

## Sequence Diversity of Genes Encoding Botulinum Neurotoxin Type F<sup>∇†</sup>

Brian H. Raphael,<sup>1\*</sup> Mallory J. Choudoir,<sup>1</sup> Carolina Lúquez,<sup>1</sup>  
Rafael Fernández,<sup>2</sup> and Susan E. Maslanka<sup>1</sup>

Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, Georgia 30329,<sup>1</sup> and  
Area Microbiología, Universidad Nacional de Cuyo, Mendoza, Argentina<sup>2</sup>

Received 23 December 2009/Accepted 17 May 2010

**Botulism due to type F botulinum neurotoxin (BoNT/F) is rare (<1% of cases), and only a limited number of clostridial strains producing this toxin type have been isolated. As a result, analysis of the diversity of genes encoding BoNT/F has been challenging. In this study, the entire *bont/F* nucleotide sequences were determined from 33 type F botulinum toxin-producing clostridial strains isolated from environmental sources and botulism outbreak investigations. We examined proteolytic and nonproteolytic *Clostridium botulinum* type F strains, bivalent strains, including Bf and Af, and *Clostridium baratii* type F strains. Phylogenetic analysis revealed that the *bont/F* genes examined formed 7 subtypes (F1 to F7) and that the nucleotide sequence identities of these subtypes differed by up to 25%. The genes from proteolytic (group I) *C. botulinum* strains formed subtypes F1 through F5, while the genes from nonproteolytic (group II) *C. botulinum* strains formed subtype F6. Subtype F7 was composed exclusively of *bont/F* genes from *C. baratii* strains. The region of the *bont/F5* gene encoding the neurotoxin light chain was found to be highly divergent compared to the other subtypes. Although the *bont/F5* nucleotide sequences were found to be identical in strains harboring this gene, the gene located directly upstream (*ntnh/F*) demonstrated sequence variation among representative strains of this subtype. These results demonstrate that extensive nucleotide diversity exists among genes encoding type F neurotoxins from strains with different phylogenetic backgrounds and from various geographical sources.**

Botulism is a potentially fatal disease caused solely by the action of serologically distinct neurotoxins (BoNT/A, -B, -C, -D, -E, -F, or -G) which prevent acetylcholine release at neuromuscular junctions, resulting in paralysis. Food-borne botulism may result from the ingestion of a preformed toxin that is produced in inadequately preserved food. Under certain conditions, botulinum neurotoxin-producing *Clostridium* sp. may colonize and produce toxin in wounds (wound botulism) or in the intestine (infant botulism or adult colonization). Globally, human botulism cases are associated with botulinum neurotoxin serotypes A, B, E, and rarely F. The Centers for Disease Control and Prevention (CDC) maintains active surveillance for botulism cases in the United States. Of 1,269 U.S. cases of botulism reported to the CDC between 1981 and 2002, approximately 1% were due to type F toxin (13). An additional 10 cases of type F botulism were reported to the CDC from 2003 to 2007 ([http://www.cdc.gov/national-surveillance/botulism\\_surveillance.html](http://www.cdc.gov/national-surveillance/botulism_surveillance.html)).

Type F botulism was first described in 1960 following an outbreak occurring in Denmark involving liver paste (30). The organism isolated in this outbreak metabolically resembled proteolytic *Clostridium botulinum* strains of types A and B. In a subsequent outbreak, type F toxin was found to be produced by a nonproteolytic *C. botulinum* strain isolated from venison jerky (29). Bivalent toxin-producing strains have been described, including Bf strains isolated from infants in the United

States and England (1, 16, 17, 35) and an Af strain isolated from individuals in Argentina with food-borne botulism (11). Bivalent strains may produce higher titers of one toxin type, which are denoted with a capital letter. The only reported organism isolated from infants with botulism due to type F toxin alone (i.e., not associated with additional serotypes as in bivalent strains) is *Clostridium baratii* (2, 14, 24). In addition, *C. baratii* type F has been isolated from adults with botulism (28) as well as suspect foods associated with botulism cases (15; CDC, unpublished data).

Botulinum neurotoxin genes (*bont*) are typically found within toxin gene clusters that include other genes encoding components of the toxin complex (*ha70*, *ha17*, *ha33*, *ntnh*), regulatory proteins (*botR*), or proteins with unknown functions (*p47*, *orfX1*, *orfX2*, *orfX3*). Two general toxin gene cluster arrangements have been described, including the *orfX* cluster (*orfX3-orfX2-orfX1-botR-p47-ntnh-bont*) and the *ha* cluster (*ha70-ha17-ha33-botR-ntnh-bont*) (21, 22). The *bont/F* genes of type F and type Bf strains examined by Hill et al. (21) were found in an *orfX* cluster.

The amino acid sequence identities of the BoNT serotypes A to G range from approximately 35 to 70% (36). In addition, within nearly all toxin serotypes, various levels of amino acid sequence variation have been observed, resulting in the identification of toxin subtypes (20, 36, 37).

Although a limited number of genes encoding type F botulinum neurotoxin have been sequenced, a comparison of sequences available in public databases indicates that significant diversity exists within this serotype. The nucleotide sequence identity of the type F neurotoxin gene from the proteolytic strain Langeland differs from that of the gene in the nonproteolytic strain 202F by 7%. The type F gene from *C. baratii* strain ATCC 43756 differs from those of Langeland and 202F

\* Corresponding author. Mailing address: Centers for Disease Control and Prevention, 1600 Clifton Road, MS G-29, Atlanta, GA 30329. Phone: (404) 639-4292. Fax: (404) 639-4290. E-mail: BRaphael@cdc.gov.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

∇ Published ahead of print on 28 May 2010.

