_1505	38	CPF_1834	37	CPF_2074	40	CPF_2290	51	CPF_2541	56	CPF_2785	50
CPF_0032	50	CPF_0345	40	CPF_0858	15	CPF_1506	5	CPF_1836	18	CPF_2076	43	CPF_2291	24	CPF_2542	58	CPF_2786	46
CPF_0033	48	CPF_0346	42	CPF_0861	50	CPF_1507	48	CPF_1842	39	CPF_2077	37	CPF_2292	40	CPF_2546	42	CPF_2787	23
CPF_0034	34	CPF_0350	28	CPF_0869	36	CPF_1508	41	CPF_1843	34	CPF_2078	16	CPF_2293	45	CPF_2548	45	CPF_2788	37
CPF_0036	48	CPF_0351	34	CPF_0877	26	CPF_1509	22	CPF_1851	38	CPF_2079	5	CPF_2294	29	CPF_2549	47	CPF_2789	60
CPF_0037	38	CPF_0352	33	CPF_0880	43	CPF_1510	32	CPF_1852	46	CPF_2080	19	CPF_2296	36	CPF_2550	6	CPF_2803	48
CPF_0038	19	CPF_0353	44	CPF_0908	61	CPF_1511	25	CPF_1853	55	CPF_2081	34	CPF_2297	45	CPF_2551	47	CPF_2804	46
CPF_0040	48	CPF_0356	53	CPF_0909	42	CPF_1512	32	CPF_1854	46	CPF_2082	34	CPF_2298	48	CPF_2552	47	CPF_2805	29
CPF_0042	55	CPF_0365	57	CPF_1166	29	CPF_1513	47	CPF_1855	43	CPF_2083	46	CPF_2300	47	CPF_2553	41	CPF_2807	61
CPF_0043	48	CPF_0367	51	CPF_1168	37	CPF_1514	41	CPF_1856	22	CPF_2084	33	CPF_2302	32	CPF_2554	42	CPF_2808	30

CPF_0044	48	CPF_0368	29	CPF_1170	20	CPF_1516	38	CPF_1859	23	CPF_2085	13	CPF_2303	49	CPF_2557	51	CPF_2811	42
CPF_0045	48	CPF_0374	46	CPF_1171	46	CPF_1517	30	CPF_1860	31	CPF_2086	37	CPF_2304	31	CPF_2558	57	CPF_2812	28
CPF_0046	38	CPF_0375	33	CPF_1173	40	CPF_1518	23	CPF_1861	46	CPF_2087	47	CPF_2305	40	CPF_2563	60	CPF_2813	43
CPF_0047	44	CPF_0376	51	CPF_1174	22	CPF_1519	35	CPF_1862	25	CPF_2088	9	CPF_2306	19	CPF_2565	58	CPF_2814	17
CPF_0050	51	CPF_0377	48	CPF_1175	15	CPF_1522	35	CPF_1864	12	CPF_2089	15	CPF_2307	10	CPF_2569	56	CPF_2815	30
CPF_0051	45	CPF_0378	51	CPF_1176	52	CPF_1528	28	CPF_1865	35	CPF_2098	51	CPF_2308	59	CPF_2570	57	CPF_2816	40
CPF_0052	53	CPF_0379	42	CPF_1178	26	CPF_1529	41	CPF_1866	23	CPF_2099	47	CPF_2309	31	CPF_2571	39	CPF_2818	23
CPF_0053	7	CPF_0380	29	CPF_1179	28	CPF_1536	16	CPF_1868	25	CPF_2100	42	CPF_2310	13	CPF_2573	50	CPF_2820	41
CPF_0054	41	CPF_0382	48	CPF_1180	37	CPF_1537	24	CPF_1869	13	CPF_2101	50	CPF_2311	50	CPF_2574	17	CPF_2821	31
CPF_0055	21	CPF_0385	46	CPF_1183	45	CPF_1538	41	CPF_1870	29	CPF_2102	40	CPF_2312	45	CPF_2576	22	CPF_2823	42
CPF_0056	26	CPF_0389	35	CPF_1189	41	CPF_1539	45	CPF_1871	15	CPF_2103	28	CPF_2314	17	CPF_2578	25	CPF_2824	48
CPF_0058	11	CPF_0392	44	CPF_1190	40	CPF_1541	49	CPF_1872	55	CPF_2106	30	CPF_2317	34	CPF_2580	52	CPF_2825	16
CPF_0059	14	CPF_0393	41	CPF_1191	53	CPF_1542	56	CPF_1873	37	CPF_2108	24	CPF_2318	58	CPF_2581	39	CPF_2826	29
CPF_0063	35	CPF_0415	40	CPF_1192	47	CPF_1544	40	CPF_1876	49	CPF_2109	38	CPF_2319	14	CPF_2582	37	CPF_2827	44
CPF_0064	46	CPF_0416	40	CPF_1193	43	CPF_1545	43	CPF_1878	58	CPF_2110	23	CPF_2320	39	CPF_2583	23	CPF_2829	51
CPF_0065	51	CPF_0417	61	CPF_1195	46	CPF_1546	34	CPF_1882	50	CPF_2112	37	CPF_2321	7	CPF_2584	32	CPF_2830	52
CPF_0066	42	CPF_0420	50	CPF_1196	15	CPF_1551	49	CPF_1883	27	CPF_2113	36	CPF_2322	49	CPF_2585	34	CPF_2831	48
CPF_0067	41	CPF_0421	60	CPF_1197	49	CPF_1552	48	CPF_1884	40	CPF_2114	31	CPF_2323	50	CPF_2586	21	CPF_2832	41
CPF_0069	33	CPF_0422	63	CPF_1198	36	CPF_1553	54	CPF_1885	25	CPF_2115	53	CPF_2324	45	CPF_2587	54	CPF_2833	23
CPF_0070	24	CPF_0426	66	CPF_1203	34	CPF_1554	54	CPF_1886	32	CPF_2116	49	CPF_2325	64	CPF_2588	18	CPF_2834	15
CPF_0071	46	CPF_0427	47	CPF_1208	28	CPF_1555	42	CPF_1887	18	CPF_2117	57	CPF_2326	48	CPF_2590	31	CPF_2835	41
CPF_0072	31	CPF_0428	60	CPF_1217	42	CPF_1556	30	CPF_1888	20	CPF_2118	18	CPF_2328	40	CPF_2591	20	CPF_2836	44
CPF_0074	51	CPF_0429	42	CPF_1227	54	CPF_1557	16	CPF_1890	42	CPF_2119	39	CPF_2329	53	CPF_2609	55	CPF_2838	32
CPF_0075	14	CPF_0430	14	CPF_1232	49	CPF_1558	46	CPF_1892	32	CPF_2121	40	CPF_2330	41	CPF_2610	50	CPF_2840	36
CPF_0076	25	CPF_0431	43	CPF_1233	58	CPF_1559	42	CPF_1893	24	CPF_2125	36	CPF_2331	55	CPF_2611	49	CPF_2841	37
CPF_0078	30	CPF_0433	52	CPF_1234	33	CPF_1563	35	CPF_1894	15	CPF_2126	35	CPF_2332	50	CPF_2612	53	CPF_2842	48
CPF_0097	31	CPF_0434	45	CPF_1238	47	CPF_1564	38	CPF_1895	40	CPF_2127	50	CPF_2334	48	CPF_2613	53	CPF_2843	44
CPF_0098	42	CPF_0437	14	CPF_1239	28	CPF_1612	43	CPF_1897	50	CPF_2132	45	CPF_2336	31	CPF_2614	44	CPF_2844	55
CPF_0104	19	CPF_0438	23	CPF_1241	46	CPF_1613	42	CPF_1899	53	CPF_2133	43	CPF_2337	70	CPF_2616	41	CPF_2845	38

