Genetic characteristics of toxigenic Clostridia and toxin gene evolution

Michel Popoff, Philippe Bouvet

Elsevier Editorial System(tm) for Toxicon
Manuscript Draft

Manuscript Number:

Title: GENETIC CHARACTERISTICS OF TOXIGENIC CLOSTRIDIA AND TOXIN GENE EVOLUTION

Article Type: From Threats to Benefits

Keywords: Clostridium, toxin, evolution, genetics

Corresponding Author: Dr Michel Popoff, Dr

Corresponding Author's Institution: Institut Pasteur

First Author: Michel Popoff, Dr

Order of Authors: Michel Popoff, Dr; Philippe Bouvet

Abstract: Clostridia form a heterogeneous group of environmental bacteria containing 15 species, which produce the most potent toxins. The origin of toxins is still enigmatic. It is hypothesized that toxins exhibiting an enzymatic activity have derived from hydrolytic enzymes, which are abundantly secreted by these bacteria, and that pore-forming toxins have evolved from an ancestor transmembrane protein. The presence of related toxin genes in distinct Clostridium species and the variability of some toxin genes support horizontal toxin gene transfer and subsequent independent evolution from strain to strain. C. perfringens toxin genes involved in myonecrosis, mainly alpha toxin and perfringolysin genes, are chromosomally located, whereas toxin genes responsible for intestinal and food borne diseases are localized on plasmids except the enterotoxin gene which can be located either on chromosome or plasmid. The distribution of these plasmids containing one or several toxin genes accounts for the diverse C. perfringens toxinotypes. C. difficile strains show a high genetic variability. But in contrast to C. perfringens, toxin genes are clustered in pathogenicity locus located on chromosome. The presence of related toxin genes in distinct clostridial species like C. sordellii, C. novyi, and C. perfringens supports interspecies mobilization of this locus. The multiple C. difficile toxinotypes based on toxin gene variants possibly reflect strain adaptation to the intestinal environment. Botulinum toxin genes also show a high level of genetic variation. They have a diverse genetic localization including chromosome, plasmid or phage, and are spread in various Clostridium species (C. botulinum groups, C. argentinense, C. butyricum, C. baratii). Exchange of toxin genes not only include transfers between Clostridium species but also between Clostridium and other bacterial species as well as eukaryotic cells as supported by the wide distribution of related pore-forming toxins of the aerolysin family in various clostridial and non-clostridial species, animal, mushroom and plant.

Suggested Reviewers: Holger Brueggemann
brueggemann@mikrobiologi.au.dk
Expert in genetics of Clostridia

Niel Fearweather
n.fairweather@imperial.ac.uk
Expert in Clostridia

Opposed Reviewers:

Highlights

- Clostridia are the bacteria which produce the most numerous and most potent toxins
- Genetic characteristics have been mainly unraveled in *C. perfringens*, *C. difficile* and *C. botulinum*
- Horizontal toxin gene transfer and subsequent evolution are the main mechanisms of toxin diversity

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

GENETIC CHARACTERISTICS OF TOXIGENIC CLOSTRIDIA AND TOXIN GENE EVOLUTION

Michel R. Popoff * and Philippe Bouvet

Institut Pasteur, Unité des Bactéries anaérobies et Toxines, Paris, France

key words: *Clostridium*, toxin, evolution, genetics

corresponding author
Unité des Bactéries anaérobies et Toxines
Institut Pasteur
28 rue du Dr Roux
75724 Paris cedex 15, France
Ph 33 1 45688307
Fax 33 1 40613123
mpopoff@pateur.fr

ABSTRACT

Clostridia form a heterogeneous group of environmental bacteria containing 15 species, which produce the most potent toxins. The origin of toxins is still enigmatic. It is hypothesized that toxins exhibiting an enzymatic activity have derived from hydrolytic enzymes, which are abundantly secreted by these bacteria, and that pore-forming toxins have evolved from an ancestor transmembrane protein. The presence of related toxin genes in distinct Clostridium species and the variability of some toxin genes support horizontal toxin gene transfer and subsequent independent evolution from strain to strain. *C. perfringens* toxin genes involved in myonecrosis, mainly alpha toxin and perfringolysin genes, are chromosomally located, whereas toxin genes responsible for intestinal and food borne diseases are localized on plasmids except the enterotoxin gene which can be located either on chromosome or plasmid. The distribution of these plasmids containing one or several toxin genes accounts for the diverse *C. perfringens* toxinotypes. *C. difficile* strains show a high genetic variability. But in contrast to *C. perfringens*, toxin genes are clustered in pathogenicity locus located on chromosome. The presence of related toxin genes in distinct clostridial species like *C. sordellii, C. novyi*, and *C. perfringens* supports interspecies mobilization of this locus. The multiple *C. difficile* toxinotypes based on toxin gene variants possibly reflect strain adaptation to the intestinal environment. Botulinum toxin genes also show a high level of genetic variation. They have a diverse genetic localization including chromosome, plasmid or phage, and are spread in various Clostridium species (*C. botulinum* groups, *C. argentinense, C. butyricum, C. baratii*). Exchange of toxin genes not only include transfers between Clostridium species but also between Clostridium and other bacterial species as well as eukaryotic cells as supported by the wide distribution of related pore-forming toxins of the aerolysin family in various clostridial and non-clostridial species, animal, mushroom and plant.

1 - INTRODUCTION

Toxins are the main virulence factors of Clostridia and are responsible for severe diseases in man and animals. Clostridia are the group of bacteria, which produces the largest number of toxins. About 20% of bacterial toxins are from *Clostridium*. The genus *Clostridium* is a vast and heterogeneous group that contains more than 150 species. These anaerobic, fermentative spore-formers are however generally regarded as better adapted to life in the environment as per the production of spores. Metabolically, *Clostridia* are quite versatile and degrade an extremely wide range of organic materials that include carbohydrates, organic

acids, alcohols, aromatic compounds, peptides, amino acids, amines, purines and pyrimidines. Thereby, they effectively participate in an important aspect of ecology involving biomass renewal. Therefore, the *Clostridia* synthesize and secrete numerous hydrolytic enzymes that degrade organic molecules in the microenvironment into more easily "digested" or assimilated compounds. The resulting monomeric compounds required for growth are brought into their cytoplasm by various transporters. Perhaps the toxins produced by some *Clostridium* species evolved from "ancestral' hydrolytic enzymes by acquisition of novel specific properties that include pore formation, translocation across lipid membranes, and/or recognition of crucial eukaryotic targets. So, one might ask whether bacterial toxins, like those produced by the clostridia, represent an evolutionary protein "tool" derived from less toxic, food-gathering origins ? But among the large number of Clostriidum species, only 15 produce potent protein toxins (less than 10%). This raises the questions which is the advantage for these toxigenic species to produce so potent toxins and how such few environmental bacterial species have acquired the ability to produce such extremely active toxins directed towards eukaryotic cells. Perhaps this phenomenon is linked to acquisition of toxin genes from other organisms and/or an omnipresent and dynamic evolutionary process in some clostridial species?

Based upon DNA alignment of ribosomal RNA genes, *Clostridium* species belong to 16 different clusters that further illustrate the phylogenetic heterogeneity of this genus. Most of the toxigenic *Clostridium* species (n=10) are classified into cluster I, which is considered as the only "true" representative of the genus *Clostridium*, and the other toxigenic species are scattered into 3 other clusters (Stackebrandt and Hippe, 2001). This indicates that all of the toxigenic *Clostridia* are not related phylogenetically.

Genetic studies have been mainly focused in three toxigenic *Clostridium* species: *Clostridium perfringens*, *Clostridium difficile*, and *Clostridium botulinum*. This review concerns the main features of toxin genes in these three *Clostridium* species, and then the evolution of some clostridial toxin genes will be discussed.

2 - Toxin genes in *Clostridium perfringens*

2 -1 *Toxin gene localization and variability*

C. perfringens is the most prolific toxin-producer among known microorganisms via toxin genes that appear either on the chromosome or plasmids. The genome sequence of *C. perfringens* type A strains reveals a low G+C content (28.6%) equally distributed throughout the chromosome without any region exhibiting a remarkably higher, or lower, G+C content.

The genes for toxins and extracellular enzymes do not form pathogencity islands or possess insertion, transposon, or phage-related sequences with one known exception being the enterotoxin gene (Shimizu et al., 2002).

The gene for alpha toxin, which is the main virulence factor responsible for a life-threatening form of myonecrosis often associated with soiled wounds and commonly known as gangrene, is by classic definition produced by all *C. perfringens* strains and localized at the, same site on a variable region of the chromosome near the origin of replication (Justin et al., 2002; Rood, 1998; Tsutsui et al., 1995). Sequencing of the alpha toxin gene from different strains reveals variations, which more often are conservative and do not affect enzymatic activity, but can impact on certain biological properties like resistance to degradation (Titball et al., 1999). The genes of other toxins, such as perfringolysin O (also known as θ–toxin), collagenase (κ-toxin), and extracellular enzymes (i.e. hyaluronidase and neuraminidase) are also localized within a variable region near the alpha toxin gene locus.

In contrast, the other toxin genes (*cpb* for Beta toxin*, cpb2* for Beta2*, etx* for epsilon*, ia* and *ib* for iota*, netB* for NetB*, tpeL* for TpeL) are localized on plasmids of varying sizes. These plasmids can be lost or transferred, thus accounting for the various *C. perfringens* toxinotypes as classically defined by the production of one or several toxins. Therefore, *C. perfringens* type A that produce only one major toxin (alpha) represent the basic toxinotype for this species, which upon acquisition of a plasmid encoding for another specific toxin (beta, epsilon, or iota) yields another distinct toxinotype (B, C, D or E) (Petit et al., 1999; Tsutsui et al., 1995). The same toxin gene can be distributed on distinct plasmids and a plasmid can harbor several toxin genes. Indeed, *etx* can be found in at least five different plasmids ranging from 48 to 110 kb (Sayeed et al., 2007), and *cpb* is localized in a 65, 90, or 110 kb plasmid (Gurjar et al., 2010; Sayeed et al., 2010). Although *etx* and *cpb* are harbored by different plasmids, *etx* plasmids can contain *cpe* and *cpb2*, and *cpb* plasmids can also carry *cpe* and/or *tpeL*. But, *cpb* and *cpb2* are localized on distinct plasmids, as well as urease and lambda toxin (or lambda protease) genes, which are on separate plasmids than those containing the other toxin genes. A same *C. perfringens* strain can possess several plasmid types. For example, *C. perfringens* type B strains may contain until three virulence plasmids including *etx* plasmid, *cpb* plasmid with or without *tpeL*, and urease/lambda plasmid. Similarly, a single *C. perfringens* type D strain may possess a *etx* plasmid and additional plasmids containing *cpe* and *cpb2*, together or separately, or a unique plasmid type harboring the three toxin genes (Gurjar et al., 2010; Sayeed et al., 2007, 2010). Toxin genes on a plasmid are commonly flanked by insertion sequences and indeed, IS*1151* lies upstream from the epsilon toxin gene,

and immediately upstream of IS*1151* there is a transposase-like gene. Intriguingly, insertion sequences related to *Staphylococcus* or *Lactococcus* transposases are also located downstream of the epsilon toxin gene (Rood, 1998). Perhaps such sequences are a hint of gene sharing amongst a broad array of bacteria? IS*1151* is also found in the *cpb* plasmids and is probably involved in mobilization of toxin genes. In addition, the virulence plasmids of *C. perfringens* contain *tcp* genes, which are known to mediate conjugative transfer of plasmids (Gurjar et al., 2010; Sayeed et al., 2007, 2009, 2010).

Iota toxin, like the other clostridial binary toxins, consists of two independent proteins as exemplified by the enzymatic (Ia) and binding (Ib) components of iota. The two proteins are encoded by distinct, yet adjacent, genes. The gene for Ia is located immediately upstream from the Ib gene, and both are regulated by the same promotor. The iota toxin genes of *C. perfringens* type E are localized on a large plasmid, whereas the highly similar binary toxin genes of *Clostridium spiroforme* toxin (CST) and *C. difficile* transferase (CDT) are on the chromosome (Popoff, 2000). The less related *C. botulinum* C2 toxin is also on plasmid (Sakaguchi et al., 2009).

Therefore, most of *C. perfringens* toxin genes are mobilizable and are located on various plasmids, which can be horizontally transferred by conjugation accounting for the extreme diversity of toxinotypes in this bacterial species.

Despite their localization on mobile genetic elements, the genes of epsilon, beta, beta2 and iota toxins are highly conserved, at least in the strains whose the sequences are known. Interestingly, the *cpb2* gene of *C. perfringens* strains isolated from horses differs from that of strains from pigs by an adenine deletion downstream of the start codon resulting in a premature stop codon after only nine amino acid codons. Therefore, the equine strains do not produce beta2 under standard culture conditions. However, antibiotics of the aminoglycoside family such as gentamycin and streptomycin are able to induce expression of *cpb2* through a frameshift process. This impacts the role of antibiotics which can exacerbate some clinical symptoms. This was the case of gentamycin treatment, which lead to severe *C. perfringens* beta2 typhlocolitis in horses (Vilei et al., 2005).

2 – 2 *The variable genetic localization of* C. perfringens *enterotoxin gene*

Only 6 % of *C. perfringens* isolates contain the *cpe* gene and produce the enterotoxin (CPE). *Cpe* is either located on the chromosome or a large, 100-120 kb plasmid. Most of the enterotoxigenic *C. perfringens* type A strains involved in human food poisoning carry *cpe* on the chromosome, which is located on a 6.3 kb transposon-like structure (Tn*1565*) flanked by

the insertion sequences *IS*1470 and *IS*1469 (Rood, 1998). In contrast, *C. perfringens* type A isolates from human cases of non-food borne disease, or those from animals, possess *cpe* on a plasmid and this gene is also flanked by insertion sequences either *IS*1151 or *IS*1470-like elements on the 3' end (Miyamoto et al., 2002). However, recent investigations show that plasmid-borne *cpe* strains have also been identified in food poisoning outbreaks (Lindstrom et al., 2011). In *C. perfringens* type E, a silent *cpe* associated with *IS*1151 and *IS*1469 is located on a same plasmid, and in the same vicinity, as the iota toxin genes (Rood, 1998) (Billington et al., 1998; Gibert et al., 1997). In some atypical *C. perfringens* strains, functional *cpe* gene without flanking insertion sequences is located on large plasmid which also harbors variant iota toxin genes (Miyamoto et al., 2011). Cpe is also found on plasmids from other strains of *C. perfringens* types, such as D. The CPE amino acid sequence is relatively conserved among various strains, although the *cpe* gene can be located on a chromosome or plasmid. This further suggests that *cpe* can be mobilized by transposition, and indeed a circular transposon intermediate form has been discovered (Brynestadt and Granum, 1999).

C. perfringens genes for toxins involved in gangrene (mainly alpha toxin and perfringolysin) are located on the chromosome, whereas the toxins responsible for gastrointestinal and food-borne diseases are plasmid encoded. It is evident that the *cpe* gene can be mobilized between a plasmid and chromosome. Strains with chromosomal *cpe* are mainly of human origin, and strains containing plasmidic *cpe* or other toxin genes (*beta1, beta2, epsilon* and *iota*) are most often isolated from animals. In all the strains, the *cpe* gene is under the control of regulating sporulation genes (*spoOA, sigE, sigK*) and it is tightly expressed in a sporulation association manner. Plasmids probably contain additional genes responsible for host specificity (Petit et al., 1999), which naturally raises the following question. What are the selection pressures for maintaining plasmids and toxin genes in animal strains?

2 – 3 *Chromosomal variability in* C. perfringens

Chromosomal DNA of *C. perfringens* strains also show genetic variability as monitored by genome sequencing and multilocus sequence typing (MLST) based on house-keeping genes. Genome analysis of 3 *C. perfringens* strains representative of strains involved in gangrene and food poisoning has identified 323 genomic islands of various size, the G+C content of which is the same than that of the conserved genes (28%). Therefore, the genomic islands have probably been mobilized by horizontal transfer between *C. perfringens* strains or have been acquired from other *Clostridium* or related microorganisms exhibiting similar G+C content. They represent until 20% of the whole genome. From these 3 complete genomes, the

core genome includes 2126 genes (more than 80%). The largest genomic islands (243 kb) contain genes for ferrous iron transport, fucose utilization, glycosyl hydrolases, and restriction modification system. These genome islands, which display genes involved in metabolic pathways, probably confer selective advantages to colonize specific environment such as digestive tract or wound (Myers et al., 2006). Analysis of genome variability by MLST based on housekeeping genes and chromosomal toxin genes (alpha toxin and collagenase A genes) in a large number of strains from various origins has identified different phylogenetic lineages. Interestingly, *C. perfringens* strains harboring *cpe* on the chromosome from food poisoning or non outbreak connected food form a distinct cluster distantly related to plasmid *cpe* carrying strains. Strains harboring *cpe* on chromosome show different biological properties (higher resistance to heat, cold, nitrates, broader growth temperature range, faster growth at optimal temperature) compared to plasmid *cpe* strains. The different genetic backgrounds inducing distinct biological properties, may reflect stress adaptation to specific ecological niches. However, recent investigations show that plasmid-borne *cpe* strains have also been identified in food poisoning outbreaks. In addition, *pfo* gene is missing in all strains carrying *cpe* on chromosome and in only some plasmid *cpe* strains. It is noteworthy that *cpe* is highly conserved without any sequence variation, whatever its localization either on plasmid or chromosome. This might result from frequent interchange of *cpe* between chromosome and plasmid. In contrast, the genes encoding alpha toxin and collagenase, both located on chromosome, show genetic variation, mainly through recombination processes in parallel to variability of house-keeping genes. But, the evolution of the core genome is not homogeneous. Some genes show high levels of recombination, whereas others do not. It is not clear that such recombinations provide selective advantages (Deguchi et al., 2009; Lindstrom et al., 2011; Myers et al., 2006; Rooney et al., 2006).

3 - Toxin genes in *Clostridium difficile*

C. difficile produces potent cytotoxins, toxin A (TcdA) and toxin B (TcdB), which are part of the large clostridial toxin family also called large clostridial glucosylating toxins (LCGTs), and some strains produce an additional toxin, which is the binary toxin CDT. Toxigenic *C. difficile* strains are responsible for intestinal infections in humans (pseudomembranous colitis and about 25% of antibiotic-associated diarrhea) and to a lower extent in animals.

3 – 1 *Toxin gene organization*

Albeit some *C. difficile* strains harbor plasmids, toxin and other virulence factor genes are located on the chromosome of *C. difficile,* thus contrasting with the toxin gene localization in *C. perfringens*. For example, *C. difficile* strain 630 contains a 7.8 kb plasmid encompassing 11 genes of unknown function (Sebaihia et al., 2006).

TcdA and *tcdB* genes are clustered on a 19.6 kb DNA fragment, termed the pathogenicity locus (PaLoc). In non-toxigenic strains, the PaLoc is deleted and replaced by a short 115 bp sequence containing a direct repeat. The *C. difficile* chromosome exhibits only one single integration site for PaLoc that always occurs in the same orientation (Braun et al., 1996). This suggests that PaLoc might be transferred by a mobile element. However, PaLoc does not show any homology with transposon, phage, or plasmid-like sequence and does not contain any typical direct or inverted repeats (Braun et al., 1996).

