The Type F6 Neurotoxin Gene Cluster Locus of Group II *Clostridium botulinum* Has Evolved by Successive Disruption of Two Different Ancestral Precursors

Andrew T. Carter*, Sandra C. Stringer, Martin D. Webb, and Michael W. Peck

Department of Gut Health and Food Safety, Institute of Food Research, Norwich Research Park, Norwich, United Kingdom

*Corresponding author: E-mail: andrew.carter@ifr.ac.uk.

Accepted: April 29, 2013

Data deposition: This project has been deposited at GenBank under the accession numbers KC516868 (IFR 06/001), KC516869 (IFR 06/005), KC516870 (Craig 610), KC516871 (Eklund 202F), and KC516872 (Hobbs FT10).

Abstract

Genome sequences of five different Group II (nonproteolytic) *Clostridium botulinum* type F6 strains were compared at a 50-kb locus containing the neurotoxin gene cluster. A clonal origin for these strains is indicated by the fact that sequences were identical except for strain Eklund 202F, with 10 single-nucleotide polymorphisms and a 15-bp deletion. The essential *topB* gene encoding topoisomerase III was found to have been split by the apparent insertion of 34.4 kb of foreign DNA (in a similar manner to that in Group II *C. botulinum* type E where the *rarA* gene has been disrupted by a neurotoxin gene cluster). The foreign DNA, which includes the intact 13.6-kb type F6 neurotoxin gene cluster, bears not only a newly introduced *topB* gene but also two nonfunctional botulinum neurotoxin gene remnants, a type B and a type E. This observation combined with the discovery of bacteriophage integrase genes and IS4 elements suggest that several rounds of recombination/horizontal gene transfer have occurred at this locus. The simplest explanation for the current genotype is that the ancestral bacterium, a Group II *C. botulinum* type B strain, received DNA firstly from a strain containing a type E neurotoxin gene cluster, then from a strain containing a type F6 neurotoxin gene cluster. Each event disrupted the previously functional neurotoxin gene. This degree of successive recombination at one hot spot is without precedent in *C. botulinum*, and it is also the first description of a Group II *C. botulinum* genome containing more than one neurotoxin gene sequence.

Key words: genome comparison, synteny, horizontal gene transfer, botulinum toxin gene.

Introduction

Botulinum neurotoxin is responsible for botulism, a severe and deadly neuroparalytic disease. It is the most potent toxin known, with as little as 30–100 ng potentially fatal to a human adult (Peck 2009). There are seven types of botulinum neurotoxin (types A–G), which are formed by the Gram-positive anaerobe *Clostridium botulinum* and some strains of *C. butyricum* and *C. baratii*. Types A, B, E, and F neurotoxins have been associated with human pathology (Austin and Dodds 2000; Peck 2006). *Clostridium botulinum* is a heterogeneous species that comprises four phylogenetically and physiologically distinct bacteria (*C. botulinum* Groups I–IV), with Groups I and II associated with human botulism (Austin and Dodds 2000; Peck 2009). Strains of Group I (proteolytic) *C. botulinum* form type A, B, or F neurotoxin, with many strains carrying genes for two neurotoxins, and in some cases also forming two neurotoxins. Strains of Group II (nonproteolytic) *C. botulinum* form type B, E, or F neurotoxins, but there are currently no reports of strains carrying more than one neurotoxin gene (Macdonald et al. 2011; Peck et al. 2011). One interpretation for this observation would be that horizontal transfer of the neurotoxin genes occurs less frequently in Group II *C. botulinum* than it does in Group I *C. botulinum*. Recent genetic analysis has identified a number of neurotoxin subtypes, with particular subtypes often found only in one Group of *C. botulinum* (Smith et al. 2005; Hill et al. 2007; Macdonald et al. 2011; Peck et al. 2011). For example, strains of Group II *C. botulinum* type F, which are the subject of this study, exclusively carry the type F6 neurotoxin gene, whereas strains of Group I *C. botulinum* type F carry a neurotoxin gene, which can be of types F1–F5, and strains of *C. baratii* all carry the type F7 neurotoxin gene (Raphael et al. 2010).