CPF_0106	59	CPF_0439	55	CPF_1242	46	CPF_1615	34	CPF_1900	56	CPF_2134	56	CPF_2338	31	CPF_2617	43	CPF_2846	58
CPF_0107	14	CPF_0443	56	CPF_1247	41	CPF_1616	39	CPF_1903	30	CPF_2135	55	CPF_2339	35	CPF_2618	31	CPF_2847	50
CPF_0108	57	CPF_0444	11	CPF_1248	51	CPF_1623	43	CPF_1904	46	CPF_2136	55	CPF_2340	36	CPF_2623	25	CPF_2848	27
CPF_0115	56	CPF_0451	69	CPF_1250	42	CPF_1624	43	CPF_1905	13	CPF_2138	51	CPF_2341	45	CPF_2624	33	CPF_2853	28
CPF_0116	65	CPF_0452	45	CPF_1251	53	CPF_1626	28	CPF_1906	23	CPF_2139	54	CPF_2342	43	CPF_2631	41	CPF_2854	39
CPF_0128	22	CPF_0453	42	CPF_1253	42	CPF_1627	18	CPF_1910	35	CPF_2140	39	CPF_2344	50	CPF_2632	57	CPF_2855	47
CPF_0129	66	CPF_0492	58	CPF_1260	11	CPF_1628	41	CPF_1914	25	CPF_2141	41	CPF_2345	53	CPF_2633	32	CPF_2856	33
CPF_0130	59	CPF_0500	65	CPF_1261	37	CPF_1631	34	CPF_1915	41	CPF_2142	21	CPF_2346	38	CPF_2634	46	CPF_2857	37
CPF_0131	45	CPF_0501	49	CPF_1262	41	CPF_1634	39	CPF_1916	46	CPF_2143	44	CPF_2347	46	CPF_2635	9	CPF_2858	4
CPF_0133	31	CPF_0502	40	CPF_1263	44	CPF_1635	38	CPF_1917	54	CPF_2144	5	CPF_2348	23	CPF_2636	44	CPF_2859	44
CPF_0134	48	CPF_0503	71	CPF_1265	33	CPF_1636	29	CPF_1918	27	CPF_2148	41	CPF_2349	44	CPF_2638	48	CPF_2860	36
CPF_0135	12	CPF_0505	37	CPF_1267	22	CPF_1639	6	CPF_1919	39	CPF_2149	48	CPF_2350	59	CPF_2639	51	CPF_2861	40
CPF_0150	50	CPF_0506	74	CPF_1268	14	CPF_1643	28	CPF_1920	31	CPF_2150	39	CPF_2351	40	CPF_2640	49	CPF_2862	32
CPF_0167	37	CPF_0508	53	CPF_1269	62	CPF_1644	49	CPF_1921	25	CPF_2158	41	CPF_2357	34	CPF_2642	26	CPF_2863	34
CPF_0168	41	CPF_0509	57	CPF_1274	42	CPF_1645	49	CPF_1922	46	CPF_2159	16	CPF_2358	40	CPF_2643	25	CPF_2864	22
CPF_0169	56	CPF_0510	53	CPF_1275	39	CPF_1646	43	CPF_1923	5	CPF_2160	46	CPF_2359	64	CPF_2644	44	CPF_2865	54
CPF_0171	37	CPF_0511	60	CPF_1279	37	CPF_1647	18	CPF_1924	39	CPF_2161	36	CPF_2360	56	CPF_2646	61	CPF_2866	10
CPF_0172	52	CPF_0512	43	CPF_1280	41	CPF_1648	35	CPF_1925	11	CPF_2162	41	CPF_2366	39	CPF_2648	45	CPF_2867	29
CPF_0173	46	CPF_0513	23	CPF_1286	50	CPF_1651	22	CPF_1926	31	CPF_2163	45	CPF_2367	47	CPF_2649	51	CPF_2868	54
CPF_0174	46	CPF_0514	41	CPF_1287	51	CPF_1652	27	CPF_1927	27	CPF_2164	33	CPF_2369	54	CPF_2650	42	CPF_2871	26
CPF_0175	49	CPF_0516	41	CPF_1288	39	CPF_1654	25	CPF_1929	29	CPF_2165	33	CPF_2371	23	CPF_2651	42	CPF_2872	46
CPF_0176	42	CPF_0517	32	CPF_1289	47	CPF_1656	57	CPF_1931	33	CPF_2166	26	CPF_2372	50	CPF_2652	42	CPF_2873	27
CPF_0177	35	CPF_0518	47	CPF_1291	44	CPF_1657	9	CPF_1932	36	CPF_2168	21	CPF_2373	45	CPF_2653	35	CPF_2875	47
CPF_0178	41	CPF_0520	38	CPF_1292	18	CPF_1658	45	CPF_1933	9	CPF_2171	42	CPF_2378	49	CPF_2654	44	CPF_2876	44
CPF_0179	46	CPF_0521	27	CPF_1293	46	CPF_1661	42	CPF_1934	43	CPF_2172	37	CPF_2379	34	CPF_2655	51	CPF_2877	18
CPF_0180	26	CPF_0523	48	CPF_1294	45	CPF_1662	54	CPF_1935	3	CPF_2173	50	CPF_2380	27	CPF_2656	42	CPF_2878	28
CPF_0181	50	CPF_0524	21	CPF_1295	47	CPF_1666	42	CPF_1936	31	CPF_2174	56	CPF_2381	12	CPF_2657	29	CPF_2881	42
CPF_0182	38	CPF_0526	60	CPF_1297	39	CPF_1667	45	CPF_1937	36	CPF_2175	57	CPF_2382	42	CPF_2658	55	CPF_2883	43
CPF_0183	23	CPF_0527	21	CPF_1298	42	CPF_1668	23	CPF_1938	34	CPF_2176	22	CPF_2385	3	CPF_2659	51	CPF_2885	40