TcdA and *tcdB* genes are similar in size (7 and 8 kb, respectively) and contain repetitive sequences within their 3' ends. They are transcribed in the same orientation and separated by a small gene designated as *tcdE*. The toxin genes are flanked by two other small genes, *tcdR* (formerly *tcdD* or *txeR*) in the 5' part, and *tcdC* in the 3' end, which is transcribed in an opposite direction (Rupnik, 2008; Rupnik and Just, 2006). *TcdR*, which is homologous to *botR* and *tetR* in *C. botulinum* and *C. tetani*, respectively, has been identified as an alternative sigma factor gene (Mani and Dupuy, 2001), and *tcdC* represents a putative inhibitory regulator of transcription that is perhaps an anti-sigma factor (Dupuy et al., 2008; Spigaglia and Mastrantonio, 2003). However, the role of *tcdC* in the control of toxin synthesis and strain hypervirulence (Carter et al., 2011) is controversial (Bouvet and Popoff, 2008; Cartman et al., 2012; Goldenberg and French, 2011).

The two genes (*cdtA* and *cdtB*) respectively encoding each protein of the binary toxin CDT are localized on a specific region of the chromosome called CdtLoc (4.2 kbp) (Carter et al., 2007; Perelle et al., 1997). They are preceded by a regulatory gene termed *cdtR*, which is also included in CdtLoc (Carter et al., 2007), but the regulatory function of this gene is controversial (Metcalf and Weese, 2011). Strains lacking CdtLoc contains a conserved 68 bp segment of unknown function (Carter et al., 2007). The mechanism of insertion/deletion of CdtLoc as well as PaLoc is still mysterious, since no genes or nucleotide sequences related to mobilization of conventional transposon, phage or plasmid lie in the proximity of these pathogenicity loci.

3 – 2 *Toxin gene variability and toxinotypes*

PaLoc possesses a great level of genetic variation, as toxigenic *C. difficile* strains produce either both toxins (A^+B^+), or only TcdB (A^-B^+). Deletions of varying lengths in the 3' part of *tcdA* gene account for the A^-B^+ strains. In contrast, no significant deletions have been found in *tcdB*. The 3' end of *tcdA* contains 30 conserved repeats divided into 5 classes, whereas the corresponding *tcdB* region is shorter and possesses less repeats (24) which can form homology groups only at the amino acid level. The small repeated DNA stretches in *tcdA* 3' part possibly support recombination events and subsequent deletions or insertions, while the non conserved DNA repeats in *tcdB* do not facilitate such recombinations (Rupnik, 2008). Various point mutations in *tcdA* and *tcdB* genes have been identified by restriction endonuclease analysis of PCR amplified fragments and DNA sequencing. These mutations are more frequent in *tcdB* than in *tcdA,* and in both genes they are mainly localized in the region encoding the catalytic domain. Genetic variation extends also to the accessory genes *tcdC, tcdE,* and *tcdR*. In particular, *tcdC* is highly variable and some genetic variants have been reported to be associated with increased toxin gene expression (Carter et al., 2011; Spigaglia and Mastrantonio, 2003). But this was not confirmed by other studies (Bouvet and Popoff, 2008; Cartman et al., 2012; Goldenberg and French, 2011). Four types of deletions as well as nonsense and conservative nucleotide mutations have been identified in *tcdC* resulting in 17 different *tcdC* genotypes (Curry et al., 2007). Therefore, *C. difficile* strains are classified into at least 24 toxinotypes, based on the genetic variations in the PaLoc region. Some toxinotypes are considered as majors since they show changes in almost all parts of both toxin genes *tcdA* and *tcdB*, whereas minor toxinotypes exhibit nucleotide modification limited to only one part of toxin genes (Rupnik, 2008). In addition, the insertion of a 1975 bp element with a combined feature of group I introns and insertion sequences containing two transposase genes has been evidenced in the *tcdA* gene. This element has the ability to splice, but not impair, the translation of functional *tcdA* (Braun et al., 2000).

Mutations in *tcdA* and *tcdB* genes lead to toxin variants, which differ in their biological activity. TcdB variants show 93-96% homology at the amino acid level through the whole toxin molecule. However, sequence variations can be higher in some regions impacting on the toxin activity. Thereby, $TcdB_{VPI10463}$ and $TcdB_{1470}$ share 95% homology in the entire molecule, but 78.6% in the catalytic domain. Both toxins recognize distinct patterns of intracellular targets. $TcdB_{VPI10463}$ glucosylates the main partners of the Rho family (Rho, Rac and Cdc42), and $TcdB_{8864}$ modifies Rac and Cdc42 (but not Rho) as well as Ras proteins (Ras, Rap, Ral, R-Ras). The mouse lethal activity of $TcdB_{8864}$ (6 ng/g) is higher than that of $TcdB_{VPI10463}$ (50 ng/g) (Chaves-Olarte et al., 2003; Chaves-Olarte et al., 1999). TcdB from

historical and recent epidemic PCR ribotype (RT)-027 varies from TcdB of other virulent strains such as $TcdB_{630}$ mainly on the C-terminal domain, which is involved in toxin binding to cellular receptor. TcdB from RT027 strains have more potent cytotoxic activity than their counterparts (Stabler et al., 2009).

Combinations of PaLoc and CdtLoc yield more variable genetic profiles of *C. difficile* strains: A+B+CDT-, A+B+CDT+, A-B+CDT-, A-B+CDT+, A+B-CDT-, A+B-CDT+, A-B-CDT-, A-B-CDT+ (Rupnik, 2008). The CdtLoc is most often associated with strains producing variant TcdA and or TcdB. Only rare A-B- strains produce the binary toxin. Numerous A+B+ strains contain a 2 kbp deletion in CdtLoc overlapping the 3' end of *cdtA* and the 5' part of *cdtB* and thus do not produce CDT (Rupnik et al., 2003; Stare et al., 2007; Stubbs et al., 2000). Except the 2 kbp deletion, only limited genetic variability by point mutations has been observed in CdtLoc in contrast to PaLoc, albeit both loci are localized on chromosome but in distinct regions. Indeed, single nucleotide polymorphism has been identified in CdtLoc, mainly in *cdtR*, and resulting in only few amino acid changes in CDTa and CDTb (Metcalf and Weese, 2011). Full length CdtLoc and CDT production are mainly observed in major variant toxinotypes, such as those which have significant altered PaLoc (Rupnik, 2008; Stare et al., 2007).

3-3 *Chromosomal variability in* C. difficile

The *C. difficile* chromosomal genome ranges from 4.9 to 4.4 Mb with a low G+C content (29%), 80-85% of which are coding sequences. The core genome based on the comparison of 14 genomes consists in 3.4 Mb containing 3,063 genes (Forgetta et al., 2011; Sebaihia et al., 2006). The differences in genome size corresponds to acquisitions or deletions of non core genomes sequences, most often due to mobile genetic elements. Analysis of strain 630 has revealed that high proportion of the genome (11%) consists of mobile elements including 7 putative conjugative transposons, one mobilisable transposon, and two prophage insertion sequenced (Sebaihia et al., 2006). Conjugative transposons have also been identified in other sequences *C. difficile* strains supporting the high plasticity of *C. difficile* genome (Brouwer et al., 2011). Numerous DNA regions show a high G+C content and mainly consist of mobile elements suggesting horizontal transfer from organisms phylogenetically distant. Indeed, many of these mobile genetic elements comprise antibiotic resistance genes some of them exhibiting similarity with sequences from *Enterococcus fecalis, Streptococcus pyogenes,* or *Thermoanaerobacter* species (He et al., 2010; Sebaihia et al., 2006).

Thereby, the *C. difficile* chromosomal genome is highly mobile and variable, through acquisition/loss of mobile genetic elements and also point mutations, which occur mainly by recombination rather than by *de novo* mutations (Dingle et al., 2011; He et al., 2010; Stabler et al., 2012). Indeed, *C. difficile* strains differ by multiple single nucleotide polymorphisms (SNPs). More than 120,000 SNPs have been identified through the core genome of 14 strains (Forgetta et al., 2011). Genetic analysis by MLST and typing methods (PCR-ribotyping, restriction endonuclease assay (REA), and pulse field typing) cluster the *C. difficile* isolates from various origin in 6 major phylogenetic lineages with multiple variations in each lineage (He et al., 2010; Knetsch et al., 2012; Stabler et al., 2012). The diverse disease-associated strains do not evolve from a unique ancestor but have rather emerged from multiple lineages, which have subsequently evolved independently. Analysis of collection strains indicate that variant *C. difficile* strains probably exist since a long period and have evolved independently. Sequence and typing methods indicate that the *C. difficile* population has a clonal structure and that PaLoc variability is largely associated with lineages based on house-keeping gene sequences (MLST). This suggests that PaLoc has been inserted into the genome once, since a unique genome location of PaLoc has been identified in all tested strains. Then, the strains of each lineage have diverged independently (Dingle et al., 2011). Coexistence of toxigenic and non-toxigenic strains in same lineages supports this genetic evolution of *C. difficile*. In addition, toxigenic and severe-disease-associated strains do not cluster in specific lineages, but have emerged from diverse lineages. Indeed, epidemic RT027 strains belong to lineage 2, RT078 strains to lineage 5, RT017 strains to lineage 4, and RT014 to lineage 1, which is the most heterogeneous, contains various severe-disease associated types (Dingle et al., 2011; Stabler et al., 2012). RT027 strains caused major outbreaks in North America around 2003 and in the UK in 2006. This clone (CD196), which produces the three toxins, TcdA, TcdB and CDT, was first isolated from a patient with pseudomembranous colitis in 1982 in France (Popoff et al., 1988). CD196, termed historic strain, and subsequent RT027 strains are genetically very similar and are identical by MLST (Stabler et al., 2012). They share 3,247 core genes with strain 630, which represents a classic toxigenic strain producing only TcdA and TcdB, but have 234 unique genes. Multiple copies of transposons spread in the genomes of RT027 and 630 strains with intergenic or intragenic localization, which can result in inactivation of the corresponding genes. Point mutations resulting in frame shifts or premature stop codons as well as inversions located upstream of coding sequences, which have been identified in these strains, might also affect gene expression (Stabler et al., 2009; Stabler et al., 2010). A sequenced epidemic RT027 strain (R20291) contains additional five genetic

regions compared to the historic CD196 strain including a unique ~20 kb phage island of high G+C content which carries a two-component response regulator, a putative lantibiotic ABC transporter, a putative cell surface protein and numerous other genes. A unique region in epidemic RT027 contains 6 genes of the multiple antimicrobial extrusion family drug/sodium antiporters with a high G+C content (Stabler et al., 2009). The additional genes in epidemic RT027 are supposed to contribute to the virulence of these strains. RT027 strains show higher motility than strain 630, which probably contributes to the intestinal colonization (Stabler et al., 2009; Valiente et al., 2012). Moreover, mutations in DNA gyrase gene (*gyrA)* have been identified in both the historic CD196 and epidemic RT027 strains. But only the epidemic RT027 strain exhibits a non-conservative mutation and is highly resistant to fluoroquinolones, in contrast to the historic strain (Stabler et al., 2009). Acquisition and loss of transposons as well as accumulation of mutations over time represent the main mechanisms of genetic changes in the emergence of *C. difficile* toxinotype variants.

Acquisition of virulence genes by horizontal transfer including genes encoding toxins, certain surface proteins and also enzymes involved in certain metabolism pathways allowing intestinal colonization, as well as subsequent genetic variation result in diverse pathogenic strains considered as epidemic strains or opportunistic pathogenic strains.

4 - Toxin genes in *Clostridium botulinum* and *Clostridium tetani*

4-1 - *The various botulinum neurotoxin-producing* Clostridium *species*

Clostridia capable of producing BoNT display heterogeneous bacteriological characters and are divided into several species and groups. The taxonomic position of the *C. botulinum* species was originally based on only one phenotype, the production of a BoNT, and non-toxic variant strains, although genetically related to *C. botulinum*, were assigned to different species such as *Clostridium sporogenes* and *Clostridium subterminale*. It appeared soon that BoNTs are seven different protein neurotoxins which are immunologically distinct and are termed with letters from A to G. More recently, sequence analysis permits to distinguish subtypes within BoNT types. However, all of them cause the typical flaccid paralysis of botulism in experimental animals, similar to that observed in human patients suffering from botulism. Later on, physiological differences between *C. botulinum* strains were identified, but the production of the different BoNT types does not necessarily correlate with the *C. botulinum* strain phenotypes. The species was divided into 4 physiological groups (Smith, 1975).

- group I : *C. botulinum* A, and proteolytic strains of *C. botulinum* B and F.
- group II : *C. botulinum* E, and glucidolytic strains of *C. botulinum* B and F.
- group III : *C. botulinum* C and D.
- group IV : *C. botulinum* G or *C. argentinense.*

The latter group which also includes non-toxic strains previously identified as *C. subterminale* and *C. hastiforme*, is metabolically distinct from the other groups and has been assigned to a different species called *C. argentinense* (Suen et al., 1988).

The taxonomic position of *C. botulinum* became more ambiguous since it has been found that BoNTs can be produced by *Clostridium* strains clearly distinct from already defined *C. botulinum,* and biochemically and genetically related to different species such as *C. butyricum* and *C. baratii* (Hall et al., 1985; McCroskey et al., 1986; McCroskey et al., 1991). In each group, all the strains regardless of toxin types, are closely related according to their phenotypic properties, DNA/DNA homology and 16s rRNA analysis. All the BoNT producing *Clostridia* have a G+C content between 26 and 28%. The atypical toxigenic *C. butyricum* and *C. baratii* strains are phenotypically and genetically related to the type strains of these species and not to the other BoNT producing *Clostridia* (review in (Minton, 1995; Oguma et al., 1999; Peck, 2009; Poulain et al., 2006).

4-2 - *Neurotoxin gene organization*

4-2-1 - The botulinum locus

BoNTs are produced by neurotoxigenic strains of *Clostridium* together with several associated non-toxic proteins (ANTPs). BoNTs and ANTPs associate to form large complexes, also known as progenitor toxins. ANTPs encompass a non-toxic non-hemagglutinin component (NTNH), and several hemagglutinin components (HAs) or OrfX proteins (Oguma et al., 1999; Popoff and Marvaud, 1999; Sharma et al., 2003). In botulinum complex, the proteins are not covalently linked but their association occurs in cultures and naturally contaminated food. The complex is stable at acidic pH, but dissociates at alkaline pH (≥ pH 7) (Eisele et al., 2011).

The neurotoxin and ANTP genes are clustered in close vicinity and constitute the botulinum locus. They are organized in two operons. The *bont* gene is immediately preceded by the gene of the NTNH component. Both genes form an operon located in the 3' part of the botulinum locus, which is well conserved in the different types of BoNT-producing Clostridia. The HA or Orfx genes, which are upstream to the *ntnh-bon*t genes, are transcribed in the opposite orientation and form the second operon. The HA operon (*ha33, ha17, ha70*

genes) is associated with *bontB, bontC, bontD, bontG,* whereas the OrfX operon (*orfX1, orfX2 and orfX3*) is linked to *bontA2, A3, A4, E* and *F* and a gene (*p*47) encoding a 47 kDa protein lies immediately upstream of the *ntnh* gene (Dineen et al., 2004; Jacobson et al., 2008a). *bont/A1* is the only gene which has been found in either HA or Orfx botulinum locus.

A gene, *bot*R, which encodes for an alternative sigma factor involved in the regulation of botulinum locus gene expression (Raffestin et al., 2005), is present in different positions in different strains of *C. botulinum*. In *C. tetani*, a gene *(tet*R) equivalent to *bot*R is present upstream of the tetanus toxin (*tent*) gene (Marvaud et al., 1998; Raffestin et al., 2005). No *antp*-like genes have been identified in *C. tetani* (Brüggemann et al., 2003). It is noteworthy that *bot*R is not present in botulinum locus containing the *bont/E* gene, in *C. botulinum* or *C. butyricum* type E.

Usually, one clostridial strain produces only one type of neurotoxin and the botulinum locus is present in a single copy on the genome. However, some strains synthesize two different BoNTs such as BoNT/A-BoNT/B, BoNT/A-BoNT/F, and BoNT/B-BoNT/F. The bivalent strains contain two botulinum loci. For example, the AB strain contains two *botulinum* loci exhibit OrfX-BoNT/A1 and HA-BoNT/B loci, which are 40 kbp distant on the chromosome. In such strains, the two neurotoxins are usually produced in different proportions. Thus, in Ba and Bf strains, BoNT/B is produced ten times more, than BoNT/A and BoNT/F (Henderson et al., 1997; Hutson et al., 1996). Some clostridial strains contain silent neurotoxin genes. Indeed, several *C. botulinum* A strains isolated from food borne and infant botulism contain a silent *bont/*B gene and are noted A(b). The *bontB* nucleotide sequence is related to that of *C. botulinum* B strains (97% identity), but it has a stop mutation in position 128 and base deletions resulting in reading frameshifts and multiple stop codons (Dineen et al., 2004; Hutson et al., 1996; Jovita et al., 1998). Silent *bont*/B has also been evidenced in non toxigenic *C. subterminale* strains (Franciosa et al., 1994; Jovita et al., 1998).

4-2-2 - Genomic localization of the botulinum locus

The genes encoding for the different types of BoNT are present on different genetic elements, including phages, plasmid or chromosomes depending on the species and strain of Clostridia. In *C. tetani* and *C. argentinense*, the neurotoxin gene is present within a large plasmid (51 and 76 MDa respectively). Plasmids of various sizes and bacteriophages have been found in *C. botulinum* A, B, E, and F and previous works have shown that toxigenicity was not associated with the presence of these genetic elements (Eklund et al., 1989; Strom et al., 1984; Weickert et al., 1986). Therefore, the genes encoding for these neurotoxins were

assumed to be located on the chromosome. However, it has been recently found that in some strains such as Loch Maree strain (subtype A3), 657Ba (type Ba and subtype A4), Okra (type B1), and Eklund 17B (type Bnp) the botulinum neurotoxin genes are harbored by large plasmids (47 to 270 kb) (Marshall et al., 2007; Smith et al., 2007). Plasmid location of neurotoxin genes seems common in *C. botulinum* type B strains, mainly in subtype B1, bivalent, and non proteolytic strains (Franciosa et al., 2009). In the bivalent strain Ba657 the two botulinum locus, locus A and locus B, are located on the same plasmid and are separated by approximately 97 kbp. Similarly, the neurotoxin genes, *bont/B* and *bont/f*, from one Bf strain are located on a same plasmid (pBf), which is very related to pCLJ from 657Ba strain (Hill et al., 2009). Interestingly, none of botulinum plasmids show synteny to *C. tetani* plasmid pE88, which contains the *tent* gene (Hill et al., 2009). In *C. botulinum* type E and neurotoxigenic *C. butyricum* strains, the location of the BoNT/E encoding gene is located on the chromosme. In *C. botulinum* C and D, it has been clearly evidenced that BoNT is encoded by bacteriophages (reviewed in (Poulain et al., 2006)).

