JOURNAL OF BACTERIOLOGY, Feb. 2007, p. 818–832 Vol. 189, No. 3
0021-9193/07/$08.00+0 doi:10.1128/JB.01180-06

Genetic Diversity among Botulinum Neurotoxin-Producing Clostridial Strains▽

K. K. Hill,[1] T. J. Smith,[2] C. H. Helma,[1] L. O. Ticknor,[3] B. T. Foley,[4] R. T. Svensson,[1] J. L. Brown,[2] E. A. Johnson,[5] L. A. Smith,[2] R. T. Okinaka,[1] P. J. Jackson,[6] and J. D. Marks[7]*

Bioscience,[1] Computing, Computational, and Statistical Sciences,[3] and Theoretical[4] Divisions, Los Alamos National Laboratory, Los Alamos, New Mexico 87545; Integrated Toxicology Division, United States Army Medical Institute of Infectious Diseases, Fort Detrick, Maryland 21702[2]; Department of Food Microbiology and Toxicology, Food Research Institute, University of Wisconsin, Madison, Wisconsin 53706[5]; Defense Biology Division, Lawrence Livermore National Laboratory, Livermore, California 94551[6]; and Department of Anesthesia and Pharmaceutical Chemistry, University of California, San Francisco, Rm. 3C-38, San Francisco General Hospital, 1001 Potrero Ave., San Francisco, California 94110[7]

Received 31 July 2006/Accepted 21 October 2006

Clostridium botulinum **is a taxonomic designation for many diverse anaerobic spore-forming rod-shaped bacteria that have the common property of producing botulinum neurotoxins (BoNTs). The BoNTs are exoneurotoxins that can cause severe paralysis and death in humans and other animal species. A collection of 174 *C. botulinum* strains was examined by amplified fragment length polymorphism (AFLP) analysis and by sequencing of the 16S rRNA gene and BoNT genes to examine the genetic diversity within this species. This collection contained representatives of each of the seven different serotypes of botulinum neurotoxins (BoNT/A to BoNT/G). Analysis of the16S rRNA gene sequences confirmed previous identifications of at least four distinct genomic backgrounds (groups I to IV), each of which has independently acquired one or more BoNT genes through horizontal gene transfer. AFLP analysis provided higher resolution and could be used to further subdivide the four groups into subgroups. Sequencing of the BoNT genes from multiple strains of serotypes A, B, and E confirmed significant sequence variation within each serotype. Four distinct lineages within each of the BoNT A and B serotypes and five distinct lineages of serotype E strains were identified. The nucleotide sequences of the seven toxin genes of the serotypes were compared and showed various degrees of interrelatedness and recombination, as was previously noted for the nontoxic nonhemagglutinin gene, which is linked to the BoNT gene. These analyses contribute to the understanding of the evolution and phylogeny within this species and assist in the development of improved diagnostics and therapeutics for the treatment of botulism.**

Clostridium botulinum is a taxonomic collection of several distinct species of anaerobic gram-positive spore-forming bacteria that produce the most poisonous substance known, botulinum neurotoxin (BoNT) (1, 8). These organisms, along with related neurotoxin-producing species that, for a variety of reasons, were not included under the *C. botulinum* taxon, pose global health problems that affect both infant and adult humans and can also affect wildlife, waterfowl, and domestic animals. They cause intoxication through ingestion of the neurotoxin in contaminated foods. Toxicoinfections can also occur after contact with bacteria or bacterial spores (6, 17). These pathogens are ubiquitous and can be found in soils and sediments in freshwater and marine environments (47).

BoNTs are classified by the Centers for Disease Control and Prevention (CDC) as one of the six highest-risk threat agents for bioterrorism (the "category A agents") due to their extreme potency and lethality, the ease of production and transport, and the need for prolonged hospital intensive care for those exposed (1). Multiple countries have produced BoNT for use as weapons (5, 45), and the Japanese cult Aum Shinrikyo attempted to use BoNT for bioterrorism (1). Since the terrorist events of 11 September 2001 and the subsequent intentional release of anthrax spores, the development of environmental toxin sensors, diagnostic tests for botulism, and specific countermeasures for the prevention and treatment of intoxication have become a high priority. The first step in such research is to define the spectrum of diversity of BoNT-producing clostridial species and the toxins that they produce.

C. botulinum strains are usually described as belonging to one of four different groups (groups I, II, III, and IV) based on physiologic characteristics (18, 38). The toxins produced are categorized into seven serologically distinct groups (serotypes A through G), based on recognition by polyclonal serum (17). Each BoNT is encoded by an approximately 3.8-kb gene, which is preceded by a nontoxic nonhemagglutinin gene and several other genes that encode toxin-associated proteins (HA-17, HA-33, HA-70, p21, and/or p47) (3, 8, 11, 12, 34). The BoNT gene for strains of serotypes A, B, E, and F can be found within the bacterial chromosome. Serotype C and D strains produce toxin from a phage genome, and serotype G strains contain a plasmid containing the toxin operon (34). Strains producing interserotype recombinant toxins, primarily the C/D and D/C phage-encoded serotypes, have been reported (31, 32). Several strains produce multiple toxins. Bivalent *C. botulinum* strains, each producing two toxins of serotypes Ab, Ba, Af, and Bf, have been reported (4, 15, 37).

The genomic background containing these BoNT genes

* Corresponding author. Mailing address: Department of Anesthesia and Pharmaceutical Chemistry, University of California, San Francisco, Rm. 3C-38, San Francisco General Hospital, 1001 Potrero Ave., San Francisco, CA 94110. Phone: (415) 206-3256. Fax: (415) 206-3253. E-mail: marksj@anesthesia.ucsf.edu.

▽ Published ahead of print on 17 November 2006.

within *C. botulinum* has been characterized as being very diverse. Moreover, other species are known to harbor BoNT genes, such as *Clostridium butyricum* (BoNT/E) (2, 30), *Clostridium baratii* (BoNT/F) (16), and *Clostridium argentinense* (BoNT/G) (43). Previous 16S rRNA gene analysis of many different *Clostridium* species has shown that *C. botulinum* strains form four distinct clusters, with each cluster representing one of the four different physiological groups (groups I to IV) (8, 22). Previous amplified fragment length polymorphism (AFLP) analysis of 70 *C. botulinum* BoNT/A, B, E, and F strains showed that this technique could also successfully differentiate strains into the distinct group I and group II clusters (25). Like the 16S rRNA gene analysis, the AFLP results show that the group I cluster included BoNT/A, B, and F proteolytic strains, while group II contained BoNT/E and nonproteolytic B and F strains (25). Thus, the phylogeny of these species based on molecular analyses has supported the current taxonomy, which has been based on the physiologic attributes of the species and the toxins produced. Such analyses have contributed to the understanding of the diversity of the genomic backgrounds that contain the very different BoNT genes.

Recently, it has become evident that there is significant sequence diversity (subtypes) within the BoNT genes and toxins of at least six of the seven serotypes (39). The relationship between toxin gene diversity and clostridial genomic diversity is unknown. Such subtypes can differ by 2.6% to 31.6% at the amino acid level, and these differences can affect the binding and neutralization by monoclonal and polyclonal antibodies (13, 28, 39). Since an analysis of only 48 published full-length toxin gene sequences revealed the presence of 18 different subtypes, it is likely that additional subtypes might exist (39). Defining the extent of such toxin diversity is a first step in the development of detection systems and countermeasures for the prevention and treatment of botulism (33, 39). In addition, analysis of a large population of strains can be used to better understand the evolutionary relationship between the toxin moieties and the genomic backgrounds that contain these toxins.