CPF_0185	32	CPF_0528	50	CPF_1299	27	CPF_1670	49	CPF_1939	12	CPF_2177	48	CPF_2386	54	CPF_2661	47	CPF_2889	42
CPF_0187	51	CPF_0536	9	CPF_1300	51	CPF_1671	34	CPF_1940	34	CPF_2178	36	CPF_2388	48	CPF_2662	30	CPF_2890	22
CPF_0188	48	CPF_0561	39	CPF_1304	50	CPF_1672	48	CPF_1941	19	CPF_2179	44	CPF_2389	49	CPF_2664	29	CPF_2891	33
CPF_0190	41	CPF_0563	43	CPF_1308	53	CPF_1674	36	CPF_1943	34	CPF_2180	31	CPF_2390	46	CPF_2665	44	CPF_2892	16
CPF_0191	41	CPF_0564	52	CPF_1309	35	CPF_1675	48	CPF_1945	56	CPF_2181	48	CPF_2391	42	CPF_2666	49	CPF_2893	39
CPF_0192	37	CPF_0565	41	CPF_1310	18	CPF_1676	9	CPF_1946	47	CPF_2182	38	CPF_2392	15	CPF_2667	54	CPF_2894	48
CPF_0197	32	CPF_0566	41	CPF_1312	45	CPF_1677	28	CPF_1947	46	CPF_2183	47	CPF_2393	31	CPF_2668	24	CPF_2895	21
CPF_0198	55	CPF_0568	47	CPF_1315	33	CPF_1678	35	CPF_1948	43	CPF_2184	40	CPF_2394	20	CPF_2670	42	CPF_2896	39
CPF_0199	40	CPF_0569	47	CPF_1316	46	CPF_1679	46	CPF_1949	26	CPF_2185	35	CPF_2395	55	CPF_2671	54	CPF_2897	63
CPF_0201	43	CPF_0571	45	CPF_1322	32	CPF_1680	37	CPF_1950	23	CPF_2186	30	CPF_2396	19	CPF_2672	31	CPF_2898	43
CPF_0202	57	CPF_0574	46	CPF_1323	22	CPF_1681	56	CPF_1951	10	CPF_2187	14	CPF_2397	38	CPF_2673	52	CPF_2899	37
CPF_0204	43	CPF_0575	44	CPF_1324	44	CPF_1683	44	CPF_1952	22	CPF_2188	52	CPF_2398	37	CPF_2674	31	CPF_2900	50
CPF_0205	40	CPF_0576	45	CPF_1325	44	CPF_1685	45	CPF_1955	21	CPF_2189	44	CPF_2399	29	CPF_2675	37	CPF_2901	35
CPF_0207	58	CPF_0577	13	CPF_1326	32	CPF_1686	43	CPF_1956	50	CPF_2190	46	CPF_2400	33	CPF_2676	37	CPF_2902	33
CPF_0208	54	CPF_0579	53	CPF_1327	45	CPF_1687	47	CPF_1958	53	CPF_2192	29	CPF_2401	35	CPF_2681	37	CPF_2903	47
CPF_0209	54	CPF_0580	61	CPF_1328	25	CPF_1688	44	CPF_1959	40	CPF_2193	51	CPF_2402	40	CPF_2682	32	CPF_2904	33
CPF_0211	42	CPF_0581	51	CPF_1329	22	CPF_1689	31	CPF_1960	43	CPF_2194	18	CPF_2403	46	CPF_2684	6	CPF_2906	41
CPF_0212	36	CPF_0582	42	CPF_1332	34	CPF_1690	37	CPF_1961	42	CPF_2195	28	CPF_2404	48	CPF_2685	22	CPF_2907	33
CPF_0214	37	CPF_0583	57	CPF_1337	34	CPF_1691	42	CPF_1970	51	CPF_2197	27	CPF_2405	52	CPF_2686	13	CPF_2909	43
CPF_0215	47	CPF_0586	56	CPF_1338	44	CPF_1693	34	CPF_1973	5	CPF_2198	39	CPF_2406	13	CPF_2687	3	CPF_2911	37
CPF_0216	57	CPF_0588	48	CPF_1339	50	CPF_1694	44	CPF_1974	37	CPF_2199	2	CPF_2407	41	CPF_2688	4	CPF_2912	41
CPF_0217	58	CPF_0589	28	CPF_1340	35	CPF_1696	46	CPF_1975	2	CPF_2200	12	CPF_2408	10	CPF_2689	2	CPF_2913	30
CPF_0219	32	CPF_0590	51	CPF_1341	37	CPF_1698	45	CPF_1976	7	CPF_2201	45	CPF_2409	46	CPF_2690	3	CPF_2914	50
CPF_0220	28	CPF_0591	72	CPF_1342	38	CPF_1699	41	CPF_1977	36	CPF_2202	45	CPF_2410	52	CPF_2691	13	CPF_2915	57
CPF_0222	43	CPF_0592	52	CPF_1343	15	CPF_1702	10	CPF_1978	36	CPF_2203	47	CPF_2411	48	CPF_2692	14	CPF_2916	60
CPF_0223	53	CPF_0594	52	CPF_1350	58	CPF_1703	43	CPF_1979	45	CPF_2204	32	CPF_2412	51	CPF_2693	19	CPF_2917	18
CPF_0225	51	CPF_0613	52	CPF_1356	51	CPF_1704	46	CPF_1981	41	CPF_2205	36	CPF_2413	41	CPF_2694	26	CPF_2919	53
CPF_0227	49	CPF_0614	52	CPF_1357	47	CPF_1705	42	CPF_1982	13	CPF_2210	45	CPF_2417	8	CPF_2695	5	CPF_2921	22
CPF_0230	33	CPF_0623	45	CPF_1360	47	CPF_1706	51	CPF_1983	24	CPF_2211	47	CPF_2421	37	CPF_2696	4	CPF_2927	51

