MAJOR ARTICLE

Molecular Characterization of a Novel Botulinum Neurotoxin Type H Gene

Nir Dover,[1] Jason R. Barash,[1] Karen K. Hill,[2] Gary Xie,[2] and Stephen S. Arnon[1]

[1]Infant Botulism Treatment and Prevention Program, California Department of Public Health, Richmond, California; and [2]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico

(See the major article by Barash and Arnon on pages 183–91, and the editorial commentaries by Popoff on pages 168–9 and Hooper and Hirsch on page 167 and Relman on pages 170–2.)

We sequenced the 2 botulinum toxin gene clusters of *Clostridium botulinum* strain IBCA10-7060 type Bh. The sequence of *bont/H* differed substantially from the sequences of the 7 known *bont* genes for toxin types A–G. The 5′ one-third terminus of *bont/H* that codes for the botulinum toxin light chain differed markedly from the light chain coding sequences of toxin types A–G. The 3′ two-thirds terminus of *bont/H* that codes for the botulinum toxin heavy chain contained a novel H_N translocation domain coding sequence and a nonneutralizing type A–like H_C binding domain coding sequence. *bont/H* was part of an *orfX* toxin gene cluster that was located at a unique chromosomal site distant from those used by other botulinum toxin gene clusters. The *bont/B* sequence was similar to that of subtype *bont/B2* and was located within its *ha* toxin gene cluster at the *oppA/brnQ* site. Our findings further establish that *C. botulinum* IBCA10-7060 produces novel BoNT/H.

Keywords. *Clostridium botulinum*; botulinum toxin; botulinum toxin type H; botulism.

Use of molecular techniques to investigate the genome and toxin gene clusters of *Clostridium botulinum* has enormously advanced understanding of the bacterium and its ability to produce botulinum toxin [1–3]. Botulinum neurotoxin (BoNT) exists in 7 toxin types (designated A–G) that are distinguished by the inability of polyclonal antibodies that neutralize one toxin type to neutralize any of the other 6 toxin types [4]. Some *C. botulinum* strains produce 2 botulinum toxins, designated Ab, Ba, Af, and Bf, with the uppercase letter identifying the predominant neurotoxin [5–9], while 1 strain (Af84) contains 3 neurotoxin genes [10]. Additionally, some type A strains contain a complete but inactive type B neurotoxin gene (A[B]) [11], while others contain a partial type B toxin gene (A[B′])] [12]. Variants (ie, subtypes) within BoNTs have been identified and designated with the serotype followed by an Arabic number (eg, A1, A2, etc.) [4, 10]. The extent of subtype nucleotide sequence diversity varies among the 7 toxin types. Type F is the most diverse, with sequence diversity as high as 25% among its 7 subtypes, while type B has only up to 4% sequence diversity among its 7 subtypes [3].

The *bont* gene that encodes botulinum neurotoxin is part of a toxin gene cluster that includes several nontoxin accessory genes. Two main *bont* gene cluster organizations are known. The hemagglutinin (*ha*) toxin gene cluster is found in *C. botulinum* types A1, A5, B, C, D, and G strains, while the *orfX* toxin gene cluster is found in *C. botulinum* types A1-A4, E, and F and in *Clostridium butyricum* type E and *Clostridium baratii* type F strains [13–15]. The toxin gene cluster is a mobile genetic element that may reside either in the chromosome, in a plasmid, or in a bacteriophage [1, 14, 16–18]. Specific insertion sites of the toxin gene cluster within the chromosome and within certain plasmids were recently identified [10, 19]. In *C. botulinum* group I strains, all chromosomal *ha* toxin gene clusters are located in the *oppA/brnQ* locus, and all chromosomal *orfX* toxin gene clusters are located either in the *arsC* or the *pulE* loci [10, 19].

The isolation of proteolytic *C. botulinum* strain IBCA10-7060 toxin type Bh from an infant botulism

Received 17 May 2013; accepted 5 August 2013; electronically published 7 October 2013.

Correspondence: Stephen S. Arnon, MD, Infant Botulism Treatment and Prevention Program, California Department of Public Health, 850 Marina Bay Pkwy, Rm E361, Richmond, CA 94804 (stephen.arnon@cdph.ca.gov).

The Journal of Infectious Diseases 2014;209:192–202

DOI: 10.1093/infdis/jit450

patient brought to recognition botulinum toxin type H, the first new botulinum toxin type to be identified in almost half a century [20]. We report here genomic studies of *C. botulinum* strain IBCA10-7060. We used gene sequencing, comparative genomics, and phylogenetic analyses to identify and characterize the novel BoNT type H gene (*bont/H)* and determined that it differs substantially from all other *bont* genes. We additionally determined the gene content and organization of the type H toxin gene cluster and found that it has a unique chromosomal location. However, the type B toxin gene cluster was in the same locus as other chromosomal *ha* toxin gene clusters in group I *C. botulinum* strains [19].

METHODS

Bacterial Strains and Culture Conditions

C. botulinum type Bh strain IBCA10-7060, originally isolated from the feces of an infant botulism patient, was subcultured anaerobically from a frozen stock culture on 4% egg yolk agar plates and incubated for 48 hours at 35°C. Isolated colonies were inoculated into 20 mL of prereduced trypticase-peptone-glucose-yeast extract broth, incubated anaerobically for 24–48 hours at 35°C, and harvested by centrifugation at 3450 g. The cell pellets were stored at −75°C.

DNA Extraction

Genomic DNA for Sanger sequencing was extracted using the MagNA Pure Compact instrument with Nucleic Acid Isolation Kit I and the Bacteria Purification Protocol (Roche Applied Science) according to the manufacturer's instructions. Genomic DNA for next-generation sequencing was extracted using the Qiagen DNeasy Blood and Tissue Kit according to the protocol for Gram-positive bacteria.

DNA Sequencing

The genome of *C. botulinum* strain IBCA10-7060 was sequenced by combining Sanger sequencing of short PCR products and next-generation sequencing. The type H and type B toxin gene clusters were initially sequenced with the Sanger method. Polymerase chain reaction (PCR) was performed with Ex Taq HS DNA Polymerase (Takara Bio) with thermocycling conditions of initial heating of 95°C for 2 minutes followed by 30 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1 minute. The PCR amplicons were purified with the QIAquick PCR purification kit (Qiagen) and sequenced using an Applied Biosystems 3730XL DNA Analyzer. The resulting amplicon sequences were assembled with Sequencher (Gene Codes) and annotated with vector NTI (Invitrogen) to create the complete *bont/H* and *bont/B* toxin gene cluster sequences.