To better understand the extent of toxin gene diversity and the relationship between genomic diversity among *C. botulinum* serotypes and subtypes and other toxin-producing species of *Clostridium*, 174 toxin-producing strains from a collection that included representatives of all neurotoxin serotypes (BoNT/A to BoNT/G) were analyzed. Several methods were used to examine the strains, including sequencing of the 16S rRNA gene, analysis of the genome by AFLP, and sequencing of BoNT/A, B, and E neurotoxin genes. Nucleotide sequences of the 16S rRNA and BoNT genes from these and other previously sequenced *Clostridium* strains were analyzed by phylogenetic and recombination detection methods. The phylogenetic relationships among these strains based on all of these methods as well as the extent of toxin gene diversity and the relationship between toxin types, subtypes, and genomic differences are presented.

MATERIALS AND METHODS

Strains. The 174 strains examined in this study are listed in Table 1 and included 59 BoNT/A, 56 BoNT/B, 19 BoNT/C, 6 BoNT/D, 21 BoNT/E, 6 BoNT/F, 7 BoNT/G, and 5 bivalent strains (Af695, Bf698, Bf258, Ba207, and Ab149). For the purposes of the above-described classification, bivalent strains were classified based on the predominant toxin produced. Strains of BoNT-producing clostridia were obtained from USAMRIID, Frederick, MD, and the Department of Food Microbiology and Toxicology, University of Wisconsin, Madison, WI. Many of these strains were part of the Virginia Polytechnic Institute Anaerobe Laboratory collection. Strains were serotyped using an antibody capture enzyme-linked immunosorbent assay with serotype-specific monoclonal antibodies. In some cases, serotypes were confirmed using mouse neutralization (19). Silent (not expressed) BoNT/B genes were detected using real-time PCR (7).

DNA isolation and purification. Individual bacterial colonies of each of the *C. botulinum* strains were removed from anaerobic CDC blood agar plates and used to inoculate 100 ml TPGY broth (Difco, Becton Dickinson and Co., Franklin Lakes, NJ). The broth cultures were incubated anaerobically for 48 h at 35°C and then harvested by low-speed centrifugation. The pellets were resuspended in 8.5 ml TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and then quickly frozen in a dry ice-ethanol bath and stored at −70°C until further processing. Upon removal from the freezer, the resuspended pellets underwent three successive cycles of freezing in a dry ice-ethanol bath followed by melting at 65°C. Sodium dodecyl sulfate (450 μl of a 10% [wt/vol] solution) and 45 μl of proteinase K (10 mg/ml) were added, mixed, and incubated at 42°C for 1 h. After incubation, 1.5 ml of 5 M NaCl solution and 1.4 ml of a 10% (wt/vol) cetyltrimethylammonium bromide solution were added, mixed thoroughly, and incubated at 65°C for 10 min. Following this incubation, three organic extractions of the mixture were performed. The initial extraction involved the addition of an equal volume of chloroform-isoamyl alcohol (24:1) with incubation at room temperature while gently rocking for 10 min. Following low-speed centrifugation, the aqueous phase was removed and extracted again by adding an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1). Incubation was done for 10 min with gentle rocking as described above. For the final extraction, an equal volume of chloroform-isoamyl alcohol (24:1) was added. After low-speed centrifugation, the nucleic acids in the upper phase were precipitated with isopropanol. After centrifugation, the pellet was washed with a 70% ethanol solution and then resuspended in 2.0 ml TE buffer. The DNA preparation was quantified with a spectrophotometer, diluted to 25 μg/ml, and analyzed on an agarose gel to determine quality.

AFLP analysis of DNA samples. The selection of the two restriction endonucleases to digest the genomic DNAs and the single nucleotide added for subsequent selective amplification of the resulting fragments was based on the low G+C content (28.2%) of the *C. botulinum* genome (http://www.sanger.ac.uk/Projects/C_botulinum/). EcoRI and MseI, with recognition sites of GAATTC and TTAA, respectively, were used to digest 100 ng of DNA from each sample. The resulting fragments were ligated into double-stranded adapters. The digested and ligated DNA was then amplified by PCR using EcoRI and MseI +0/+0 primers (5′-GTAGACTGCGTACCAATTC-3′ and 5′-GACGATGAGTCCTGAGTAA-3′, respectively). Five microliters of each product was used as a template in subsequent selective amplifications using the +1/+1 primer combination of 6-carboxyfluorescein-labeled EcoRI-T (5′-GTAGACTGCGTACCAATTCT-3′) and MseI-T (5′-GACGATGAGTCCTGAGTAAT-3′) (underlining shows difference from the +0/+0 primers). Selective amplifications were performed in 20-μl reaction mixtures. The resulting products (0.5 to 1.0 μl) were mixed with a solution containing DNA size standards, Genescan-500 (Applied Biosystems Inc., Foster City, CA) labeled with *N,N,N,N*-tetramethyl-6-carboxyrhodamine. Following a 5-min heat denaturation step at 95°C, the products were loaded onto an ABI 3100 automated fluorescent sequencer. Each set of AFLP reaction mixtures also contained a control DNA as a template. Inclusion of such a reaction mixture in each run-and-analysis set allowed a comparison of results from previously archived analysis sets that were run at different times. Genescan analysis software (Applied Biosystems, Inc., Foster City, CA) was used to determine the lengths of the sample fragments by comparison to the DNA fragment length size standards included with each sample. To minimize capillary gel electrophoresis artifacts, each labeling reaction product was run in triplicate. Samples were loaded into a 96-well plate in a random order.

AFLP data analysis was performed as described previously by Ticknor et al. (45). Sample fragments between 100 and 500 bp and with fluorescence above 50 arbitrary units in all three runs on the ABI sequencer were used in the analysis. Similarities among samples were determined using three separate methods to allow comparisons between methods. First, the Jaccard coefficient, which compares the presence and absence of fragments of a given length, was used. Second, Euclidean distance with the relative abundance values was used, so that both presence and abundance are compared. Third, a Manhattan distance was used, which is similar to Euclidean distances except that the absolute value instead of the squared value is reported. The 40 tallest peaks for each sample fingerprint