TABLE 1. Strains examined in this study

Strain <sup>a</sup>	Source	Origin <sup>d</sup>	Yr	Toxin serotype	Toxin gene(s)	<i>bont</i> / <i>F</i> subtype	<i>bont</i> / <i>F</i> accession no.	Reference <sup>b</sup>
<i>C. botulinum</i>								
Langeland	Liver paste	Denmark	1958	F	F	F1	GU213203	30
12F	Unknown	Unknown	Unknown	F	F	F1	GU213204	This study
4VI	Crab	Virginia	~1968	F	F	F1	GU213205	39
4VII <sup>c</sup>	Unknown	Unknown	Unknown	F	F	F1	GU213206	38
8g	Crab	Virginia	~1968	F	F	F1	GU213207	39
Pasteurized crab	Crab	Atlantic coast	~1974	F	F	F1	GU213208	23
CDC4013	Stool	Texas	1981	Bf	B, F	F2	GU213209	16, 17
CDC49930	Honey	Ohio	1994	F	F	F4	GU213210	This study
CDC54074	Soil	Mendoza	1978	F	F	F5	GU213211	This study
CDC54075	Soil	Mendoza	1978	F	F	F5	GU213212	This study
CDC54076	Soil	Mendoza	1978	Af	A, F	F4	GU213213	This study
CDC54078	Chamomile	Argentina	Unknown	F	F	F4	GU213214	This study
CDC54079	Soil	Mendoza	1978	Af	A, F	F5	GU213215	This study
CDC54084	Soil	Mendoza	1978	Af	A, F	F5	GU213216	This study
CDC54085	Soil	Tucuman	1987	F	F	F5	GU213217	This study
CDC54086	Soil	Buenos Aires	2003	F	F	F3	GU213218	This study
CDC54087	Dust	Mendoza	2001	F	F	F4	GU213219	This study
CDC54088	Dust	Mendoza	2001	F	F	F4	GU213220	This study
CDC54089	Anchovies	Argentina	1984	F	F	F4	GU213221	This study
CDC54090	Stool	Mendoza	1981	F	F	F5	GU213222	This study
CDC54091	Soil	Mendoza	1978	Af	A, F	F4	GU213223	This study
CDC54093	Soil	Mendoza	1978	Af	A, F	F4	GU213224	This study
CDC54096	Stool	Mendoza	1982	Af	A, F	F5	GU213225	This study
VPI2382 (ATCC 27321)	Marine sediment	Unknown	Unknown	F	F	F6	GU213226	This study
VPI4257 (F160)	Soil	Argentina	~1968	F	F	F3	GU213227	12
VPI7943 (5436)	Venison jerky	California	1966	F	F	F6	GU213228	29
83F	Marine sediment	Pacific coast	~1967	F	F	F6	GU213229	9
KA-173 (610B 8-11)	Salmon	Columbia River	~1966	F	F	F6	GU213230	This study
<i>C. baratii</i>								
CDC59837	Stool	Oregon	2004	F	F	F7	GU213231	This study
CDC51192	Stool	Iowa	2006	F	F	F7	GU213232	This study
CDC35112	Stool	New Mexico	1980	F	F	F7	GU213233	This study
CDC32356	Stool	Georgia	1987	F	F	F7	GU213234	This study
CDC51267	Deer meat	Thailand	2006	F	F	F7	GU213235	This study

<sup>a</sup> Other strain identifiers shown in parentheses.

<sup>b</sup> References are shown for previously published reports indicating the origin and/or source of a specific strain.

<sup>c</sup> Presumably isolated from crab similar to strain 4VI.

<sup>d</sup> Mendoza, Tucuman, and Buenos Aires are provinces within Argentina.

by 18% and 20%, respectively. Although the bivalent (Bf) strain CDC3281 is phenotypically proteolytic, the toxin gene shows greater similarity to those from nonproteolytic strains (34). In addition to metabolic differences observed between proteolytic and nonproteolytic *C. botulinum* strains as well as *C. baratii*, these organisms are phylogenetically distinct based on differences among their 16S rRNA sequences (5, 20).

In order to define the degree of genetic diversity among strains encoding botulinum neurotoxin type F, we sequenced the *bont*/*F* gene and partially characterized the toxin gene cluster by using a panel of 33 strains with diverse origins. These strains were selected from those available in the CDC culture collection as well as several isolated in Argentina. The only reported Af strains have been isolated in Argentina. Among 68 outbreaks of serotype-confirmed food-borne botulism in Argentina between 1922 and 2007, type F was isolated in two outbreaks, and type Af was isolated in one outbreak. In addition, Lúquez et al. (26) reported isolation of type F and Af strains from Argentine soils.

Here, we report that analysis of the *bont*/*F* genes from the strains examined in this study revealed a high degree of nucle-

otide sequence heterogeneity and the identification of seven type F subtypes (F1 to F7). In addition, the nucleotide sequence of one subtype (F5) has not been previously reported and contains evidence of recombination compared to the other subtypes.

#### MATERIALS AND METHODS

**Bacterial strains used in this study.** Clostridial strains were grown anaerobically at 35°C either on egg yolk agar (EYA) medium (7) or in Trypticase-peptone-glucose-yeast extract (TPGY) medium (Remel, Lenexa, KS). Stock cultures were stored in bovine brain medium (7) at 4°C. Strains used in this study are identified in Table 1.

**Characterization of strains.** Neurotoxin genes *bont*/*A*, *bont*/*B*, and *bont*/*F* were detected using PCR. Previously reported primers A2Fa and A2Ra were used to amplify *bont*/*A* (33), and primers B2Fa and B3Ra were used to amplify *bont*/*B* (27). Real-time PCR primers and probes reported by Fach et al. (10) were used to detect *bont*/*F* with the following modifications. The assay was performed using an ABI 7500 Fast Real-Time PCR system (Applied Biosystems, Foster City, CA) and FastStart DNA Master HybProbe (Roche, Indianapolis, IN) PCR reagents. The PCR conditions included an initial 95°C cycle for 10 min followed by 40 cycles of 95°C for 10 s and 50°C for 70 s.

TPGY cultures were inoculated with clostridial strains and incubated for 5

days. Botulinum toxins were detected in culture supernatants by the mouse bioassay, using standard methods (4).

**Isolation of genomic DNA.** TPGY cultures (10 ml) were incubated for 24 h and then centrifuged for 10 min at 4,000 rpm in a swinging-bucket centrifuge to pellet the bacteria. The pellet was resuspended in 300 µl of Tris-EDTA (TE) containing 30 mg/ml lysozyme (Sigma, St. Louis, MO) and incubated for 30 min at 37°C. The MasterPure DNA purification kit (Epicentre, Madison, WI) was used with some modifications to isolate genomic DNA. Specifically, 2× tissue and cell lysis buffer (300 µl) containing 200 µg RNase A (Qiagen, Valencia, CA) was added to the resuspended bacterial pellet followed by 350 µl of protein precipitation buffer. The suspension was centrifuged for 10 min at 4°C, and genomic DNA was precipitated by addition of 1 ml of isopropanol. The DNA was resuspended in TE buffer, passed through a 0.1-µm centrifuge filter (Millipore, Billerica, MA), and stored at 4°C until analysis (<6 months).

**Botulinum neurotoxin gene sequencing.** Overlapping fragments of the entire *bont/A* and *bont/B* genes were amplified and sequenced using previously reported primers (27, 33). Primers used to amplify and sequence *bont/F* genes were selected from various previously reported sequences and from unique *bont/F* sequences determined during the course of this study (see Table S1 in the supplemental material). PCR was performed with Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA), and thermocycling conditions included an initial 95°C cycle for 10 min followed by 40 cycles of 95°C for 30 s, 50°C for 30 s, and 68°C for 1.5 to 3 min. DNA sequencing was performed using an Applied Biosystems 3730 DNA analyzer.