For next-generation sequencing, a high-quality draft genome sequence was generated by combining both standard and long-insert 454 Titanium (656K reads yielding 60-fold genome coverage) and shotgun Illumina GAiix (30.3 M reads yielding 379-fold genome coverage) data. The 454 data were assembled with Newbler, version 2.3, and the Illumina data were assembled with Velvet, version 0.7.63. These initial assemblies were integrated using parallel Phrap version SPS 4.24, and the integrated assembly was examined with Consed. The final de novo genome assembly included 51 contigs with a total genome size of 4.1 Mbp and an average G + C content of 28.0%.

Phylogenetic and Similarity Analyses

The gene sequences used in this study were aligned with ClustalW and then hand edited and analyzed using Mega 5 software [21]. Pairwise sequence identities were computed with the Mega 5 software and the Emboss pairwise sequence alignment algorithm. Possible recombination events were explored using SimPlot [22]. The SimPlot analyses were generated with a sliding window of 200 bp that was moved 20 bp between each data point.

RESULTS

The Novel *bont/H* Neurotoxin Gene

The initial Sanger sequencing of the botulinum toxin gene clusters of strain IBCA10-7060 identified a novel *bont* gene within an *orfX* neurotoxin gene cluster and a type B *bont* gene within a hemagglutinin neurotoxin gene cluster. Subsequent Illumina and 454 next-generation sequencing of the entire genome of strain IBCA10-7060 generated 51 contigs, with a total genome size of 4.1 Mbp. The novel *bont/H* and its toxin gene cluster were contained in a contig of 918 184 bp. The *bont/B* toxin gene cluster was contained in a separate contig of 49 646 bp.

BoNT/H was found by pairwise sequence alignment analysis to differ substantially from all other toxin types and subtypes (Table 1). The nearest sequence identity was with BoNT/F5 (76.4% and 63.5% nucleotide and amino acid identity, respectively). The other BoNT/F subtypes shared no more than 69.0% and 52.8% nucleotide and amino acid identity, respectively, with BoNT/H. Notably, BoNT/H was only 53.7%–66.0% and 32.3%–50.9% identical to the nucleotide and amino acid sequences, respectively, of the other 6 BoNT types (ie, A, B, C, D, E, and G).

A phylogenetic tree compared the full-length coding regions of *bont/H* and the 7 known *bont/A-G* gene types and their subtypes (Figure 1). The *bont/A-G* gene sequences aggregated into 7 monophyletic clades. In the phylogenetic tree, the novel *bont/H* gene of strain IBCA10-7060 formed a distinct lineage that was separate from the clades of the 7 known *bont* types, with the type F clade as its nearest neighbor.

A sequence identity comparison of BoNT/H to all known BoNT toxin types with 1 representative from each further distinguished the novel type H sequence (Figure 2). The 7 known BoNT types had nucleotide and amino acid sequence identities of 53.9%–75.1% and 32.6%–64.1%, respectively (Figure 2).

Table 1. Comparison of Nucleotide and Amino Acid Pairwise Sequence Identities of Botulinum Neurotoxin (BoNT) Type H to Known BoNT Subtypes

	Sequence Identity, %	
BoNT Subtype	Nucleotides	Amino Acids
A1	66.0	50.9
A2	65.2	49.9
A3	65.7	49.1
A4	65.0	49.7
A5	65.5	50.3
B1	59.9	40.6
B2	60.5	41.3
B3	60.2	41.3
B4	59.4	41.7
B5	60.4	41.4
B6	60.2	41.2
B7	60.6	41.5
C	53.7	32.8
C/D	56.1	32.3
D	56.9	33.8
D/C	54.6	33.3
E1	65.6	48.7
E2	65.8	48.7
E3	65.5	48.5
E4	65.5	48.4
E5	65.6	48.4
E6	65.6	48.8
E7	65.3	48.7
E8	65.6	48.7
E9	66.2	48.9
F1	69.0	52.8
F2	67.1	50.5
F3	67.7	50.3
F4	67.9	51.7
F5	76.4	63.5
F6	67.7	50.8
F7	67.6	52.5
G	59.8	40.6

One representative of each subtype is shown. See Figure 1 for GenBank accession numbers.

Notably, the identity between BoNT/H and BoNT/F1 (69.0% and 52.8% for nucleotide and amino acid sequences, respectively) falls within the ranges of identities among the 7 A–G BoNT types. BoNT/H and BoNT/F1 were less identical to each other (69.0% and 52.8% for nucleotide and amino acid sequences, respectively) than were BoNT/B1 and BoNT/G (72.0% and 57.7%, nucleotide and amino acid sequences, respectively) and BoNT/E1 and BoNT/F1 (75.1% and 64.1%, nucleotide and amino acid sequences, respectively) (Figure 2).

Because the novel *bont/H* gene of strain IBCA10-7060 was most closely related to *bont/F*, we compared the identities of the 7 BoNT/F subtypes to each other and to BoNT/H (Figure 3). Most BoNT/F subtypes (F1–4 and F6) shared 90.9%–98.9% nucleotide identity and 83.4%–97.1% amino acid identity. BoNT/F5 and BoNT/F7 were less identical to the other 5 BoNT/F subtypes and shared only 79.4%–83.1% and 69.3%–74.1% identity (nucleotide and amino acid sequences, respectively), with the other 5 subtypes. BoNT/F5 and BoNT/F7 were the least identical subtypes and shared only 74.5% and 63.9% sequence identity (nucleotides and amino acids, respectively). Notably, BoNT/H shared only 67.1%–76.4% and 50.3%–63.5% sequence identity (nucleotides and amino acids, respectively), with the 7 BoNT/F subtypes. BoNT/H was most closely related to BoNT/F5 (76.4% and 63.5% nucleotides and amino acid sequence identity, respectively).

Similarity analysis demonstrated that *bont/H* can be divided into 3 parts (Figure 4). The approximately 1500 nucleotides at its 5′ terminus that code for the BoNT light chain (LC) were most related to the 5′ terminus of *bont/F5* (nucleotide and amino acid identity of approximately 86% and 80%, respectively). The middle section of the *bont/H* gene (nucleotides 1500–2700, approximately) did not share significant identity with any *bont* gene sequence; it was most similar to the middle section of *bont/F1* and *bont/F5* (nucleotide and amino acid identity of approximately 77% and 64%, respectively). This section of *bont* codes for the translocation domain (HN), which constitutes approximately one-half of the heavy chain. The 3′ terminus of *bont* codes for the binding domain (HC), which constitutes approximately the other half of the heavy chain. This segment showed high identity to the equivalent region of the *bont/A* gene, with no particular preference for any *bont/A* subtype (Figure 4*A*). The 764 nucleotides at the 3′ end (nucleotides 3104–3867) of the *bont/H* gene were 95.9%–97.4% identical to the 5 known *bont/A* subtypes (Figure 4*A*). However, similarity analysis comparison of *bont/H* to *bont/A1* and *bont/F5* showed that these 2 subtypes, although containing gene sections similar to *bont/H*, were nevertheless quite different (Figure 4*B*).