CPF_0231	52	CPF_0624	35	CPF_1361	9	CPF_1707	23	CPF_1985	43	CPF_2212	53	CPF_2423	51	CPF_2697	3	CPF_2928	27
CPF_0232	44	CPF_0628	45	CPF_1362	51	CPF_1709	29	CPF_1986	6	CPF_2213	42	CPF_2424	50	CPF_2698	4	CPF_2929	40
CPF_0238	58	CPF_0629	51	CPF_1367	16	CPF_1710	36	CPF_1987	5	CPF_2215	55	CPF_2425	43	CPF_2699	11	CPF_2930	23
CPF_0239	52	CPF_0632	27	CPF_1373	16	CPF_1712	51	CPF_1988	35	CPF_2216	26	CPF_2426	62	CPF_2700	4	CPF_2931	28
CPF_0241	55	CPF_0633	41	CPF_1374	21	CPF_1713	20	CPF_1989	35	CPF_2217	39	CPF_2427	54	CPF_2701	2	CPF_2932	37
CPF_0242	48	CPF_0634	47	CPF_1375	30	CPF_1718	42	CPF_1990	35	CPF_2220	29	CPF_2429	57	CPF_2702	3	CPF_2933	41
CPF_0243	43	CPF_0635	27	CPF_1376	31	CPF_1719	18	CPF_1992	27	CPF_2221	38	CPF_2432	44	CPF_2703	3	CPF_2934	27
CPF_0245	29	CPF_0637	44	CPF_1377	44	CPF_1720	39	CPF_1993	36	CPF_2222	15	CPF_2433	50	CPF_2704	4	CPF_2935	14
CPF_0246	9	CPF_0638	52	CPF_1379	35	CPF_1721	48	CPF_1994	42	CPF_2223	38	CPF_2434	46	CPF_2705	2	CPF_2936	31
CPF_0247	18	CPF_0639	51	CPF_1381	37	CPF_1722	33	CPF_1995	36	CPF_2224	34	CPF_2435	35	CPF_2706	2	CPF_2937	46
CPF_0252	29	CPF_0650	47	CPF_1382	46	CPF_1724	28	CPF_1996	36	CPF_2226	46	CPF_2444	21	CPF_2707	2	CPF_2938	44
CPF_0253	46	CPF_0652	56	CPF_1383	43	CPF_1727	12	CPF_1997	26	CPF_2227	20	CPF_2445	34	CPF_2708	8	CPF_2939	38
CPF_0254	34	CPF_0654	18	CPF_1384	46	CPF_1728	37	CPF_1998	54	CPF_2228	25	CPF_2446	11	CPF_2709	5	CPF_2940	44
CPF_0258	33	CPF_0655	42	CPF_1385	46	CPF_1729	32	CPF_1999	41	CPF_2229	16	CPF_2447	34	CPF_2710	1	CPF_2953	53
CPF_0261	24	CPF_0656	40	CPF_1387	36	CPF_1730	36	CPF_2001	19	CPF_2230	37	CPF_2448	32	CPF_2711	15	CPF_2954	42
CPF_0262	24	CPF_0657	59	CPF_1388	44	CPF_1732	6	CPF_2003	25	CPF_2231	51	CPF_2449	42	CPF_2712	5	CPF_2956	10
CPF_0263	29	CPF_0658	20	CPF_1389	36	CPF_1733	37	CPF_2004	39	CPF_2232	44	CPF_2450	32	CPF_2713	8	CPF_2957	43
CPF_0264	36	CPF_0664	46	CPF_1390	40	CPF_1734	21	CPF_2005	38	CPF_2233	42	CPF_2451	13	CPF_2717	30	CPF_2958	49
CPF_0265	42	CPF_0667	49	CPF_1391	38	CPF_1736	35	CPF_2006	33	CPF_2234	40	CPF_2452	32	CPF_2718	9	CPF_2959	26
CPF_0266	44	CPF_0668	29	CPF_1392	49	CPF_1738	45	CPF_2007	33	CPF_2236	36	CPF_2453	20	CPF_2720	8	CPF_2960	40
CPF_0267	39	CPF_0669	49	CPF_1393	34	CPF_1739	9	CPF_2008	42	CPF_2237	38	CPF_2454	33	CPF_2721	46	CPF_2961	55
CPF_0268	52	CPF_0670	44	CPF_1401	32	CPF_1740	17	CPF_2009	39	CPF_2238	55	CPF_2455	20	CPF_2722	50	CPF_2962	41
CPF_0269	37	CPF_0671	63	CPF_1407	48	CPF_1741	37	CPF_2010	46	CPF_2239	53	CPF_2456	8	CPF_2733	13	CPF_2963	37
CPF_0271	11	CPF_0672	38	CPF_1409	49	CPF_1745	25	CPF_2011	36	CPF_2240	35	CPF_2457	4	CPF_2734	27	CPF_2969	46
CPF_0273	18	CPF_0673	44	CPF_1413	39	CPF_1746	39	CPF_2012	16	CPF_2243	49	CPF_2458	15	CPF_2735	31	CPF_2970	46
CPF_0276	41	CPF_0674	47	CPF_1414	36	CPF_1756	46	CPF_2013	16	CPF_2244	45	CPF_2459	15	CPF_2736	22	CPF_2971	55
CPF_0278	22	CPF_0675	43	CPF_1415	46	CPF_1757	39	CPF_2014	26	CPF_2245	41	CPF_2460	34	CPF_2737	38	CPF_2972	20
CPF_0279	39	CPF_0677	46	CPF_1416	54	CPF_1758	44	CPF_2015	33	CPF_2247	50	CPF_2461	38	CPF_2738	39	CPF_2973	55
CPF_0280	45	CPF_0678	39	CPF_1417	42	CPF_1760	51	CPF_2016	31	CPF_2248	30	CPF_2462	24	CPF_2739	23	CPF_2975	19

CPF_0281	39	CPF_0680	27	CPF_1418	9	CPF_1761	51	CPF_2020	39	CPF_2249	43	CPF_2463	26	CPF_2740	32	CPF_2976	4
CPF_0283	23	CPF_0695	39	CPF_1419	12	CPF_1762	29	CPF_2022	41	CPF_2250	64	CPF_2464	49	CPF_2741	21	CPF_2978	4
CPF_0284	34	CPF_0698	42	CPF_1420	11	CPF_1763	25	CPF_2023	24	CPF_2251	39	CPF_2465	25	CPF_2743	43	CPF_2979	13
CPF_0288	51	CPF_0703	50	CPF_1421	42	CPF_1765	48	CPF_2024	32	CPF_2252	41	CPF_2466	31	CPF_2744	32	CPF_2980	40
CPF_0289	35	CPF_0706	35	CPF_1422	41	CPF_1766	49	CPF_2025	17	CPF_2253	44	CPF_2467	53	CPF_2745	31	CPF_2981	10
CPF_0290	10	CPF_0707	51	CPF_1427	41	CPF_1767	37	CPF_2026	36	CPF_2254	40	CPF_2468	22	CPF_2746	17	CPF_2982	19
CPF_0291	14	CPF_0708	38	CPF_1428	37	CPF_1768	45	CPF_2027	39	CPF_2255	31	CPF_2469	3	CPF_2747	6	CPF_2983	45
CPF_0292	24	CPF_0709	26	CPF_1429	45	CPF_1769	22	CPF_2028	32	CPF_2256	9	CPF_2470	42	CPF_2748	36	CPF_2984	46
CPF_0293	23	CPF_0714	43	CPF_1430	40	CPF_1770	48	CPF_2029	9	CPF_2257	15	CPF_2472	48	CPF_2749	16	CPF_2985	33
CPF_0294	6	CPF_0716	52	CPF_1431	39	CPF_1771	36	CPF_2030	10	CPF_2258	42	CPF_2473	28	CPF_2750	38	CPF_2986	26
CPF_0295	46	CPF_0718	42	CPF_1432	26	CPF_1774	36	CPF_2031	14	CPF_2259	35	CPF_2474	42	CPF_2751	50	CPF_2987	34
CPF_0297	38	CPF_0719	42	CPF_1433	30	CPF_1779	22	CPF_2032	4	CPF_2260	49	CPF_2475	31	CPF_2752	44	CPF_2988	28
CPF_0298	44	CPF_0739	55	CPF_1434	46	CPF_1780	50	CPF_2033	46	CPF_2261	43	CPF_2476	31	CPF_2754	18	CPF_2989	31
CPF_0299	50	CPF_0743	25	CPF_1435	23	CPF_1781	44	CPF_2037	34	CPF_2262	42	CPF_2477	61	CPF_2755	35	CPF_2990	10
CPF_0300	58	CPF_0744	46	CPF_1436	29	CPF_1783	52	CPF_2038	20	CPF_2263	25	CPF_2479	53	CPF_2757	5	CPF_2991	47
CPF_0301	41	CPF_0745	43	CPF_1438	23	CPF_1784	43	CPF_2044	37	CPF_2264	37	CPF_2480	7	CPF_2758	27	CPF_2992	44
CPF_0304	46	CPF_0746	69	CPF_1440	33	CPF_1786	41	CPF_2045	41	CPF_2265	47	CPF_2481	48	CPF_2759	35	CPF_2993	34
CPF_0305	47	CPF_0747	44	CPF_1441	8	CPF_1799	56	CPF_2046	40	CPF_2266	46	CPF_2484	57	CPF_2760	26	CPF_2994	21
CPF_0306	40	CPF_0748	54	CPF_1446	41	CPF_1800	36	CPF_2047	40	CPF_2268	54	CPF_2485	50	CPF_2761	19	CPF_2996	11
																CPF_2997	2