**DNA sequence and phylogenetic analysis.** DNA sequences were assembled using Sequencher (Gene Codes, Ann Arbor, MI). BIOEDIT (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) was used to edit sequences and perform predicted amino acid translation. Gene sequences determined in this study and from previously published sequences were aligned using CLUSTALW, and phylogenetic analysis was performed using MEGA 4 (<http://www.megasoftware.net>).

Additional comparative analysis among *bont/F* genes was performed using SimPlot (<http://sray.med.som.jhmi.edu/SCoftware/simplot/>). Pairwise comparisons of nucleotide and predicted amino acid sequences were performed using the EMBOSS pairwise alignment algorithm (<http://www.ebi.ac.uk/Tools/emboss/align/index.html>).

An approximately 1.3-kb region of the 16S rRNA gene of each strain examined in this study was amplified and sequenced using primers and conditions reported previously (20).

**Analysis of partial neurotoxin gene clusters.** A PCR assay targeting an internal fragment of *p47* by the use of previously reported primers (33) was used to detect the presence of this gene in genomic DNA of the strains examined in this study.

Partial neurotoxin gene clusters present in representative F5 strains CDC54079, CDC54085, CDC54090, and CDC54096 were amplified and sequenced using primers shown in Table S2 in the supplemental material. The thermocycling conditions included an initial 95°C cycle for 10 min followed by 40 cycles of 95°C for 30 s, 50°C for 30 s, and 68°C for 1.5 to 3 min.

**Nucleotide sequence accession numbers.** The GenBank accession numbers for the *bont/F* genes are shown in Table 1. The *bont/A* sequences of Af strains sequenced in this study were deposited under accession numbers GU206530 to GU206535. The *bont/B* sequence of Bf strain CDC4013 was deposited under accession number GU271943.

The *ntnh/F* sequences present in strains CDC54079, CDC54085, CDC54090, and CDC54096 were deposited under accession numbers GU221928, GU221929, GU299676, and GU299675, respectively. The *ntnh/A* sequences present in strains CDC54079 and CDC54096 were deposited under accession numbers GU221927 and GU299674, respectively.

**RESULTS**

**Characterization of type F strains.** As shown in Table 1, a panel of type F botulinum neurotoxin-producing strains representing various geographical regions and sources was assembled. The panel consisted of 21 *C. botulinum* type F strains, 5 *C. baratii* type F strains, 1 *C. botulinum* type Bf strain, and 6 *C. botulinum* type Af strains.

As shown in Fig. 1, 16S rRNA gene sequencing revealed that 24 of the strains examined had sequences corresponding to *C. botulinum* group I, four of the strains were *C. botulinum* group II, and five of the strains clustered with the *C. baratii* sequence.

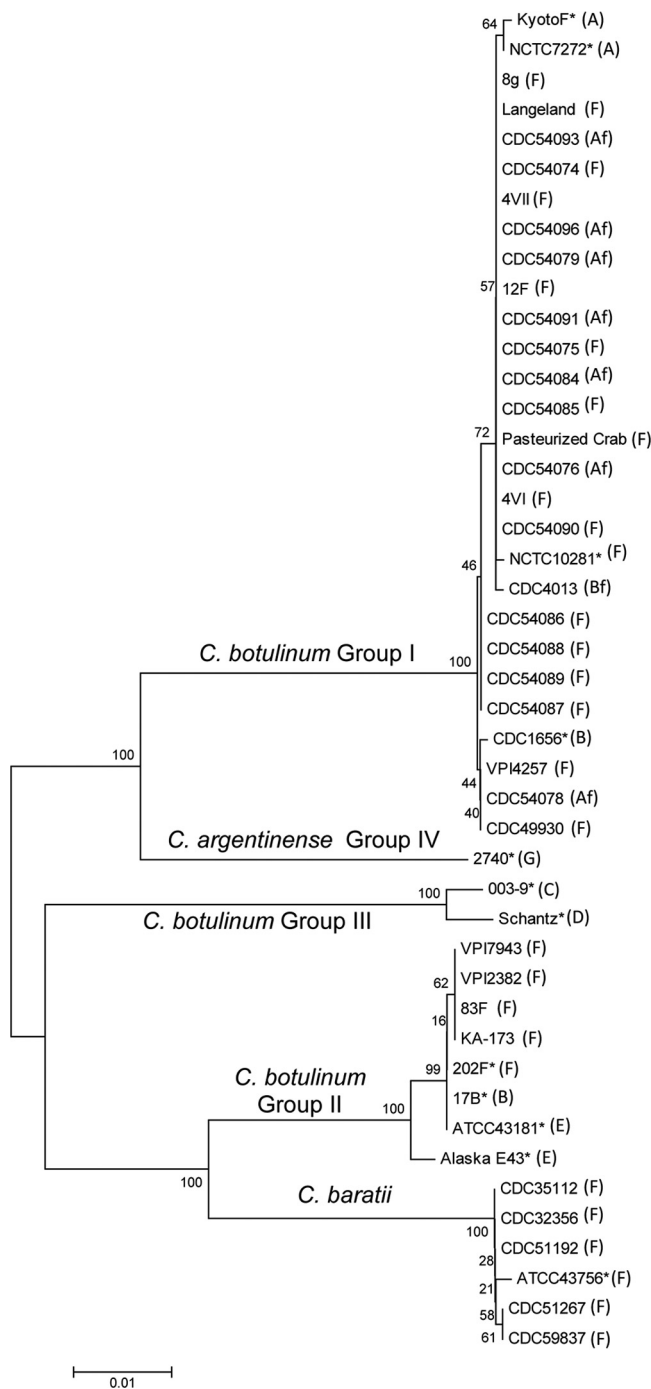


FIG. 1. Phylogenetic analysis of partial 16S rRNA gene nucleotide sequences (~1.3 kb) from various botulinum toxin-producing clostridial strains. The phylogenetic tree was generated using the neighbor-joining method. Bootstrap values and genetic distance (bar) are shown. Clusters corresponding to different phylogenetic groups are labeled according to previous reports (5, 20). The toxin serotype of each strain is shown in parentheses. An asterisk indicates that the sequence of the specified strain was obtained from GenBank (accession numbers are as follows: KyotoF, X73844; NCTC7272, X68185; NCTC10281, X68172; CDC1656, EF051572; 2740, EF030542; 003-9, EF030537; Schantz, EF030538; 202F, EF030541; 17B, EF030536; ATCC 43181, EF030540; Alaska E43, EF030539; ATCC 43756, X68175).