TABLE 1. *C. botulinum* strains analyzed

Strain	Serotype	Description	Strain	Serotype	Description
A142	A	Schantz	B492	B	VPI 3,801
A143	A	ATCC 3502 (Hall 174)	B493	B	CA SHD
A144	A	ATCC 17862	B494	B	Smith K. Llbke
A146	A	ATCC 25763	B495	B	Smith L-590
A147	A	CDC 1757	B496	B	ATCC 8083
A148	A	CDC 1744	B497	B	Hall 80
Ab149	Ab	CDC 1436	B498	B	Hall 178
A150	A	Hall	B499	B	Hall 6517(B)
A254	A	Loch Maree	B500	B	Hall 6560
A256	A	Hall 5675	B501L	B	Hall 6707
A312	A	Prevot Ppois	B502	B	Hall 10,007
A384	A	CDC 297	B506	B	CDC 795
A385	A	Prevot P146	B507	B	CDC 8188
A386	A	VPI 7124	B508	B	CDC 6242
A387	A	ATCC 4894	B509	B	CDC 6291
A388	A	CDC 4997	B512	B	Prevot 1687
A389	A	ATCC 449	B513	B	Prevot 1662
A391	A	Hall 183	B514	B	Prevot 1490
A393	A	Hall 3676	B515	B	Prevot 1542
A394	A	Hall 3685a	B516	B	Prevot 1552
A395	A	Hall 4934Aa	B517	B	Prevot 2345
A396	A	Hall 4834	B518	B	Prevot 1837
A397	A	Hall 8388A	B519	B	Prevot B“B”
A398	A	Hall 8857Ab	B520	B	Prevot 1962“B”
A401	A	Hall 11481	B521	B	Prevot PP
A402	A	Hall 11569	B696	B	Eklund 2B
A403	A	Hall 17544	B697	B	10068
A404	A	Hall 6581Ae	Bf698	Bf	CDC 3281
A405	A	McClung 447	C167	C	Stockholm
A406	A	McClung 452	C169	C	2048-Mich
A407	A	CDC 2084	C173	C	ATCC 17849 (nontoxic variant)
A408	A	CDC 7243	C174	C	ATCC 17784
A410	A	CDC 8701	C209	C	003-9
A411	A	CDC 2357	C210	C	468
A412	A	McClung 465	C522	C	Prevot 526
A413	A	Prevot 792	C523	C	Copenhagen 41/59-60
A414	A	Prevot 910	C525	C	6812
A415	A	Prevot 969	C526	C	Smith 6813
A416	A	Prevot 62NCA	C527	C	Smith 6814
A417	A	Prevot P179	C528	C	6816
A418	A	Prevot 878	C529	C	9,846C
A419	A	Prevot 62	C530	C	Prevot 571Y
A420	A	Prevot 865	C531	C	Prevot 2233
A421	A	Prevot F18	C532	C	Prevot 2266
A422	A	Prevot F16	C659	C	Copenhagen 41/59-60
A423	A	Prevot F57	C699	C	Brazil
A424	A	Prevot Dewping	C700	C	South Africa
A425	A	Prevot F60	D175	D	1873
A427	A	Prevot 697B	D177	D	Schantz
A428	A	Prevot F5G	D211	D	ATCC 11873
A429	A	Prevot 892	D534	D	ATCC 2751
A487	A	ATCC 17916	D535	D	M’Bour
A503	A	McClung 844	D701	D	CB-16 (nontoxic variant)
A504	A	McClung 450	E182	E	ATCC 17852
A505	A	McClung 457	E183	E	ATCC 17854
A674	A	ATCC 7948	E184	E	ATCC 17855
A693	A	FRI honey	E185	E	Alaska E43
A694	A	Kyoto-F	E213	E	Beluga (ATCC 43181)
Af695	Af	Strain 84	E216	E	EF4
B152	B	NCTC 7273	E536	E	CDC KA-95B
B155	B	Okra	E537	E	Tenno
B159	B	ATCC 17843	E538	E	Beluga (ATCC 43181)
B160	B	ATCC 17844	E539	E	Hobbs FT18
B161	B	ATCC 17845	E540	E	FDA066B
B162	B	213B (ATCC 7949)	E541	E	L-572
B163	B	CDC 1656	E542	E	Beluga (ATCC 43181)
B164	B	CDC 1828	E543	E	BL5262 (*C. butyricum*)
B165	B	CDC 1758	E544	E	CDC 5247
B170	B	ATCC 17783	E545	E	CDC 5258
B192	B	Contaminant in CDC 714	E546	E	CDC5906
Ba207	Ba	657	E547	E	Prevot Ped 1
B257	B	Eklund 17B	E548	E	Prevot Ped 4
Bf258	Bf	An436	E549	E	Prevot R81-3A
B259	B	ATCC 51386	E675	E	Hazen 36208E (ATCC 9564)
B260	B	ATCC 51387	F187	F	CDC 2821
B305	B	Prevot 59	F188	F	Langeland
B306	B	Prevot 25 NCASE	F189	F	6/14
B307	B	Prevot 1740	F550	F	Eklund202F
B308	B	Prevot 1504	F552	F	Wall strain 8-G (ATCC 25764)
B309	B	Prevot CM	F658	F	Langeland
B310	B	Prevot 594	G190	G	5/18/78
B311	B	Prevot B	G193	G	2738
B313	B	Prevot F11	G194	G	1353
B426	B	Prevot P64	G195	G	2739
B488	B	VPI 558	G196	G	2740
B489	B	VPI 560	G197	G	2741
B491	B	Prevot 1884BA	G198	G	2742

were used to calculate the distance coefficients among samples. Dendrograms were produced using each of the three similarity matrices using the unweighted-pair group average agglomerative hierarchical clustering method (24). All statistical data manipulations were done using codes developed using S-Plus (Data Analysis Products Division, MathSoft, Seattle, WA). The dendrograms using the Euclidian and Manhattan distances, which include relative fragment abundance values, were compared to the Jaccard distance dendrogram, and there were no differences in the groupings. This shows that these groupings are robust and are not artifacts of the data analysis methods. The dendrogram using the Jaccard distances is presented. Replicates have Jaccard distance measures at the 0.20 level or below. No differences below the 0.20 level on the Jaccard dendrogram are presented since it cannot be determined if the differences are due to variability in the assay or actual sample differences.

16S rRNA gene sequencing of *C. botulinum* samples. Representatives of the different BoNT-producing *Clostridium* strains were selected for 16S rRNA gene sequencing. Primers 1492R (5′-GGTTACCTTGTTACGACTT-3′) and 27F (5′-AGAGTTTGATCMTGGCTCAG-3′) were used to PCR amplify approximately 1,400 bases of this 1.5-kb gene. The purified PCR template was then sequenced using these primers and internal primers 533Fb (5′-GCCA GCAGCNGCGGTAA-3′), 940Fb (5′-CGGGGGYCCGCACAAGC-3′), and 910Rb (5′-GCCCCCGTCAATTYHTTTGAG-3′). The 16S rRNA gene phylogenetic dendrogram was created from an alignment of 16S rRNA gene sequences, some new to this study and others obtained from GenBank entries of previously sequenced genes. It should be noted that *Clostridium* genomes each contain more than one copy or allele of the 16S rRNA gene. After multiple sequence alignment with MUSCLE (http://www.drive5.com/muscle/), columns in the alignment in which more than 80% of the sequences were represented by a gap character were removed, leaving an alignment of 1,329 bases for phylogenetic analysis. The phylogenetic dendrogram was calculated using PHYLIP dnadist and neighbor programs with the F84 model of evolution and four sequences from the genus *Alkaliphilus* (GenBank accession numbers AY554415, AB037677, AF467248, and AJ630291) to serve as the outgroup to the *Clostridium* genus sequences. The resulting tree was rendered with TreeTool (http://packages.debian.org/unstable/science/TreeTool/), and the outgroup was removed to produce the final figure.