The neurotoxin has several accessory proteins, the genes for which are adjacent and form the so-called neurotoxin gene cluster. There are two classes; the *ha*-cluster encodes the neurotoxin, a nontoxic-nonhaemagglutinin protein, and three hemagglutinins; and the *orf-x*-cluster encodes the neurotoxin, a nontoxic-nonhaemagglutinin protein, and four other proteins (Orf-X1, Orf-X2, Orf-X3, and P47) of unknown function (Hill et al. 2009; Peck 2009; Peck et al. 2011). In addition, clusters often contain the gene for a regulatory sigma factor, *botR*. When chromosomally located, botulinum neurotoxin gene clusters are usually flanked by fragments of mobile elements, an observation which has raised the possibility that they may have been acquired by horizontal gene transfer (Sebaihia et al. 2007). Neurotoxin gene clusters can also be located on bacteriophages (Group III *C. botulinum*) and plasmids (Groups I, II, and IV). Horizontal acquisition of neurotoxin gene clusters must occur relatively rarely on an evolutionary timescale, based on the knowledge that in Group I, *C. botulinum* neurotoxin gene clusters are found at only three different chromosomal locations, and more dramatically in both Group II *C. botulinum* type E and in *C. butyricum* type E, the neurotoxin gene cluster has been inserted directly into *rarA*, a recombination-associated gene (Hill et al. 2009).

Recent work in our laboratory studied the genetic relationship between members of Group II *C. botulinum* using a whole-genome DNA microarray. This study extended the observations of others that strains of Group II *C. botulinum* type B and type F are closely related but that Group II *C. botulinum* type E strains form a distinct clade (Stringer et al. 2013). Reference genome sequences were already available for Group II *C. botulinum* type B (Eklund17B) and type E (strains Beluga and Alaska), so we decided to sequence the genome of each type F strain from our microarray study to explain this close relationship with type B strains.

Materials and Methods

Bacterial Strains and Growth

Five strains of Group II (nonproteolytic) *C. botulinum* type F were used in this study (table 1). Three historical strains have been widely studied by previous authors, whereas two strains were recently isolated from scallops (Peck et al. 2010). Before use, all strains were checked for purity by growth on peptone-yeast-glucose-starch (PYGS) and reinforced clostridial medium with 5% (w/v) skim milk agar plates incubated under aerobic and anaerobic atmospheres (Carter et al. 2009).

DNA Preparation and Sequencing

Genomic DNA was purified from exponentially growing cells cultured in PYGS medium as described previously (Carter et al. 2011). Library construction and genome sequencing was carried out by The Genome Analysis Centre (TGAC), Norwich Research Park, UK, on the Illumina MiSeq platform, which produces longer sequencing reads than previous Illumina platforms (MiSeq Personal Sequencer, Illumina at http://www.illumina.com/systems/miseq.ilmn, last accessed May 21, 2013). Reads were assembled using the short-read assembler ABySS (Simpson et al. 2009) with custom in house refinements. All sequencing reads were assembled to generate an improved high-quality draft sequence (Chain et al. 2009).

Sequence Analysis

Sequences were loaded into Vector NTI Advance 11 (Invitrogen), and for each genome, the contig that contained the type F6 toxin gene cluster was identified. Pairwise alignments of homologous contig regions were performed using the Vector NTI program AlignX, which uses the ClustalW algorithm. The Artemis Comparison Tool (ACT) was used for whole-genome comparisons (Carver et al. 2005). Open reading frames (ORFs) of minimum 150-bp length, identified using standard Vector NTI software set to use ATG, GTG, and TTG start codons were annotated using BLASTn, BLASTp, and domain enhanced lookup time accelerated (DELTA)-Basic Local Alignment Search Tool (BLAST) programs from the National Center for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed May 21, 2013). Predicted coding sequences (CDSs) were not included in the annotation if they showed no obvious homology to known CDSs or gene products and had no obvious ribosome binding sites upstream of any start codon. GenBank accession numbers for each 50-kb region are KC516868 (IFR 06/001), KC516869 (IFR 06/005), KC516870 (Craig 610), KC516871 (Eklund 202F), and KC516872 (Hobbs FT10).