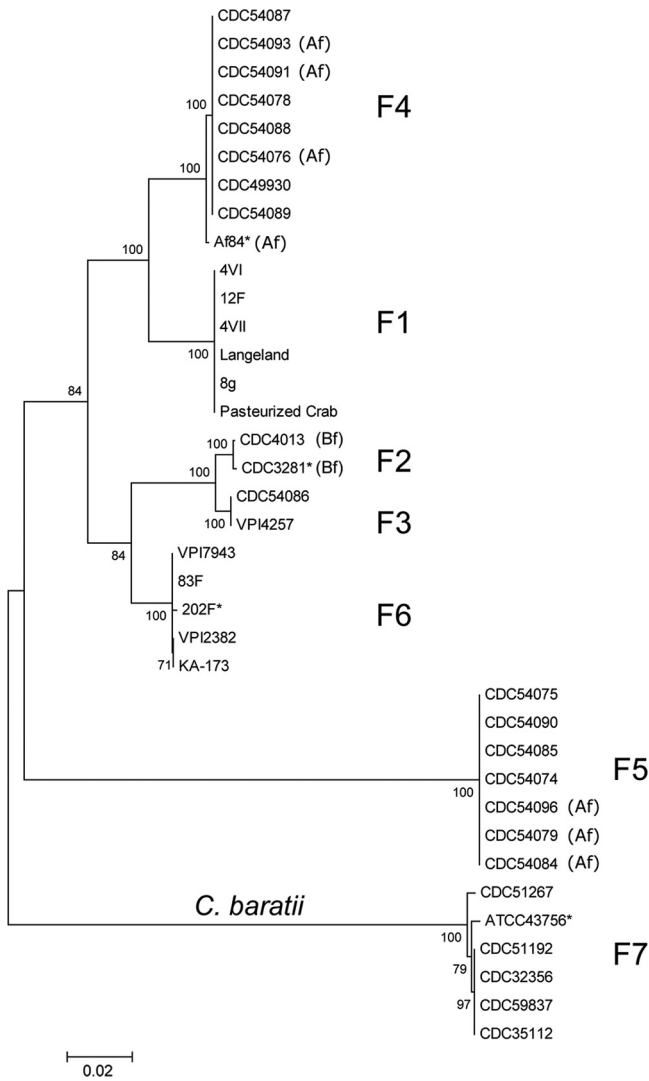


FIG. 2. Phylogenetic analysis of *bont/F* nucleotide sequences. The phylogenetic tree was generated using the neighbor-joining method. Bootstrap values and genetic distance (bar) are shown. Clusters designated subtypes F1 to F7 are labeled on the right. The toxin serotypes of bivalent strains are shown in parentheses. An asterisk indicates that the *bont/F* sequence of the specified strain was obtained from GenBank (accession numbers for the strains are as follows: CDC3281, Y13631; 202F, M92906; Af84, FJ968748; ATCC 43756, X68262).

**Neurotoxin gene sequences.** The *bont/F* gene sequence was determined for each of the strains in the panel. As shown in Fig. 2, phylogenetic analysis of the *bont/F* genes and additional sequences available in public databases revealed a high degree of nucleotide sequence variation. The *bont/F* genes formed seven clusters termed subtypes F1 to F7. The nucleotide sequence identities of these subtypes differed by up to 25% (Table 2).

Subtypes F1 to F5 consisted of *C. botulinum* group I strains, while subtype F6 consisted of *C. botulinum* group II strains. The *bont/F* genes from *C. baratii* type F strains were >99% identical to each other and formed subtype F7. Notably, the F7 subtype sequences were the most divergent compared to the other subtypes. The type F and Af strains from Argentina

TABLE 2. Nucleotide and amino acid identities of representative BoNT/F subtypes

Subtype <sup>a</sup>	% Nucleotide sequence identity, % amino acid sequence identity with:					
	F2	F3	F4	F5	F6	F7
F1	90.9, 83.5	91.2, 83.9	96.0, 92.3	80.3, 69.9	93.2, 87.7	82.0, 73.7
F2		98.9, 97.1	91.8, 83.5	82.9, 74.2	95.2, 90.2	79.5, 68.9
F3			91.8, 83.8	83.0, 74.0	95.1, 90.1	79.2, 69.2
F4				81.3, 69.5	93.6, 87.1	81.1, 72.2
F5					83.3, 74.0	74.5, 63.9
F6						80.0, 69.9

<sup>a</sup> Representative strains of each BoNT/F subtype are as follows: F1, Langeland; F2, CDC4013; F3, CDC54086; F4, CDC54078; F5, CDC54074; F6, VPI7943; F7, CDC59837.

were divided among subtypes F3, F4, and F5. Interestingly, bivalent Af strains harbored *bont/F* genes from subtype F4 or F5, while Bf strains harbored *bont/F* genes only from subtype F2.

All of the strains harboring *bont/F5* originated in Argentina and were 100% identical to each other despite being isolated from distinct sources (i.e., soil and stool). Strains harboring *bont/F4* included several Argentine strains from distinct sources sequenced in this study; strain Af84 (GenBank accession number FJ968748), which also originated in Argentina; and a type F strain (CDC49930) isolated from honey in the United States.

As shown in Fig. 3, all of the F subtypes showed poor similarity with *bont/F5* for approximately the first 1,500 nucleotides of the gene. BLAST analysis revealed that while the full-length *bont/F5* sequence has not been previously reported, nucleotides 1319 to 3834 of *bont/F5* showed the highest degree of sequence identity (94%) to the *bont/F2* gene from type Bf strain CDC3281 (data not shown).

We also sequenced the *bont/A* and *bont/B* genes in the bivalent Af and Bf strains, respectively. Consistent with previous observations (20), all of the type Af strains examined harbored an A2 subtype neurotoxin gene, and the Bf strain (CDC4013) harbored the *bont/B* gene associated with bivalent strains (subtype bvB).

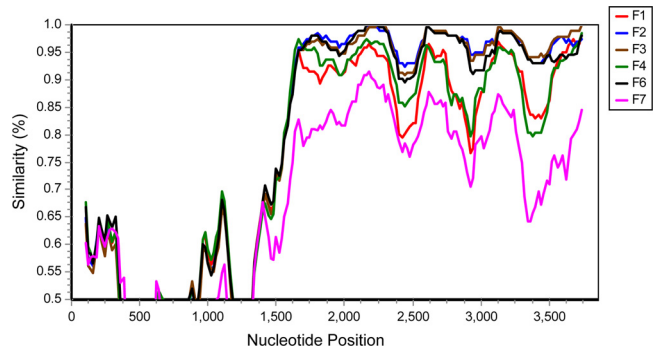


FIG. 3. Similarity plot comparing representative *bont/F* nucleotide sequences to *bont/F5* from strain CDC54074. Sequences from representative strains are compared in the plot as follows: F1, Langeland; F2, CDC4013; F3, CDC54086; F4, CDC54078; F6, VPI7943; F7, CDC59837. Nucleotide sequences were aligned using CLUSTALW, and the plot was rendered with Simplot using a 200-bp window and a 20-bp step.