The location of botulinum locus within chromosome or plasmid seems to occur not at random but at specific sites. Indeed, in strains from group I or II, whose genome sequencing is available, three specific sites of botulinum locus integration have been identified. OrfX-A2, orfX'-A1, and orfX-F locus are located in the *ars* operon, which contains 3 to 5 genes involved in arsenic reduction. OrfX'-A1 and orfX-F locus share a similar integration site at the 5' end of the *ars* operon, whereas OrfX-A2 locus is inserted between two copies of *arsC* gene. HA-A1 and HA-B locus, which contain a recombinant *ntnh* gene type A and type B strains, are found in the *oppA/brnQ* operon encoding for extracellular solute binding protein and branched chain amino acid transport proteins, respectively. This operon is lacking in non-proteolytic *C. botulinum* type B, *C. botulinum* type E, and *C. butyricum* type E strains. The third integration site is the *rarA* gene, which contains the OrfX-E locus in *C. botulinum* type E and *C. butyricum* type E strains. *rarA* encodes a resolvase protein involved in recombination or insertion events of transposons. Interestingly, the botulinum E locus is inserted in the same codon (102) of *rarA* gene in both *C. botulinum* type E and *C. butyricum* type E strains, and the inserted botulinum locus contains an additional intact *rarA* gene (Hill et al., 2009)

Two specific sites of botulinum locus location have been identified on plasmids from group I strains, one contains OrfX-A3, OrfX-A4 from Ba strain, or OrfX-F from Bf strain, and the second harbors the HA-B locus from *C. botulinum* B1 strain or bivalent Ba4 or Bf strains. The HA-non-proteolytic B locus is located on a plasmid different from those of group

I strains. However, the downstream flanking region of the HA-npB locus contains an IS element, a transposon-associated resolvase, and a site specific recombinase (Hill et al., 2009).

4-3 - *Genome characteristics of botulinum neurotoxin-producing* Clostridia

The complete genome sequences of 10 *C. botulinum* strains from group I and II and one from group III have been solved so far. *C. botulinum* genomes from group I and II consist of a circular DNA chromosome, the size of which (3 659 644 to 4 155 278 bp) is in the same range than that of other known *Clostridium* genomes (*C. perfringens* 3 031 430 bp; *C. difficile* 4 290 252 bp; *C. acetobutylicum* 3 940 252 bp) except those of one type C/D strain from group III (2 773 191 bp) and *C. tetani* (2 799 250 bp) and which are considerably smaller. The G+C content (27-28%) is similar to that of most clostridial species. Most strains also possess plasmids which vary in size (16 kbp to 270 kbp).

The *C. botulinum* A (strain Hall) genome (Sebaihia et al., 2007) shares 1126 (31%) to 1565 (43%) genes with the genomes of *C. acetobutylicum*, *C. perfringens*, *C. difficile*, and *C. tetani*, confirming the heterogeneity of the *Clostridium* genus. *C. botulinum* A shares a larger number of orthologous genes with *C. acetobutylicum* (43%), a non-toxigenic *Clostridium*, than with the other toxigenic Clostridia. Among the sequenced toxigenic Clostridia, the most related to *C. botulinum* A is *C. tetani* (40% orthologous genes) which also synthesizes a neurotoxin related to botulinum neurotoxin; the most unrelated is *C. perfringens* (31% orthologous genes).

The shared genes between *C. botulinum* and other Clostridia mainly encode basic functions such as cell division, macromolecule biosynthesis, central/intermediary metabolism and germination/sporulation. The *C. botulinum* unique genes are spread all around the genome and are not clustered in specific regions. They seem to be involved in accessory functions including transport/binding proteins, energy metabolism, cell surface proteins, and regulation (Sebaihia et al., 2007).

In contrast to the *C. difficile* genome (strain 630) which shows a mosaic organization with multiple horizontal gene acquisitions mainly via mobile elements such as conjugative transposons (11% of the genome) (Sebaihia et al., 2006), *C. botulinum* genomes have a more homogeneous composition without evidence of recent gene acquisition. Only two prophages, two prophage remnants, one functional and 11 non-functional transposase genes were found in the genome of *C. botulinum* A ATCC3502 (Sebaihia et al., 2007), and two prophages were evidenced in *C. botulinum* A ATCC19397. A relatively low proportion (0.8 to 6.8%) of genes are predicted (Colombo prediction) to be acquired horizontally in *C. botulinum* genomes.

However, this portion is significantly higher in group I strains (4.5 to 6.8% of all genes) than in group II strains (0.8 to 1.2%).

An overall comparison between complete *C. botulinum* genomes revealed high similarity between genomes from group I strains, and their strong distance to group II strains. Four to six large variable genomic regions exist which are scattered around the genomes. These are absent in some strains or show decreased similarity. Whereas most proteins (81--86%) of group I strains share a protein identity of over 90%, only 2-3% of proteins from group II genomes have orthologs encoded in group I genomes with such a high protein identity. Applying a less stringent cutoff (10% protein identity), group I and group II strains share only 48% of their proteome. In addition, there is almost no genome-wide synteny between group I and group II strains, except a region around the origin of replication. The fact that the genomes of 6 strains of group I (three *C. botulinum* A1, one *C. botulinum* A3, one proteolytic *C. botulinum* B1, and one proteolytic *C. botulinum* F) organisms are highly similar confirms previous genetic analysis indicating that group I *C. botulinum* strains form a homogeneous bacterial species distinct from the other *C. botulinum* types (reviewed in (Brüggemann et al., 2011)). This was further confirmed in a set of 5 *C. botulinum* A strains by comparative genomic hybridization microarrays and pulse-field gel electrophoresis (Raphael et al., 2008). However, these strains show limited genomic rearrangements in the botulinum locus. A multiple-locus variable-number tandem-repeat analysis was designed in order to differentiate *C. botulinum* A strains within subtypes. Ten variable-number tandem-repeat (VNTR) regions dispersed throughout the *C. botulinum* A ATCC3502 genome, which are distant from the botulinum locus, permitted to differentiate the strains into subtypes A1-A4 and to distinguish 30 genotypes within 53 strains of subtype A1 (Macdonald et al., 2008). A multiple sequence typing (MLST) based on 7 housekeeping genes, has also been proposed for *C. botulinum* A subtyping and phylogenetic analysis. Thus, 24 lineages were identified from 73 *C. botulinum* A strains (Jacobson et al., 2008b). Chromosome organization of group I *C. botulinum* A, B, and F strains is also shared by *C. sporogenes* (Hill et al., 2009), confirming that this *Clostridium* species is very close from proteolytic *C. botulinum* strains except the absence of neurotoxin gene.

Two *C. botulinum* E strains (Beluga E1 and Alaska E3) and the non-proteolytic *C. botulinum* B Eklund17B share chromosome synteny, indicating that strains from group II contain a related chromosomal background, which is distinct from that of *C. butyricum* type E (Hill et al., 2009).

Plasmids are common in *C. botulinum* with sizes ranging from 16344 to 270346 bp and containing 19 to 329 genes. Plasmids from strains Loch Maree, Okra, and Ba657, although of different sizes, share large regions that are very similar to each other (Smith et al., 2007). However, they are poorly related to plasmids that do not contain neurotoxin genes. They can replicate independently, since they contain complete DNA polymerase III complex enzymes and DNA helicase II (Smith et al., 2007).

Genomic synteny of one representative type C strain and one representative type D strain indicates a related chromosome organization between these two strains from group III (Brüggemann et al., 2011). Chromosome organization is conserved in a type C/D strain, the complete genome sequence of which is available, and in group III isolates. Interestingly, the genome sequences of group III strains are very different from those of group I and II strains, but they are closely related to that of *C. novyi* and *C. haemolyticum* (Skarin et al., 2011). Botulinum C2 toxin, which is a binary toxin involved in actin filament depolymerization and produced by *C. botulinum* of group III (Barth et al., 2004), is encoded by genes located on a large plasmid in *C. botulinum* type D or C (Sakaguchi et al., 2009). Genomes sequencing of *C. botulinum* type D strain 1873 shows that this strain contains two plasmids, one pCLG1 harboring the genes for the enzymatic component and binding components (BCs) of C2 toxin, and the other, pCLG2, containing genes for clostripain (a cysteine protease) and a thermolabile hemolysin. Five different plasmids have been characterized in a type C/D strain and its genome contains numerous mobile elements indicating an important genome remodeling by transposition (Skarin et al., 2011). In addition, *C. botulinum* strains from group III contain prophages, which harbor the neurotoxin gene and the C3 exoenzyme gene.

4-4 - *Genetic diversity of botulinum neurotoxin-producing* Clostridium *strains and botulinum neurotoxin gene variation*

Genetic analysis by 16s RNA gene sequence comparison or DNA/DNA homology have shown that *C. botulinum* strains form four distinct clusters which correspond to the physiological groups I to IV (Collins and East, 1997; Hutson et al., 1994). Amplified fragment length polymorphism (AFLP) and pulsed-field gel electrophoresis (PFGE) analysis also confirms the classification of proteolytic types A, B and F strains in group I and the non proteolytic types B, E and F strains in group II, but can differentiate individual strains into each group (Hielm et al., 1998a; Hielm et al., 1998b; Hill et al., 2007; Hyytia et al., 1999a; Hyytia et al., 1999b; Keto-Timonen et al., 2005; Nevas et al., 2005). These methods have been used in epidemiological studies and are useful tools to investigate relatedness between

strains isolated from patients and food. For example, among proteolytic *C. botulinum* strains, PFGE analysis differentiates the toxinotypes A, B, and F at a 83-86% similarity level, and enable discrimination of most of individual strains. A greater diversity was observed between type A strains than in type B strains (Nevas et al., 2005). These studies also indicate that each *C. botulinum* group is heterogeneous at the genome level.

A high level of similarity was observed between strains from group I by using DNA hybridization with a DNA microarray including 94% of the coding sequences from strain Hall. Two type A strains share 95-96% of the strain Hall coding sequences, and seven other proteolytic strains have 87-91% common coding sequences (Sebaihia et al., 2007). A larger investigation reports that 58 *C. botulinum* strains from group I share 63% of coding sequences with those of strain ATCC3502 (Carter et al., 2009). Interestingly, two *C. sporogenes* strains (physiologically related to *C. botulinum* group I but non-toxigenic) are significantly similar to Hall strain and share 84-87% of the coding sequences (Sebaihia et al., 2007). In another microarray study, three *C. sporogenes* strains show approximately 63% common coding sequences with *C. botulinum* A ATCC3502 (Carter et al., 2009).

BoNT gene has been sequenced from a large number of strains and sequence comparison has permitted to identify sequence variations in each toxinotype. Thereby, botulinum toxinotypes are divided into subtypes, which are defined as toxin sequences differing by at least 2.6% identity at the amino acid level (Smith et al., 2005). BoNT genes from type A strains show 92 to 95% nucleotide identities corresponding to 84-90% amino acid identities and are divided into four subtypes termed A1, A2, A3, A4, and A5. Subtypes A1 to A5 also differ in the botulinum locus composition. Type B genes differ from 2 to 4% at the nucleotide level and 3 to 6% at the amino acid level. They are classified in five subtypes, B1, B2, B3, bivalent B, and non proteolytic B. BoNT genes from non proteolytic type B strains form only one subtype, whereas those from proteolytic strains show a greater variation leading to a four subtype division. Sequences of neurotoxin genes type B show an overall less variation than those of type A, but a more sequence variation is observed within members of each type B subtype compared to bont/A. BoNT/E sequences from *C. botulinum* type E (group II) fit in 9 subtypes (E1, E2, E3, E6, E7, E8, E9) sharing 99% nucleotide identity and 97-99% amino acid identity, and are more distantly related to BoNT/E sequences from *C. butyricum* strains which are into two subtypes (E4, E5) with 97-98% nucleotide and 95-96% amino acid identities between sequences from both *Clostridium* species. Gene diversity has also been evidenced in the other parts of the genome as tested by MLST and AFLP analysis, but most of *C. botulinum* E strains are conserved in a same clade. Subtype variation in *C.*

botulinum E srains seems to result from recombination events rather tha random mutations. High differences (up to 25%) have been found in nucleotide sequences of BoNT/F mainly in the region coding the light chain, and 5 subtypes have been identified in proteolytic *C. botulinum* F. The low number of strains type C, D, F and G which were analyzed does not permit to significantly evaluate the neurotoxin gene diversity. In group III, mosaic genes between BoNT genes types C and D can be distinguished from classical types C and D strains. BoNT/F sequences from *C. botulinum* type F form a different cluster of those from *C. baratii* (Carter et al., 2009; Chen et al., 2007; Hill et al., 2007; Macdonald et al., 2011; Raphael et al., 2010; Raphael et al., 2012).

Signification of sequence diversity in each toxinotype is not yet well known, but could be important in diagnostic tests and development of therapeutic agents such as those based on immunotherapy. Thereby, BoNT/A1 and BoNT/A2, which differs by 10% at the amino acid sequence level, show large differences in monoclonal antibody-binding affinity. Among six monoclonal antibodies, which bind to BoNT/A1 with high affinity, three show a marked decrease in binding affinity (500 to more than 1000 fold) to BoNT/A2. Only combinations of monoclonal antibodies, which tightly bind to toxin subtype, potently neutralize the corresponding toxin *in vivo.* Association of the three monoclonal antibodies with high affinity binding to subtypes A1 and A2, completely neutralizes A1 or A2 toxin, while replacement of two from three monoclonal antibodies by two having a low binding affinity to BoNT/A2 induces a decrease in BoNT/A2 neutralization (50 fold less) (Smith et al., 2005). The impact of subtype variation in binding and neutralization potency of polyclonal antibodies remains to be determined. Thus, development of therapeutic polyclonal or monoclonal antibodies as well as vaccines based on single toxin subtype, needs to be evaluated for their protection ability with the other related subtypes. Although two toxins show a low level of sequence difference, they can have marked difference in activity if amino acid variation occurs in strategic toxin sites. Subtypes A1, A2, A3, and A4 of BoNT/A have been analyzed by sequence comparison, as well as molecular modeling and structure comparison with the crystal structure of subtypes, the impact of which is not known. Ganglioside binding site is conserved in all subtypes of BoNT/A. The greatest variability was found in the light (L) chain, mainly between subtypes A3 and A4 (76% identity). The enzymatic site of L chain is conserved, but non-conservative mutations are observed in domains involved in substrate (SNAP25) recognition. When compared to subtypes A1 and A2, subtypes A3 and A4 show sequence variation in α-exosite and S1' subsite recognition, respectively, suggesting that these subtypes have a decreased affinity and catalytic efficiency for their substrate (Arndt et al., 2006). Indeed, L chains from

subtypes A3 and A4 show different catalytic properties of the substrate SNAP25 compared to L chain from subtypes A1 and A2, which show the same catalytic activity, although all L chain isoforms bind SNAP25 with similar affinity. L chain from subtype A4 and to a lower extent from subtype A3, cleaves less efficiently SNAP25 than L chain subtype A1 (2 and 23 fold less, respectively) (Ahmed et al., 2001; Henkel et al., 2009). Another example of gene variation and toxin activity difference is given by neurotoxins type B. BoNT/B from strain 111 (subtype B2) isolated from infant botulism differs from strain Okra/NT associated with food-borne botulism in Japan by 56 amino acid changes (95.7% identity), from which most occur in the half C-terminal part of the toxin (Ihara et al., 2003). BoNT/B from strain 111 shows an about 10 fold lower specific activity than that of strain Okra/NT, and most of monoclonal antibodies which recognize the C-terminus of Okra/NT BoNT/B do not react with BoNT/B of strain 111. Binding affinity of BoNT/B of strain 111 to the receptor synatotagmin II in the presence of ganglioside GT1b is 4.2 lower than that of Okra/NT BoNT/B. Mutations of 23 residues in the C-terminus of BoNT/B of strain 111 have been attributed to the lower binding affinity of the toxin to its receptor and thus to the lower specific toxicity (Ihara et al., 2003; Kozaki et al., 1998).

Sequence comparison of *bont* genes suggests that they have evolved separately in different genomic backgrounds (Hill et al., 2007). BoNT genetic diversity could also reflect a different geographical distribution of strains or their involvement in different epidemiological situations. *C. botulinum* subtype A2 was first identified in infant botulism in Japan and was found to differ from strains involved in food-borne botulism in adults referred as subtype A1 (Tabita et al., 1991; Willems et al., 1993). However, no correlation was evidenced between strains subtype A1 and A2 isolated from the United States and UK and their clinical origin, food-borne or infant botulism (Cordoba et al., 1995; Johnson et al., 2005). But, strains A1 are more prevalent in the United States, whereas subtype A2 strains are commonly isolated in Europe. Indeed, all the strains from food-borne botulism in the United States which have been analyzed fall into subtype A1, and all 33 *C. botulinum* type A isolated from Italy belong to subtype A2 as well as two strains from infant botulism in United Kingdon (Franciosa et al., 2004; Hill et al., 2007; Johnson et al., 2005). But 18 *C. botulinum* strains type A isolated in France or Europe by Prevot during the period approximately 1950-1960 are of subtype A1 (Hill et al., 2007). Divergent strains of subtype A2 characterized by 5 amino acid differences in BoNT/A2 and a slightly different botulinum locus organization (locus A2-OrfX') with a shorter intergenic region between *orfX1* and *botR/A* genes (77 versus 1228 nucleotides) when compared to strain A2 Kyoto-F, have been identified in Italy such as strain associated with

consumption of contaminated cheese (Mascarpone) (Franciosa et al., 2006). Organization of botulinum locus of strain Mascarpone is closely related to that of locus containing *bont/A1* in strain type A(B) NCTC2916 (Dineen et al., 2003; Henderson et al., 1996). Strains Mascarpone and Kyoto-F have probably a common origin and then a distinct evolution including a gene rearrangement in strain Mascarpone with an ancestor of strain NCT2916. Four *C. botulinum* A5(B) strains were isolated from wound botulism in heroin users in UK and one from infant botulism in California (USA) supporting a *bont* gene evolution independent of the geographical location and epidemiological situation (Carter et al., 2009; Dover et al., 2009).

Genetic diversity is also observed in the two FGI regions. Six profiles have been evidenced in 58 proteolytic strains from group I by DNA microarray which correlate with the diversity of flagellin glycan composition as determined by mass spectrometry. The FGI genetic diversity does no match with that of the botulinum locus, indicating an independent evolution of FGI and botulinum locus genes in a relatively stable genomic background of group I *C. botulinum* strains (Carter et al., 2009). In addition to cell wall and surface structure variations, a marked difference in proteolytic strains of group I consists in resistance to toxic compounds. Thereby, group I *C. botulinum* type B strains representative of strains found in North Europe are divided in two clusters BI and BII which differ by 413 coding sequences but contain a same neurotoxin gene of B2 subtype in a HA locus. In contrast to cluster BI strains, cluster BII strains are more resistant to arsenic and more sensitive to cadmiun. Moreover, strains from the two clusters show other differences in metabolism, such as cluster BII strains growing at lower temperature than cluster BI strains (Hinderink et al., 2009; Lindstrom et al., 2009). This suggests a differential evolution of these environmental clostridia in response to adaptation to distinct ecological niches.

Another differential genetic evolution is illustrated by neurotoxigenic *C. butyricum* strains. *C. butyricum* strains producing type E neurotoxin have been isolated from infant botulism and young people in Italy (Fenicia et al., 1999; McCroskey et al., 1986), as well as from botulism associated with consumption of fermented soybean in China (Meng et al., 1997; Meng et al., 1999; Wang et al., 2000). Based on toxin gene sequence, toxigenic *C. butyricum* strains from Italy and China are divided in two distinct subtypes, termed E4 and E5, respectively, indicating an independent evolution of *bont/E* gene after transfer in *C. butyricum* (Hill et al., 2007).