Table 1
Strains of Group II (Nonproteolytic) *Clostridium botulinum* Type F6 Used in This Study

Strain	Material from Which Strain Isolated (Country, Year)	Received from
Eklund 202F	Pacific sediments (CA, the United States, 1965)	J. Crowther (Unilever Research)
Hobbs FT10	Atlantic Herring from Moray Firth (Scotland, UK, 1970s)	J. Crowther (Unilever Research)
Craig 610	Salmon from Columbia River (OR, the United States, 1965)	J. Crowther (Unilever Research)
IFR 06/001	Scallops (Canada, 2006)	Isolated at IFR
IFR 06/005	Scallops (Canada, 2006)	Isolated at IFR

Analysis of Group II *C. botulinum* Type F6 Neurotoxin Gene Cluster Locus

DNA sequences of five different strains of Group II *C. botulinum* type F6 (table 1) were assembled into contigs from Illumina MiSeq data. Contig analysis using the ACT confirmed our microarray results (Stringer et al. 2013) by showing that the majority of the Group II *C. botulinum* type F6 genome is very similar in content and organization to that of the Group II *C. botulinum* type B strain Eklund17B and rather less similar to that of Group II *C. botulinum* type E strains Alaska and Beluga. This is exemplified in figure 1, in which an ACT comparison between regions of chromosomal DNA of strains Eklund17B, IFR 06/001, and Alaska demonstrates the reduced synteny with the latter strain. There was no evidence for the type F neurotoxin gene cluster being carried by a plasmid.

A 50-kb region from each *C. botulinum* type F6 genome containing the neurotoxin gene cluster was compared. Apart from 10 single-nucleotide polymorphisms (SNPs) and one 15-bp deletion, all 50-kb sequences were identical, suggesting that Group II *C. botulinum* type F strains are clonal, deriving from one rare event. All sequence differences mapped to the Eklund 202F genome (table 2); this agrees with observations made by Hielm et al. (1998) using pulsed-field gel electrophoresis, also Raphael et al. (2010) using 16S rRNA sequencing that Eklund 202F differs slightly from other Group II *C. botulinum* type F strains. In our recent comparative genomic microarray analysis, Eklund 202F was also found to be slightly divergent from other Group II *C. botulinum* type F strains (Stringer et al. 2013).

Each of these 50-kb regions contains an apparently foreign insertion of 34.4-kb DNA sequence, which has disrupted the *topB* gene, encoding DNA topoisomerase III (fig. 1*B*). This enzyme decatenates the newly replicated bacterial chromosome and is essential for survival. In the other Group II *C. botulinum* genomes currently available in databases (Eklund17B, Beluga, Alaska), only one copy of *topB*