TABLE 3. Comparison of light-chain and heavy-chain amino acid sequences of BoNT/F subtypes

Subtype <sup>a</sup>	% Amino acid identity with <sup>b</sup> :											
	F2		F3		F4		F5		F6		F7	
	LC	HC	LC	HC	LC	HC	LC	HC	LC	HC	LC	HC
F1	81.8	84.3	82.9	84.4	96.4	90.1	47.3	81.8	94.3	84.2	63.3	79.2
F2			97.9	96.7	82.7	83.9	46.3	89.1	81.5	94.8	59.9	73.5
F3					83.8	83.8	46.3	88.8	82.9	93.8	60.8	73.6
F4							47.6	81.1	93.6	83.7	64.2	76.4
F5									48.3	87.6	46.9	72.7
F6											63.3	73.4

<sup>a</sup> Representative strains of each BoNT/F subtype are as follows: F1, Langeland; F2, CDC4013; F3, CDC54086; F4, CDC54078; F5, CDC54074; F6, VPI7943; F7, CDC59837.

<sup>b</sup> LC, light chain; HC, heavy chain.

### Comparison of amino acid sequences in subtypes F1 to F7.

The predicted BoNT/F amino acid sequences from representative strains of each BoNT/F subtype were compared in pairwise alignments (Table 2). While subtypes F2 and F3 shared the greatest amino acid sequence identity (97.1%), subtypes F5 and F7 shared the least sequence identity (63.9%).

Botulinum neurotoxins contain a zinc endopeptidase domain (HExxH) found in the light chain and a membrane-spanning domain (PYxGxAL) found in the heavy chain (18, 25). The amino acid sequences associated with these functional domains are conserved in each of the representative type F subtypes (data not shown). Moreover, residues C429 and C445 (positions based on the Langeland sequence), which form a disulfide linkage between the light and heavy chains, are also conserved in each type F subtype representative (data not shown).

The light chain region of subtype F5 demonstrated the lowest level of amino acid sequence identity (range, 46.3% to 48.3%) to any other subtype (Table 3). However, the heavy chain region of subtype F5 demonstrated significantly high amino acid sequence identity compared to the other subtypes (range, 72.7% to 89.1%).

The light chain region of subtype F7 also displayed low sequence identity (range, 46.9% to 64.2%) with the other subtypes. Interestingly, the heavy chain of this subtype shared the lowest amino acid sequence identity with the other subtypes (range, 72.7% to 79.2%).

**Neurotoxin gene cluster characterization.** The gene *p47* is a component of a class of neurotoxin gene clusters that typically also harbors *orfX1-orfX3*. All of the strains examined in this study harbored *p47* as demonstrated by PCR targeting this gene (data not shown).

Because of the presence of a unique 5' region in the *bont/F5* sequence, the *ntnh/F* gene, which is located directly downstream of *p47* and upstream of the neurotoxin gene, was sequenced to determine if there was any evidence of recombination in *ntnh/F*. Since the F5 subtype contained both monovalent type F strains and bivalent Af strains, we also wanted to determine (i) the similarity of the *ntnh/F* sequences between monovalent and bivalent strains and (ii) whether bivalent Af strains harbored separate toxin gene clusters associated with each neurotoxin gene or if both neurotoxin genes were associated with the same toxin gene cluster.

The *ntnh/F* genes from representative subtype F5 strains

CDC54079, CDC54085, CDC54090, and CDC54096 are compared in Fig. 4A. The genes found in strains CDC54085, CDC54090, and CDC54096 are identical and share approximately 96% nucleotide sequence identity with the gene found in strain CDC54079.

In bivalent (Af) strains CDC54079 and CDC54096, separate PCR products were obtained when the region between *p47* and either *bont/A* or *bont/F5* was amplified, indicating that the neurotoxin genes are arranged in different gene clusters (Fig. 4B).

The *ntnh/A* genes in strains CDC54079 and CDC54096 differed by a single nucleotide. Furthermore, the *ntnh/A* gene found in strain CDC54079 was identical with that found in the previously sequenced subtype A2 strain Kyoto (Fig. 4A).

The *ntnh/F* gene in the type Af strain CDC54079 differs from the *ntnh/F* genes in strains CDC54085, CDC54090, and CDC54096 by 124 nucleotides. Moreover, the *ntnh/F* gene in strain CDC54079 shared greater sequence identity (>99%) with *ntnh/A* in the bivalent strains examined than with *ntnh/F* from the remaining representative subtype F5 strains examined (Fig. 4A). The *ntnh/F* gene in strain CDC54079 is nearly identical to the *ntnh/A* gene in this strain until nucleotide position 3418 (Fig. 4C); however, from this position, the *ntnh/F* genes from CDC54079 and CDC54085 are identical through the remainder of the gene. Notably, all of the *ntnh/F* genes are 12 nucleotides longer than the *ntnh/A* gene found in strain CDC54079 or CDC54096.

## DISCUSSION

The term "toxin subtype" has been applied to strains producing toxins with multiple antigenic components (e.g., Af) and more recently to different clades of highly similar neurotoxin gene sequences encoding specific toxin serotypes (e.g., A1 to A5) (11, 20, 37). In 2005, Smith et al. (37) reported that among the 49 complete neurotoxin sequences (A through G) available at that time, the lowest level of amino acid sequence variability within a serotype was 2.6%. In that study, the minimum level of BoNT/F amino acid sequence differences among type F strains (including proteolytic, nonproteolytic, and bivalent *C. botulinum* as well as *C. baratii* strains) was 10.7%, indicating that all of the sequences were different subtypes. Hill et al. (20) examined the diversity of neurotoxin genes within serotypes A, B, and E and clustered toxin genes into