4-5 *Neurotoxin gene transfer*

The similarity among the different BoNTs as well as with tetanus toxin (TeNT), the fact that different *Clostridium* species can produce BoNT and that some strains contain combinations of *bont* genes (Franciosa et al., 1997; Henderson et al., 1997), strongly suggest that *bont* and *tent* genes derive from a common ancestor and have been transferred between *Clostridium* strains. Bivalent strains producing Ab, Ba, Af, or Bf toxins (Barash and Arnon, 2004; Franciosa et al., 1994; Henderson et al., 1997; Santos-Buelga et al., 1998), as well as strains from type A containing an additional botulinum locus with silent *bont/B* gene, indicate a horizontal gene transfer. Indeed, strains subtypes A2 or A4 with silent *bont/B* gene have the same genomic background, as tested by amplified length polymorphism, suggesting either a horizontal transfer of *bont/B* gene in strains already having *bont/A2* or *bont/A4* genes or that two strains with *bont/B* gene acquired *bont/A2* or *bont/A4* horizontally (Hill et al., 2007). In addition, non toxigenic derivatives are present in certain toxinotypes such as in *C. botulinum* B (Yamakawa et al., 1997), indicating an instability of a DNA fragment harboring the *bont* genes. Genetics in *Clostridium* is yet poorly understood, but it can be assumed that toxin gene transfer has been probably mediated by genetically mobile elements (Minton, 1995). Horizontal gene transfer also concerns other genes than neurotoxin genes. Indeed, analysis of chromosomal sequencing indicates that 4.5 to 6.8% of genes have been probably acquired by horizontal gene transfer in group I *C. botulinum* strains, whereas these genetic events seem less frequent in group II strains.

4.5.1 Plasmids

Since the *bont*/G genes have been localized on large plasmids in *C. argentinense*, as well as *bont/*A and *bont/*B in some *C. botulinum* A and B strains, their transfer can be achieved by mobilization of the corresponding plasmids. Conjugation and mobilization of large plasmids in *Clostridium* such as *C. perfringens*, have already been reported (Brefort et al., 1977). Conjugative plasmids between strains from group I and II have been evidenced (Marshall et al., 2010). Multiple plasmids have been characterized in *C. botulinum* group III strains and some of them contain related toxin genes from other Clostridia such as alpha-clostripain, alpha toxin of aerolysin family, binary toxin, and epsilon toxin (Skarin and Segerman, 2011). Therefore, plasmids can be mobilized between strains of intra- or inter-*Clostridium* species and can be rearranged (deletion, insertion, recombination, transposition). They represent a major horizontal transfer vehicle of toxin genes.

4.5.2 Bacteriophages

Bacteriophages mediate the neurotoxin gene transfer in group III of *C. botulinum*. A pseudolysogenic relationship corresponding to the presence of bacteriophages free within the bacterial cytoplasm, exists between these phages and hosts. Thus, variants free of bacteriophages can be obtained with high frequency using curing reagents such as acridine orange and U.V. light. Under laboratory culture conditions, a proportion of the bacteria, which depends on the strain and growth conditions (temperature, salinity), are lysed and lose free bacteriophages which can reinfect them. Such lysogeny and reinfection cycles occur probably in the environment (soil, intestinal tract of birds and animals) and account for isolation of non toxigenic or low toxin producer variants (Eklund and Dowell, 1987). *C. botulinum* C and D strains cured of their phages do not produce BoNT/C or BoNT/D, respectively, while they continue to produce another botulinum toxin, termed C2 toxin, which is not neurospecific and whose encoding gene is chromosomally located. Such bacteria could be converted into neurotoxigenic strains C or D by reinfection with phages obtained from toxigenic *C. botulinum* C or D strains. In addition, BoNT/C1 and BoNT/D genes can be transferred in different *Clostridium* host strains and they determine the toxinotype of these strains (Eklund and Poysky, 1974; Eklund et al., 1974; Eklund et al., 1971).

The genome sequence of *C. botulinum* type C1 neurotoxin-converting phage from strain C-ST was determined (Sakaguchi et al., 2005). It consists on 185682 bp with 404 bp terminal repeats, and 198 potential coding sequences including BoNT/C1 and exoenzyme C3. Among the 198 coding sequences, 57 are related to genes encoding known functional proteins, 30 to genes with unknown function, and 102 have no sequence homology with any proteins in databases. Two genes have been identified to code the main structural proteins of phage sheath and head, which are conserved in two other type C phages and one type D phage but not in a phage lacking the botulinum locus. Two other genes also produce structural proteins (Hwang et al., 2007).

Molecular analysis of the phage genome revealed that although this phage is linear DNA with 404 bp tandem repeats, it exists as a circular plasmid prophage in the lysogenic bacteria and does not integrate into the host chromosome. This mode of lysogenization appears to be related to the unstable lysogeny of BoNT phages.

A notable feature of C-ST phage is the abundance of IS elements, 12 copies 7 of which structurally intact, representing about 10% of the phage genome. The presence of IS elements is uncommon in bacteriophages from other bacteria, and can be partially involved in the prophage instability in *C. botulinum* (Sakaguchi et al., 2005). Comparison of C-ST phage

with those from other *C. botulinum* C and D strains by PCR analysis shows a high level of diversity and a mosaic structure of these phage DNAs. IS elements might have been partially responsible for this genetic diversification. Four regions containing the putative replication origin, the terminus replication origin, the botulinum type C locus and C3 enzyme, respectively, are well conserved in three type C and D phages (Sakaguchi et al., 2005). Interestingly, the flanking regions of the botulinum locus are also conserved and suggest that they have been included in DNA transfer between phages by a yet undefined mechanism.

In *C. botulinum* A and F, the involvement of bacteriophages has been suggested on the basis of the identification of a gene (*lyc*) in the vicinity of the *bont* genes. *lyc*/A and *lyc*/F genes have been mapped approximately 1 kb downstream from the corresponding *bont* genes, and are partially related to bacteriophage genes encoding lytic enzymes in *Lactobacillus* and *Streptococcus pneumoniae* (Henderson et al., 1997). Since lytic enzymes participate in the bacteriophage life cycle, the presence of *lyc* gene in the vicinity of *bont*/A gene in NCTC2916 could indicate that the botulinum locus is part of an integrated prophage. Genes encoding two prophages have been identified in the genome of strain Hall, but they are not conserved in two other *C. botulinum* type A1 strains (Sebaihia et al., 2007).

4.5.3 *Transposons, insertion sequences, recombination*

Transposable elements have been identified in *C. botulinum* C and D. The exoenzyme C3 gene which codes for an ADP-ribosyltransferase specific of the eukaryotic Rho protein, is also harboured by the same bacteriophages carrying *bont*/C1 and *bont*/D genes. The transformation defective mutant phage CN shows a deletion of a 21.5 kb fragment containing the C3 gene. This fragment is present in several bacteriophages isolated from type C and type D bacteria, and it is marked by a 6 bp core motif AAGGAG. The DNA sequence on its left end is homologous among the C and D types, whereas the sequences diverge immediately downstream of the core motif. The 21.5 kb fragment appears to be a mobile DNA element responsible for the spreading of the C3 gene in *C. botulinum* C and D, and it has features similar to the site specific Tnl554 transposon family, including (i) asymmetric ends, (ii) absence of either inverted or terminal repeats, and (iii) presence of a 6 bp core motif sequence at both insertion junctions and the insertion site itself (Hauser et al., 1993).

Transposable elements analogous to that containing the C3 gene, could account for the different localization of the neurotoxin genes (chromosomal, plasmid, bacteriophage) and subsequently for the gene transfer between *Clostridium* strains. Such elements have not yet been clearly identified. However, nucleotide sequence analysis in *C. botulinum* A suggests

their possible existence. A 97 nucleotide stretch downstream of the stop codon of the *bont*/A gene is identical in *C. botulinum* A strains 62 and NTCT2916, whereas the following nucleotides are totally unrelated between both strains. The 97 nucleotide stretch could be part of a mobile DNA element encompassing the *bont*/A and *antp* genes. In addition, the different surrounding sequences indicate that *bont*/A gene has different localizations on the chromosome of strains 62 and of strain NCTC2916. This is confirmed by a different restriction profile of the 3' part of *bont*/A gene in the two strains (Popoff and Marvaud, 1999). The genomic sequence of *C. botulinum* A strain Hall shows that the *botulinum* locus is flanked by two insertion sequences in the 5' part and by a transposase gene in the 3' part (Bennik et al., 2003). In addition, the upstream regions of *bont*/A and silent *bont*/B loci in strain NCTC2916 share marked homology indicating a similar evolutionary origin, and insertion sequences with multiple internal mutations were identified downstream of the *bont*/A locus (Dineen et al., 2003). In plasmids of strains Loch Maree, Okra and Ba657, the botulinum locus seem to have been acquired or deleted by means of insertion sequences flanking these DNA regions (Smith et al., 2007). However, most of insertion sequences found in the flanking regions of certain botulinum loci are incomplete, indicating that gene mobilization by these elements is a very old event.

Neurotoxigenic *C. butyricum* strains probably originated from non toxigenic *C. butyricum* strains by acquisition of *bont/E* gene from *C. botulinum* E (Poulet et al., 1992). The *bont/E* gene and its flanking regions are absent in non toxigenic *C. butyricum* strains, suggesting a possible gene transfer mediated by a mobile DNA element (Hauser et al., 1992). This gene has been transferred from a neurotoxigenic *C. butyricum* strain to a non-toxigenic *C. botulinum* E strain following a protocol resembling transduction with a defective phage (Zhou et al., 1993). The precise mechanism of molecular transfer has not been elucidated since DNA/DNA hybridization studies suggest that the *bont/E* gene is localized on chromosomal DNA and not on phage DNA in toxigenic *C. butyricum* (Zhou et al., 1993). Recently, genome sequencing of three *C. botulinum* type E strains and one toxigenic *C. butyricum* strain shows that the *orfX-E* locus is inserted in the chromosome of both bacterial species and inside a same gene, *rarA* gene, and at the same site (codon 102). But the split *rarA* gene in *C. botulinum* type E and *C. butyricum* type E are not identical, suggesting separate events of neurotoxin gene insertion in different genomic backgrounds. The presence of an additional intact *rarA* gene encoding a resolvase involved in recombination or insertion events of transposon and the presence of insertion sequences flanking the *orfX-E* locus support a horizontal gene transfer by IS elements and/or transposon (Hill et al., 2009).

In addition, rearrangement, probably by homologous recombination, appears to have occurred between genes of the botulinum locus of different types of neurotoxigenic bacteria. This is suggested by the identification of atypical strains carrying mosaic genes containing elements derived from *C. botulinum* C and D (Moriishi et al., 1996a), or mosaic *ntnh* genes from *C. botulinum* A and C (Kubota et al., 1996), or from proteolytic *C. botulinum* A, B, and F (East et al., 1996; Hutson et al., 1996). Indeed, BoNT/C and BoNT/D are distinct proteins except the N-terminal part of H chain (amino acids 522 to 944) which shows 75% identity between both toxins. A variant type C strain, called C-6813, produces a C/D mosaic neurotoxin the two third N-terminal part of which shares 95% identity with BoNT/C and the one third C-terminal shares 95% identity with BoNT/D (Moriishi et al., 1996b). An additional mosaic neurotoxin (D/C) produced by strain D-SA, can be divided in three homology regions: the N-terminal part (amino acids 1-521) which shares 96% identity with BoNT/D, a central core (amino acids 522-944) having 81 and 90% identity with BoNT/C and BoNT/D, respectively, and the C-terminal part (amino acids 945-1285), which is related to BoNT/C (72 % identity) (Moriishi et al., 1996a). In Japan, the C/D variant strains are mainly associated with avian botulism, whereas the D/C variants are mainly found from bovine (Nakamura et al., 2010; Takeda et al., 2005).

Subtype BoNT/A2 probably results from a recombination event between the 5' end of the *bont/A1* L chain gene and the 3' end of the *bont/A3* gene (Hill et al., 2007). Thereby, the N-terminal part of BoNT/A2 (amino acids 1 to 382) is related to the corresponding sequence of BoNT/A1 (97.6% identity, 99% homology), versus that of BoNT/A3 (83.2% identity, 89% homology), and the C-terminal part (amino acids 383 to 1296) is closer to the corresponding region from BoNT/A3 (96.0% identity, 97.4% homology) than from BoNT/A1 (86.8% identity, 93.3% homology) (Hill et al., 2007).

The *ntnh* gene is the most conserved gene in the botulinum locus in addition to *bont* gene. This gene is located in all botulinum loci directly upstream of *bont* gene and in the same orientation. Sequence variations are observed in *ntnh* genes, notably between *ntnh* from HA-BoNT and OrfX-BoNT loci (Hill et al., 2009). A recombination event in *ntnh* gene seems to be involved in the genealogy of HA-BoNT in group I *C. botulinum* strains and in the evolution of the OrfX-BoNT/A locus from HA-BoNT/A locus. Indeed, the N-terminal parts of NTNH (1-613 amino acids) from group I *C. botulinum* strains type A and B with an HA-BoNT locus are almost identical (99.5% identity), whereas the sequences from 614 to the C-terminus show a reduced relatedness (63.8% identity). Thus, HA-BoNT/A and HA-BoNT/B loci have probably evolved by a recombination event in the *ntnh* gene. In contrast, the NTNH

C-terminal parts (490 C-terminal amino acids) from HA-BoNT/A1 (strains 62A, Hall, or ATCC3502) and HA-BoNT/A2 (strain Kyoto-F) are highly related (94% identity), whereas the N-terminal parts (amino acids 1 to 669) are more divergent (63% identity) (East et al., 1996). Similarly, NTNH from HA-BoNT/(B) locus of bivalent AB strains NCT2916 and 667 exhibits a mosaic organization with amino acids 1-628 and 995-1198 highly related to NTNH from proteolytic *C. botulinum* type B strains and a central region 551-1021 highly homologous of that of NTNH from the NCTC2916 OrfX-BoNT/A1 locus (Jovita et al., 1998). In addition, the high homology level between NTNHs from OrfX-BoNT/A2-A4, OrfX-BoNT/A1 (NCTC2916), and proteolytic BoNT/F (97% identity at the nucleotide level) except in the 58 final nucleotides (51% identity) supports a recombination event in the 3' end of *ntnh* gene (Hill et al., 2009). Another example of recombination event in *ntnh* gene is supported by the mosaic organization of NTNH from an infant botulism *C. botulinum* strain A2 (7103-H) between NTNH from *C. botulinum* type C and type A. The N-terminal part (691 amino acids) of NTNH 7103-H is almost identical (98% identity) to that of NTNH from *C. botulinum* type C (strain 468), whereas the C-terminal region (692 to 1193) is related to the corresponding sequence of NTNH from *C. botulinum* A2 (Kubota et al., 1996). Interestingly, a deletion of 42 amino acids corresponding to amino acids 108 to 150 of NTNH in HA-BoNT locus of group I and II strains is observed in all NTNHs located in an OrfX-BoNT locus (East et al., 1996) (and unpublished), possibly as a consequence of a genetic rearrangement by a recombination mechanism.

5 – Evolution and horizontal transfer of Clostridium toxin genes

Whereas plasmid mobilization harboring toxin genes is the main source of diversification of toxigenic *C. perfringens* strain variants, chromosomal acquisition of virulence genes by mobile genetic elements, mainly transposons or prophages, and subsequent evolutions (point mutations, recombination, inversions) account for the diverse *C. difficile* toxinotypes. The botulinum locus shows a complex genetic diversity and is localized on various genetic elements such as chromosome, plasmid or phage supporting horizontal transfer between various strains of *Clostridium* species leading to multiple botulinum toxinotypes and subtypes. In contrast, *C. tetani* strains show limited genetic variations and only one tetanus locus is known. Although all the *Clostridium* species are bacteria from the environment, which accidentally develop in human and animals, they have distinct evolution which probably reflect adaptation to specific ecological niches. Interactions with other microorganisms from their common habitat likely influence their genetic development. The

intriguing question is where genes encoding the most potent toxins come from and how they have evolved? Why environmental bacteria have acquired such as potent toxins able to rapidly kill occasional hosts not essential for their survival?

5-1 *Gene evolution of enzymatically active toxins*

Clostridia are fermentative bacteria, which secrete numerous hydrolytic enzymes able to degrade various environmental substrates. These bacteria have multiple transporter systems, which allow the internalization of degradation products required for their metabolism. Numerous bacterial toxins are active through an enzymatic mechanism and thus probably derive from ancestral hydrolytic enzymes or enzyme precursors. Indeed, a likely evolution mechanism might include mutations in enzyme ancestor gene(s) leading to modification of substrate specificity from non-essential to highly critical molecules in cell survival or physiology, which results in the emergence of a novel enzymatic toxin or toxin domain. Subsequent combination or fusion of the new enzyme domain with a delivery system able to transport and internalize the catalytic domain into specific target cells might yield highly potent toxin(s). For example, BoNTs are the most potent toxins and are zinc-dependent metalloproteases. All BoNT toxinotypes and subtypes retain the classical zinc-coordinating HExxH motif and have enzymatic kinetics similar to other metalloproteases. But, BoNTs contain a receptor binding domain which drive the toxin molecules specifically to motor-neurons where they enter and cleave specific proteins (SNAREs) involved in the evoked release of acetylcholine (Poulain et al., 2008). Therefore, BoNT trafficking to only target neurons and specific impairment of a crucial step in motor-neuron function account for the extreme *in vivo* potency of these toxins, albeit they have standard *in vitro* enzymatic efficiency. As mentioned above, the botulinum locus consists in two operons, one of which contains *bont* and *ntnh* genes having similar size. Interestingly albeit BoNT and NTNH from *C. botulinum* type A share a weak amino acid sequence identity (~20%), they retain a similar structure. NTNH associate with BoNT by non-covalent bonds in a pH-dependent manner to form medium size botulinum complex, which is resistant to acidic pH and protease degradation (Gu et al., 2012). Thereby, NTNH is a non-toxic protein which acts as a chaperone protein to protect BoNT. NTNH does not contain the catalytic HExxH motif, but another zinc binding motif, KCLIK, at the same position. Indeed, NTNH binds one zinc atom per each molecule but exhibits no proteolytic activity (Inui et al., 2012). This strongly supports that all NTNH and BoNT variants derive from a common ancestor gene by duplication and subsequent independent reshuffling. Moreover, another gene in the vicinity of

bont possesses a zinc-binding motif. Indeed, a flagellin gene located immediately upstream of the botulinum locus in *C. botulinum* type A exhibits a sequence and structure related to the collagenase domain including a HExxH motif (Doxey et al., 2008). This further suggests that *bont* and adjacent genes have evolved from a common ancestor collagenase gene. Collagenases are produced by various Clostridia such as *C. perfringens* collagenase, ColA or κ-toxin, *C. histolyticum* collagenases, *C. tetani* collagenase, which contains the characteristic zinc-binding motif (Popoff and Bouvet, 2009). These collagenases have a broader spectrum of activity towards diverse substrates including collagen and gelatin. But, how and from which source BoNTs have acquired a binding domain specific of neuronal cell surface receptor which mediates their entry in only certain neurons and a highly specific proteolytic activity towards SNARE proteins? It has been hypothesized that the catalytic domain has derived from a viral protease ancestor, which has been subsequently fused to a delivery system consisting in the heavy chain which contains the receptor-binding and translocation domains (DasGupta, 2006). The localization of botulinum locus genes on genetic mobile elements (plasmid, phage, transposon) has allowed horizontal transfer and subsequent independent evolution in different *Clostridium* species: the three *C. botulinum* groups, which are bacteriologically distinct bacterial species, *C. argentinense*, *C. baratii*, and *C. butyricum*.