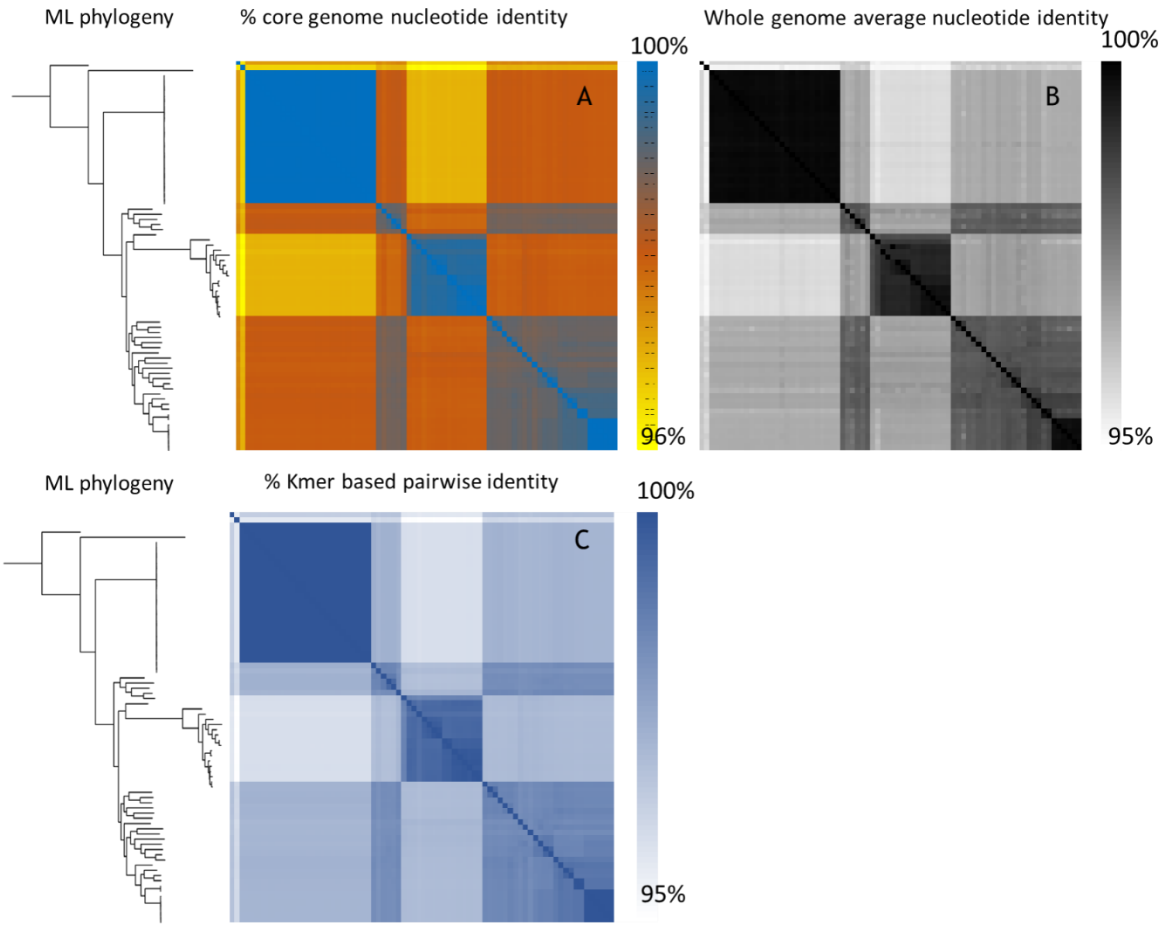


Figure S1: Nucleotide identity between the investigated 76 *C. perfringens* genomes aligned against the core genome ML phylogenetic tree (as in Figure 5). Heat maps showing **(A)** the nucleotide identity percentage calculated within the core genome (1,207 genes), **(B)** Average Nucleotide Identity (ANI) calculated pairwise based on the matching regions across the whole genome, **(C)** pairwise nucleotide identity calculated based on k-mer analysis (alignment independent).