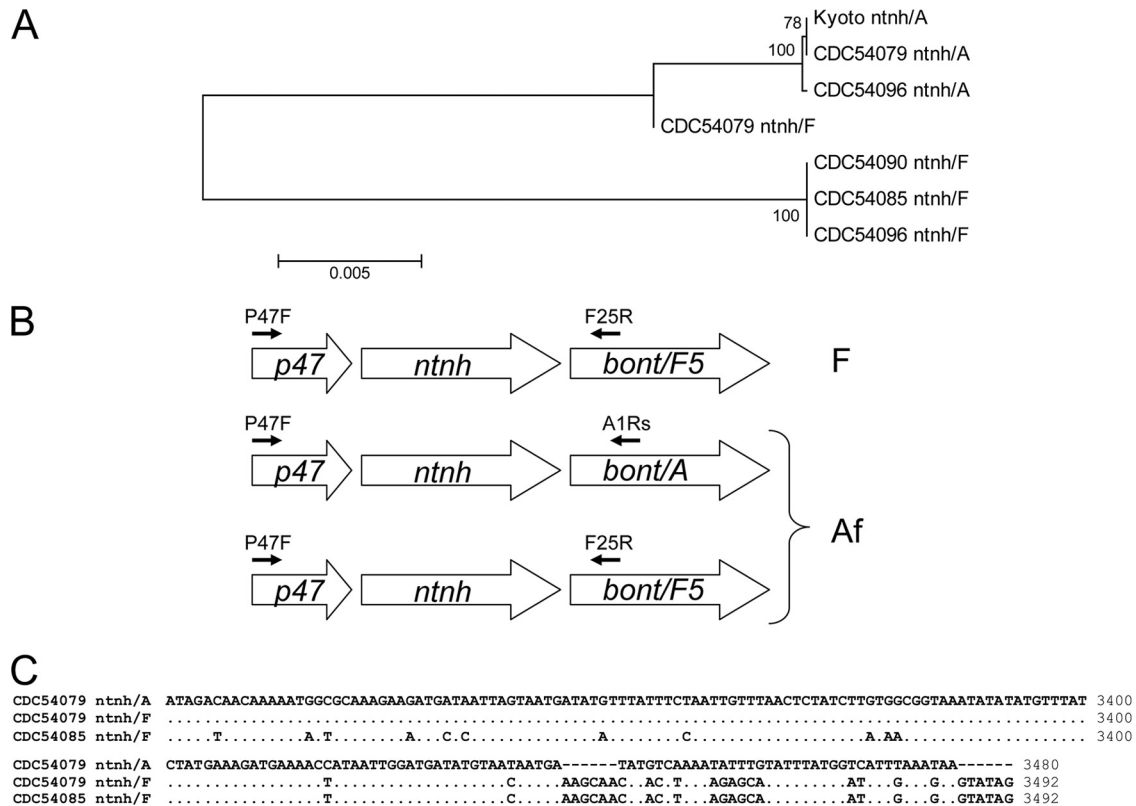


FIG. 4. Analysis of the locations and nucleotide sequences of the *ntnh* gene in representative subtype F5 strains. (A) Phylogenetic tree of the *ntnh* nucleotide sequences found in monovalent subtype F5 strains (CDC54085 and CDC54090) and bivalent (Af) subtype F5 strains (CDC54079 and CDC54096). The *ntnh* gene associated with *bont/F5* is indicated as "*ntnh/F*," and the *ntnh* gene associated with a *bont/A* gene is indicated as "*ntnh/A*." *ntnh/A* from the subtype A2 strain Kyoto (accession number CP001581) is shown as a reference. The tree was generated using the neighbor-joining method. Bootstrap values and genetic distance (bar) are shown. (B) Schematic representation of the arrangement of partial toxin gene clusters in monovalent (labeled "F") and bivalent (labeled "Af") subtype F5 strains. The locations of primers used to amplify the *ntnh* genes for sequencing are shown with arrows. (C) Alignment of the 3' ends of the *ntnh/A* and *ntnh/F* nucleotide sequences found in representative subtype F5 strains CDC54079 and CDC54085.

subtypes by using a phylogenetic approach. The levels of subtype nucleotide sequence variation differ among serotypes. For instance, the nucleotide sequences of subtypes E1 to E5 differ by up to 3%, while subtypes A1 to A5 vary by up to 8% (3, 6, 20).

The objective of this study was to define the sequence diversity of type F botulinum neurotoxin genes found in strains isolated from different sources and geographical origins. Although type F strains are infrequently isolated from individuals with botulism and have a low prevalence in the environment, a surprisingly high level of diversity was found among *bont/F* genes, resulting in the identification of seven subtypes (F1 to F7). These subtypes differed from each other by up to 25% at the nucleotide level, indicating that serotype F strains contain the highest level of nucleotide sequence diversity reported to date. Although a high level of genetic diversity exists among *bont/F* genes, the phylogenetic diversity of organisms harboring subtypes F1 to F5 belonged to *C. botulinum* group I, while subtypes F6 and F7 were found in *C. botulinum* group II and *C. baratii* strains, respectively.

Subtypes F1 and F6 contain strains isolated from food associated with human botulism. Strain Langeland was isolated

from liver paste, and VPI7943 was isolated from venison jerky (29, 30). Remarkably, both subtypes F1 and F6 contain several strains isolated from marine environments.

The F2 and F3 subtypes share the highest level of nucleotide sequence identity. Subtype F2 contains nearly identical *bont/F* genes from bivalent (Bf) strains (CDC4013 and CDC3281) which were isolated from different infants with botulism (16, 17, 34). Subtype F3 strains are monovalent and contain *bont/F* genes from two different strains (CDC54086 and VPI4257) isolated from soil separated by 400 miles in Buenos Aires province (Argentina). The BoNT/F amino acid sequences of the representative subtype F2 strain (CDC4013) differed from those of the subtype F3 strains (CDC54086 and VPI4257) by 2.9%.

All but one of the subtype F4 strains was isolated from Argentina. Strain CDC49930 was isolated from honey that was examined during an investigation of a laboratory-confirmed (type A) infant botulism case in Ohio. Since the toxin serotypes of the organism isolated from the infant and the isolate from the honey were different, the honey was not the source of infant botulism in this case. Subtype F4 also includes strains isolated from soil, dust, chamomile, and canned anchovies (not associated with a botulism case). The strains isolated from soil

(CDC54076, CDC54091, CDC54093) originated in Mendoza province, which included locations with different types of land use (cultivated, uncultivated, etc.).

All of the subtype F5 strains were isolated from Argentina. These strains included those isolated from the stools of different individuals with botulism (CDC54090, CDC54096) as well as from soil. Notably, the strains isolated from soil originated in different locations in Mendoza and Tucuman provinces.

We examined five *C. baratii* strains isolated from separate botulism cases. Four of the *bont/F7* genes which were isolated from *C. baratii* cases occurring in the United States had identical nucleotide sequences. One isolate from food associated with a botulism outbreak occurring in Thailand harbored a *bont/F7* gene that differed from the others by <1%.