Another example of a likely toxin gene evolution from ancestor enzyme and spreading in Clostridia concerns LCGTs (TcdA, TcdB, TcsL, TcsH, TcnA, TpeL), which are produced by different species (*C. difficile*, *C. sordellii*, *C. novyi, C. perfringens*). These toxins share sequence and structure similarities, and they all contain the catalytic motif DxD in their N-terminal domain. The adjacent *tcdA* and *tcdB* genes in *C. difficile* PaLoc strongly suggests a duplication of the original gene leading to the two toxin variants. *C. novyi*, *C. perfringens*, and most strains of *C. sordellii* synthesize only one type of toxins TcnA, TpeL and TcsL, respectively, indicating that these strains have received only one gene copy, and that eventual duplication and subsequent gene evolution have occurred differently in each *Clostridium* strain. The phage localization of *tcnA* in *C. novyi* (Eklund et al., 1976; Schallehn et al., 1980), and possibly that of *tcsL* in *C. sordellii* (data not shown), as well as the plasmid localization of *tpeL* in *C. perfringens* (Gurjar et al., 2010; Sayeed et al., 2010) have facilitated horizontal transfer of toxin genes between strains of these *Clostridium* species. The chromosomal localization of PaLoc in *C. difficile* probably emerged from a transposition event since transposase-like genes lie in the flanking regions of PaLoc (Sebaihia et al., 2006). Interestingly, interchangeability of phages harboring either *bont/C/bont/D* or *tcnA* between *C. botulinum* type C or D and *C. novyi* (Eklund et al., 1974) further shows that toxin genes can

be horizontally transferred between *Clostridium* species. LCGTs recognize ubiquitous cell surface receptor(s), which remain to be characterized, since these toxins are active on most of the eukaryotic cell lines, and enzymatically modify specific intracellular targets consisting in Rho/Ras family proteins. Again, how and from which source, LCGTs have acquired specific receptor-binding domains associated to catalytic domains targeting a particular class of G-proteins? LCGTs share similar overall structure of the catalytic domain with other bacterial and eukaryotic glucosyltransferases such as *Neisseria meningitides* galactosyltransferase LgtC (lipooligosaccharide glycosyl transferase), *Bacillus subtilis* glycosyltransferase, bovine galactosyltransferase α3GalT (Jank and Aktories, 2008). In addition, at the amino acid sequence level, the LCGT C-terminal domain is similar to various saccharide- or glucan-binding proteins and the catalytic domain containing the DxD motif shows similarity with other bacterial glucosyltransferases, notably from diverse *Streptococcus* species (Wren, 1991). Moreover, *E.* coli proteins called *E. coli* ToxB and Efa1 produced by some enterohemorrhagic strains as well as LifA synthesized by enteropathogenic strains are related to TcdB, but do not exhibit enzymatic activity and do not alter the cytoskeleton. Related genes to LCGTs have also been identified in genomes of other microorganisms such as *Chlamidophyla caviae, C. muridarum,* and *Chlamidia trachomatis* with still unknown function (review in (Rupnik and Just, 2006)). Thereby, LCGTs likely derive from a common glucosyltransferase or saccharide-binding protein ancestor, which has been spread in numerous microrganisms. However, in contrast to the other bacterial and eukaryotic enzymes, they have retained a unique catalytic activity towards Rho/Ras proteins yielding mysterious their origin and evolution.

Gene duplication seems tobe a common mechanism in toxin gene evolution as shown above in *C. botulinum* and *C. difficile*. An additional example is provided by C3 enzyme and C2 toxin, which are produced by *C. botulinum* type C and D. C3 is a 25 kDa protein containing the ADP-ribosylation site characterized by the QLE motif (Han et al., 2001; Ménétrey et al., 2002). In contrast, the enzymatic component C2-I is a 50 kDa protein encompassing two structurally related domains of same size. The C-terminal domain contains the biglutamic EYE enzymatic site, whereas the N-terminal domain has not enzymatic activity but interacts with the binding component C2-II, which facilitates its entry into the cells. C3 enzyme and the two C2-I domains retain a similar folding of a core β–strands surrounded by α-helices albeit a low amino acid sequence identity (Han et al., 2001; Schleberger et al., 2006). It is tempting to speculate that C2-I derives from a C3 gene

duplication. The resulting protein has a double size with two related domains, one retaining a modified enzymatic activity (C2-I ADP-ribosylates monomeric actin instead of RhoA, B, C proteins which are the substrates of C3) and the other one, which has acquired a new function consisting in interaction with a delivery molecule to enter the cells (Figure 1). It is noteworthy that gene duplication leads to structurally related proteins or protein domains, which share moderate or low amino acid sequence identity.

5-2 *Pore-forming toxin gene evolution*

Many clostridial toxins (almost one third) as well as a large number of other bacterial toxins are pore-forming toxins. Most of them form pores through insertion of amphipatic β-hairpins organized in a β-barrel into the cell membrane and are called β-pore-forming toxins (β-PFTs). They are divided into several families but all β-PFTs retain conserved structural domains including a receptor-binding domain, one or two domains involved in oligomerization, and a domain containing α-helices which unfold forming amphipathic β-sheets (Figure 2). The largest β-PFT family is the cholesterol-dependent cytoslysin (CDC) family, which encompasses toxins from at least 9 *Clostridium* species such as *C. perfringens* perfringolysin (PFO), *C. botulinum* botunolysin, and *C. tetani* tetanolysin (Popoff and Bouvet, 2009), toxins from more than 15 other Gram positive bacterial species (*Streptococcus*, *Bacillus*, *Listeria* …,) and a few Gram negative bacteria (Hotze et al., 2012; Tweten, 2005). PFO is the prototype of the CDC family. PFO has an unusual elongated rod shape containing four structural domains. Domain 4 binds to cell membrane cholesterol inducing a conformation change of domain 1 permitting the oligomerization via domain1-domain1 interaction. Domain 3 rotates from domain 2 and unfolds three α-helices in two amphipathic β-sheets. The particularity of CDC is that each monomer contributes two amphipathic β-hairpins to the formation of the transmembrane β-barrel and that they form large pores including 40 to 50 monomers (Heuck et al., 2000; Popoff and Bouvet, 2009; Shatursky et al., 1999). CDCs show structural relatedness with eukaryotic membrane attack complex/perforin (MACPF) family, which encompasses pore-forming proteins involved in immunity, host defence, venom toxicity and pathogenicity. Mouse perforin exhibits a remarkable similar overall structure with that of PFO, whereas the other MACPF proteins are more distantly related but retain a protein fold similar to domain 3 of CDCs. MACPF and CDCs seem to use a common mode of pore formation by unfolding two α-helices into membrane spanning β-strands but with differences in these mechanisms (Anderluh and Lakey, 2008; Dunstone and

Tweten, 2012). Another clostridial *C. perfringens* β-PFT family encompasses epsilon toxin (ETX), *C. perfringens* enterotoxin (CPE), and *C. septicum* alpha-toxin are structurally related to aerolysin produced by Gram-negative bacteria of *Aeromonas sp.,* although ETX and CPE show no significant homology with aerolysin at the amino acid level (Briggs et al., 2011; Cole et al., 2004). The β-PFT aerolysin family also contains toxins from diverse origin, bacteria, animal, plant, like mosquitocidal toxins (Mtxs) from the Gram-positive bacteria *Bacillus sphaericus,* hydralysins from the animal *Chlorohydra viridis,* enterolobin from the Brazilian tree *Enterolobium contortisiliquum, Laetiporus sulphurous* lectin (LSL) from the mushroom *Laetiporus sulphurous* (Knapp et al., 2010), and lysenin from the eartworm *Eisenia fetida* (De Colibus et al., 2012). PFTs from aerolysin family exhibit a more elongated shape than PFO (Figure 2). They consist in 3 to 4 domains and form heptamers. The domain interacting with the receptor is the N-terminal domain 1, while in PFO it is the C-terminal domain 4. β-PFTs of aerolysin family recognize GPI-anchored proteins or membrane proteins as receptors. Each monomer deploys only one β-hairpin forming the transmembrane β-barrel (Knapp et al., 2010). *C. perfringens* Delta toxin and NetB toxin constitute a third β–PFT family structurally related to staphylococcal β–PFTs, the prototype of which is the alpha-hemolysin (Savva et al., 2012) (submitted). This family also includes *C. perfringens* Beta toxin and Beta2 toxin which share significant sequence homology with Delta toxin and staphylococcal β–PFTs and likely related structure (Manich et al., 2008) and in press). Alpha-hemolysin contains three main domains a N-terminal β–sandwich domain, a central or stem domain, and a rim domain. PFTs of this family have a more globular conformation than those of the aerolysin family. A hallmark of alpha-hemolysin and related β–PFTs is that the central stem domain of monomers contains a β–hairpin packed against the β–sandwich domain. Upon heptamerization, the β–hairpin moves from the β–sandwich to form the β-barrel (Song et al., 1996). Despite localized structural differences, β-PFTs share a global common organization and a β-barrel based mechanism of pore-formation suggesting that they all derive from a common ancestor with subsequent distinct evolution through the different β-PFT families in each bacterial and eukaryotic lineages.

Interestingly, BCs of clostridial binary toxins (C2-II of *C. botulinum* C2 toxin, CDTb of *C. difficile* CDT, CSTb of *C. spiroforme* toxin, and Ib of *C. perfringens* Iota toxin) as well as those of *Bacillus* binary toxins (the protective antigen (PA) of *Bacillus* anthrax toxins, VIP-1 of *Bacillus cereus* and *Bacillus thuringiensis* vegetative insecticidal proteins) also form pores permitting the passage of enzymatic components into the cytosol of targeted cells. Each

BC specifically mediates internalization of the corresponding enzymatic component. However, certain BCs have a broader specificity. Indeed, those (CDTb, CSTb, and Ib) of the iota toxin family indifferently mediate translocation of the respective enzymatic components CDTa, CSTa, and Ia (Barth et al., 2004; Stiles et al., 2011). In addition, PA promotes the entry of the corresponding enzymatic components of anthrax toxins (edema and lethal factors), but also C2-I (Kronhardt et al., 2011). The BCs of clostridial and *Bacillus* binary toxins share a similar structural organization. PA is the prototype of BCs and consists in 4 domains highly related to those of PFO. However, the mechanisms of oligomerization and insertion into the membrane, as well as the shape of PA oligomers are more similar to those of aerolysin and related toxins such as staphylococcal leucocidins, which are also binary toxins, and *Staphylococcus* alpha-hemolysin than to those of PFO (Petosa et al., 1997; Song et al., 1996). Indeed, BCs form heptamers with only one β-hairpin from each monomer forming the β-barrel thus leading to small pores.

This strongly suggests that all β-PFTs and BCs have evolved from a common ancestor, possibly a transmembrane protein. BCs have retained a core structure of β-PFTs, and they have acquired the ability to recognize and translocate specific enzymatic components, whereas β-PFTs form pores in plasma membrane of eukaryotic cells leading to drastic cellular effects. It is intriguing that homologues of aerolysin and CDC families are wide spread in all the kingdoms of life. As mentioned above, structural homologous proteins of aerolysin (hydralysins, enterolobin, LSL, …) are produced by plants, fungi and animals (Knapp et al., 2010). In addition, more than 300 proteins from diverse groups of organisms share aerolysin domain similarity based on local sequence alignment and phylogenetic analysis. It is hypothesized that these proteins derive from a common ancestor probably in early bacterial lineages, which has been transmitted between organisms of different phylum by horizontal gene transfer. This analysis suggests that at least six independent transfer events have occurred between distantly related organisms including between bacteria and eukaryotic cells (Moran et al., 2012). The structural homology between MACPF and CDC proteins restricted to the pore-forming domain, whereas the other domains are distantly related (Dunstone and Tweten, 2012), rather suggests a convergent evolution of eukaryote and prokaryote proteins of these families to interact with the lipid bilayer.

6 - Concluding remarks

It is intriguing why some Clostridia species, which are essentially environmental bacteria and which occasionally or accidentally develop in human and animal tissues, produce so numerous and potent toxins. Two main classes of toxins can be distinguished, toxins which retain an enzymatic activity and β–PFTs. A likely hypothesis is that toxins with an enzymatic activity have evolved from hydrolytic enzymes or enzyme precursors, which are secreted by these bacteria for their metabolism, and that β–PFTs might have derived from ancestor transmembrane protein(s). Interestingly, binary toxins seem to result from a cross evolution between enzymatic and transmembrane protein ancestors (Figure 3). The main benefit of toxins is lkely to participate in nutrient supply from eukaryotic cells with sibsequent uptake into bacterial cell. However, the advantage of certain toxins, which have acquired a high substrate specificity, is not obvious. Indeed, botulinum toxins target the neuroexocytosis machinery and induce dramatic effects at very low dose but without evident benefit for the bacteria. *C. botulinum* can grow in certain foods producing a sufficient level of toxin able to kill a man or animal but without further bacterial development in the host, since ingestion of a too low number or non viable bacteria. Some toxigenic Clostridium species show a high genetic diversity including toxin gene variation, whereas others are more genetically stable. For example, *C. botulinum* consists in numerous genetically diverse strains whereas *C. tetani* forms a more homogenous group of bacteria. Which are the selection pressure or the environmental factors controlling the genetic evolution of toxigenic bacteria? The digestive tract, which contains the densest population of bacteria, represents an ecological niche favorable to genetic exchanges between bacteria. However, a precise adaptation is required for each bacterial strain to survive in this environment. Toxigenic Clostridia which transit or develop in the digestive tract such as *C. perfringens, C. difficile or C. botulinum*, show a high level of genetic variability and toxin gene mobilization, in contrast to *C. tetani* which has a more restricted environmental localization. But why the genetic variability of *C. perfringens* strains involved in gastrointestinal diseases is mainly based on exchange of plasmids harboring conserved toxin genes, whereas *C. difficile* strains show a high level of chromosomal genetic diversity including toxin genes? Do these two modes of genetic variation reflect distinct types of adaptation to the intestinal ecosystem?

Genetic exchanges not only concern gene transfer between bacteria but also between bacteria and eukaryotic cells in both directions. If most bacterial toxins derive from ancestor bacterial genes, it is not excluded that certain bacterial toxin genes have been acquired from higher organisms. Inversely, bacterial toxin genes might be transferred to eukaryotic cells. This raises the questions which benefits provide bacterial toxin genes in eukaryotic

organisms? β–PFT homologues in eukaryotic cells are supposed to be involved in uptake of nutriments. But, how certain bacterial proteins have evolved in highly potent and specific toxins?

REFERENCES

Ahmed, S.A., Byrne, M.P., Jensen, M., Hines, H.B., Brueggemann, E., Smith, L.A., 2001. Enzymatic autocatalysis of botulinum A neurotoxin light chain. J Protein Chem 20, 221-231.

Alouf, J.E., 2003. Molecular features of the cytolytic pore forming bacterial protein toxins. Fol. Microbiol. 48, 5-16.

Anderluh, G., Lakey, J.H., 2008. Disparate proteins use similar architectures to damage membranes. Trends Biochem Sci 33, 482-490.

Arndt, E.R., Jacobson, M.J., Abola, E.E., Forsyth, C.M., Tepp, W.H., Marks, J.D., Johnson, E.A., Stevens, E.S., 2006. A structural perspective of the sequence variability within botulinum neurotoxin subtypes A1-A4. J. Mol. Biol. 362, 733-742.

Barash, J.R., Arnon, S.S., 2004. Dual toxin-producing strain of Clostridium botulinum type Bf isolated from a California patient with infant botulism. J Clin Microbiol 42, 1713-1715.

Barth, H., Aktories, K., Popoff, M.R., Stiles, B.G., 2004. Binary bacterial toxins: biochemistry, biology, and applications of common *Clostridium* and *Bacillus* proteins. Microbiol. Mol. Biol. Rev. 68, 373-402.

Bennik, M.H.J., Mauchline, M., Bosveld, F., Elmore, M., Minton, N.P., Parkhill, J., Peck, M.W., 2003. The genome sequence of proteolytic *Clostridium botulinum* ATCC3502 (Hall A): some highlights, in: Sonenshein, L. (Ed.), Clostridial 03 Pathogenesis, Woods Hole, p. 44.

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. Infect. Immun. 66, 4531-4536.

Bouvet, P.J., Popoff, M.R., 2008. Genetic relatedness of Clostridium difficile isolates from various origins determined by triple-locus sequence analysis based on toxin regulatory genes tcdC, tcdR, and cdtR. J Clin Microbiol 46, 3703-3713.

Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M., von Eichel-Streiber, C., 1996. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. Gene 181, 29-38.

Braun, V., Mehlig, M., Moos, M., Rupnik, M., Kalt, B., Mahony, D.E., von Eichel-Streiber, C., 2000. A chimeric ribozyme in *Clostridium difficile* combines features of group I introns and insertion elements. Mol. Microbiol. 36, 1447-1459.

Brefort, G., Magot, M., Ionesco, H., Sebald, M., 1977. Characterization and transferability of *Clostridium perfringens* plasmids. Plasmid 1, 52-66.

Briggs, D.C., Naylor, C.E., Smedley, J.G., 3rd, 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. J Mol Biol 413, 138-149.

Brouwer, M.S., Warburton, P.J., Roberts, A.P., Mullany, P., Allan, E., 2011. Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of Clostridium difficile. PLoS One 6, e23014.

Brüggemann, H., Bäumer, S., Fricke, W.F., Wiezr, A., Liesagang, H., Decker, I., Herzberg, C., Martinez-Arias, R., Henne, A., Gottschalk, G., 2003. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. Proc. Ntl. Acad. Sci. (USA) 100, 1316-1321.

Brüggemann, H., Wolher, A., Mazuet, C., Popoff, M.R., 2011. *Clostridium botulinum*, in: Fratamico, P., Liu, Y., Kathariou, S. (Eds.), Genomes of Food- and Water Borne Pathogens. ASM, Washington DC, pp. 185-212.

Brynestadt, S., Granum, P.E., 1999. Evidence that Tn5565, which includes the enterotoxin gene in *Clostridium perfringens*, can have a circular form, which may be a transposition intermediate. FEMS Microbiol. Lett. 170, 281-286.

Carter, A.T., Paul, C.J., Mason, D.R., Twine, S.M., Alston, M.J., Logan, S.M., Austin, J.W., Peck, M.W., 2009. Independent evolution of neurotoxin and flagellar genetic loci in proteolytic Clostridium botulinum. BMC Genomics 10, 115.

Carter, G.P., Douce, G.R., Govind, R., Howarth, P.M., Mackin, K.E., Spencer, J., Buckley, A.M., Antunes, A., Kotsanas, D., Jenkin, G.A., Dupuy, B., Rood, J.I., Lyras, D., 2011. The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of Clostridium difficile. PLoS Pathog 7, e1002317.

Carter, G.P., Lyras, D., Allen, D.L., Mackin, K.E., Howarth, P.M., O'Connor, J.R., Rood, J.I., 2007. Binary toxin production in Clostridium difficile is regulated by CdtR, a LytTR family response regulator. J Bacteriol 189, 7290-7301.