BoNT gene PCR amplification and sequencing. Overlapping primer pairs covering the coding sequence of the different BoNT genes were designed for PCR amplification using available GenBank sequences. Internal DNA oligomers were also designed within each amplicon to provide confirming sequence data in both directions. These PCR amplification and sequencing primers for each of the neurotoxin gene fragments are listed in Table 2. The initial PCR mixture contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (wt/vol) gelatin, 0.2 mM each deoxynucleotide triphosphate, 20 pmol of each primer, 2.5 U of Amplitaq DNA polymerase (Perkin-Elmer, Inc., Boston, MA), and approximately 1 ng template DNA in a 100-μl total reaction volume. Template DNA was initially denatured by heating at 94°C for 2 min. This was followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and primer extension at 72°C for 1 min. Incubation for 5 min at 72°C followed to complete the extension. PCR amplicons were analyzed by electrophoresis through a 3.0% agarose gel dissolved in a solution containing 10 mM Tris-borate (pH 8.3) and 1 mM EDTA for 1 h at 80 V. Gels were stained for 20 min with a solution containing 1 μg of ethidium bromide/ml, destained in distilled water, and then visualized and photographed under UV light. PCR amplicons were purified using a QIAGEN PCR purification kit (QIAGEN Inc., Valencia, CA) and then sequenced using ABI Dye Terminator 3.1 chemistry with an ABI 3730 instrument.

DNA alignments were created with a combination of Sequencer software (http://www.genecodes.com/), PAUP (http://paup.csit.fsu.edu/), MUSCLE (http://www.drive5.com/muscle/), and CLUSTAL-W (http://www.ebi.ac.uk/clustalw/) and hand editing with BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) software and were gap stripped and then analyzed using PHYLIP (http://evolution.genetics.washington.edu/phylip.html) with dnadist with the F84 model of evolution and a transition-to-transversion ratio of 2.0 (default) and neighbor-joining algorithms. Phylogenetic dendrograms were rendered with TreeTool (http://packages.debian.org/unstable/science/TreeTool/). Intra- and interserotype BoNT gene recombination was explored with SimPlot (http://sray.med.som.jhmi.edu/SCRoftware/simplot/) and BioEdit. The SimPlot analysis shown in Fig. 4 is a comparison of different BoNT sequences and the BoNT/A2 sequence reported under GenBank accession number X73423. It was generated with a sliding window of 200 bp, and the percent similarity between the two sequences was plotted at the center of the window. The window was moved 20 bp between each point.

RESULTS

The *Clostridium* strains used in this study encompass strains collected by different investigators over many years. Table 1 lists the strains used in this study and includes the strain identification numbers from the original researchers or institutions if known. This collection of 174 strains included 59 BoNT/A, 56 BoNT/B, 19 BoNT/C, 6 BoNT/D, 21 BoNT/E, 6 BoNT/F, and 7 BoNT/G strains. Five bivalent strains (Af695, Bf698, Bf258, Ba207, and Ab149) were also included. All strains were tested to confirm that they produced toxin by an enzyme-linked immunosorbent assay and/or by a mouse neutralization assay. The collection contains strains from diverse geographic locations and from various sources, including specimens from food, adults, infants, animals, birds, soil, and marine sediments.

16S rRNA gene analysis. Comparative analysis of the nucleotide sequences of the conserved 16S rRNA gene using a subset of 109 strains representing the different serotypes in this collection and those from other *Clostridium* species illustrates that the genetic distances between the toxin-producing groups are typical of the distances between other species within this genus. Figure 1 illustrates that the toxin-producing clostridia are comprised of the four previously characterized distinct phylogenetic clusters, which would logically be defined as discrete species within the *Clostridium* genus. These results confirm and extend previous work of many others (8, 21, 22, 42). The majority of the strains in this collection are tightly clustered with 16S profiles that are identical to or nearly identical to many sequences that have previously been designated as belonging to the proteolytic group I *C. botulinum* strains that contain all of the A and most of the B and F neurotoxin genes (22). The remaining strains form phylogenetically distant species clusters that define the remaining physiological groups, groups II, III, and IV (21). The group II strains include all of the neurotoxin E strains, nonproteolytic B strains, and nonproteolytic F strains and are most closely related to *Clostridium beijerinckii* and *C. butyricum*. Group III (closely related to *Clostridium novyi*) strains encode BoNT/C and D and C/D recombinant and D/C recombinant serotypes encoded by toxin operons on a bacteriophage. The 16S sequences of group IV strains are nearly identical to those of *Clostridium subterminale* and *C. argentinense* and belong to BoNT serotype G, which is encoded by a plasmid.

AFLP analysis. The dendrogram of the *Clostridium* DNAs generated by AFLP analysis is shown in Fig. 2. The large branching patterns within this tree further resolve the tree obtained with the 16S rRNA gene. The AFLP results are clearly consistent with all other methods for classifying *C. botulinum* strains, including both genetic sequence analyses and the characterization of physiological differences used in classical bacteriology. These physiological differences have historically been used to categorize *C. botulinum* strains into groups I to IV. The AFLP dendrogram shows a large separation between the proteolytic (group I) and nonproteolytic (groups II, III, and IV) strains and forms distinct branches representing groups I to IV separated by distances greater than 0.75. In general, the AFLP dendrogram also groups the strains by toxin serotype; i.e., most of those strains producing either BoNT/C or BoNT/D, BoNT/E, proteolytic BoNT/F, and BoNT/G cluster by toxin serotype within distinct branches. However, the

TABLE 2. Primers used for PCR amplification and sequencing of BoNT/A, B, and E genes