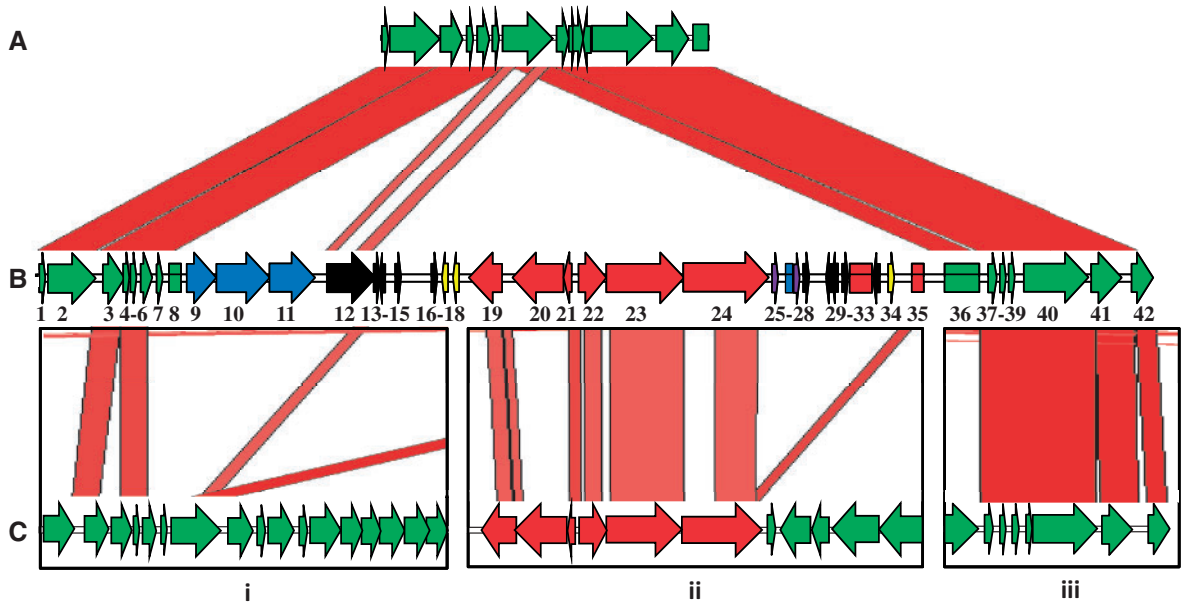

Fig. 1.—Group II *Clostridium botulinum* chromosomal DNA ACT comparisons of the 50-kb neurotoxin cluster locus of type F6 strain IFR 06/001 (*B*) with type B strain Eklund17B (*A*) and type E strain Alaska (*C*). Red bars denote regions sharing a high degree of sequence identity; the two thinner red bars linking section A with B are due to homology between the two versions of the *topB* gene, one present as a newly acquired gene, one as the original disrupted version. A similar duplication of homology is seen for the *topB* gene (CLH_1538) of Alaska (*C*, box i). Color code for features: 1) red, neurotoxin gene/cluster sequence; 2) blue, bacteriophage DNA; 3) purple, plasmid DNA; 4) yellow, IS4 elements; and 5) green and black, sequences unrelated to mobile DNA elements or toxin cluster genes, with black features apparently acquired horizontally. Boxes lacking arrowheads represent features with disrupted ORFs. Greatest synteny is seen between strain 06/001 (*B*) and Eklund17B (*A*): CDS/features 1–8 and 36–42 are homologs of Eklund17B (NRP [Norwich Research Park]) CDSs CB17B1604-1610 and CB17B1610-1617 (all sharing >90% identity). CDS 12 (in black): newly introduced *topB* gene. CDSs 19–24: neurotoxin cluster genes *orf-x3*, *orf-x2*, *orf-x1*, *p47*, nontoxic–nonhemagglutinin gene, and type F6 neurotoxin gene, respectively. Features 32 and 35 (red boxes): fragments of type B neurotoxin gene and type E neurotoxin gene, respectively. Chromosomal regions of type E strain Alaska homologous to the 50-kb toxin cluster locus of strain 06/001 are scattered due to poorer synteny; three main regions are depicted (section C). Box i: genes CLH_1532-1548, (88–96% identity). Box ii: CLH_1115-1105 (79–90% identity, note lack of homology with *orf-x2*, CLH_1114). Box iii: CLH_1908-1901 (95–96% identity). The type E3 neurotoxin gene of Alaska, CLH_1110 shares homology with strain 06/001 at two locations; the 3′-end of the type F6 neurotoxin gene (CDS 24) plus the nonfunctional type E neurotoxin gene fragment (feature 35).