Figure 5 presents a phylogenetic analysis comparing the predicted amino acid sequences of BoNT/H and representative subtypes of BoNT/A and BoNT/F divided into their light and heavy chain sequences. As mentioned above, the light chain sequences of BoNT/H and BoNT/F5 were the most similar to each other and formed a separate and unique clade (Figure 5*A*). The predicted amino acid sequence of the heavy chain of BoNT/H formed its own separate and unique lineage, while the predicted amino acid sequence of the heavy chain of BoNT/F5 was located in the BoNT/F clade (Figure 5*B*) [23].

The BoNT/H predicted heavy chain amino acid sequence was further divided into its translocation and binding domains and compared to the homologue domains of BoNT types A and F (Figure 5*B1* and 5*B2*, respectively). The translocation domain of BoNT/H had a novel sequence that was more closely

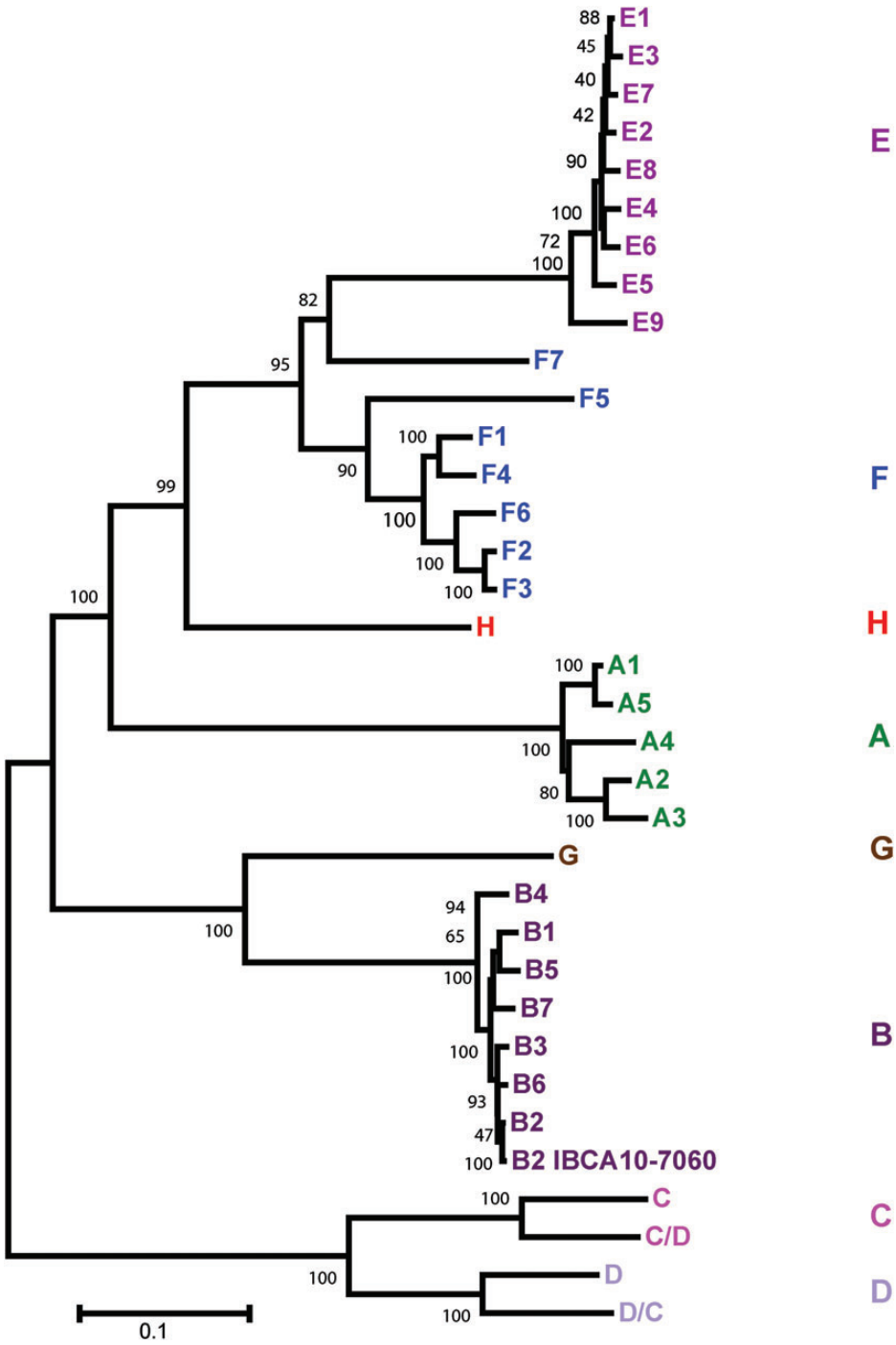

Figure 1. Phylogenetic analysis of nucleotide sequences representing all 7 *bont* serotypes and their subtypes. One representative of each toxin subtype is shown. The phylogenetic tree was generated using the neighbor-joining method and the Kimura 2-parameter model. Genetic distance and bootstrap values are shown. Note that the novel *bont/H* creates a distinct lineage separated from the 7 *bont/A-G* types. Botulinum toxin subtypes, and strains and GenBank accession numbers used in the analysis, are as follows: A1 = NCTC2916, X52066; A2 = Kyoto, CP001581; A3 = Loch Maree, CP000963; A4 = 657Ba, CP001081; A5 = IBCA94-0216, FJ959094; B1 = Okra, CP000940; B2 = Prevot 25 NCASE, EF033129; B3 = CDC795, EF028400; B4 = Eklund 17B, CP001057; B5 = 657Ba, CP001081; B6 = Osaka05, AB302852; B7 = NCTC3807, JN120760; C = Strain 573, AB200359; C/D = Strain S19, FN436022; D = BVD/−3, X54254; D/C = Strain OFD17, AB461921; E1 = NCTC 11219, X62683; E2 = CDC5147, EF028404; E3 = Alaska E43, CP001078; E4 = ATCC 43755, X62088; E5 = LCL 095, AB037714; E6 = K35, AM695752; E7 = IBCA97-0192, JN695729; E8 = E134, JN695730; E9 = CDC66177, JX424534; F1 = Langeland, CP000728; F2 = CDC4013 Bf, GU213209; F3 = CDC54086, GU213218; F4 = CDC54078, GU213214; F5 = CDC54096, GU213225; F6 = IBCA66-5436, HQ441176; F7 = CDC59837, GU213231; and G = NCFB 3012, X74162.