Primer	Type[a]	Sequence	Location[b] (positions)
BoNT A-1F	Amp/Seq	TTTATGGTCATTTAAATAATTAATA	35–59
BoNT A-1R	Amp/Seq	AATGTTCTAAGTTCCTCAAAG	873–893
BoNT A-1Fs	Seq	GGTGGAAGTACAATAGATACAG	451–472
BoNT A-1Rs	Seq	TGTATCTATTGTACTTCCACCC	450–471
BoNT A-2F	Amp/Seq	AGATCCAGCAGTAACATTAGC	741–761
BoNT A-2R	Amp/Seq	TCCCAATTATTAACTTTGATACATA	1454–1478
BoNT A-2Fs	Seq	GAGATTTACACAGAGGATAATT	1135–1156
BoNT A-2Rs	Seq	AATTATCCTCTGTGTAAATCTC	1135–1156
BoNT A-3F	Amp/Seq	TGCTATGTGTAAGAGGGATAATA	1379–1401
BoNT A-3R	Amp/Seq	ATCCCATTTTTCATTTCTTTTACTT	2193–2217
BoNT A-3Fs	Seq	ATACTATGTTCCATTATCTTCG	1739–1760
BoNT A-3Rs	Seq	CGAAGATAATGGAACATAGTAT	1739–1760
BoNT A-4F	Amp/Seq	GCTTTAAGTAAAAGAAATGA	2188–2207
BoNT A-4R	Amp/Seq	CCAGATTATTTCACCATAAT	3032–3051
BoNT A-4Fs	Seq	ATCAATGCTCTGTTTCATATT	2462–2482
BoNT A-4Rs	Seq	AATATGAAACAGAGCATTGAT	2462–2482
BoNT A-5F	Amp/Seq	TGCTATTGTATATAATAGTATG	2886–2907
BoNT A-5R	Amp/Seq	TTGACTTCATTACTACTACTT	3752–3772
BoNT A-5Fs	Seq	TTAGGTAATATTCATGCTAGTAA	3226–3248
BoNT A-5Rs	Seq	TTACTAGCATGAATATTACCTAA	3226–3248
BoNT A-6F	Amp/Seq	ATATTGTTAGAAATAATGATCG	3611–3632
BoNT A-6R	Amp/Seq	TAGTTTGAGATTAATTACAGTG	3980–4001
BoNT B-1F	Amp/Seq	CAATATACCTAAAGCTGCACA	26–46
BoNT B-1R	Amp/Seq	TACTTTAATGCCATATAATCCA	807–828
BoNT B-1Fs	Seq	CATTGGGTGAAAAGTTATTAGA	404–425
BoNT B-1Rs	Seq	TCTAATAACTTTTCACCCAATG	404–425
BoNT B-2F	Amp/Seq	CAGAATATGTAAGCGTATTTA	686–706
BoNT B-2R	Amp/Seq	ATCAGTAAGTGATTCTGTATTT	1614–1635
BoNT B-2Fs	Seq	TATAGCAGAAAATTATAAAATAAA	1176–1199
BoNT B-2Rs	Seq	TTTATTTTATAATTTTCTGCTATA	1176–1199
BoNT B-3F	Amp/Seq	AGGAGCATTTGGCTGTATAT	1373–1392
BoNT B-3R	Amp/Seq	AATGCTTGTGCTTGATAATTTA	2264–2285
BoNT B-3Fs	Seq	GGATTATATTAAAACTGCTAAT	1821–1842
BoNT B-3Rs	Seq	ATTAGCAGTTTTAATATAATCC	1821–1842
BoNT B-4F	Amp/Seq	ATATGTACGGATTAATAGTAGC	2180–2201
BoNT B-4R	Amp/Seq	TTACCCCTAATAGATATTTTCC	2981–3002
BoNT B-4Fs	Seq	AGATGAAAATAAATTATATTTAA	2526–2548
BoNT B-4Rs	Seq	TTAAATATAATTTATTTTCATCT	2526–2548
BoNT B-5F	Amp/Seq	TATACAAAATTATATTCATAATGA	2919–2942
BoNT B-5R	Amp/Seq	ATCTTCTTTTCTAACTATATCATC	3568–3591
BoNT B-5Fs	Seq	TATAAAATTCAATCATATAGCG	3319–3340
BoNT B-5Rs	Seq	CGCTATATGATTGAATTTTATA	3319–3340
BoNT B-6F	Amp/Seq	GAGAAAAATTTATTATAAGAAG	3521–3542
BoNT B-6R	Amp/Seq	TAGCTACATCCTTAAACTTAAGAT	4028–4051
BoNT B-6Fs	Seq	TAAAAGAATATGATGAACAGCC	3725–3746
BoNT B-6Rs	Seq	GGCTGTTCATCATATTCTTTTA	3725–3746
BoNT E-1F	Amp/Seq	GTGATCTTAATCATGATATACC	145–166
BoNT E-1R	Amp/Seq	TTAATGTAAGAGCAGGATCT	839–858
BoNT E-1Fs	Seq	TAGTCACAAAAATATTTAATAGAA	484–507
BoNT E-1Rs	Seq	TTCTATTAAATATTTTTGTGACTA	484–507
BoNT E-2F	Amp/Seq	TGGATCAATAGCTATAGTAACA	761–782
BoNT E-2R	Amp/Seq	GGTGCTGATTCACTATTAAA	1650–1669
BoNT E-2Fs	Seq	TTATACAGCTTTACGGAATTTG	1218–1239
BoNT E-2Rs	Seq	CAAATTCCGTAAAGCTGTATAA	1218–1239
BoNT E-3F	Amp/Seq	TGGCTTCCGAGAATAGTT	1537–1554
BoNT E-3R	Amp/Seq	ATTCATTGCTATAGAAACCTTT	2474–2495
BoNT E-3Fs	Seq	GTACTGTTGATAAAATTGCAGA	1999–2020
BoNT E-3Rs	Seq	TCTGCAATTTTATCAACAGTAC	1999–2020
BoNT E-4F	Amp/Seq	GAACAAATGTATCAAGCTTT	2331–2350
BoNT E-4R	Amp/Seq	TGTCACTAACATGAATATTACC	3288–3309
BoNT E-4Fs	Seq	ACTTCAGGATATGATTCAAATA	2820–2841
BoNT E-4Rs	Seq	TATTTGAATCATATCCTGAAGT	2820–2841
BoNT E-5F	Amp/Seq	TGATTATATAAATAAGTGGATT	3179–3200
BoNT E-5R	Amp/Seq	TGGAATTTATGACTTTAGCC	4001–4020
BoNT E-5Fs	Seq	GCTAATAGATTATATAGTGGAA	3579–3600
BoNT E-5Rs	Seq	TTCCACTATATAATCTATTAGC	3579–3600

[a] Amp, amplification; Seq, sequencing.

[b] Location of primer sequence of BoNT/A within the sequence reported under GenBank accession number X73423, location of primer sequence of BoNT/B within the sequence reported under GenBank accession number X71343, and location of primer sequence of BoNT/E within the sequence reported under GenBank accession number X62683.