Table 2
Sequence Differences within Strains of Group II *Clostridium botulinum* Type F6

SNP/Deletion	Co-Ordinate	Gene
T-C	68	Intergenic
G-A	529	Sensory transduction protein kinase
T-A	1,855	Sensory transduction protein kinase
C-T	1,938	Sensory transduction protein kinase
A-G	5,582	Intergenic
A-G	29,314	Type F6 neurotoxin
ATGATTTTGAAGAAG (deletion)	36,850–36,864	5′-flank of type B neurotoxin fragment
G-A	40,549	Intergenic
C-T	47,283	Exonuclease/helicase
A-C	48,699	MATE efflux family protein
A-G	49,015	MATE efflux family protein

Note.—All sequence differences were found in the genomic DNA of strain Eklund 202F. Numbering refers to the 50-kb region of all strains except for Eklund 202F, the coordinates of which differ after the 15-bp deletion at position 36850.

exists. Comparison with closely related *topB* sequences showed that although the type F6 *topB* is split by the 34.4-kb inserted DNA, the full CDS is still present; the N-terminal 109 codons border the 5′-flank of the DNA insertion, whereas its 3′-flank is defined by the remaining 620 codons. Thus, the original foreign DNA acquisition event did not rearrange or duplicate the sequence of the target site. The disrupted *topB* was found to have been replaced by a new copy, carried within the inserted DNA (see later). Percent G + C was analyzed and did not vary any more within the 34.4-kb region than it did in the flanking regions (data not shown). This suggests that the new DNA is the result of homeologous horizontal gene transfer(s), that is, it has been acquired from an organism closely related to the recipient rather than from a completely unrelated bacterium (Fall et al. 2007).

Predicted CDSs within the 34.4-kb foreign DNA were examined for evidence of the history of the insertion event (fig. 1*B*). Prime suspects at its 5′-end were three contiguous regions encoding site-specific recombinases/bacteriophage integrases or remnants thereof (CDSs 9–11, fig. 1*B*). Significantly, after a short interval (529 bp) is an intact *topB* gene (CDS 12, fig. 1*B*), presumably replacing the function of the disrupted copy. Most interestingly, this *topB* gene is more closely related to those of *C. beijerinckii* (76% amino acid identity) and *C. butyricum* (75% amino acid identity) than to those of Group II *C. botulinum*. The insertion of neurotoxin cluster DNA into a gene together with replacement of the same gene by another copy has precedent; as mentioned earlier in Group II *C. botulinum* type E and *C. butyricum* type E strains, the location of the type E neurotoxin gene cluster appears to be the result of insertion events that split the *rarA* gene at codon 102 in both species (Hill et al. 2009). The *rarA* gene product is a resolvase involved in recombination or transposon insertion events, and Hill et al. (2009) speculated that this might have played a part in these two independent insertion events. Is it possible that the *topB* gene, whose product is also involved in DNA manipulation, may have acted as a target for foreign DNA insertion via a similar mechanism? Intriguingly, BLAST results for the most 5′ of the three integrase-like CDSs (CDS 9, fig. 1*B*) of the newly acquired piece of DNA suggest that its gene product is a member of the XerD family, which in *Escherichia coli* acts in concert with XerC to perform a similar function to that of TopB, that is, conversion of dimeric chromosomes into monomers (Blakely et al. 1993).

Located 4.2 kb after the intact *topB* gene is a type F6 neurotoxin gene in a complete *orf-x*-cluster (CDSs 19–24, fig. 1*B*), identical in organization to that described by Dover et al. (2011) for the Group II *C. botulinum* type F6 strain IBCA66-5436. All the type F6 strains described in this study, together with IBCA66-5436 and all Group II *C. botulinum* type E neurotoxin clusters lack the sigma factor *botR* gene. Interestingly, *botR* is present in the *orf-x*-cluster of the Group I *C. botulinum* type F strains Langeland and 230613 (Hill et al. 2009; Tian et al. 2011), but its presence and position in the *orf-x*-cluster of *C. baratii* type F has yet to be established (Raphael et al. 2010). The neurotoxin gene cluster in the Group II *C. botulinum* type F described in this study spans 13.55 kb from the 3′-end of the *orf-x3* gene to the 3′-end of the type F6 neurotoxin gene. It is flanked at its 5′-end by two IS4 element-related regions (CDSs 17, 18) and at its 3′-end by remnants of a bacteriophage integrase/recombinase (CDS 26, fig. 1*B*). The type F6 neurotoxin genes in strains Hobbs FT10, Craig 610, IFR 06/001, and IFR 06/005 were identical, whereas strains Eklund 202F and IBCA66-5436 share a single SNP in the 27th codon where there is an A–G transition changing an AAA (lysine) codon to a GAA (glutamate) codon (table 2). Except for this one SNP, the remainder of the neurotoxin gene cluster is identical for the five type F6 strains examined in this study and for IBCA66-5436. This high similarity of the neurotoxin gene cluster implies that strain IBCA66-5436 will also fall into this same clonal group. Indeed, 16S rRNA sequence data places IBCA66-5436 in the same clade as Hobbs FT10 and Craig 610 (Raphael et al. 2010).