% nucleotide sequence identity								
	A1	B1	C	D	E1	F1	G	H
A1	----	59.1	**53.9**	56.5	60.8	61.5	59.9	66.0
B1	38.8	----	55.6	56.5	59.2	59.0	72.0	59.9
C	**32.6**	34.0	----	67.0	54.1	56.0	56.3	53.7
D	33.0	35.9	52.5	----	56.2	55.9	57.0	56.9
E1	39.9	37.4	33.5	34.5	----	**75.1**	58.5	65.6
F1	40.3	38.3	33.9	33.9	**64.1**	----	58.9	69.0
G	39.6	57.7	34.0	35.4	38.2	38.1	----	59.8
H	50.9	40.6	32.8	33.8	48.7	52.8	40.6	----
% amino acid sequence identity								

Figure 2. Comparison of nucleotide and amino acid sequence identities of representatives of BoNT types. Maximum and minimum sequence identities are in bold. Maximum sequence identities to type H are highlighted in grey. One representative of each type is shown. See Figure 1 for GenBank accession numbers.

related to the translocation domain of BoNT/F than to the translocation domain of BoNT/A (Figure 5*B1*). In contrast, the binding domain of BoNT/H was more closely related to the binding domain of BoNT/A than to the binding domain of BoNT/F (Figure 5*B2*).

Like all BoNTs [23–25], the predicted amino acid sequence of BoNT/H contained the conserved zinc endopeptidase domain HEXXH (HELIH in BoNT/H) in the light chain at residues 226–230. Interestingly, unlike all known BoNTs, the second histidine residue in the zinc endopeptidase domain of BoNT/H was coded by the codon CAC, rather than by the codon CAT.

The heavy chain membrane spanning domain PYxGxAL is conserved within the BoNTs [23–25]. BoNT/H contained the amino acid sequence PYIGLAL at residues 617–623, which is identical to the membrane spanning domains of BoNT/B and BoNT/E. In contrast, all BoNT/F toxin subtypes have the amino acid sequence PYVGLAL [23], and all BoNT/A subtypes have the amino acid sequence PYIGPAL for this domain.

bont/H contains 3867 nucleotides. This number of nucleotides differs from the numbers for both *bont/A* (3891 nucleotides, except for *bont/A3*, which contains 3879 nucleotides) and *bont/F* (3807–3843 nucleotides).

The *bont/H* Toxin Gene Cluster and Its Genomic Location

Annotation of the genome sequence of strain IBCA10-7060 determined that *bont/H* was part of an *orfX* toxin gene cluster (Figure 6*A*). The arrangement of accessory genes in the IBCA10-7060 *orfX* toxin gene cluster was identical to that in the *orfX* clusters of *bont/A*1 in strains NCTC2916 and CDC5328A (GenBank accession nos. AY497357 and EU341305, respectively) [12, 15, 26]. This arrangement differed from all type F and subtypes A2–A4 *orfX* toxin gene clusters. The *bont/H* toxin gene cluster contained an *orfX2-orfX3* intergenic spacing of 0.65 kb that was >99% identical to the *orfX2-orfX3* intergenic spacing in the 2 *bont/A1 orfX* clusters.

Supplementary Table 1 compares the pairwise identities of nucleotide and predicted amino acid sequences of the genes of the *bont/H orfX* toxin gene cluster to those of the types A and F *orfX* gene clusters available in GenBank. The *ntnh* of the *bont/H* toxin gene cluster was most similar to the *ntnh* of the *bont/F1* toxin gene cluster (95.8% and 92.8% nucleotide and amino acid identity, respectively), while the *p47*, *botR*, *orfX1*, *orfX2*, and *orfX3* genes were most similar to their homologue genes in the type A1 toxin gene cluster in strain NCTC2916 (Supplementary Table 1).

% nucleotide sequence identity								
	F1	F2	F3	F4	F5	F6	F7	H
F1	----	90.9	91.2	96.2	80.2	93.2	82.0	69.0
F2	83.4	----	**98.9**	91.8	82.9	95.2	79.7	67.1
F3	83.9	**97.1**	----	91.8	83.0	95.1	79.4	67.7
F4	92.5	83.5	83.8	----	81.2	93.7	81.1	67.9
F5	69.8	74.4	74.0	69.7	----	83.1	**74.5**	76.4
F6	87.7	90.2	90.1	87.1	74.1	----	80.0	67.7
F7	73.7	68.9	69.3	72.2	**63.9**	69.9	----	67.6
H	52.8	50.5	50.3	51.7	63.5	50.8	52.5	----
% amino acid sequence identity								

Figure 3. Comparison of nucleotide and amino acid sequence identities of *bont/H* and representatives of *bont/F* subtypes. Maximum and minimum type F sequence identities are in bold. Maximum sequence identities to type H are highlighted in grey. One representative of each subtype is shown. See Figure 1 for GenBank accesion numbers.

BLAST search upstream of the *bont/H* toxin gene cluster (ie, upstream of *orfX3*) identified overall synteny and approximately 93% sequence homology to the upstream sequence of the *bont/A4* toxin gene cluster in plasmid pCLJ that resides in *C. botulinum* strain 657 type Ba (GenBank accession no. CP001081). This homology extended approximately 31 kb upstream, where a putative gene coding for an IS110 transposase was located. Interestingly, a second copy of the same putative transposase-coding gene (99.8% nucleic acid identity) was located 1570 bp downstream of *bont/H* (Figure 6*B*). Alignment of the sequences upstream and downstream of these 2 IS110 transposases with the genome sequence of strain ATCC 3502 revealed that the *bont/H* toxin gene cluster is part of an insert of 54 464 bp (flanked by the 2 IS110 transposases) located in the chromosome between nucleotides 2 308 161 and 2 307 116 (strain ATCC 3502 numbering), in place of deleted gene *cbo2164* (Figure 6*B*).

The *bont/B* Neurotoxin Gene Cluster

BLAST search of the nucleic acid and predicted amino acid sequences of *bont/B* of strain IBCA10-7060 found only a 4-nucleotide and 1–amino acid difference from the *bont/B2* of *C. botulinum* strain Smith L-590 (GenBank accession no. EF028398). Phylogenetic analysis of all known *bont* subtypes further confirmed that the *bont/B* gene of strain IBCA10-7060 should be classified as subtype B2 (Figure 1).

As with all characterized *bont/B* toxin gene clusters, *bont/B2* of strain IBCA10-7060 was part of a hemagglutinin toxin gene cluster that contained the genes *ntnh*, *botR*, *ha33*, *ha17*, and *ha70* [4]. The nucleotide sequence of the IBCA10-7060 *bont/B2* gene cluster (excluding *bont/B2* itself) was most similar (99.2% identity) to the *bont/B6* gene cluster (excluding *bont/B6*) of strain Osaka05 (GenBank accession no. AB302852) [27].