Botulinum neurotoxins are composed of a 50-kDa light chain and a 100-kDa heavy chain. The heavy chain possesses a binding domain that enables binding and uptake of the toxin into the neuronal cell. The light chain, which possesses endopeptidase activity, cleaves specific proteins associated with synaptic vesicle docking at the membrane of the host cell. Comparison of the amino acid sequences of the light chain and heavy chain of BoNT/F1 to -F7 revealed that the light chain of BoNT/F5 was highly divergent compared to those of the other subtypes. Although known functional domains of BoNT/F5 are intact, the activity of the light chain in this subtype compared to that of the others remains to be determined. Two lines of evidence indicate that the light chain is functional: (i) the F5 strain CDC54090 was isolated from an individual with botulism, and (ii) F5 strains produced culture supernatants that required neutralization with F antitoxins in the mouse bioassay.

The heavy chain of BoNT/F5 is most similar to that of BoNT/F2 (89.1% identity). As a result, BoNT/F5 appears to be a hybrid composed of a unique light chain and an F2-like heavy chain. Therefore, we sequenced the *ntnh/F* genes in representative monovalent and bivalent (Af) subtype F5 strains to determine if there was evidence of recombination in this gene, which is located directly upstream of *bont/F*. While *ntnh/F* genes found in three subtype F5 strains (CDC54085, CDC54090, and CDC54096) were identical, the gene in strain CDC54079 differs from the others by approximately 4%. This finding is surprising as the *bont/F5* genes are 100% identical to each other.

We also sequenced the *ntnh/A* gene from Af strains CDC54079 and CDC54096. The *ntnh/F* gene in strain CDC54079 was nearly identical to the *ntnh/A* gene until nucleotide position 3418. At this position, the *ntnh/F* sequence in strain CDC54079 becomes 100% identical with the gene found in strains CDC54085, CDC54090, and CDC54096 through the stop codon. These results indicate that a similar recombination event likely occurred in the 3' end of the *ntnh/F* genes in all of the subtype F5 strains examined. Similarly, the *ntnh/F* and *ntnh/A* genes from strains Langeland and Kyoto, respectively, share a high degree of nucleotide identity but have divergent 3' termini followed by different neurotoxin genes (8, 21).

Further work is needed to define the genotypic relationships of strains within and among the F subtypes identified. Nevas et al. (31) demonstrated that three distinct proteolytic (group I) type F strains had indistinguishable SacII pulsed-field gel electrophoresis (PFGE) patterns, while another study (19) indi-

cated that two out of three nonproteolytic (group II) strains had indistinguishable SmaI PFGE profiles. It is possible that the level of genotypic diversity among strains harboring different *bont/F* subtypes varies.

The extent of the diversity in *bont/F* nucleotide sequences determined in this study underscores the importance of examining *C. botulinum* strains isolated from multiple sources and geographical locations. These findings have particular impact for the design of detection assays for both the *bont/F* gene and type F botulinum neurotoxin.

#### ACKNOWLEDGMENTS

We thank Lavin Joseph (CDC) for critical review of the manuscript. DNA sequencing was performed at the Division of Food-borne, Bacterial, and Mycotic Diseases Genomics Unit (CDC).

This publication was supported by funds made available from the Centers for Disease Control and Prevention, Coordinating Office for Terrorism Preparedness and Emergency Response.

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

#### REFERENCES

1. Barash, J. R., and S. S. Arnon. 2004. Dual toxin-producing strain of *Clostridium botulinum* type Bf isolated from a California patient with infant botulism. *J. Clin. Microbiol.* **42**:1713-1715.
2. Barash, J. R., T. W. H. Tang, and S. S. Arnon. 2005. First case of infant botulism caused by *Clostridium baratii* type F in California. *J. Clin. Microbiol.* **43**:4280-4282.
3. Carter, A. T., C. J. Paul, D. R. Mason, S. M. Twine, M. J. Alston, S. M. Logan, J. W. Austin, and M. W. Peck. 2009. Independent evolution of neurotoxin and flagellar genetic loci in proteolytic *Clostridium botulinum*. *BMC Genomics* **10**:115. doi:10.1186/1471-2164-10-115.
4. Centers for Disease Control and Prevention. 1998. Botulism in the United States (1899-1996). Handbook for epidemiologists, clinicians, and laboratory workers. U.S. Department of Health and Human Services, CDC, Atlanta, GA.
5. Collins, M. D., and A. K. East. 1998. Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *J. Appl. Microbiol.* **84**:5-17.
6. Dover, N., J. R. Barash, and S. S. Arnon. 2009. Novel *Clostridium botulinum* toxin gene arrangement with subtype A5 and partial subtype B3 botulinum neurotoxin genes. *J. Clin. Microbiol.* **47**:2349-2350.
7. Dowell, V. R., and T. M. Hawkins. 1990. Media for isolation and characterization of anaerobic bacteria, p 53-59. In V. R. Dowell and T. M. Hawkins (ed.), Laboratory methods in anaerobic bacteriology: CDC laboratory manual. U.S. Department of Health and Human Services, CDC, Atlanta, GA.
8. East, A. K., M. Bhandari, S. Hieml, and M. D. Collins. 1998. Analysis of the botulinum neurotoxin type F gene clusters in proteolytic and nonproteolytic *Clostridium botulinum* and *Clostridium baratii*. *Curr. Microbiol.* **37**:262-268.
9. Eklund, M. W., R. T. Poysky, and D. I. Wieler. 1967. Characteristics of *Clostridium botulinum* type F isolated from the Pacific Coast of the United States. *Appl. Microbiol.* **15**:1316-1323.
10. Fach, P., P. Micheau, C. Mazuet, S. Perelle, and M. Popoff. 2009. Development of real-time PCR tests for detecting botulinum neurotoxins A, B, E, F producing *Clostridium botulinum*, *Clostridium baratii* and *Clostridium butyricum*. *J. Appl. Microbiol.* **107**:465-473.
11. Fernández, R. A., A. S. Ciccarelli, G. N. Arenas, and D. F. Giménez. 1986. First outbreak of botulism caused by *Clostridium botulinum* subtype Af. *Rev. Argent. Microbiol.* **18**:29-31.
12. Giménez, D. F., and A. S. Ciccarelli. 1968. *Clostridium botulinum* type F in the soil of Argentina. *Appl. Microbiol.* **16**:732-734.
13. Gupta, A., C. J. Sumner, M. Castor, S. Maslanka, and J. Sobel. 2005. Adult botulism type F in the United States, 1981-2002. *Neurology* **65**:1694-1700.
14. Hall, J. D., L. M. McCroskey, B. J. Pincomb, and C. L. Hatheway. 1985. Isolation of an organism resembling *Clostridium baratii* which produces type F botulinum toxin from an infant with botulism. *J. Clin. Microbiol.* **21**:654-655.
15. Harvey, S. M., J. Sturgeon, and D. E. Dassey. 2002. Botulism due to *Clostridium baratii* type F toxin. *J. Clin. Microbiol.* **40**:2260-2262.
16. Hatheway, C. L., and L. M. McCroskey. 1987. Examination of feces and serum for diagnosis of infant botulism in 336 patients. *J. Clin. Microbiol.* **25**:2334-2338.
17. Hatheway, C. L., and L. M. McCroskey. 1989. Unusual neurotoxicogenic clostridia recovered from human fecal specimens in the investigation of botu-