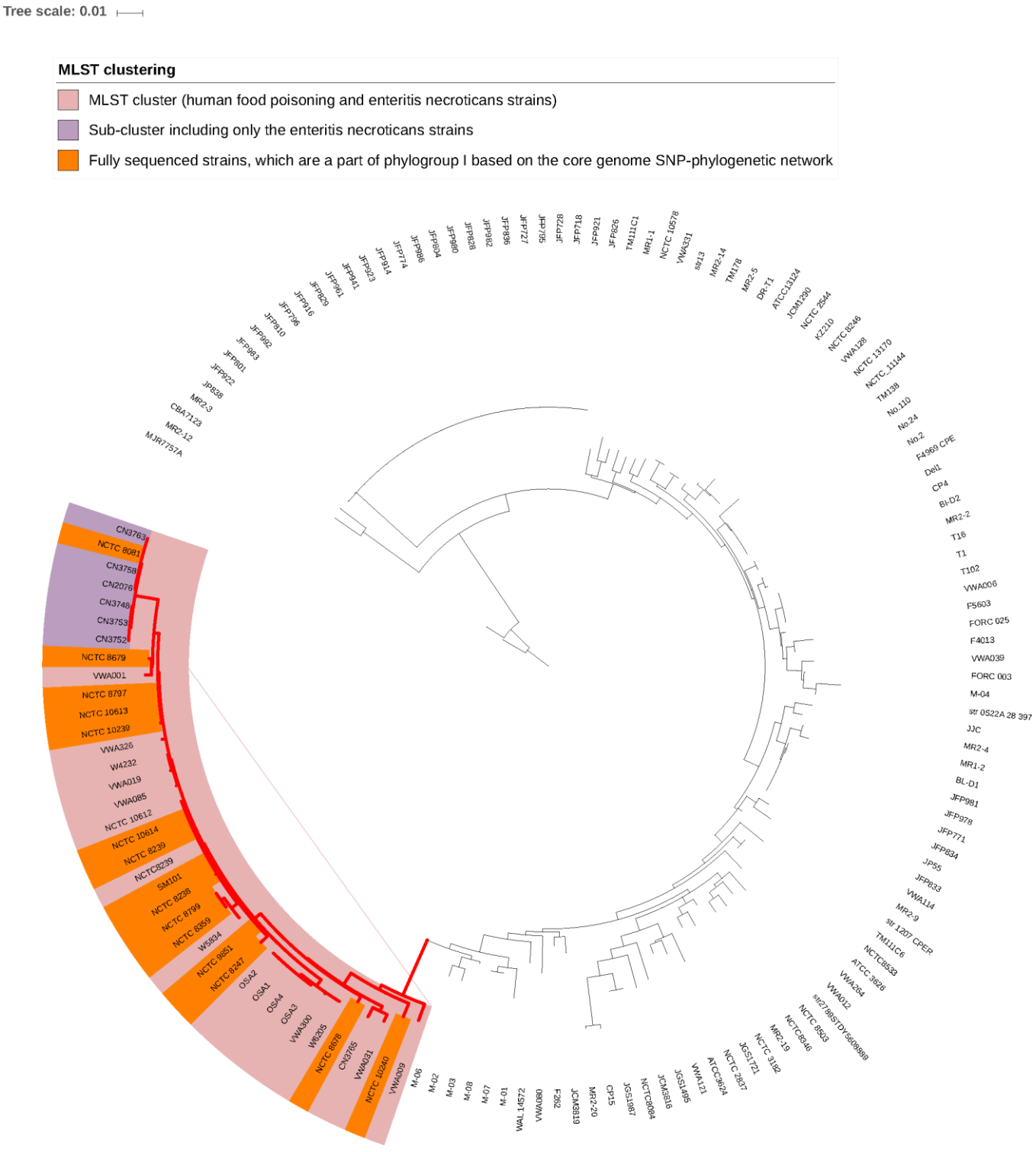


Figure S2: MLST phylogenetic tree of 147 *C. perfringens* strains based on eight MLST housekeeping genes. MLST data from strains investigated in prior studies (Deguchi et al., 2009, Ma et al., 2012 and Xiao et al., 2012) were compared to the 76 strains analyzed in this study. The food poisoning strains that carry chromosomal *cpe* and the human enteritis necroticans strains were grouped together (38 strains) and formed a distinct phylogenetic branch (colored red). For a scalable image, please see additional data 10.

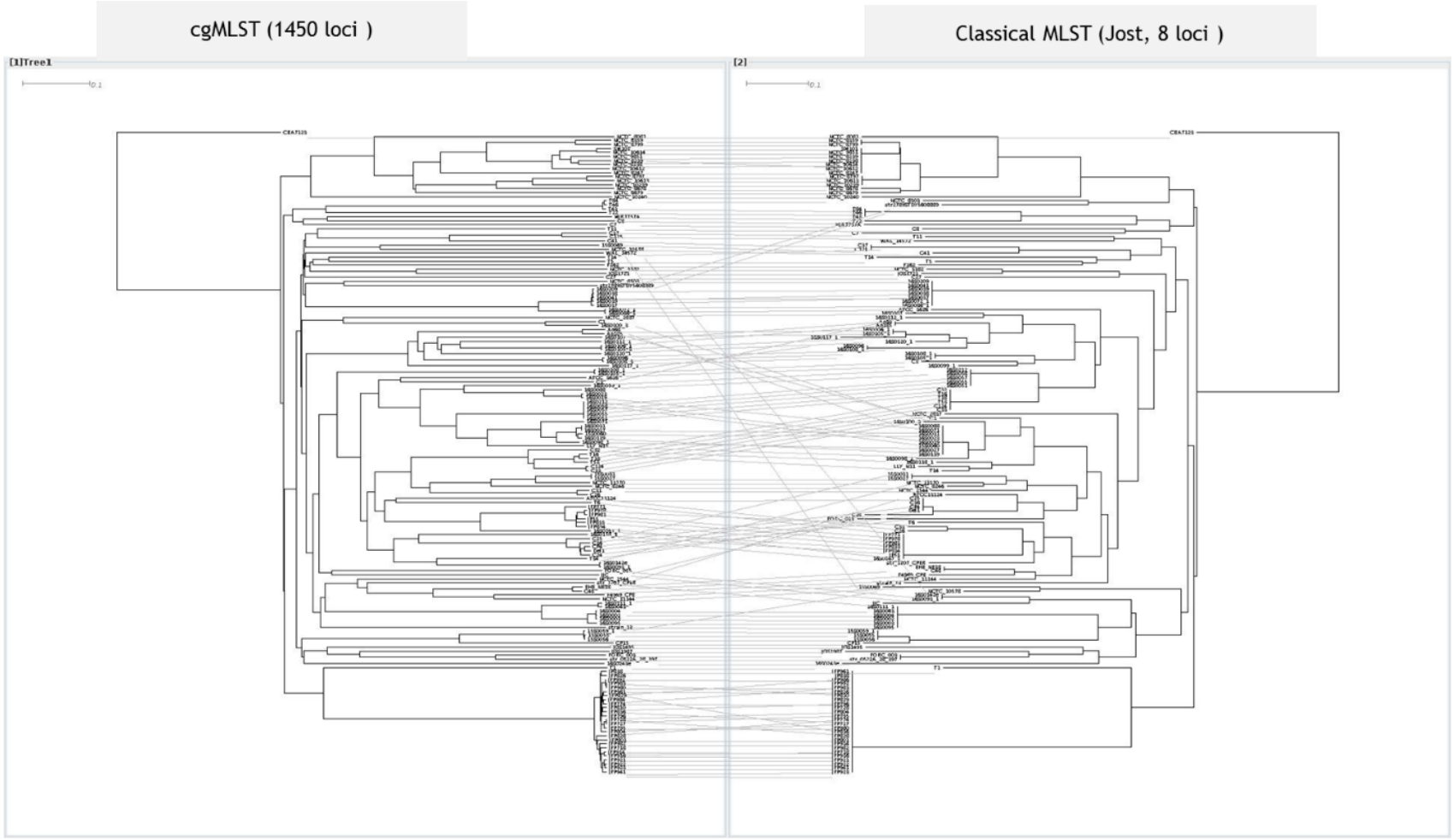


Figure S3A: Topological concordance between the neighbor-joining (NJ) trees from cgMLST (left) and classical MLST methods for *C. perfringens* (right) is represented as tanglegram. The NJ trees from the classical MLST scheme of Jost et al., (2006) was compared to the NJ tree calculated from the cgMLST (left). For a scalable overview, please see additional data 27.