Alignment of sequences flanking the IBCA10-7060 *bont/B2* gene cluster revealed that it is located in the *oppA/brnQ* locus. This locus was previously identified in group I *C. botulinum* as the chromosomal location of all known *ha*-containing toxin gene clusters [3, 19].

Phylogenetic Characterization of *C. botulinum* Strain IBCA10-7060

Examination of the draft genome of *C. botulinum* strain IBCA10-7060 identified a contig that contained the 16S ribosomal RNA (rRNA) gene. BLAST search of this 16S rRNA

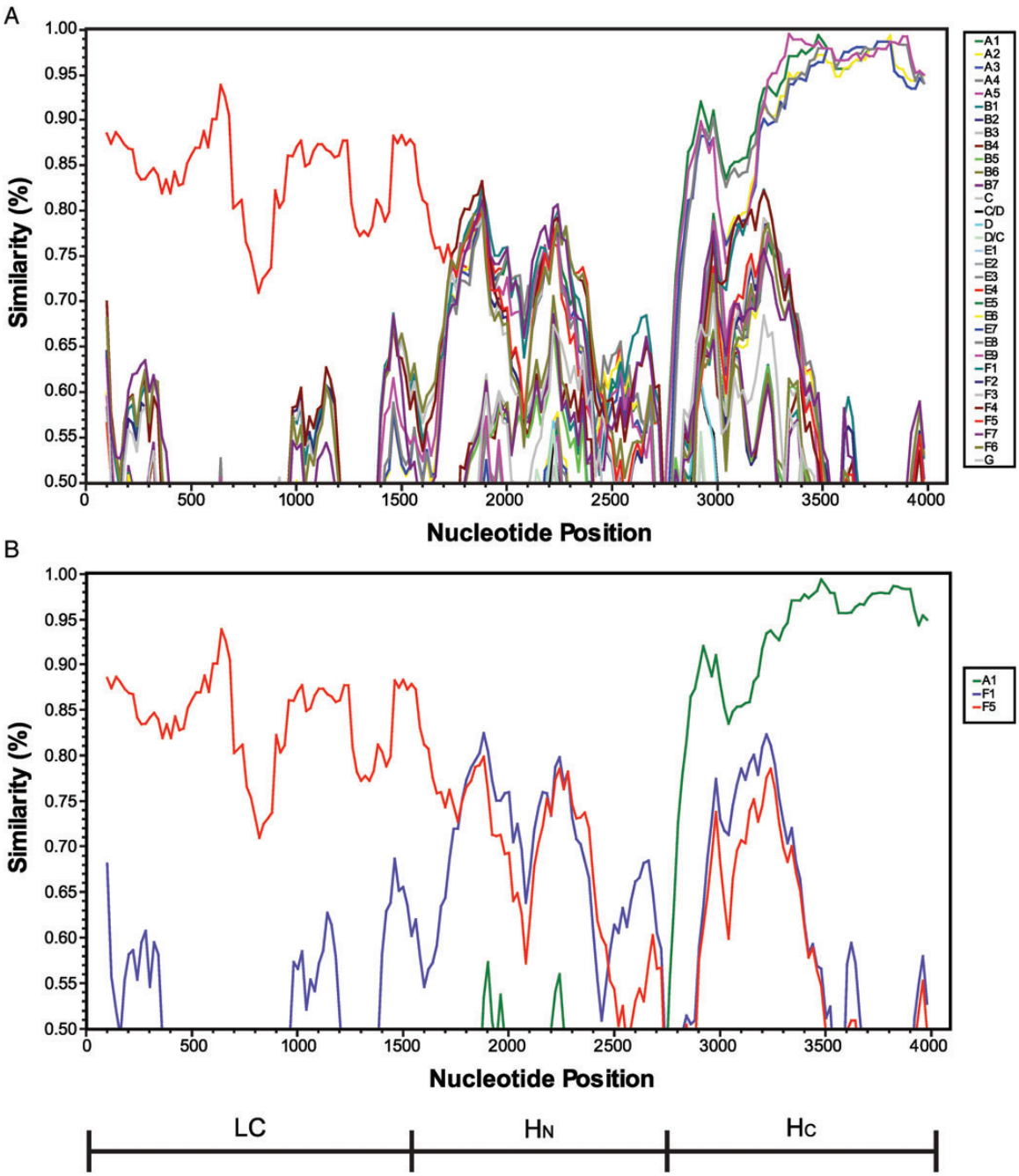

Figure 4. Similarity plots comparing *bont/H* nucleotides sequence to representatives of all *bont* subtypes (*A*) and to *bont/A1*, *bont/F1*, and *bont/F5* subtypes (*B*). The similarity plots show that the *bont/H* sequence is approximately 86% identical to *bont/F5* through nucleotides 1–1500 and approximately 85% identical to *bont/A* through nucleotides 2700–3867. The middle section of the gene (nucleotides 1500–2700, approximately) does not share significant identity with any *bont/A-G* gene sequence but is more closely related to *bont/F* than to any of the other 6 toxin types. A schematic drawing of the *bont* gene segments coding for the botulinum toxin functional domains is presented below the plots. See Figure 1 for GenBank accession numbers. Abbreviations: LC, light chain (catalytic domain); HN, heavy chain translocation domain; Hc, heavy chain binding domain.

gene sequence identified it as belonging to *C. botulinum* group I. The variable region of *flaA* (flaVR) may serve for further genotyping of *C. botulinum* [28, 29]. The IBCA10-7060 flaVR sequences were identical to type 1 flaVR, a group type that contains multiple type A, B, and AB strains [29].

DISCUSSION

C. botulinum strain IBCA10-7060 was found to produce BoNT type B and the novel BoNT type H [20]. We used a combination of Sanger and next-generation sequencing methods to