- lism, p. 477–481. In T. Hattori, Y. Ishida, Y. Maruyama, R. Y. Morita, and A. Uchida (ed.), Recent advances in microbial ecology. Proceedings of the 5th International Symposium on Microbial Ecology. Scientific Societies Press, Tokyo, Japan.
18. Henderson, I., T. Davis, M. Elmore, and N. P. Minton. 1997. The genetic basis of toxin production in *Clostridium botulinum* and *Clostridium tetani*, p. 261–294. In J. I. Rood, B. A. McClane, J. G. Songer, and R. W. Titball (ed.), The clostridia: molecular biology of pathogenesis. Academic Press, San Diego, CA.
  19. Hielm, S., J. Bjorkroth, E. Hyttia, and H. Korkeala. 1998. Genomic analysis of *Clostridium botulinum* group II by pulsed-field gel electrophoresis. Appl. Environ. Microbiol. **64**:703–708.
  20. Hill, K. K., T. J. Smith, C. H. Helma, L. O. Ticknor, B. T. Foley, R. T. Svensson, J. L. Brown, E. A. Johnson, L. A. Smith, R. T. Okinaka, P. J. Jackson, and J. D. Marks. 2007. Genetic diversity among botulinum neurotoxin-producing clostridial strains. J. Bacteriol. **189**:818–832.
  21. Hill, K. K., G. Xie, B. T. Foley, T. J. Smith, A. C. Munk, D. Bruce, L. A. Smith, T. S. Brettin, and J. C. Detter. 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. doi:10.1186/1741-7007-7-66.
  22. Jacobson, M. J., G. Lin, B. Raphael, J. Andreadis, and E. A. Johnson. 2008. Analysis of neurotoxin cluster genes in *Clostridium botulinum* strains producing botulinum neurotoxin serotype A subtypes. Appl. Environ. Microbiol. **74**:2778–2786.
  23. Kautter, D. A., T. Lilly, A. J. LeBlanc, and R. K. Lynt. 1974. Incidence of *Clostridium botulinum* in crabmeat from the blue crab. Appl. Microbiol. **28**:722.
  24. Keet, C. A., C. K. Fox, M. Margeta, E. Marco, A. L. Shane, S. J. Dearmond, J. B. Strober, and S. P. Miller. 2005. Infant botulism, type F, presenting at 54 hours of life. Pediatr. Neurol. **32**:193–196.
  25. Lacy, D. B., and R. C. Stevens. 1999. Sequence homology and structural analysis of the clostridial neurotoxins. J. Mol. Biol. **291**:1091–1104.
  26. Lúquez, C., M. I. Bianco, L. I. de Jong, M. D. Sagua, G. N. Arenas, A. S. Ciccarelli, and R. A. Fernández. 2005. Distribution of botulinum toxin-producing clostridia in soils of Argentina. Appl. Environ. Microbiol. **71**:4137–4139.
  27. Lúquez, C., B. H. Raphael, and S. E. Maslanka. 2009. Neurotoxin gene clusters in *Clostridium botulinum* type Ab strains. Appl. Environ. Microbiol. **75**:6094–6101.
  28. McCroskey, L. M., C. L. Hatheway, B. A. Woodruff, J. A. Greenberg, and P. Jurgenson. 1991. Type F botulism due to neurotoxicogenic *Clostridium baratii* from an unknown source in an adult. J. Clin. Microbiol. **29**:2618–2620.
  29. Midura, T. F., G. S. Nygaard, R. M. Wood, and H. L. Bodily. 1972. *Clostridium botulinum* type F: isolation from venison jerky. Appl. Microbiol. **24**:165–167.
  30. Moller, V., and I. Scheibel. 1960. Preliminary report on the isolation of an apparently new type of *Clostridium botulinum*. Acta Pathol. Microbiol. Scand. **48**:80.
  31. Nevas, M., M. Lindstrom, S. Hielm, K. J. Bjorkroth, M. W. Peck, and H. Korkeala. 2005. Diversity of proteolytic *Clostridium botulinum* strains, determined by a pulsed-field gel electrophoresis approach. Appl. Environ. Microbiol. **71**:1311–1317.
  32. Reference deleted.
  33. Raphael, B. H., C. Luquez, L. M. McCroskey, L. A. Joseph, M. J. Jacobson, E. A. Johnson, S. E. Maslanka, and J. D. Andreadis. 2008. Genetic homogeneity of *Clostridium botulinum* type A1 strains with unique toxin gene clusters. Appl. Environ. Microbiol. **74**:4390–4397.
  34. Santos-Buelga, J. A., M. D. Collins, and A. K. East. 1998. Characterization of the genes encoding the botulinum neurotoxin complex in a strain of *Clostridium botulinum* producing type B and F neurotoxins. Curr. Microbiol. **37**:312–318.
  35. Smith, G. E., F. Hinde, D. Westmoreland, P. R. Berry, and R. J. Gilbert. 1989. Infantile botulism. Arch. Dis. Child. **64**:871–872.
  36. Smith, T. J., K. K. Hill, B. T. Foley, J. C. Detter, A. C. Munk, D. C. Bruce, N. A. Doggett, L. A. Smith, J. D. Marks, G. Xie, and T. S. Brettin. 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. doi:10.1371/journal.pone.0001271.
  37. Smith, T. J., J. Lou, I. N. Geren, C. M. Forsyth, R. Tsai, S. L. Laporte, W. H. Tepp, M. Bradshaw, E. A. Johnson, L. A. Smith, and J. D. Marks. 2005. Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization. Infect. Immun. **73**:5450–5457.
  38. Solomon, H. M., R. K. Lynt, D. A. Kautter, and T. Lilly. 1971. Antigenic relationships among the proteolytic and nonproteolytic strains of *Clostridium botulinum*. Appl. Microbiol. **21**:295–299.
  39. Williams-Walls, N. J. 1968. *Clostridium botulinum* type F: isolation from crabs. Science **162**:375–376.