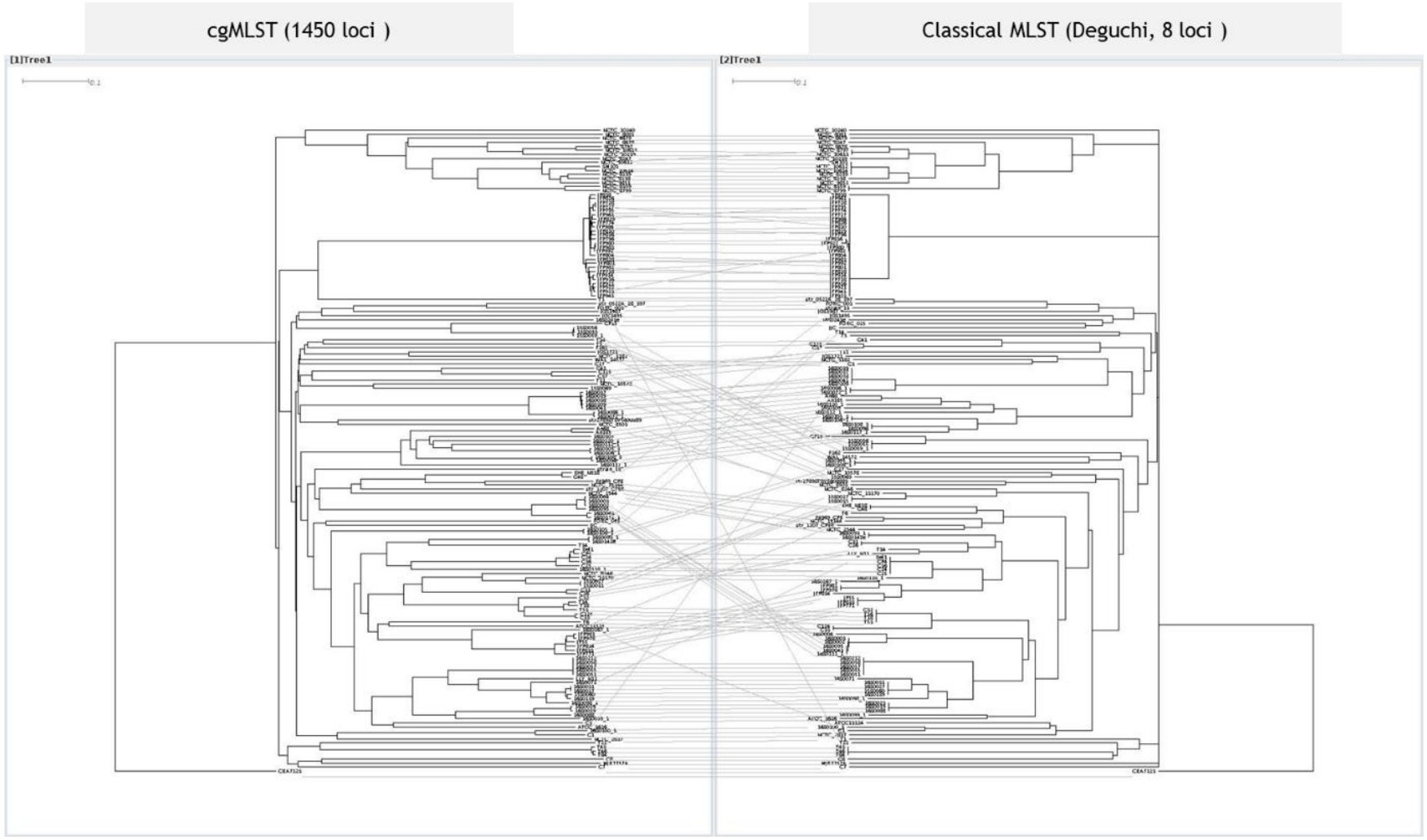


Figure S3B: Topological concordance between the neighbor-joining (NJ) trees from cgMLST (left) and classical MLST methods for *C. perfringens* (right) represented as tanglegram. The NJ trees from the classical MLST scheme of Deguchi et al., (2009) (right) was compared to the NJ tree calculated from the cgMLST (left). For a scalable overview, please see additional data 27.