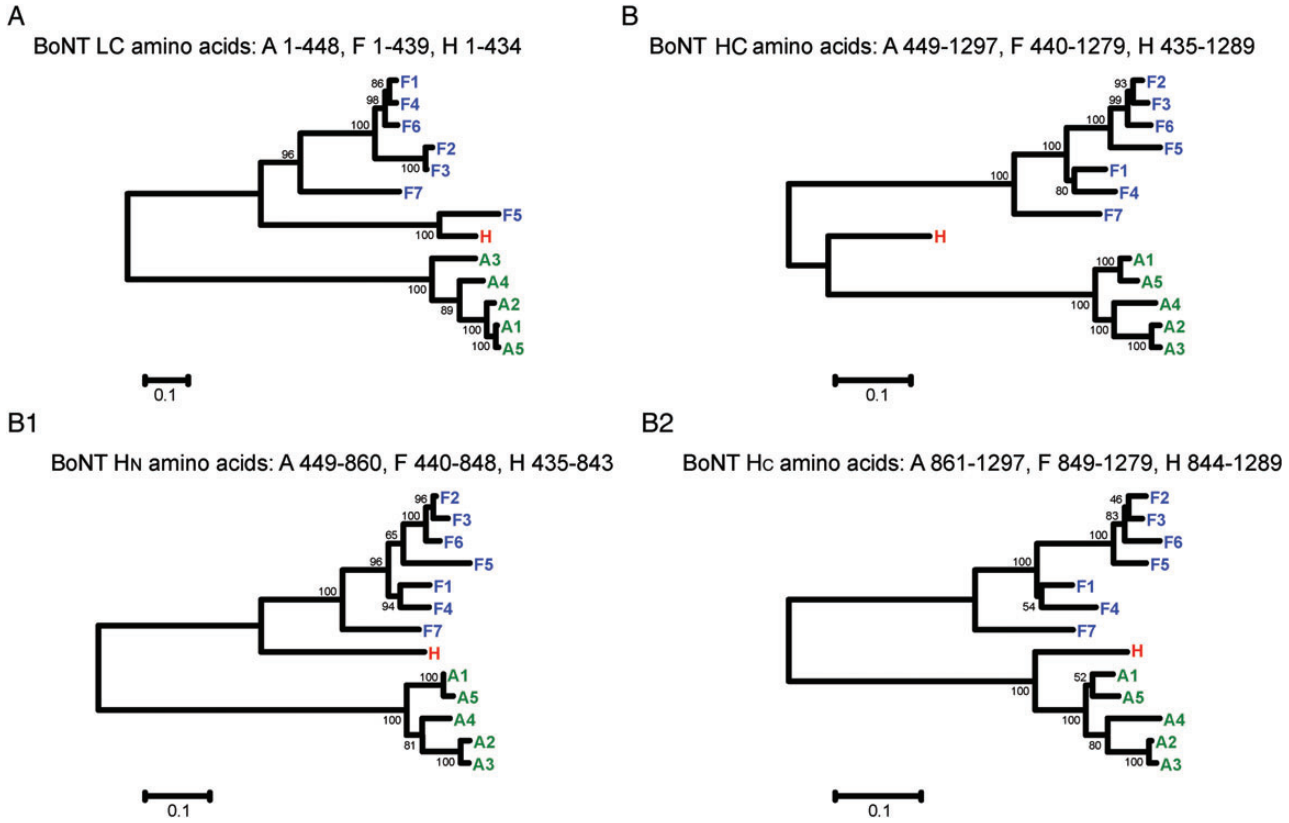

Figure 5. Phylogenetic analysis of amino acid sequences of botulinum neurotoxin (BoNT) types A, F, and H domains. Shown are the light chain catalytic domain; (*A*), the heavy chain (*B*), translocation domain (*B1*), and binding domain (*B2*). One representative of each subtype is shown. All sequences except *bont/H* were obtained from GenBank. The phylogenetic tree was generated using the neighbor-joining method and the JTT matrix model. Genetic distance and bootstrap values are shown. Amino acid positions are based on strain ATCC 1937 for BoNT/A and on strain Langeland for BoNT/F. Note that for light chain amino acid sequences, type H is most closely related to subtype F5; for complete heavy chain amino acid sequences, type H forms its own lineage separated from both the type A and type F clades; for the heavy chain H_N domain amino acid sequences, type H is more closely related to the type F clade than to the type A clade; and for the heavy chain H_C domain amino acid sequences, the situation is reversed, and type H is more closely related to the type A clade than to the type F clade. See Figure 1 for GenBank accession numbers.

identify and characterize both toxin gene clusters and found that one was *bont/B2*. The second toxin gene cluster contained a novel *bont* nucleotide sequence that did not match any of the 7 known *bont* types (ie, A–G). Thus, our sequence characterization provides a molecular explanation for the inability of either a botulinum antitoxins ABF mixture or a US Army heptavalent $F(ab')_2$ botulinum antitoxin A–G to neutralize IBCA10-7060 culture filtrate [20].

BoNT/H is most closely related to BoNT/F5 (Table 1). The BoNT/F5 LC amino acid sequence was 80.0% identical to the LC amino acid sequence of type H, while the other 6 subtype F LC amino acid sequences were <48.9% identical (Figures 4 and 5*A*). However, the LC amino acid sequence of BoNT/F5 had just 46.3%–48.3% identity with the other 6 toxin F subtypes. BoNT/F5 was initially considered to be a hybrid toxin composed of a unique LC and an F2-like HC because its HC amino acid sequence had substantially higher identity (72.7%–89.1%) to the other 6 type F subtypes than its LC did [23]. In contrast, the BoNT/H HC amino acid sequence is only 52.4%–60.2% identical to the HC amino acid sequences of the 7 type F subtypes (Figure 5*B*). These molecular analyses demonstrate that BoNT/H is not simply another subtype of BoNT/F and that BoNT/F5 may need to be redefined as a hybrid of a type H–like LC and a type F2–like HC (Figure 5).

The possible origins of *bont/H* are uncertain. *bont/H* did not result from a simple recombination event between *bont/F5* and *bont/A* because the approximately middle third of *bont/H* is <78% identical to the homologous section of any known *bont/A-G* nucleotide sequence (Figure 4). The 3′ one-third terminus of *bont/H* is closely related to the various toxin type A subtypes, with the percentage of sequence identity highest (96.0%–97.4%) for the 764 nucleotides at the 3′ terminus itself (Figure 4 and 5*B2*). We speculate that *bont/H* resulted from either: (1) multiple recombination events between a *bont/F5*-like gene, an as