Only 3.7 kb from the type F6 gene is a 1 kb CDS that begins with sequence distantly related to a phage integrase; unexpectedly, BLASTp hits for the predicted product of a 270-codon region further downstream were to the receptor binding domain of the botulinum neurotoxin heavy chain (Feature 32, fig. 1*B*). The top three hits were to type G neurotoxin, associated with strains of Group IV *C. botulinum* (*C. argentinense*) (Terilli et al. 2011); however, the amino acid identity was only 35%, with 22 gaps. Because the query sequence contained conserved domains, DELTA-BLAST was used. This makes use of a subset of the NCBI's Conserved Domain Database and should produce better quality alignments (Boratyn et al. 2012). Using DELTA-BLAST, all the top hits (30% identity but only 18 gaps) were to the heavy chain of type B neurotoxins, at residue positions approximately 1009–1291. Based on this evidence, the DNA sequence is annotated as deriving from a type B neurotoxin gene.

The 3′-end of the foreign DNA insertion bears further interesting evidence of recombination events. Located 242 bp after the type B neurotoxin gene fragment is a truncated (130 of the expected 170–190 codons) CDS for signal peptidase I, closely followed by a partially deleted IS4 sequence containing reading frame errors (CDSs 33, 34, fig. 1*B*). The reconstructed IS4 ORF encodes 311 amino acid residues, for which the best BLASTp hits (52% identity) were to IS4 elements of *C. cellulovorans* 743B. The remains of IS mobile elements are often found associated with *C. botulinum* neurotoxin gene clusters, usually not only at their 5′- and 3′-flanks but also sometimes within the gene cluster itself (Hill et al. 2009), and are thought to indicate horizontal gene transfer activity. The fact that there are phage integrase genes at one end of the DNA insertion, and transposable element DNA at the other end, as is also the case for the intact type F6 neurotoxin cluster, makes it difficult to speculate on the exact mechanism for insertion of the foreign DNA as these two different mobile elements and the CDSs contained within their borders provide evidence for several such events in this region, not merely one.

Adjacent to the inactive IS4 element is a 736-bp fragment of DNA containing a disrupted ORF (Feature 35, fig. 1*B*). After correction of reading frame errors, the original ORF encodes 247 amino acid residues with 81% identity to a type E neurotoxin (DELTA-BLAST alignment to residues 884–1,130 of most type E neurotoxins). The closest relatives to this type E fragment are the recently described type E9 neurotoxin of Group II *C. botulinum*, followed by the type E5 neurotoxin of *C. butyricum.* Interestingly, as with Group II *C. botulinum* type F6, the chromosomal DNA of the strain that forms type E9 neurotoxin was reported as being more closely related (99%) to that of Eklund17B than to other Group II *C. botulinum* type E strains (94%) (Raphael et al. 2012).