Cartman, S.T., Kelly, M.L., Heeg, D., Heap, J.T., Minton, N.P., 2012. Precise Manipulation of the Clostridium difficile Chromosome Reveals a Lack of Association between the tcdC Genotype and Toxin Production. Appl Environ Microbiol 78, 4683-4690.

Chaves-Olarte, E., Freer, E., Parra, A., Guzman-Verri, C., Moreno, E., Thelestam, M., 2003. R-Ras glucosylation and transient RhoA activation determine the cytopathic effect

produced by toxin B variants from toxin A-negative strains of *Clostridium difficile*. J. Biolo. Chem. 278, 7956-7963.

Chaves-Olarte, E., Low, P., Freer, P., Norlin, T., Weidmann, M., von Eichel-Streiber, C., Thelestam, M., 1999. A novel cytotoxin from *Clostridium difficile* serogroup F is a functional hybrid between two tother large clostridial toxins. J. Biol. Chem. 274, 11046-11052.

Chen, Y., Korkeala, H., Aarnikunnas, J., Lindstrom, M., 2007. Sequencing the botulinum neurotoxin gene and related genes in Clostridium botulinum type E strains reveals orfx3 and a novel type E neurotoxin subtype. J Bacteriol 189, 8643-8650.

Cole, A.R., Gibert, M., Popoff, M.R., Moss, D.S., Titball, R.W., Basak, A., 2004. *Clostridium perfringens* ε-toxin shows structural similarity to the pore-forming toxin aerolysin. Nat. Struct. Mol. Biol. 11, 797-798.

Collins, M.D., East, A.K., 1997. Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. J. Appl. Microbiol. 84, 5-17.

Cordoba, J.J., Collins, M.D., East, A.K., 1995. Studies on the genes encoding botulinum neurotoxin type A of *Clostridium botulinum* from a variety of sources. System. Appl. Microbiol. 18, 13-22.

Curry, S.R., Marsh, J.W., Muto, C.A., O'Leary, M.M., Pasculle, A.W., Harrison, L.H., 2007. tcdC genotypes associated with severe TcdC truncation in an epidemic clone and other strains of Clostridium difficile. J Clin Microbiol 45, 215-221.

DasGupta, B.R., 2006. Botulinum neurotoxins: perspective on their existence and as polyproteins harboring viral proteases. J Gen Appl Microbiol 52, 1-8.

De Colibus, L., Sonnen, A.F., Morris, K.J., Siebert, C.A., Abrusci, P., Plitzko, J., Hodnik, V., Leippe, M., Volpi, E., Anderluh, G., Gilbert, R.J., 2012. Structures of lysenin reveal a shared evolutionary origin for pore-forming proteins and its mode of sphingomyelin recognition. Structure 20, 1498-1507.

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.

Dineen, S.S., Bradshaw, M., Johnson, E.A., 2003. Neurotoxin gene clusters in *Clostridium botulinum* type A strains: sequence comparison and evolutionary implications. Cur. Microbiol. 46, 342-352.

Dineen, S.S., Bradshaw, M., Karasek, C.E., Johnson, E.A., 2004. Nucleotide sequence and transcriptional analysis of the type A2 neurotoxin gene cluster in *Clostridium botulinum*. FEMS Microbiol. Lett. 235, 9-16.

Dingle, K.E., Griffiths, D., Didelot, X., Evans, J., Vaughan, A., Kachrimanidou, M., Stoesser, N., Jolley, K.A., Golubchik, T., Harding, R.M., Peto, T.E., Fawley, W., Walker, A.S., Wilcox, M., Crook, D.W., 2011. Clinical Clostridium difficile: clonality and pathogenicity locus diversity. PLoS One 6, e19993.

Dover, N., Barash, J.R., Arnon, S.S., 2009. Novel Clostridium botulinum toxin gene arrangement with subtype A5 and partial subtype B3 botulinum neurotoxin genes. J Clin Microbiol 47, 2349-2350.

Doxey, A.C., Lynch, M.D., Muller, K.M., Meiering, E.M., McConkey, B.J., 2008. Insights into the evolutionary origins of clostridial neurotoxins from analysis of the Clostridium botulinum strain A neurotoxin gene cluster. BMC Evol Biol 8, 316.

Dunstone, M.A., Tweten, R.K., 2012. Packing a punch: the mechanism of pore formation by cholesterol dependent cytolysins and membrane attack complex/perforin-like proteins. Curr Opin Struct Biol 22, 342-349.

Dupuy, B., Govind, R., Antunes, A., Matamouros, S., 2008. Clostridium difficile toxin synthesis is negatively regulated by TcdC. J Med Microbiol 57, 685-689.

East, A.K., Bhandari, M., Stacey, J.M., Campbell, K.D., Collins, M.D., 1996. Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic *Clostridium botulinum* types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component. Int. J. Syst. Bacteriol. 46, 1105-1112.

Eisele, K.H., Fink, K., Vey, M., Taylor, H.V., 2011. Studies on the dissociation of botulinum neurotoxin type A complexes. Toxicon 57, 555-565.

Eklund, M.W., Dowell, J., 1987. Avian botulism. Charles C. Thomas, Springfield, Ill.

Eklund, M.W., Poysky, F.T., 1974. Interconversion of type C and D strains of *Clostridium botulinum* by specific bacteriohages. Appl. Microbiol. 27, 251-258.

Eklund, M.W., Poysky, F.T., Habig, W.H., 1989. Bacteriophages and plasmids in *Clostridium botulinum* and *Clostridium tetani* and their relationship to production of toxins., in: Simpson, L.L. (Ed.), Botulinum Neurotoxin and Tetanus Toxin. Academic Press, San Diego, pp. 25-51.

Eklund, M.W., Poysky, F.T., Meyers, J.A., Pelroy, G.A., 1974. Interspecies conversion of *Clostridium botulinum* type C to *Clostridium novyi* type A by bacteriophage. Science 172, 480-482.

Eklund, M.W., Poysky, F.T., Peterson, M.E., Meyers, J.A., 1976. Relationship of bacteriophage to alpha toxin production in *Clostridium novyi* types A and B. Infect. Immun. 14, 798-803.

Eklund, M.W., Poysky, F.T., Reed, S.M., Smith, C.A., 1971. Bacteriophage and the toxigenicity of *Clostridium botulinum* type C. Science 172, 480-482.

Fenicia, L., Franciosa, G., Pourshaban, M., Aureli, P., 1999. Intestinal toxemia botulism in two young people, caused by *Clostridium butyricum* Type E. Clin. Infect. Dis. 29, 381-387.

Forgetta, V., Oughton, M.T., Marquis, P., Brukner, I., Blanchette, R., Haub, K., Magrini, V., Mardis, E.R., Gerding, D.N., Loo, V.G., Miller, M.A., Mulvey, M.R., Rupnik, M., Dascal, A., Dewar, K., 2011. Fourteen-genome comparison identifies DNA markers for severe-disease-associated strains of Clostridium difficile. J Clin Microbiol 49, 2230-2238.

Franciosa, G., Fenicia, L., Pourshaban, M., Aureli, P., 1997. Recovery of a strain of *Clostridium botulinum* producing both neurotoxin A and neurotoxin B from canned macrobiotic food. Appl. Environ. Microbiol. 63.

Franciosa, G., Ferreira, J.L., Hatheway, C.L., 1994. Detection of type A, B, and E botulism neurotoxin genes in *Clostridium botulinum* and other *Clostridium* species by PCR: evidence of unexpressed type B toxin genes in type A toxigenic organisms. J. Clin. Microbiol. 32, 1911-1917.

Franciosa, G., Floridi, F., Maugliani, A., Aureli, P., 2004. Differentitation of the gene clusters encoding botulinum neurotoxin type A complexes in *Clostridium botulinum* type A, Ab, and A(B) strains. Appl Environ Microbiol 70, 7192-7199.

Franciosa, G., Maugliani, A., Floridi, F., Aureli, P., 2006. A novel type A2 neurotoxin gene cluster in *Clostridium botulinum* strain Mascarpone. FEMS Microbiol. Lett. 261, 88-94.

Franciosa, G., Maugliani, A., Scalfaro, C., Aureli, P., 2009. Evidence that plasmid-borne botulinum neurotoxin type B genes are widespread among Clostridium botulinum serotype B strains. PLoS One 4, e4829.

Gibert, M., Perelle, S., Daube, G., Popoff, M.R., 1997. *Clostridium spiroforme* toxin genes are related to *C. perfringens* iota toxin genes but have a different genomic localization. Syst. Appl. Microbiol. 20, 337-347.

Goldenberg, S.D., French, G.L., 2011. Lack of association of tcdC type and binary toxin status with disease severity and outcome in toxigenic Clostridium difficile. J Infect 62, 355-362.

Gu, S., Rumpel, S., Zhou, J., Strotmeier, J., Bigalke, H., Perry, K., Shoemaker, C.B., Rummel, A., Jin, R., 2012. Botulinum neurotoxin is shielded by NTNHA in an interlocked complex. Science 335, 977-981.

Gurjar, A., Li, J., McClane, B.A., 2010. Characterization of toxin plasmids in Clostridium perfringens type C isolates. Infect Immun 78, 4860-4869.

Hall, J.D., McCroskey, L.M., Pincomb, B.J., Hatheway, C.L., 1985. Isolation of an organism resembling *Clostridium barati* which produces type F botulinal toxin from an infant with botulism. J. Clin. Microbiol. 21, 654-655.

Han, S., Arvai, A.S., Clancy, S.B., Tainer, J.A., 2001. Crystal structure and novel recognition motif of Rho ADP-ribosylating C3 exoenzyme from *Clostridium botulinum*: structural insights for recognition specificity and catalysis. J. Mol. Biol. 305, 95-107.

Hauser, D., Gibert, M., Boquet, P., R., P.M., 1992. Plasmid localization of a type E botulinal neurotoxin gene homologue in toxigenic *Clostridium butyricum* strains, and absence of this gene in non-toxigenic *C. butyricum* strains. FEMS Microbiol. Let. 99, 251-256.

Hauser, D., Gibert, M., Eklund, M.W., Boquet, P., Popoff, M.R., 1993. Comparative analysis of C3 and botulinal neurotoxin genes and their environment in *Clostridium botulinum* C and D. J. Bacteriol. 175, 7260-7268.

He, M., Sebaihia, M., Lawley, T.D., Stabler, R.A., Dawson, L.F., Martin, M.J., Holt, K.E., Seth-Smith, H.M., Quail, M.A., Rance, R., Brooks, K., Churcher, C., Harris, D., Bentley, S.D., Burrows, C., Clark, L., Corton, C., Murray, V., Rose, G., Thurston, S., van Tonder, A., Walker, D., Wren, B.W., Dougan, G., Parkhill, J., 2010. Evolutionary dynamics of Clostridium difficile over short and long time scales. Proc Natl Acad Sci U S A 107, 7527-7532.

Henderson, I., Davis, T., Elmore, M., Minton, N., 1997. The genetic basis of toxin production in *Clostridium botulinum* and *Clostridium tetani*, in: Rood, I. (Ed.), The *Clostridia*: Molecular Biology and Pathogenesis. Academic Press, New York, pp. 261-294.

Henderson, I., Whelan, S.M., Davis, T.O., Minton, N.P., 1996. Genetic characterization of the botulinum toxin complex of *Clostridium botulinum* strain NCTC2916. FEMS Microbiol. Lett. 140, 151-158.

Henkel, J.S., Jacobson, M., Tepp, W., Pier, C., Johnson, E.A., Barbieri, J.T., 2009. Catalytic Properties of Botulinum Neurotoxin Subtypes A3 and A4. Biochemistry 48, 2522-2528.

Heuck, A.P., Hotze, E.M., Tweten, R.K., Johnson, A.E., 2000. Mechanism of membrane insertion of a multimeric beta-barrel protein: perfringolysin O creates a pore using ordered and coupled conformational changes. Mol Cell 6, 1233-1242.

Hielm, S., Bjorkroth, J., Hyytia, E., Korkeala, H., 1998a. Genomic analysis of Clostridium botulinum group II by pulsed-field gel electrophoresis. Appl Environ Microbiol 64, 703-708.

Hielm, S., Björkroth, J., Hyytiä, E., Korkeala, H., 1998b. Prevalence of *Clostridium botulinum* in Finnish trout farms: pulse-field gel electrophoresis typing reveals extensive genetic diversity among Type E isolates. Apl. Environ. Microbiol. 64, 4161-4167.

Hill, K.K., Smith, T.J., Helma, C.H., Ticknor, L.O., Foley, B.T., Svensson, R.T., Brown, J.L., Johnson, E.A., Smith, L.A., Okinaka, R.T., Jackson, P.J., Marks, J.D., 2007. Genetic diversity among Botulinum Neurotoxin-producing clostridial strains. J Bacteriol 189, 818-832.

Hill, K.K., Xie, G., Foley, B.T., Smith, T.J., Munk, A.C., Bruce, D., Smith, L.A., Brettin, T.S., Detter, J.C., 2009. Recombination and insertion events involving the botulinum neurotoxin complex genes in Clostridium botulinum types A, B, E and F and Clostridium butyricum type E strains. BMC Biol 7, 66.

Hinderink, K., Lindstrom, M., Korkeala, H., 2009. Group I Clostridium botulinum strains show significant variation in growth at low and high temperatures. J Food Prot 72, 375-383.

Hotze, E.M., Le, H.M., Sieber, J.R., Bruxvoort, C., McInerney, M.J., Tweten, R.K., 2012. Identification and characterization of the first cholesterol-dependent cytolysins from Gram-negative bacteria. Infect Immun.

Hutson, R.A., Collins, M.D., East, A.K., Thompson, D.E., 1994. Nucleotide sequence of the gene coding for non-proteolytic *Clostridium botulinum* type B neurotoxin: comparison with other Clostridial neeurotoxins. Cur. Microbiol. 28, 101-110.

Hutson, R.A., Zhou, Y., Collins, M.D., Johnson, E.A., Hatheway, C.L., Sugiyama, H., 1996. Genetic characterization of *Clostridium botulinum* type A containing silent type B neurotoxin gene sequences. J. Biol. Chem. 271, 10786-10792.

Hwang, H.J., Lee, J.C., Yamamoto, Y., Sarker, M.R., Tsuchiya, T., Oguma, K., 2007. Identification of structural genes for Clostridium botulinum type C neurotoxin-converting phage particles. FEMS Microbiol Lett 270, 82-89.

Hyytia, E., Bjorkroth, J., Hielm, S., Korkeala, H., 1999a. Characterisation of Clostridium botulinum groups I and II by randomly amplified polymorphic DNA analysis and repetitive element sequence-based PCR. Int J Food Microbiol 48, 179-189.

Hyytia, E., Hielm, S., Björkroth, J., Korkeala, H., 1999b. Biodiversity of *Clostridium botulinum* Type E strains isolated from fish and fishery products. Appl. Environ. Microbiol. 65, 2057-2064.

Ihara, H., Kohda, T., Morimoto, F., Tsukamoto, K., Karasawa, T., Nakamura, S., Mukamoto, M., Kozaki, S., 2003. Sequence of the gene for Clostridium botulinum type B neurotoxin associated with infant botulism, expression of the C-terminal half of heavy chain and its binding activity. Biochim Biophys Acta 1625, 19-26.

Inui, K., Sagane, Y., Miyata, K., Miyashita, S., Suzuki, T., Shikamori, Y., Ohyama, T., Niwa, K., Watanabe, T., 2012. Toxic and nontoxic components of botulinum neurotoxin complex are evolved from a common ancestral zinc protein. Biochem Biophys Res Commun 419, 500-504.

Jacobson, M.J., Lin, G., Raphael, B., Andreadis, J., Johnson, E.A., 2008a. Analysis of neurotoxin cluster genes in Clostridium botulinum strains producing botulinum neurotoxin serotype A subtypes. Appl Environ Microbiol 74, 2778-2786.

Jacobson, M.J., Lin, G., Whittam, T.S., Johnson, E.A., 2008b. Phylogenetic analysis of Clostridium botulinum type A by multi-locus sequence typing. Microbiology 154, 2408-2415.

Jank, T., Aktories, K., 2008. Structure and mode of action of clostridial glucosylating toxins: the ABCD model. Trends Microbiol 16, 222-229.

Johnson, E.A., Tepp, W., Bradshaw, M., Gilbert, R.J., Cook, P.E., McIntosh, E.D., 2005. Characterization of *Clostridium botulinum* strains associated with an infant botulism case in the United Kingdom. J. Clin. Microbiol. 43, 2602-2607.

Jovita, M.R., Collins, M.D., East, A.K., 1998. Gene organization and sequence determination of the two botulinum neurotoxin gene clusters in *Clostridium botulinum*. Cur. Microbiol. 36, 226-231.

Justin, N., Walker, N., Bullifent, H.L., Songer, G., Bueschel, D.M., Jost, H., Naylor, C., Miller, J., Moss, D.S., Titball, R.W., Basak, A.K., 2002. The first strain of *Clostridium perfringens* isolated from an avian source has an alpha-toxin with divergent structural and kinetics properties. Biochem. 41, 6253-6262.

Keto-Timonen, R., Nevas, M., Korkeala, H., 2005. Efficient DNA fingerprint of *Clostridium botulinum* types A, B, E, and F by amplified fragment length polymorphism analysis. Appl Environ Microbiol 71, 1148-1154.

Kitadokoro, K., Nishimura, K., Kamitani, S., Fukui-Miyazaki, A., Toshima, H., Abe, H., Kamata, Y., Sugita-Konishi, Y., Yamamoto, S., Karatani, H., Horiguchi, Y., 2011. Crystal structure of Clostridium perfringens enterotoxin displays features of beta-pore-forming toxins. J Biol Chem 286, 19549-19555.

Knapp, O., Stiles, B.G., Popoff, M.R., 2010. The aerolysin-like toxin family of cytolytic, pore-forming toxins. Open Toxinol. J. 3, 53-68.

Knetsch, C.W., Terveer, E.M., Lauber, C., Gorbalenya, A.E., Harmanus, C., Kuijper, E.J., Corver, J., van Leeuwen, H.C., 2012. Comparative analysis of an expanded Clostridium difficile reference strain collection reveals genetic diversity and evolution through six lineages. Infect Genet Evol 12, 1577-1585.

Kozaki, S., Kamata, Y., Nishiki, T., Kakinuma, H., Maruyama, H., Takahashi, H., Karasawa, T., Yamakawa, K., Nakamura, S., 1998. Characterization of Clostridium botulinum type B neurotoxin associated with infant botulism in Japan. Infect Immun 66, 4811-4816.

Kronhardt, A., Rolando, M., Beitzinger, C., Stefani, C., Leuber, M., Flatau, G., Popoff, M.R., Benz, R., Lemichez, E., 2011. Cross-reactivity of anthrax and C2 toxin: protective antigen promotes the uptake of botulinum C2I toxin into human endothelial cells. PLoS One 6, e23133.

Kubota, T., Shirakawa, S., Kozaki, S., Isogai, E., Isogai, H., Kimura, K., Fujii, N., 1996. Mosaic structure of the nontoxic-nonhemagglutinating component gene in *Clostridium botulinum* type A strain isolated from infant botulism in Japan. Biochem. Biophys. Res. Commun. 224, 843-848.