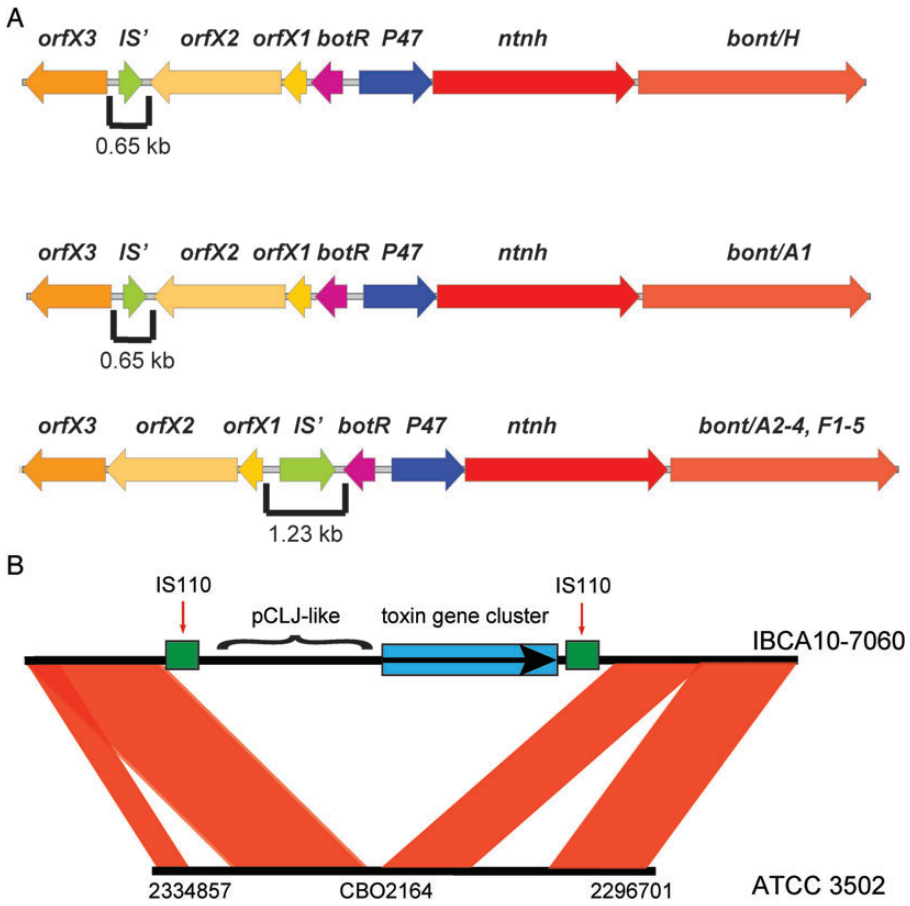

Figure 6. The botulinum neurotoxin gene cluster of *bont/H* (*A*) and its location within strain IBCA10-7060, compared with strain ATCC 3502 (*B*). *A*, *orfX* neurotoxin gene clusters in group I (proteolytic) *Clostridium botulinum* strains were drawn to scale. Gene names and the 0.65-kb *orfX3-orfX2* and 1.23-kb *orfX1-botR* intergenic sequences are indicated. The gene arrangement within the *bont/H* toxin gene cluster is identical to the gene arrangement of the *bont/A*1 *orfX* toxin gene clusters (in *C. botulinum* strains NCTC2916 A(B) and 5328A) but differs from all type F (and subtypes A2–A4) *orfX* toxin gene clusters. *B*, A portion of the serotype A *C. botulinum* ATCC 3502 chromosome (GenBank accession no. NC_009495.1) is shown in comparison to IBCA10-7060. The regions in red indicate similarity between the 2 isolates, and the regions in white are unique. In IBCA10-7060, the *cbo2164* (*araC* family transcriptional regulator) is not present, but instead 2 direct repeat insertion sequence (*IS*) elements, *IS110* (in green), flank an inserted region that contains the toxin gene cluster of *bont/H*, as well as a pCLJ-like sequence. The black arrow indicates the direction of the toxin gene cluster (*orfX3* to *bont/H*).

yet-undiscovered second *bont* gene, and a third *bont/A* gene, each contributing approximately one-third of its gene sequence; or (2) a primordial *bont* gene that also contributed its light chain coding sequence to *bont/F5* and its 3′ one-third terminus (Hc) to *bont/A*.

The 80% amino acid identity of the BoNT/H and BoNT/F5 light chains suggests, that like BoNT/F5, the BoNT/H LC may also cleave synaptobrevin at the unusual 54L and 55E site [30]. The approximately 93% amino acid identity of the BoNT/H Hc binding domain and the binding domains of various BoNT/A subtypes (Figure 4) suggests that BoNT/H may use presynaptic neuronal membrane receptors similar to those used by BoNT/A. Both possibilities will need to be tested experimentally.

The *bont/H orfX* toxin gene cluster has the same gene arrangement as the *bont/A1 orfX* toxin gene clusters of strains NCTC2916 and CDC5328A. These 2 type A toxin gene clusters are uniquely characterized by an intergenic insertion of 0.65 kb between the genes *orfX3* and *orfX2* [15, 31]. Moreover, the *orfX3-orfX2* 0.65 kb intergenic insertion sequences of the *bont/H* and of the *bontA1 orfX* toxin gene clusters were almost identical, suggesting a common origin. The *ntnh* gene aside, we found that the other genes composing the *bont/H* toxin gene cluster were most similar to the genes of the *bont/A1* toxin gene cluster in strain NCTC2916 (Supplementary Table 1), further supporting their possible common origin. Notably in this context, the *bont/H orfX* toxin gene cluster gene arrangement differs from all known type F (including its closest *bont*, subtype F5) *orfX* toxin gene cluster gene arrangements (Figure 6).

The *bont/H* toxin gene cluster had a unique chromosomal location different from those used by other chromosomally

located toxin gene clusters [3, 10, 19]. This novel chromosomal location was between the genes *cbo2163* and *cbo2165*, where an insertion event appears to have deleted gene *cbo2164* and replaced it with the *bont/H* toxin gene cluster and other genetic material between nucleotides 2 308 161 and 2 307 116 (strain ATCC 3502 gene names and numbering; GenBank accession no. AM412317; Figure 6*B*). *cbo2164* codes for an AraC family transcriptional regulator, and its sequence is conserved among group I *C. botulinum* strains. We do not know how the lack of *cbo2164* affects the strain IBCA10-7060 cells. In contrast, all other chromosomally located botulinum toxin gene clusters (including the *bont/B2* toxin gene cluster of strain IBCA10-7060) in *C. botulinum* group I strains are found within the *arsC*, *oppA/brnQ*, and *pulE* loci and in group II strains within the *rarA* locus [3, 10, 19].

The sequence of the *bont/H* toxin gene cluster upstream of its *orfX3* is similar to the upstream sequence of the *bont/A4* toxin gene cluster in plasmid pCLJ. Also, the sequences upstream and downstream of the IS110 elements that flank the *bont/H* insert are similar to chromosomal sequences in several *C. botulinum* strains (Figure 6*B*). These genomic arrangements suggest that *bont/H* may have originated from a pCLJ-like plasmid that integrated into the chromosome. Alternatively, the current chromosomal location of *bont/H* could have resulted from horizontal gene transfer between a plasmid and the chromosome.

The molecular studies reported here complement and reinforce the mouse bioassay studies with botulinum antitoxins A–G that identified the novel botulinum toxin type H produced by *C. botulinum* strain IBCA10-7060 [20]. The *bont/H* gene was in an *orfX* toxin gene cluster that had a novel chromosomal location. Comparison of the *bont/H* nucleotide and predicted amino acid sequences to those of *bont/A-G* by percentage identity, similarity plots, and phylogenetic analyses identified substantial differences and some similarities that distinguished the molecular uniqueness of *bont/H* and its product, BoNT/H.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank Gary Resnick and John C. Detter, Bioscience Division, Los Alamos National Laboratory, for their project contributions and helpful discussions.