Alignment of predicted peptides for the type B and type E gene fragments with complete type B and type E neurotoxin reference sequences shows that both of these fragments partially overlap, spanning approximately the same region of the neurotoxin heavy chain. This shows that a single neurotoxin gene from the ancestral gene cluster has not been disrupted by insertion of new DNA; rather, two completely different neurotoxin genes have been disrupted in successive rounds of recombination at this locus. The observations that the rest of the chromosome is most similar to that of a Group II *C. botulinum* type B, and that the CDS for the type B neurotoxin fragment is more degraded (30% amino acid identity) than that of the type E fragment (81% amino acid identity), suggests that the original bacterium was probably a Group II *C. botulinum* type B strain.

A further 335 bp after the type E gene fragment, the C-terminal fragment of the disrupted *topB* gene marks the end of the 34.4-kb DNA insertion and the return to synteny with Eklund17B (CDSs 36–42, fig. 1*B*).

Because the CDSs that flank the 34.4-kb insertion show such close synteny with Eklund17B, it would be interesting to examine this region in other Group II *C. botulinum* strains for evidence of genetic recombination. Unfortunately, at present, nothing can be said regarding the frequency of use of this *topB* locus for insertion of neurotoxin gene clusters in Group II *C. botulinum* type B, as the only available genome is that of Eklund17B, which carries its neurotoxin gene cluster on a plasmid. In their study of plasmid borne neurotoxin gene clusters in Group I and Group II *C. botulinum* type B, Franciosa et al. (2009) identified two other Group II type B strains, which bear their neurotoxin gene cluster on a plasmid, as do Group II *C. botulinum* type B strains CDC 3875 and IFR 05/025 that we have recently sequenced (data not shown).

Conclusions

Although the type B and type E neurotoxin genes discovered in this study are incomplete, this report is the first to describe evidence of more than one neurotoxin gene in Group II *C. botulinum* and completely changes the view that members of this group are less recombinogenic than those of Group I, at least with respect to neurotoxin gene cluster sequences. We also identify the *topB* gene as a chromosomal locus for a *C. botulinum* neurotoxin gene cluster that has not been described previously.

The simplest explanation for the evolution of this type F6 clonal group is that the original ancestor was a Group II *C. botulinum* type B strain, either with a chromosomally located neurotoxin gene cluster or because all type B strains so far examined seem to possess a plasmid-based neurotoxin gene cluster, one where this cluster has integrated into the chromosome. This received DNA from a strain containing a type E gene cluster, followed by the most recent event, where a complete type F6 gene cluster was inserted. In each case, the previous neurotoxin gene was inactivated, so there still remains to be found a Group II *C. botulinum* bearing two

functional neurotoxin genes, unlike the situation in Group I *C. botulinum* where several such strains are reported.

It should be emphasized that the above is only the simplest predicted course of events. At least two different types of mobile element have been involved in these genetic exchanges on at least two different occasions. Because the newly acquired *topB* gene is more like that from *C. butyricum* or *C. beijerinckii*, it is possible that not all partners in these DNA swapping events were *C. botulinum*. Genetic exchange between bacteria must be evidence that at some stage they shared an environment; Group II *C. botulinum*, *C. butyricum,* and *C. beijerinckii* are all saccharolytic anaerobes, so this speculation has at least some biological relevance.

It has been shown that members of Group II *C. botulinum* type F6 are closely related. The series of recombination events needed to deliver them has probably occurred just once to give a clonal group that has spread around the world. It is also possible that strains with evidence of other recombinational events at the *topB* hot spot exist but have not yet been described. Perhaps, the recently described Group II *C. botulinum* type E9 strain CDC 66177 (Raphael et al. 2012), which has chromosomal DNA more closely related to that of Eklund17B than to other Group II *C. botulinum* type E strains mirrors one of the evolutionary stages of Group II *C. botulinum* type F6, although the insertion site in this case is again the *rarA* gene. To date, the only available genome sequence for Group II *C. botulinum* type E strains is for strains Alaska and Beluga, and these both lack evidence for the partial type B gene fragment described in this work. Further genetic clues to enable us to unravel the complex evolutionary route that has resulted in the present Group II *C. botulinum* type F6 strains will be uncovered as the genomes of more Group II *C. botulinum* are examined.