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

Lindstrom, M., Hinderink, K., Somervuo, P., Kiviniemi, K., Nevas, M., Chen, Y., Auvinen, P., Carter, A.T., Mason, D.R., Peck, M.W., Korkeala, H., 2009. Comparative genomic hybridization analysis of two predominant Nordic group I (proteolytic) Clostridium botulinum type B clusters. Appl Environ Microbiol 75, 2643-2651.

Macdonald, T.E., Helma, C.H., Shou, Y., Valdez, Y.E., Ticknor, L.O., Foley, B.T., Davis, S.W., Hannett, G.E., Kelly-Cirino, C.D., Barash, J.R., Arnon, S.S., Lindstrom, M., Korkeala, H., Smith, L.A., Smith, T.J., Hill, K.K., 2011. Analysis of Clostridium

botulinum Serotype E Strains by Using Multilocus Sequence Typing, Amplified Fragment Length Polymorphism, Variable-Number Tandem-Repeat Analysis, and Botulinum Neurotoxin Gene Sequencing. Appl Environ Microbiol 77, 8625-8634.

Macdonald, T.E., Helma, C.H., Ticknor, L.O., Jackson, P.J., Okinaka, R.T., Smith, L.A., Smith, T.J., Hill, K.K., 2008. Differentiation of Clostridium botulinum serotype A strains by multiple-locus variable-number tandem-repeat analysis. Appl Environ Microbiol 74, 875-882.

Mani, N., Dupuy, B., 2001. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. Proc. Ntl. Acad. Sci; (USA) 98, 5844-5849.

Manich, M., Knapp, O., Gibert, M., Maier, E., Jolivet-Reynaud, C., Geny, B., Benz, R., Popoff, M.R., 2008. Clostridium perfringens delta toxin is sequence related to beta toxin, NetB, and Staphylococcus pore-forming toxins, but shows functional differences. PLoS ONE 3, e3764.

Marshall, K.M., Bradshaw, M., Johnson, E.A., 2010. Conjugative Botulinum Neurotoxin-Encoding Plasmids in Clostridium botulinum. PLoS One 5.

Marshall, K.M., Bradshaw, M., Pellet, S., Johnson, E.A., 2007. Plasmid encoded neurotoxin genes in *Clostridium botulinum* serotype A subtypes. Biochem Biophys Res Commun 361, 49-54.

Marvaud, J.C., Eisel, U., Binz, T., Niemann, H., Popoff, M.R., 1998. *tetR* is a positive regulator of the Tetanus toxin gene in *Clostridium tetani* and is homologous to *botR*. Infect. Immun. 66, 5698-5702.

McCroskey, L.M., Hatheway, C.L., Fenicia, L., Pasolini, B., Aureli, P., 1986. Characterization of an organism that produces type E botulinal toxin but which resembles *Clostridium butyricum* from the feces of an infant with type E botulism. J. Clin. Microbiol. 23, 201-202.

McCroskey, L.M., Hatheway, C.L., Woodruff, B.A., Greenberg, J.A., Jurgenson, P., 1991. Type F botulism due to neurotoxigenic *Clostridium baratii* from an unknown source in an adult. J. Clin. Microbiol. 29, 2618-2620.

Ménétrey, J., Flatau, G., Stura, E.A., Charbonnier, J.B., Gas, F., Teulon, J.M., Le Du, M.H., Boquet, P., Ménez, A., 2002. NAD binding induces conformational changes in Rho ADP-ribosylating *Clostridium botulinum* C3 exoenzyme. J. Biol. Chem. 277, 30950-30957.

Meng, X., Karasawa, T., Zou, K., Kuang, X., Wang, X., Lu, C., Wang, C., Yamakawa, K., Nakamura, S., 1997. Characterization of a neurotoxigenic *Clostridium butyricum* strain

isolated from the food implicated in an outbreak of food-borne type E botulism. J. Clin. Microbiol. 35, 2160-2162.

Meng, X., Yamakawa, K., Zou, K., Wang, X., Kuang, X., Lu, C., Wang, C., Karasawa, T., Nakamura, S., 1999. Isolation and characterization of neurotoxigenic *Clostridium butyricum* from soil in China. J. Med. Microbiol. 48, 133-137.

Metcalf, D.S., Weese, J.S., 2011. Binary toxin locus analysis in Clostridium difficile. J Med Microbiol 60, 1137-1145.

Minton, N.P., 1995. Molecular genetics of clostridial neurotoxins. Curr. Top. Microbiol. Immunol. 195, 161-194.

Miyamoto, K., Chakrabarti, G., Morino, Y., McClane, B.A., 2002. Organization of the plasmid *cpe* locus in *Clostridium perfringens* type A isolates. Infect. Immun. 70, 4261-4272.

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.

Moran, Y., Fredman, D., Szczesny, P., Grynberg, M., Technau, U., 2012. Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. Mol Biol Evol 29, 2223-2230.

Moriishi, K., Koura, M., Aba, N., Fujii, N., Fujinaga, Y., Inoue, K., Oguma, K., 1996a. Mosaic structures of neurotoxins produced from *Clostridium botulinum* types C and D. Biochem. Biophys. Acta 1307, 123-126.

Moriishi, K., Koura, M., Fujii, N., Fujinaga, Y., Inoue, K., Syuto, B., Oguma, K., 1996b. Molecular cloning of the gene encoding the mosaic neurotoxin, composed of parts of botulinum neurotoxin types C1 and D, and PCR detection of this gene from Clostridium botulinum type C organisms. Appl Environ Microbiol 62, 662-667.

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 Res 16, 1031-1040.

Nakamura, K., Kohda, T., Umeda, K., Yamamoto, H., Mukamoto, M., Kozaki, S., 2010. Characterization of the D/C mosaic neurotoxin produced by Clostridium botulinum associated with bovine botulism in Japan. Vet Microbiol 140, 147-154.

Nevas, M., Lindström, M., Hielm, S., Björkroth, K.J., Peck, M.W., Korkeala, H., 2005. Diversity of proteolytic *Clostridium botulinum* strains, determined by a pulse-field gel electrophoresis approach. Appl Environ Microbiol 71, 1311-1317.

Oguma, K., Inoue, K., Fujinaga, Y., Yokota, K., Watanabe, T., Ohyama, T., Takeshi, K., Inoue, K., 1999. Structure and function of *Clostridium botulinum* progenitor toxin. J. Toxicol. 18, 17-34.

Peck, M.W., 2009. Biology and genomic analysis of Clostridium botulinum. Adv Microb Physiol 55, 183-265, 320.

Perelle, S., Gibert, M., Bourlioux, P., Corthier, G., Popoff, M.R., 1997. Production of a complete binary toxin (actin-specific ADP-ribosylating toxin) by *Clostridium difficile* CD196. Infect. Immun. 65, 1402-1407.

Petit, L., Gibert, M., Popoff, M.R., 1999. *Clostridium perfringens*: toxinotype and genotype. Trends Microbiol. 7, 104-110.

Petosa, C., Collier, J.R., Klimpel, K.R., Leppla, S.H., Liddington, R.C., 1997. Crystal structure of the anthrax toxin protective antigen. Nature (London) 385, 833-838.

Popoff, M.R., 2000. Molecular biology of actin-ADP-ribosylating toxins, in: Aktories, K., Just, I. (Eds.), Bacterial Protein Toxins. Springer, Berlin, pp. 275-302.

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

Popoff, M.R., Marvaud, J.C., 1999. Structural and genomic features of clostridial neurotoxins, in: Alouf, J.E., Freer, J.H. (Eds.), The Comprehensive Sourcebook of Bacterial Protein Toxins, 2 ed. Academic Press, London, pp. 174-201.

Popoff, M.R., Rubin, E.J., Gill, D.M., Boquet, P., 1988. Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. Infect. Immun. 56, 2299-2306.

Poulain, B., Popoff, M.R., Molgo, J., 2008. How do the botulinum neurotoxins block neurotransmitter release: from botulism to the molecular mechanism of action. Botulinum J. 1, 14-87.

Poulain, B., Stiles, B.G., Popoff, M.R., Molgó, J., 2006. Attack of the nervous system by clostridial toxins: Physical findings, cellular and molecular actions, in: Alouf, J.E., Popoff, M.R. (Eds.), The Sourcebook of Bacterial Protein Toxins, 3° ed. Elsevier, Academic Press, Amsterdam, pp. 348-389.

Poulet, S., Hauser, D., Quanz, M., Niemann, H., Popoff, M.R., 1992. Sequences of the botulinal neurotoxin E derived from *Clostridium botulinum* type E (strain Beluga) and

Clostridium butyricum (strains ATCC43181 and ATCC43755). Biochem. Biophys. Res. Commun. 183, 107-113.

Prévost, G., Mourey, L., Colin, D.A., Monteil, H., Dalla sera, M., Menestrina, G., 2006. Alpha-helix and beta-barrel pore-forming toxins (leucocidins, alpha-, gamma-, and delta-cytolysins) of *Staphylococcus aureus*, in: Alouf, J.E., Popoff, M.R. (Eds.), The Comprehensive Sourcebook of Bacterial Protein Toxins, 3° ed. Elsevier Academic Press, Amsterdam, pp. 590-607.

Pruitt, R.N., Chumbler, N.M., Rutherford, S.A., Farrow, M.A., Friedman, D.B., Spiller, B., Lacy, D.B., 2012. Structural determinants of Clostridium difficile toxin A glucosyltransferase activity. J Biol Chem 287, 8013-8020.

Raffestin, S., Dupuy, B., Marvaud, J.C., Popoff, M.R., 2005. BotR/A and TetR are alternative RNA polymerase sigma factors controlling the expression of the neurotoxin and associated protein genes in *Clostridium botulinum* type A and *Clostridium tetani*. Mol. Microbiol. 55, 235-249.

Raphael, B.H., Choudoir, M.J., Luquez, C., Fernandez, R., Maslanka, S.E., 2010. Sequence diversity of genes encoding botulinum neurotoxin type F. Appl Environ Microbiol 76, 4805-4812.

Raphael, B.H., Lautenschlager, M., Kalb, S.R., de Jong, L.I., Frace, M., Luquez, C., Barr, J.R., Fernandez, R.A., Maslanka, S.E., 2012. Analysis of a unique *Clostridium botulinum* strain from the Southern hemisphere producing a novel type E botulinum neurotoxin subtype. BMC Microbiol 12, 245.

Raphael, B.H., Luquez, C., McCroskey, L.M., Joseph, L.A., Jacobson, M.J., Johnson, E.A., Maslanka, S.E., Andreadis, J.D., 2008. Genetic homogeneity of Clostridium botulinum type A1 strains with unique toxin gene clusters. Appl Environ Microbiol 74, 4390-4397.

Reinert, D.J., Jank, T., Aktories, K., Schulz, G.E., 2005. Structural basis for the function of Clostridium difficile toxin B. J Mol Biol 351, 973-981.

Rood, J.I., 1998. Virulence genes of *Clostridium perfringens*. Annu. Rev. Microbiol. 52, 333-360.

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.

Rossjohn, J., Polekhina, G., Feil, S.C., Morton, C.J., Tweten, R.K., Parker, M.W., 2007. Structures of perfringolysin O suggest a pathway for activation of cholesterol-dependent cytolysins. J Mol Biol 367, 1227-1236.

Rupnik, M., 2008. Heterogeneity of large clostridial toxins: importance of Clostridium difficile toxinotypes. FEMS Microbiol Rev 32, 541-555.

Rupnik, M., Grabnar, M., Geric, B., 2003. Binary toxin producing *Clostridium difficile* strains. Anaerobe 9, 289-294.

Rupnik, M., Just, I., 2006. Large clostridial cytotoxins modifying small GTPases, in: Alouf, J.E., Popoff, M.R. (Eds.), The Sourcebook of Bacterial Protein Toxins, 3° ed. Elsevier, Academic Press, Amsterdam, pp. 409-429.

Sakaguchi, Y., Hayashi, H., Kurokawa, K., Nakayama, K., Oshima, K., Fujinaga, Y., Ohnishi, M., Ohtsubo, E., Hattori, M., Oguma, K., 2005. The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. Proc Natl Acad Sci U S A 102, 17472-17477.

Sakaguchi, Y., Hayashi, T., Yamamoto, Y., Nakayama, K., Zhang, K., Ma, S., Arimitsu, H., Oguma, K., 2009. Molecular analysis of an extrachromosomal element containing the C2 toxin gene discovered in Clostridium botulinum type C. J Bacteriol 191, 3282-3291.

Santos-Buelga, J., Collins, M.D., East, A.K., 1998. Characterization of the genes encoding the botulinum neurotoxin complex in a strain of *Clostridium botulinum* producing type B and F neurotoxins. Cur. Microbiol. 37, 312-318.

Savva, C.G., Fernandes da Costa, S.P., Bokori-Brown, M., Naylor, C.E., Cole, A.R., Moss, D.S., Titball, R.W., Basak, A.K., 2012. Molecular architecture and functional analysis of NetB, a pore-forming toxin from Clostridium perfringens. J Biol Chem.

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

Sayeed, S., Li, J., McClane, B.A., 2009. Characterization of Virulence Plasmid Diversity Amongst Clostridium perfringens Type B Isolates. Infect Immun.

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

Schallehn, G., Eklund, M.W., Brandis, H., 1980. Phage conversion of *Clostridium novyi* type A. Zentralbl. Bakteriol. 247, 95-100.

Schleberger, C., Hochmann, H., Barth, H., Aktories, K., Schulz, G.E., 2006. Structure and action of the binary C2 toxin from *Clostridium botulinum*. J. Mol. Biol. 364, 705-715.

Sebaihia, M., Peck, M.W., Minton, N.P., Thomson, N.R., Holden, M.T., Mitchell, W.J., Carter, A.T., Bentley, S.D., Mason, D.R., Crossman, L., Paul, C.J., Ivens, A., Wells-Bennik, M.H., Davis, I.J., Cerdeno-Tarraga, A.M., Churcher, C., Quail, M.A., Chillingworth, T., Feltwell, T., Fraser, A., Goodhead, I., Hance, Z., Jagels, K., Larke, N., Maddison, M., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., White, B., Whithead, S., Parkhill, J., 2007. Genome sequence of a proteolytic (Group I) Clostridium botulinum strain Hall A and comparative analysis of the clostridial genomes. Genome Res 17, 1082-1092.

Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeno-Tarraga, A.M., Wang, H., Holden, M.T., Wright, A., Churcher, C., Quail, M.A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B., Parkhill, J., 2006. The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. Nat. Genet. 38, 779-786.

Sharma, S.K., Ramzan, M.A., Singh, B.R., 2003. Separation of the components of type A botulinum neeurotoxin complex by electrophoresis. Toxicon 41, 321-331.

Shatursky, O., Heuck, A., Shepard, L., Rossjhon, J., Parker, M., Johnson, A., Tweten, R., 1999. The mechanism of membrane insertion of a cholesterol-dependent cytolysin: a novel paradigm for pore-forming toxins. Cell 99, 293-299.

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. Proc. Natl. Acad. Sci (USA) 99, 996-1001.

Skarin, H., Hafstrom, T., Westerberg, J., Segerman, B., 2011. Clostridium botulinum group III: a group with dual identity shaped by plasmids, phages and mobile elements. BMC Genomics 12, 185.

Skarin, H., Segerman, B., 2011. Horizontal gene transfer of toxin genes in Clostridium botulinum: Involvement of mobile elements and plasmids. Mob Genet Elements 1, 213-215.

Smith, L.D.S., 1975. *Clostridium tetani*, in: Smith, L.D.S. (Ed.), The Pathogenic Anaerobic Bacteria. Charles C. Thomas Publisher, Springfield Ill, pp. 177-201.

Smith, T.J., Hill, K.K., Foley, B.T., Detter, J.C., Munk, A.C., Bruce, D.C., Doggett, N.A., Smith, L.A., Marks, J.D., Xie, G., Brettin, T.S., 2007. Analysis of the Neurotoxin Complex Genes in Clostridium botulinum A1-A4 and B1 Strains: BoNT/A3, /Ba4 and /B1 Clusters Are Located within Plasmids. PLoS ONE 2, e1271.

Smith, T.J., Lou, J., Geren, I.N., Forsyth, C.M., Tsai, R., Laporte, S.L., Tepp, W.H., Bradshaw, M., Johnson, E.A., Smith, L.A., Marks, J.D., 2005. Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization. Infect Immun 73, 5450-5457.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., Gouaux, J.E., 1996. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. Science 274, 1859-1866.

Spigaglia, P., Mastrantonio, P., 2003. Molecular analysis of the pathogenicity locus and polymorphism in the putative negative regulator of toxin production (TcdC) among *Clostridium difficile* clinical isolates. J. Clin. Microbiol. 40, 3470-3475.

Stabler, R.A., Dawson, L.F., Valiente, E., Cairns, M.D., Martin, M.J., Donahue, E.H., Riley, T.V., Songer, J.G., Kuijper, E.J., Dingle, K.E., Wren, B.W., 2012. Macro and micro diversity of Clostridium difficile isolates from diverse sources and geographical locations. PLoS One 7, e31559.

Stabler, R.A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T.D., Sebaihia, M., Quail, M.A., Rose, G., Gerding, D.N., Gibert, M., Popoff, M.R., Parkhill, J., Dougan, G., Wren, B.W., 2009. Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol 10, R102.

Stabler, R.A., Valiente, E., Dawson, L.F., He, M., Parkhill, J., Wren, B.W., 2010. In-depth genetic analysis of Clostridium difficile PCR-ribotype 027 strains reveals high genome fluidity including point mutations and inversions. Gut Microbes 1, 269-276.

Stackebrandt, E., Hippe, H., 2001. Taxonomy and systematics, in: Bahl, H., Dürre, P. (Eds.), Clostridia. Willey-VCH, Weinheim, pp. 19-48.

Stare, B.G., Delmee, M., Rupnik, M., 2007. Variant forms of the binary toxin CDT locus and tcdC gene in Clostridium difficile strains. J Med Microbiol 56, 329-335.

Stiles, B.G., Wigelsworth, D.J., Popoff, M.R., Barth, H., 2011. Clostridial binary toxins: iota and c2 family portraits. Front Cell Infect Microbiol 1, 11.

Strom, M.S., Eklund, M.W., Poysky, F.T., 1984. Plasmids in *Clostridium botulinum* and related species. Appl. Environ. Microbiol. 48, 956-963.

Stubbs, S., Rupnik, M., Gibert, M., Brazier, J., Duerden, B., Popoff, M.R., 2000. Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*. FEMS Microbiol. Lett. 186, 307-312.

Suen, J.C., Hatheway, C.L., Steigerwalt, A.G., Brenner, D.J., 1988. *Clostridium argentinense* sp. nov.: a genetically homogeneous group composed of all strains of *Clostridium botulinum* toxin type G and some nontoxigenic strains previously identified as *Clostridium subterminale* or *Clostridium hastiforme*. Int. J. Syst. Bacteriol. 38, 375-381.

Tabita, K.S., Sakaguchi, S., kozaki, S., Sakaguchi, G., 1991. Distinction between *Clostridium botulinum* type A strains associated with food-borne botulism and those with infant botulism in Japan in intraintestinal toxin production in infant mice and some other properties. FEMS Microbiol. Lett. 79, 251-256.