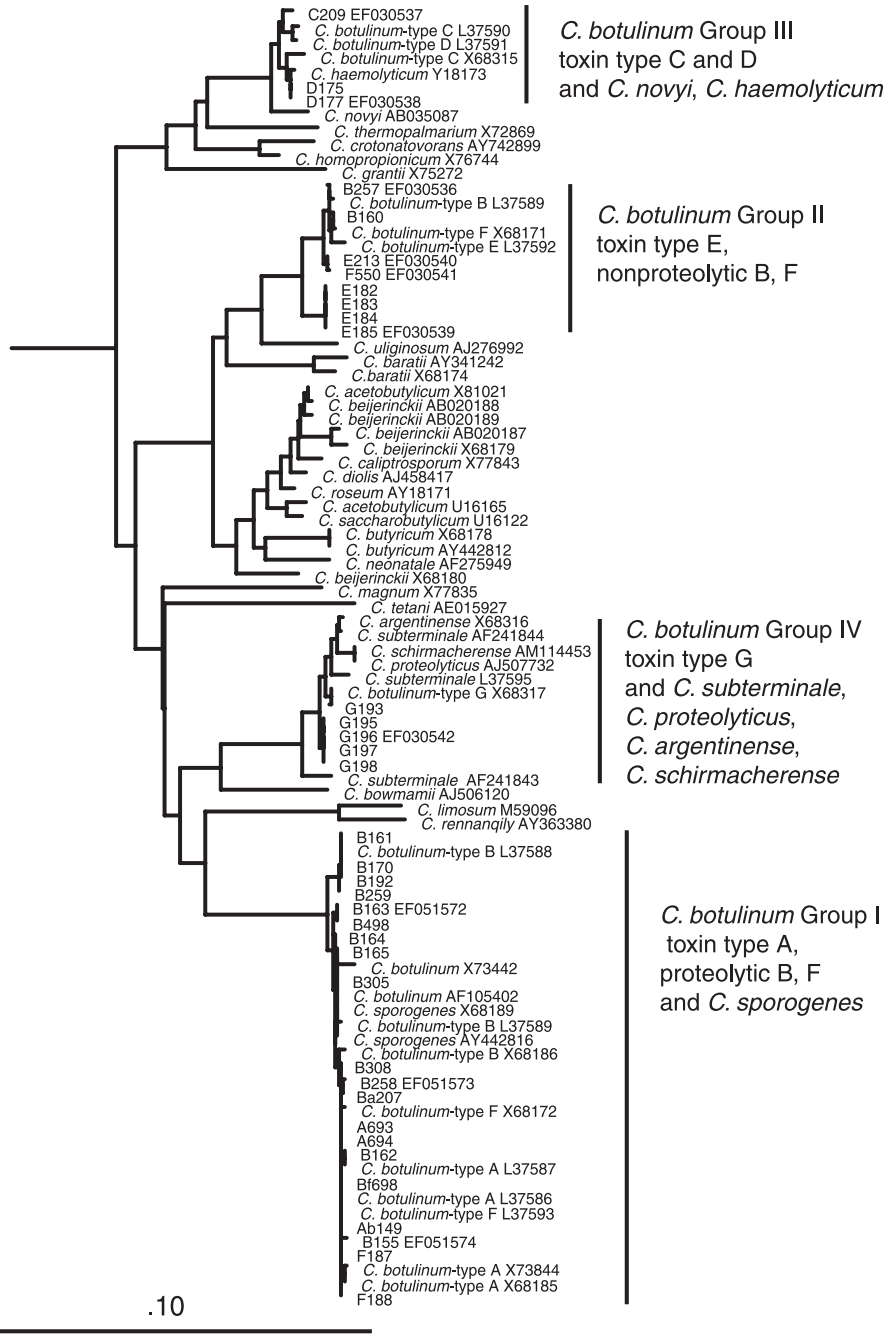
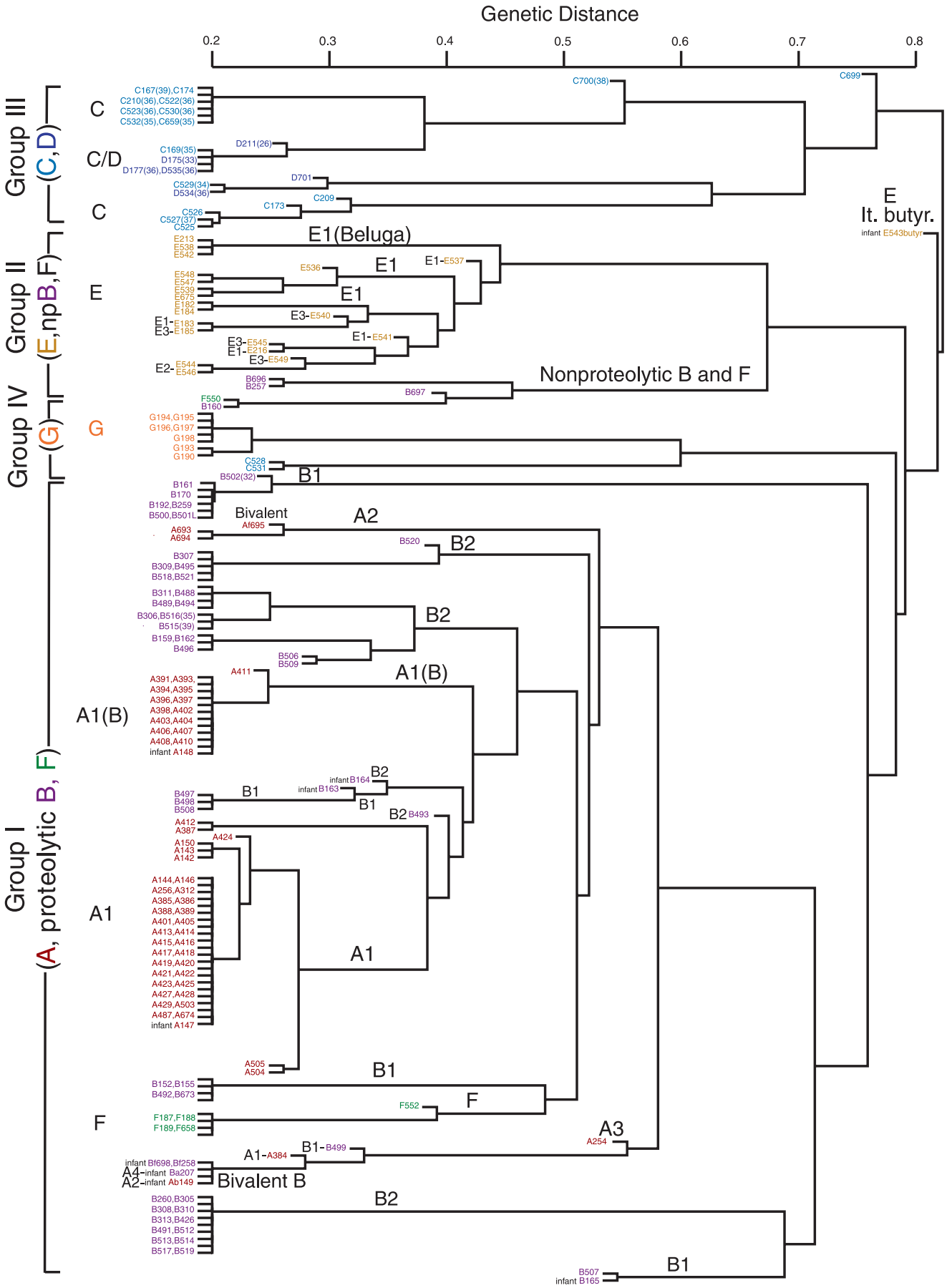


FIG. 1. Phylogenetic dendrogram of *Clostridium* species based on 16S rRNA genes. A neighbor-joining tree of 54 sequences reported in GenBank and 36 sequences representative of the strains from this collection is shown. This illustrates the genetic diversity within the clostridia. *C. botulinum* strains cluster into four distinct groups that follow the group I to group IV designation historically based on physiological characteristics. These groups are interspersed among the 27 other clostridial species in the tree. The tree was constructed using an alignment of 16S rRNA gene sequences that contained 1,329 bases after removal of columns containing more than 80% gap characters and includes sequences from bivalent, nonproteolytic, and proteolytic toxin-producing strains.

strains representing the different BoNT/A and B subtypes show more diversity and are not clearly differentiated using AFLP analysis. Four of the five bivalent strains included in this study (Bf698, Bf258, Ba207, and Ab149) also appear as a cluster, which is most closely related to one of the BoNT/B clusters. Each of these strains produces a unique BoNT/B, termed bivalent BoNT/B (see below) (37).