Acknowledgments

The authors thank Dr John Walshaw for comments on the manuscript and help in submission of sequences to GenBank. This work was supported by the BBSRC Institute Strategic Programme on Gut Health and Food Safety (grant number BB/J004529/1).

Literature Cited

Austin JW, Dodds LK. 2000. *Clostridium botulinum*. In: Hui, YH, Pierson, MD, Gorham, JR, editors. Foodborne disease handbook. New York: Marcel Dekker. p. 107–138.

Blakely G, et al. 1993. Two related recombinases are required for site-specific recombination at dif and cer in *E. coli* K12. Cell 75:351–361.

Boratyn GM, et al. 2012. Domain enhanced lookup time accelerated BLAST. Biol Direct. 7:12.

Carter AT, et al. 2009. Independent evolution of neurotoxin and flagellar genetic loci in proteolytic *Clostridium botulinum*. BMC Genomics 10:115.

Carter AT, et al. 2011. Complete genome sequence of the proteolytic *Clostridium botulinum* type A5 (B3′) strain H04402 065. J Bacteriol. 193:2351–2352.

Carver TJ, et al. 2005. ACT: the Artemis Comparison Tool. Bioinformatics 21:3422–3423.

Chain PSG, et al. 2009. Genomics. Genome project standards in a new era of sequencing. Science 326:236–237.

Dover N, Barash JR, Hill KK, Detter JC, Arnon SS. 2011. Novel structural elements within the non-proteolytic *Clostridium botulinum* type F toxin gene cluster. Appl Environ Microbiol. 77:1904–1906.

Fall S, et al. 2007. Horizontal gene transfer regulation in bacteria as a "spandrel" of DNA repair mechanisms. PLoS One 2:e1055.

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.

Hielm S, Björkroth J, Hyytiä E, Korkeala H. 1998. Genomic analysis of *Clostridium botulinum* Group II by pulsed-field gel electrophoresis. Appl Environ Microbiol. 64:703–708.

Hill KK, et al. 2007. Genetic diversity among botulinum neurotoxin-producing clostridial strains. J Bacteriol. 189:818–832.

Hill KK, et al. 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.

Macdonald TE, et al. 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.

Peck MW. 2006. *Clostridium botulinum* and the safety of minimally heated chilled foods: an emerging issue? J Appl Microbiol. 101: 556–570.

Peck MW. 2009. Biology and genomic analysis of *Clostridium botulinum*. Adv Microb Physiol. 55:183–265.

Peck MW, et al. 2010. Development and application of a new method for specific and sensitive enumeration of spores of nonproteolytic *Clostridium botulinum* types B, E and F in foods and food materials. Appl Environ Microbiol. 76:6607–6614.

Peck MW, Stringer SC, Carter AT. 2011. *Clostridium botulinum* in the post-genomic era. Food Microbiol. 28:183–191.

Raphael BH, Choudoir MJ, Lúquez C, Fernández R, Maslanka SE. 2010. Sequence diversity of genes encoding botulinum neurotoxin type F. Appl Environ Microbiol. 76:4805–4812.

Raphael BH, et al. 2012. Analysis of a unique *Clostridium botulinum* strain from the Southern hemisphere producing a novel type E botulinum neurotoxin subtype. BMC Microbiol. 12:245.

Sebaihia M, et al. 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.

Simpson JT, et al. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19:1117–11230.

Smith TJ, et al. 2005. Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization. Infect Immun. 73:5450–5457.

Stringer SC, et al. 2013. Genomic and physiological variability within Group II (non-proteolytic) *Clostridium botulinum*. BMC Genomics 14: 333.

Terilli RR, et al. 2011. A historical and proteomic analysis of botulinum neurotoxin type/G. BMC Microbiol. 11:232.

Tian RM, et al. 2011. The complete genome sequence of *Clostridium botulinum* F str. 230613, insertion sites, and recombination of BoNT gene clusters. Genome 54:546–554.

Associate editor: Soojin Yi