Financial support. This work was supported by the Infant Botulism Treatment and Prevention Fund of the California Department of Public Health.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Raffestin S, Couesnon A, Pereira Y, Mazuet C, Popoff MR. Botulinum and tetanus neurotoxins: molecular biology, toxin gene regulation and mode of action. In: Bruggermann H, Gottschalk G, eds. Clostridia: molecular biology in the post-genomic era. Norfolk, UK: Caister Academic Press, **2009**:1–28.
2. Peck MW. Biology and genomic analysis of *Clostridium botulinum*. Adv Microb Physiol **2009**; 55:183–265.
3. Hill KK, Smith TJ. Genetic diversity within *Clostridium botulinum* serotypes, botulinum neurotoxin gene clusters and toxin subtypes. Curr Top Microbiol Immunol **2013**; 364:1–20.
4. Hill KK, Smith TJ, Helma CJ, et al. Genetic diversity among botulinum neurotoxin-producing clostridial strains. J Bacteriol **2007**; 189:818–32.
5. Barash JR, Arnon SS. Dual toxin-producing strain of *Clostridium botulinum* type Bf isolated from a California patient with infant botulism. J Clin Microbiol **2004**; 42:1713–5.
6. Gimenez DF, Ciccarelli AS. New strains of *Clostridium botulinum* subtype Af. Zentralbl Bakteriol Orig A **1978**; 240:215–20.
7. Santos-Buelga JA, Collins MD, East AK. Characterization of the genes encoding the botulinum neurotoxin complex in a strain of *Clostridium botulinum* producing types B and F neurotoxins. Curr Microbiol **1998**; 37:312–8.
8. Hatheway CL, McCroskey LM. Examination of feces and serum for diagnosis of infant botulism in 336 patients. J Clin Microbiol **1987**; 25:2334–8.
9. Smith GE, Hinde F, Westmoreland D, Berry PR, Gilbert RJ. Infantile botulism. Arch Dis Child **1989**; 64:871–2.
10. Dover N, Barash JR, Hill KK, et al. *Clostridium botulinum* Strain Af84 contains three neurotoxin gene clusters: *bont/A2*, *bont/F4* and *bont/F5*. PLoS One **2013**; 8:e61205.
11. Hutson RA, Zhou Y, Collins MD, Johnson EA, Hatheway CL, Sugiyama H. Genetic characterization of *Clostridium botulinum* type A containing silent type B neurotoxin gene sequences. J Biol Chem **1996**; 271: 10786–92.
12. Dover N, Barash JR, Arnon SS. Novel *Clostridium botulinum* toxin gene arrangement with subtype A5 and partial subtype B3 botulinum neurotoxin genes. J Clin Microbiol **2009**; 47:2349–50.
13. Dover N, Barash JR, Hill KK, Detter JC, Arnon SS. Novel structural elements within the non-proteolytic *Clostridium botulinum* type F toxin gene cluster. Appl Environ Microbiol **2011**; 77:1904–6.
14. Smith TJ, Hill KK, Foley BT, et al. 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 **2007**; 2:e1271.
15. Jacobson MJ, Lin G, Raphael B, Andreadis J, Johnson EA. Analysis of neurotoxin cluster genes in *Clostridium botulinum* strains producing botulinum neurotoxin serotype A subtypes. Appl Environ Microbiol **2008**; 74:2778–86.
16. Marshall KM, Bradshaw M, Pellett S, Johnson EA. Plasmid encoded neurotoxin genes in *Clostridium botulinum* serotype A subtypes. Biochem Biophys Res Commun **2007**; 361:49–54.
17. Franciosa G, Maugliani A, Scalfaro C, Aureli P. Evidence that plasmid-borne botulinum neurotoxin type B genes are widespread among *Clostridium botulinum* serotype B strains. PLoS One **2009**; 4: e4829.
18. Eklund MW, Poysky FT, Mseitif LM, Strom MS. Evidence for plasmid-mediated toxin and bacteriocin production in *Clostridium botulinum* type G. Appl Environ Microbiol **1988**; 54:1405–8.
19. Hill KK, Xie G, Foley BT, et al. 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 **2009**; 7:e66.

20. Barash JR, Arnon SS. A novel strain of *Clostridium botulinum* that produces type B and type H botulinum toxins. J Infect Dis **2014**; 209:183–91.
21. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **2011**; 28:2731–9.
22. Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol **1999**; 73:152–60.
23. Raphael BH, Choudoir MJ, Luquez C, Fernandez R, Maslanka SE. Sequence diversity of genes encoding botulinum neurotoxin type F. Appl Environ Microbiol **2010**; 76:4805–12.
24. Henderson I, Whelan SM, Davis TO, Minton NP. Genetic characterisation of the botulinum toxin complex of *Clostridium botulinum* strain NCTC 2916. FEMS Microbiol Lett **1996**; 140:151–8.
25. Hielm S, Bjorkroth J, Hyytia E. Genomic analysis of *Clostridium botulinum* group II by pulsed-field gel electrophoresis. Appl Environ Microbiol **1998**; 64:703–8.
26. Luquez C, Raphael BH, Maslanka SE. Neurotoxin gene clusters in *Clostridium botulinum* type Ab strains. Appl Environ Microbiol **2009**; 75:6094–101.
27. Umeda K, Seto Y, Kohda T, Mukamoto M, Kozaki S. Genetic characterization of *Clostridium botulinum* associated with type B infant botulism in Japan. J Clin Microbiol **2009**; 47:2720–8.
28. Paul CJ, Twine SM, Tam KJ, et al. Flagellin diversity in *Clostridium botulinum* groups I and II: a new strategy for strain identification. Appl Environ Microbiol **2007**; 73:2963–75.
29. Woudstra C, Lambert D, Anniballi F, De Medici D, Austin J, Fach P. Genetic diversity of the flagellin genes of *Clostridium botulinum* Groups I and II. Appl Environ Microbiol **2013**; 79:3926–32.
30. Kalb SR, Baudys J, Webb RP, et al. Discovery of a novel enzymatic cleavage site for botulinum neurotoxin F5. FEBS Lett **2012**; 586: 109–15.
31. Dineen SS, Bradshaw M, Karasek CE, Johnson EA. Nucleotide sequence and transcriptional analysis of the type A2 neurotoxin gene cluster in *Clostridium botulinum*. FEMS Microbiol Lett **2004**; 235: 9–16.