The group I BoNT/A strains are found within four AFLP clusters that generally contain just one toxin serotype or a single combination of two serotypes. One branch contains the majority of the BoNT/A-producing strains examined (37/59 strains). These strains produce BoNT/A1 and include the ATCC type strain ATCC 25763 (A146) and the BoNT/A Hall 174 strain (A143) sequenced by the Sanger Institute (http:


Genetic Distance
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Group I (A, proteolytic B, F)
Group II (E,npB,F)
Group III (C,D)
Group IV (G)
C167(39),C174
C210(36),C522(36)
C523(36),C530(36)
C532(35),C659(35)
C
C169(35)
D175(33)
D177(36),D535(36)
D211(26)
C/D
C529(34)
D534(36)
D701
C526
C527(37)
C525
C173
C209
C
C700(38)
C699
E
lt. butyr.
infant E543butyr
E213
E538
E542
E1(Beluga)
E548
E547
E539
E675
E182
E184
E536
E1
E1-E537
E1
E3-E540
E1-E183
E3-E185
E1-E541
E3-E545
E1-E216
E3-E549
E2-E544
E546
E
B696
B257
B697
Nonproteolytic B and F
F550
B160
G194,G195
G196,G197
G198
G193
G190
G
C528
C531
B502(32)
B1
B161
B170
B192,B259
B500,B501L
Bivalent
Af695
A2
A693
A694
B520
B2
B307
B309,B495
B518,B521
B311,B488
B489,B494
B306,B516(35)
B515(39)
B159,B162
B496
B2
B506
B509
A411
A1(B)
A391,A393,
A394,A395
A396,A397
A398,A402
A403,A404
A406,A407
A408,A410
infant A148
A1(B)
infant B164
B2
B497
B498
B508
B1
infant B163
B1
B2 B493
A412
A387
A424
A150
A143
A142
A144,A146
A256,A312
A385,A386
A388,A389
A401,A405
A413,A414
A415,A416
A417,A418
A419,A420
A421,A422
A423,A425
A427,A428
A429,A503
A487,A674
infant A147
A1
A1
A505
A504
B1
B152,B155
B492,B673
F552
F
F187,F188
F189,F658
F
A3
A254
A1-A384
B1-B499
infant Bf698,Bf258
A4-infant Ba207
A2-infant Ab149
Bivalent B
B260,B305
B308,B310
B313,B426
B491,B512
B513,B514
B517,B519
B2
B1
B507
infant B165

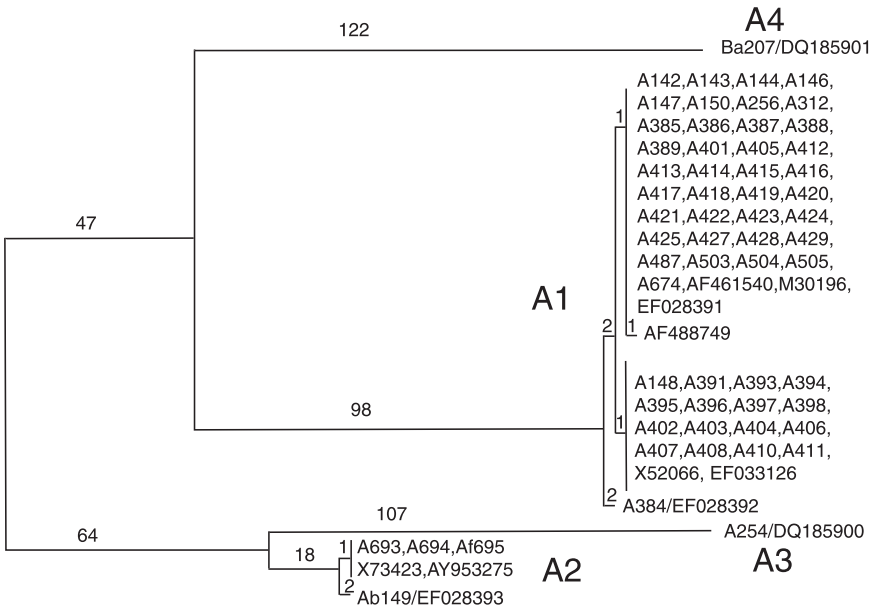


FIG. 3. Comparison of BoNT/A gene sequences. The full-length coding region of the BoNT/A gene in 60 strains and six GenBank sequences were aligned. Four distinct subtypes are apparent. Most strains (54 strains) are of the BoNT/A1 subtype, and four strains are within the BoNT/A2 subtype. Two newly identified subtypes, BoNT/A3 and BoNT/A4, each contain one member: the A254 (Loch Maree) strain and the bivalent Ba207 strain, respectively. These strains show significant sequence variations compared to BoNT/A1 and A2 subtypes.

//www.sanger.ac.uk/projects/C_botulinum/). Of the 37 BoNT/A1 strains, 29 of them form a monophyletic cluster by analysis of both the BoNT gene sequence and AFLP data, which suggests a common clonal derivation of this cluster even though these strains are from various geographic locations and sources. Another cluster within the A1 subtype includes 16 A1(B) strains described as having an A1 subtype with a silent B neurotoxin gene. Six other BoNT/A-producing strains in three different clusters are as closely related to BoNT/B-producing strains as they are to the other BoNT/A-producing strains. These three distantly related clusters include (i) a group of three of the four BoNT/A2 strains (Af695, A693, and A694); (ii) a separate cluster containing the bivalent Ab strain (Ab149), a BoNT/A1-producing strain (A384), and three BoNT/B-producing strains; and (iii) the A254 strain, also known as Loch Maree. This strain, from a 1922 botulism incident in Scotland (29), was found to produce a unique and not previously sequenced BoNT/A, which we have designated BoNT/A3 (see below). In addition, the BoNT/A produced by the bivalent strain Ba207 was also not previously described. We have termed this toxin BoNT/A4 (see below).

The AFLP analysis subdivides the group I proteolytic BoNT/B strains into smaller clusters, which include the serologically distinct BoNT/B1- and BoNT/B2-producing strains (27) and four bivalent BoNT/B-producing strains (Ba207, Ab149, Bf698, and Bf258), all of which produce bivalent BoNT/B toxins. The most common BoNT/B subtype represented here is the BoNT/B2 subtype. The BoNT/B1 strains are more likely to be of U.S. origin and associated with food-borne cases due to improperly processed vegetables, while the BoNT/B2 strains are mostly from Europe and associated with animal cases or meat. The original BoNT/B2 strain was isolated from a case of infant botulism in Japan, and two recently published sequences from BoNT/B strains isolated from Korean soil (GenBank accession numbers DQ417353 and DQ417354) are also from BoNT/B2 strains. The BoNT/A2 (Ab149) and BoNT/A3 (A254) subtype strains and proteolytic BoNT/F strains cluster in separate branches within the BoNT/B strains. The five proteolytic BoNT/F strains cluster together and are distinct from the other BoNT/B strains. These branches reveal genetic similarities of proteolytic BoNT/B strains with both BoNT/A subtypes and proteolytic BoNT/F-producing strains and support the group I designation for all of these strains. The close relationship among the BoNT/B- and BoNT/F-producing strains is also observed in the group II area of the AFLP dendrogram, where three nonproteolytic BoNT/B-producing strains (B160, B257, and B697) cluster and are most closely related to a nonproteolytic BoNT/F-producing strain (F550).