Takeda, M., Tsukamoto, K., Kohda, T., Matsui, M., Mukamoto, M., Kozaki, S., 2005. Characterization of the neurotoxin produced by isolates associated with avian botulism. Avian Dis 49, 376-381.

Titball, R.W., Naylor, C.E., Basak, A.K., 1999. The *Clostridium perfringens* α-toxin. Anaerobe 5, 51-64.

Tsutsui, K., Minami, J., Matsushita, O., Katayama, S., Taniguchi, Y., Nakamura, S., Nishioka, M., Okabe, A., 1995. Phylogenetic analysis of phospholipase C genes from *Clostridium perfringens* types A to E and *Clostridium novyi*. J. Bacteriol. 177, 7164-7170.

Tweten, R.K., 2005. Cholesterol-dependent cytolysins, a family of versatile pore-forming toxins. Infect. Immun. 73, 6199-6209.

Valiente, E., Dawson, L.F., Cairns, M.D., Stabler, R.A., Wren, B.W., 2012. Emergence of new PCR ribotypes from the hypervirulent Clostridium difficile 027 lineage. J Med Microbiol 61, 49-56.

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. Mol Microbiol 57, 1570-1581.

Wang, X., Maegawa, T., Karazawa, T., Kozaki, S., Tsukamoto, K., Gyobu, Y., Yamakawa, K., Oguma, K., Sakaguchi, Y., Nakamura, S., 2000. Genetic analysis of Type E botulism toxin-producing *Clostridium butyricum* strains. Appl. Environ. Microbiol. 66, 4992-4997.

Weickert, M.J., Chambliss, G.H., Sugiyama, H., 1986. Production of toxin by *Clostridium botulinum* type A strains cured of plasmids. Appl. Environ. Microbiol. 51, 52-56.

Willems, A., East, A.K., Lawson, P.A., Collins, M.D., 1993. Sequence of the gene coding for the neurotoxin of *Clostridium botulinum* type A associated with infant botulism: comparison with other clostridial neurotoxins. Res. Microbiol. 144, 547-556.

Wren, B.W., 1991. A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. Mol. Microbiol. 5, 797-803.

Yamakawa, K., Karasawa, T., Kakinuma, H., Maruyama, H., Takahashi, H., Nakamura, S., 1997. Emergence of *Clostridium botulinum* typeB-like nontoxigenic organisms in a patient with type B infant botulism. J. Clin. Microbiol. 35, 2163-2164.

Zhou, Y., Sugiyama, H., Johnson, E.A., 1993. Transfer of neurotoxigenicity from *Clostridium butyricum* to a nontoxigenic *Clostridium botulinum* type E-like strain. Appl. Environ. Microbiol. 59, 3825-3831.

FIGURE LEGENDS

Figure 1. Examples of toxin evolution by gene duplication.

A. *C. difficile* toxin genes (*tcdA* and *tcdB*) probably derive from a common glucosyltransferase gene by duplication. TcdA and TcdB shares 36% amino acid sequence identity, but are structurally related at least regarding their N-terminal catalytic domains (Pruitt et al., 2012; Reinert et al., 2005).

B. Duplication of an ancestor collagenase gene in *C. botulinum* possibly yielded two structurally related proteins sharing about 20% amino acid sequence, a potent *botulinum* neurotoxin (BoNT) and a non-toxic non-hemagglutinin (NTNH) protein, which assembles in a complex highly resistant to acidic pH and protease degradation (Gu et al., 2012; Inui et al., 2012).

C. The enzymatic component (C2-I) gene likely derives from C3 gene by duplication. C2-I contains two structurally related domains similar to C3 enzyme. The C-terminal domain of C2-I exhibits an ADP-ribosyltransferase activity as C3 enzyme but recognizes a different substrate, monomeric actin instead of Rho proteins. The C2-I N-termninal domain has acquired a new function consisting in the interaction with the binding component C2-II, which facilitates its transport into the cells (Han et al., 2001; Schleberger et al., 2006).

Figure 2. Representative toxin structures of the three main β-PFT families: cholesterol-dependent cytolysin (CDC), aerolysin, and alpha-hemolysin families. Perfringolysin is the protoptype of the CDC family and shows structural homology with the binding components of the binary toxins from *Bacillus* (Protective antigen) and *Clostridium* (C2-II) (Petosa et al., 1997; Rossjohn et al., 2007; Schleberger et al., 2006). The aerolysin family encompasses various toxins from diverse origin, bacteria such as *C. perfringens* epsilon toxin and *C. perfringens* enterotoxin, animal, plant like the lectin from the mushroom *Laetiporus sulphurous* (Briggs et al., 2011; Cole et al., 2004; Kitadokoro et al., 2011; Knapp et al., 2010). The domain containing the amphipatic β-hairpin that contributes to the β-barrel is in red. *C. perfringens* Delta toxin is structurally related to *S. aureus* alpha hemolysin (Song et al., 1996) (Huyet et al. submitted). The amphipatic β-hairpin (red) is packed against the β-sadnwich domain in the soluble form and unfolds to assemble in the β-barrel.

Figure 3. Hypothetical evolutionary lineages of bacterial toxin genes. Based on their biological activity and structure, two main classes of bacterial toxins can be distinguished: pore-forming toxins (PFTs) and toxins active through an enzymatic activity. PFTs likely derive from a common ancestor, probably a transmembrane protein ancestor. PFTs are the most common bacterial toxins (about on third of total bacterial toxins) and mainly belong to the β-PFT family, which is characterized by a conserved basic structure with a receptor binding domain, one or two oligomerization domain(s), and one domain containing one or two helices able to unfold in amphipatic hairpin(s) to form the β-barrel. Horizontal gene transfer and subsequent evolution in distinct bacterial species and strains account for the diversity of bacterial β-PFTs. The cholesterol dependent cytolysins (CDC), which form large pores, and PFTs related to the hexameric or heptameric staphylococcal α-toxin are largely spread in Gram-positive bacteria. Indeed, CDCs, the prototype of which being Perfringolysin from *C. perfringens*, are produced by 9 *Clostridium* species, 5 *Bacillus* species, 6 *Streptococcus* species, 3 *Listeria* species (Alouf, 2003; Tweten, 2005), and the staphylococcal α-toxin-related proteins include *Staphylococcus* leucocidins, *C. perfringens* Beta, Delta and NetB toxins, and *Bacillus cereus* hemolysin II and cytotoxin K α-toxin α-toxin (Manich et al., 2008; Prévost et al., 2006). Members of the aerolysin family contain 3 domains as staphylococcal α-toxin (except aerolysin which exhibits an additional N-terminal domain), but show a more elongated shape. They are widely spread in various organisms (Gram-positive bacteria such as *C. perfringens* epsilon and enterotoxin, *C. septicum* α-toxin, *Bacillus sphaericus* mosquitocidal toxin Mtx, *Bacillus thuringiensis* parasporin, Gram negative bacteria such as aerolysin from *Aeromonas*, *Vibrio*, *Pseudomonas*, plant like enterolobin of the Brazilian tree *Enterolobium contortisiliquum*, letosporin (or LSL) of the pathogenic fungus *Laetiporus sulphureus*, and animal like hydralysins of the cindarian genus *Hydra*, thus supporting horizontal gene transfer between bacteria and eukaryotes (Briggs et al., 2011; Knapp et al., 2010; Manich et al., 2008; Moran et al., 2012). CDCs and membrane attack complex/perforin (MACPF) proteins retain a conserved structural pore-forming domain, which possibly reflects a convergent evolution of eukaryotic and prokaryotic proteins to interact with lipid bilayers. Toxins exhibiting an enzymatic activity likely derive from enzyme precursors, which have also lead to hydrolytic enzymes, or these toxins represent a more differentiated state of hydrolytic enzymse by acquisition of a specific receptor binding domain and/or translocation domain allowing toxin internalization into cells. Since hydrolytic

enzymes are abundantly produced by Clostridia for their metabolism, this raises the question whether toxins with enzymatic activity have a clostridial origin, Interestingly, binary toxins produced by certain Clostridium and Bacillus, seem to have emerged from a convergent or cross evolution between PFTs and toxins having an enzymatic activity. Binding components, which are structurally related to β-PFTs of the CDC family and retain similarity with aerolysin toxins, have acquired have evolved to specifically internalize an enzymatic protein from a different origin into cell trough a pore-forming mechanism.

Figure
Click here to download high resolution image

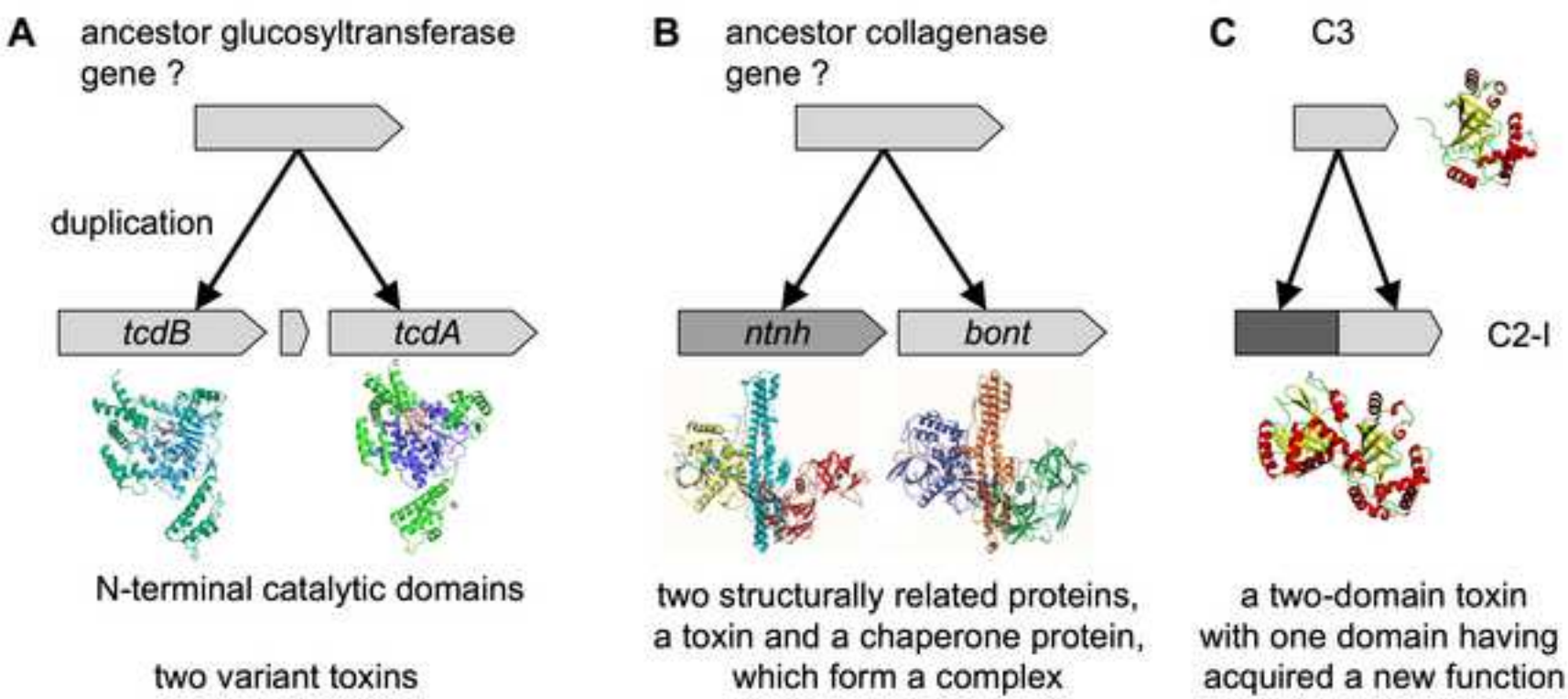

Figure 1

Figure
Click here to download high resolution image

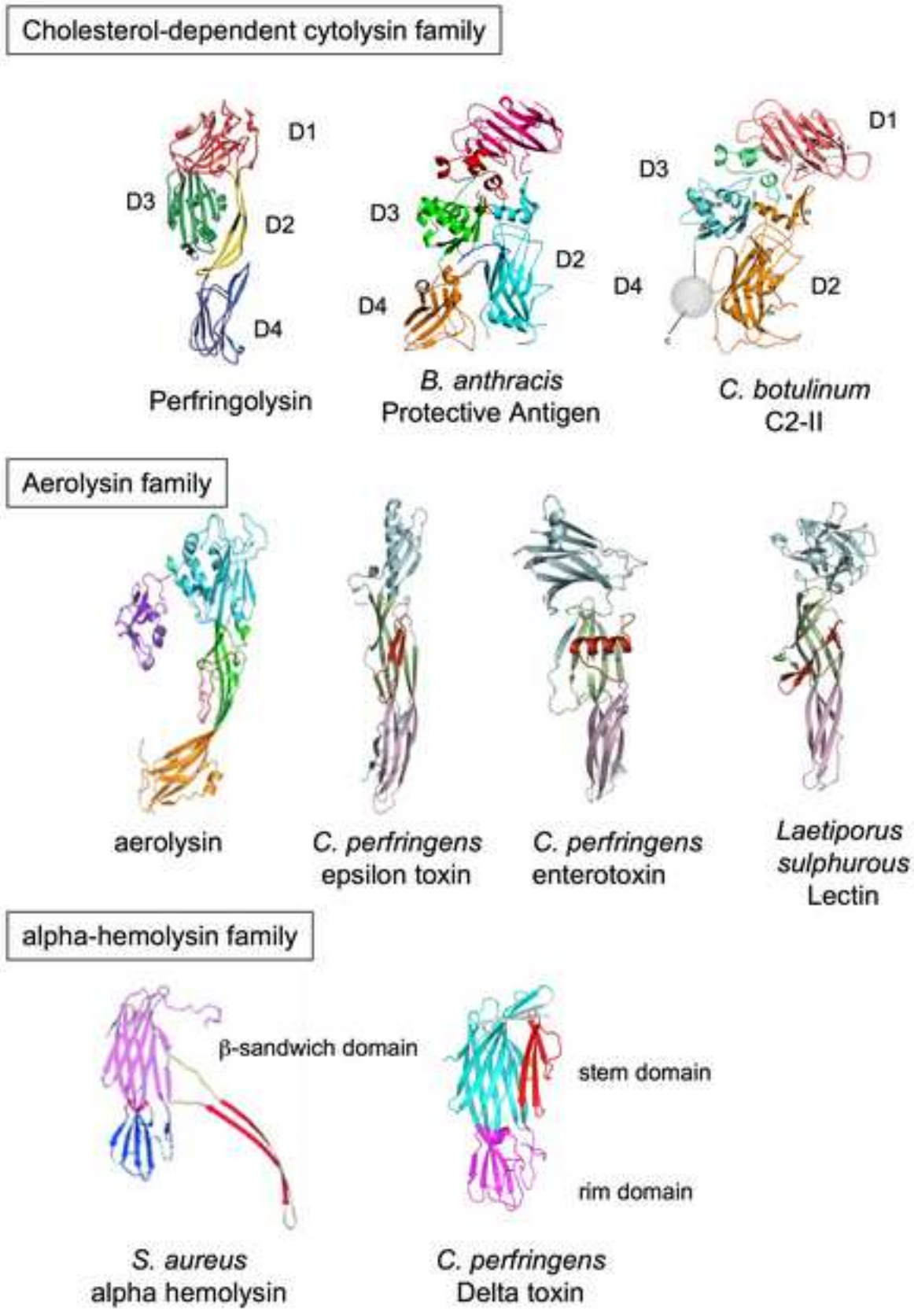

Figure 2

Figure
Click here to download high resolution image

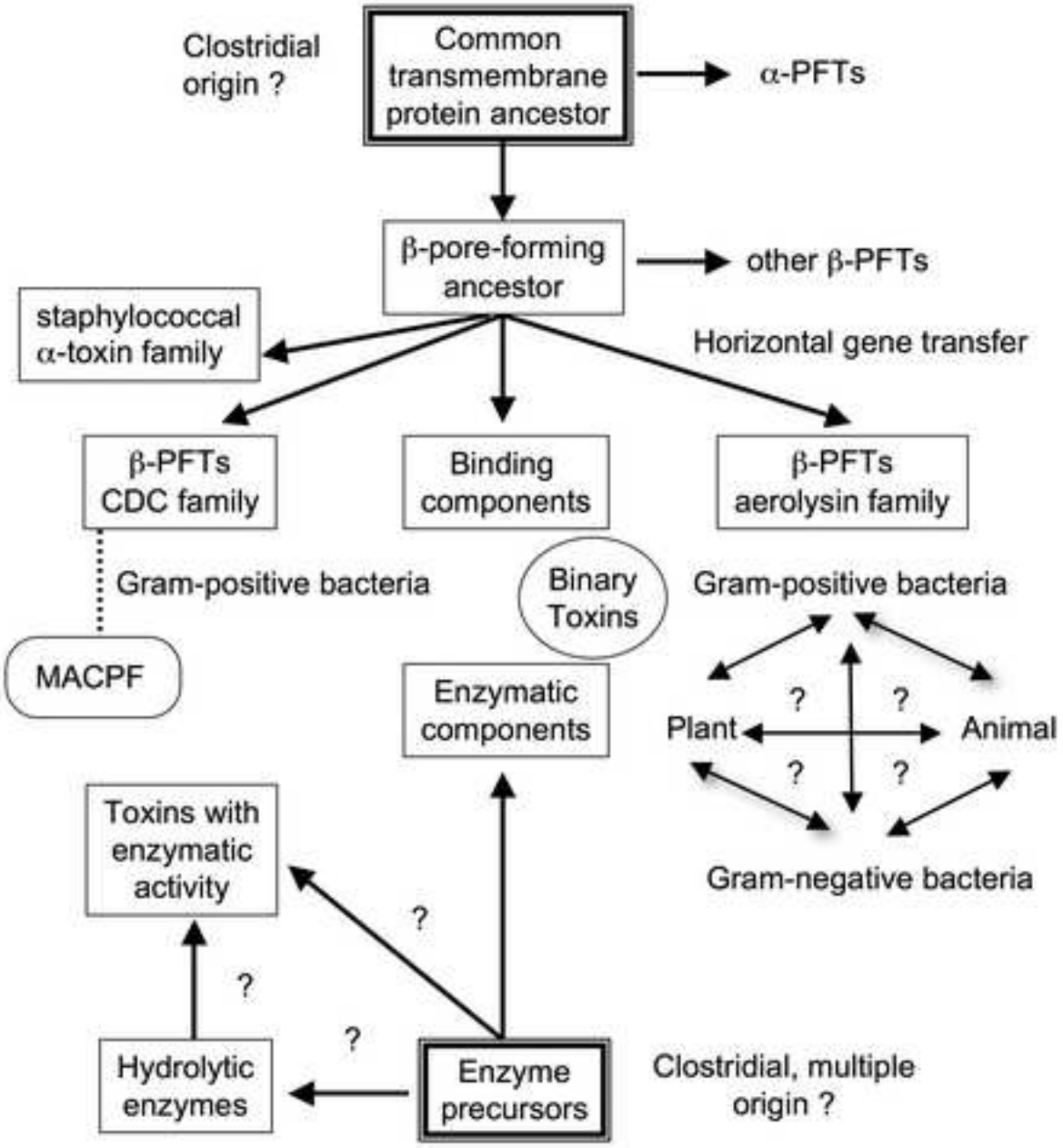

Figure 3