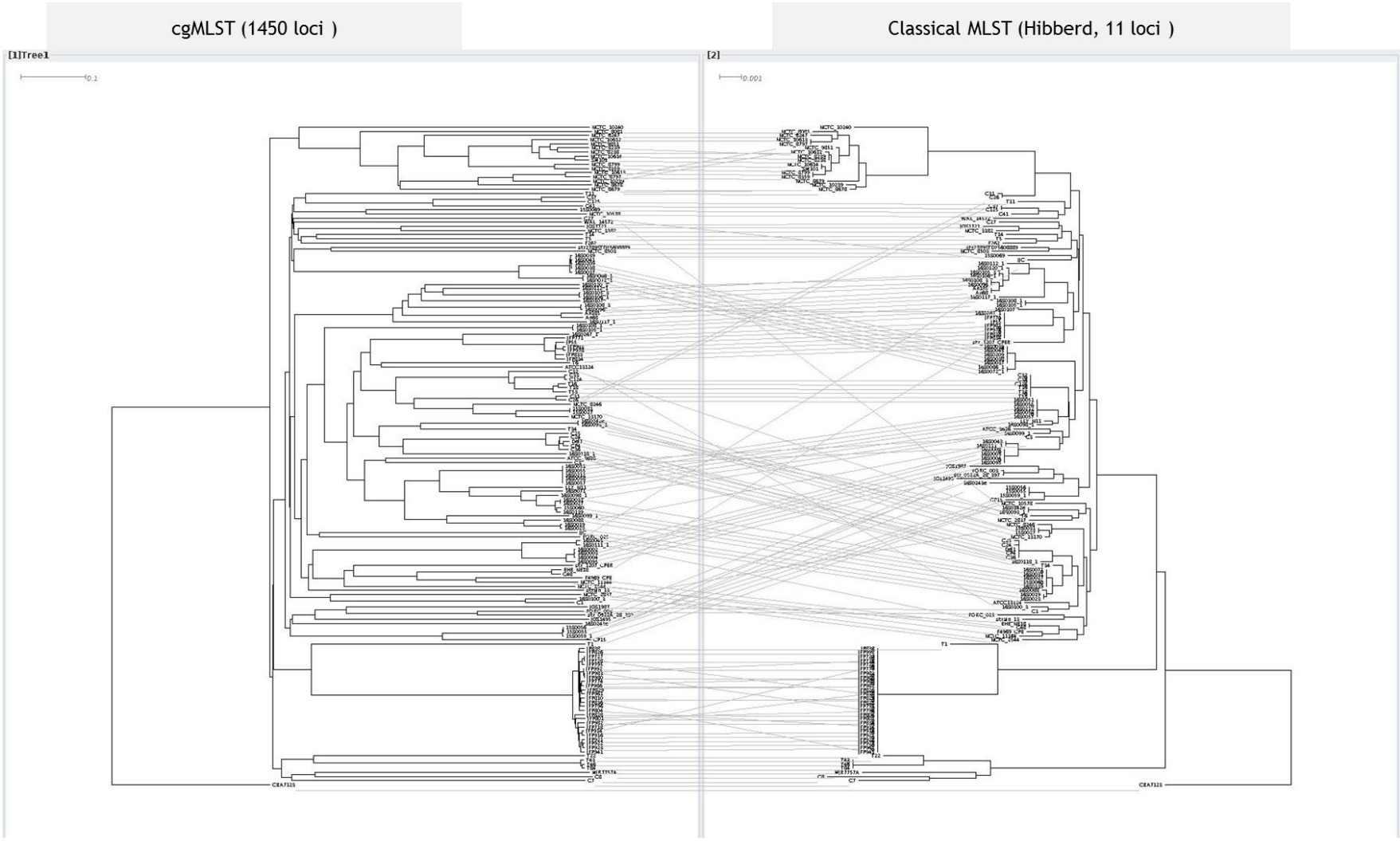


Figure S3C: Topological concordance between the neighbor-joining (NJ) trees from cgMLST (left) and classical MLST methods (right) for *C. perfringens* represented as tanglegram. The NJ trees from the classical MLST scheme of Hibberd et al., (2011) (right) was compared to the NJ tree calculated from cgMLST (left). For a scalable overview, please see additional data 27.

Acknowledgements

“Our Lord, for You is all praise, an abundant beautiful blessed praise”. O God, I thank You - so that I was able to complete the dissertation project in a way that I hope, pleases You.

This thesis would not have been possible without the inspiration and outstanding help of my supervisors, colleagues, collaborators and family - my thanks and appreciation to all of them.

My special thanks goes to **Prof. Dr. Heinrich Neubauer**, the head of the Institute of Bacterial Infection and Zoonoses at the Friedrich-Loeffler-Institut, for the opportunity he gave to me to pursue my doctorate at the FLI and to transfer the entitled research topic. His constant support significantly influenced the success of the dissertation project. Prof. Neubauer was always within easy reach despite his busy schedule and so many commitments. I thank him for the always friendly, highly competent support provided at any time during the execution and preparation of the work.

I particularly want to thank **Prof. Dr. Lothar H. Wieler**, the head of Robert Koch-Institut in Berlin, Germany for the supervision of my doctoral thesis at the FU-Berlin and for his willingness to discuss during my doctoral studies and the support in writing the doctoral thesis. He has all my thanks, appreciation and gratitude for his valuable feedback and reliable responsiveness.

In addition, I would like to express my sincere gratitude to my direct supervisor **Dr. Christian Seyboldt**, the head of our group at the Friedrich-Loeffler-Institut. Dr. Seyboldt allowed me to be a part of his research laboratory and guided me in the field of microbiology. He has also designed and supervised my dissertation, and put considerable efforts in correcting and improving my writings and presentations. His expertise, outstanding support, invaluable guidance, affectionate attitude, understanding, patience but also his constructive criticism can never be described with words. Without his continuous support and inspiration, it would have been impossible to complete this thesis. I thank him for all his efforts to bring out the best in me and shaping the doctorate research study.

I would like to warmly thank **Prof. Dr. Thomas C. Mettenleiter**, the president of the Friedrich-Loeffler-Institut, and **PD Dr. Herbert Tomaso**, the deputy of the Institute of Bacterial Infections and Zoonoses at the Friedrich-Loeffler-Institut, for organizing a job and transitional funding without which I could not have stayed at the FLI.

I am thankful to **Dr. Jörg Linde** from Friedrich-Loeffler-Institut for his inputs and good comments. Furthermore, I would like to thank my former doctoral colleague **Dr. Prasad**

Thomas for his support and encouragement. Prasad introduced me to the field of computational biology and helped me getting familiar with the Linux-based bioinformatics analysis of sequence data.

I am also grateful to **Dr. Lisa Sprague**, the leader of the working group: Bacterial Zoonoses of Companion Animals, for the effort taken by her for critical proofreading of my thesis.

Sincere thanks to **Sandra Hennig, Jutta Carmon, Renate Danner and Janin Rücknagel** for their excellent technical assistance. Sincere thanks to **Eric Zuchantke, Dr. Gernot Schmoock** and **Dr. Anne Busch** at Friedrich-Loeffler-Institut, Jena for their help.

I also gratefully thank all members of the AGr150, AGr110 and AGr180 at the Friedrich-Loeffler-Institut. **Mrs. Anja Hackbart**, **Dr. Gamal Wareth**, **Dr. Michael Böhringer**, **Dr. Belén González-Santamarina** and **Silvia García-Soto** will all be warmly remembered. Special thanks also goes to **Marcél Trautmann** and the information-technology department for their excellent technical support.

I thank the **DAAD** (Deutscher Akademischer Austauschdienst) for providing me the scholarship - so I was able to carry out this work at the Friedrich-Loeffler-Institut in Jena. In this connection, special thanks from the bottom of my heart goes to all the staff members of the Friedrich-Loeffler-Institut.

Finally, I would like to express my profound gratitude to **my parents** for their prayers and support. They were always and still the source of support and encouragement for me. I thank **my lovely wife**, **Samira**, for her patience, sacrifices, unequivocal support and understanding throughout the long journey of accomplishing my doctoral thesis. Samira, thank you for being my best friend. I owe you everything. Without you, this thesis would never have been written and I would not be the person I am today. I dedicate this thesis to you and to **our lovely children, Adam and** our new baby **Lina**, who are the pride and joy of my life, giving me unlimited pleasure and happiness. I love you more than anything.

Sponsorship

The author has been awarded a PhD scholarship from DAAD.

Selbstständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig angefertigt habe. Ich versichere, dass ich ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen habe.

Berlin, den 03.12.2019 Mostafa Y. Abdel-Glil

mbvberlin mensch und buch verlag

49,90 Euro | ISBN: 978-3-96729-030-1