In addition, four bivalent strains of serotypes Ab (Ab149),

FIG. 2. AFLP-based dendrogram of 174 *C. botulinum* strains. DNA fragments generated from restriction endonuclease digestion of each of the strain DNAs were ligated into linkers and selectively amplified. Forty DNA fragments generated by AFLP experiments were used as a fingerprint to represent each of the strains. If 40 fragments did not exist, fewer fragments were used, as noted in parentheses. The comparison of fingerprints from the 174 strains shows a large separation between the proteolytic (group I) and nonproteolytic (groups II, III, and IV) strains and distinct branches representing groups I to IV. The AFLP groups also contain generally distinct toxin serotypes. The distance measure or genetic distance is the proportion of fragments that two samples do not have in common.

TABLE 3. Nucleotide and amino acid identities in strains representing the BoNT/A, B, and E subtypes[a]

Subtype	% Identity (DNA/amino acid) with:			
BoNT/A	A2 (Ab149)	A3 (A254)	A4 (Ba207)	
A1	95/90	92/85	94/89	
A2		97/93	94/88	
A3			92/84	
BoNT/B	B2 (B162)	B3 (B506)	npB[b] (B257)	BvB[c] (Ba207)
B1	98/96	98/96	96/93	98/96
B2		99/98	96/94	97/95
B3			96/94	98/96
npB				96/93
BoNT/E	E2 (E544)	E3 (E185)	E It butyr[d] (E543)	E Ch butyr[e] (LCL155)
E1	99/99	99/98	98/97	98/97
E2		99/97	98/96	98/96
E3			98/96	98/95
E It butyr				97/95

[a] The coding sequence for a representative of each BoNT/A, B, and E subtype was compared to determine overall homology at the nucleic acid and amino acid levels.
[b] npB, nonproteolytic B.
[c] bvB, bivalent B.
[d] It butyr, Italian *C. butyricum* strain BL5262.
[e] Ch butyr, Chinese *C. butyricum* strain LCL155.

Ba (Ba207), and Bf (Bf698 and Bf258) included in this study cluster together at the 0.2 level in this portion of the AFLP dendrogram. The genetic backgrounds of these four strains cannot be distinguished by AFLP analysis, and their 16S rRNA genes were found to be more than 99.93% identical to one another. By comparison, A150 and Bf258, separated by AFLP analysis, contained 16S rRNA gene sequences that were 99.78% identical to each other. These bivalent strains with similar genetic backgrounds each contain combinations of the different toxin genes BoNT/A, B, and F expressed at different levels. This finding appears to indicate very recent horizontal transfer of these toxin genes into the same bacterial lineage. All these strains were isolated from cases of infant botulism in different geographic locations: Sweden, Texas, New Mexico, and Utah.

Group II *C. botulinum* BoNT/E-producing strains, which are usually associated with fish and marine mammals, appear within their own branch of the AFLP dendrogram. The 21 BoNT/E-producing strains include samples from salmon, whale, and soil from the Olympic National Forest. The placement of these group II BoNT/E strains within a distinct branch of the AFLP dendrogram reflects the genetic background of these strains that have evolved to include different hosts and environmental habitats occupied by this serotype. The only *C. butyricum* strain (E543) containing a BoNT/E gene in this study is distant from these other 20 *C. botulinum* type E strains and was isolated from an infant botulism case in Italy (30). A small branch within the BoNT/E-producing strains includes three isolates (E213, E538, and E542) whose differences are below the replicate variability in this AFLP analysis. These three isolates of the "Beluga" strain, which were received from two different research collections (USAMRIID and Virginia Polytechnic Institute), were intentionally included in these experiments. These Beluga isolates are indistinguishable and add confidence to the results obtained using strains collected by different investigators over many years.

The majority of the group III BoNT/C (17/19) and BoNT/D (6/6) serotypes form a distinct branch in the AFLP dendrogram. These group III strains form several clusters containing BoNT/C strains or combinations of BoNT/C and D serotypes that are not distinguishable by this method. One cluster contains eight BoNT/C strains (C167, C174, C210, C522, C523, C530, C532, and C659), seven of which are from Western Europe. These strains are linked to disease in mammals. Another cluster of five strains, shown to be C/D strains, were collected from marine or freshwater sediments. Three of the strains (C525, C526, and C527) are from marine sediments in the United States. Strain C209, which differs slightly from them, is from Japan. Other group III strains are from the United States (C529), Japan (D701), South America (C700), and Africa (C524, C699, and D535). Two of these isolates, C523 and C659, were identical strains that were intentionally included in this study, and the results show that these two strains cluster at the 0.2 level by AFLP analysis. A distant branch contains a BoNT/C strain (C531) and a BoNT/C/D strain (C528), which shows these two strains to be most similar to the BoNT/G serotypes.

The final cluster includes all seven BoNT/G strains in the AFLP dendrogram. This plasmid-encoded toxin gene was first identified in isolates from soil in Argentina (14). Two of the strains in this study are from Argentinean soil (G190 and G194), and the other five strains are from human autopsy specimens in Switzerland (40, 41). These seven samples from different sources show genetic similarity and cluster at the 0.25 level in this portion of the AFLP dendrogram. Four of the five autopsy specimens cluster together with one of the soil isolates (G194). The fifth human specimen (G193) maps closer to the second soil isolate (G190) than to the others.

Results of the AFLP analysis reported here support previous AFLP analyses that showed that this technique could differentiate group I and group II *C. botulinum* strains (25). The current work extends those findings to include strains that are representative of groups III and IV. This analysis illustrates the

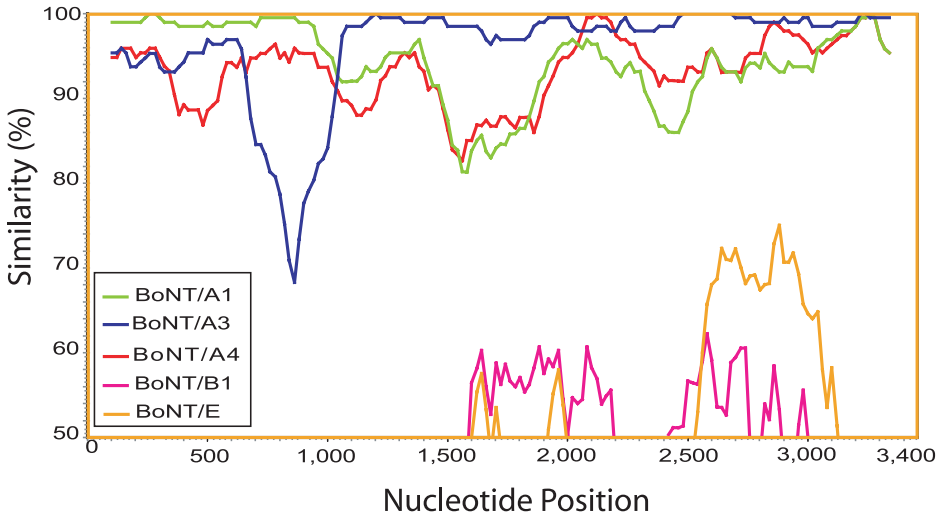


FIG. 4. Similarity plot comparing BoNT subtype sequences to the BoNT/A2 subtype. BoNT sequences of the BoNT/A1, A3, and A4 subtypes and BoNT/B1 and Chinese *C. butyricum* BoNT/E were compared to the BoNT sequence of the BoNT/A2 Kyoto-F subtype (GenBank accession number X73423). This plot illustrates that the BoNT/A2 subtype is approximately 99% identical to the BoNT/A1 subtype (A142) through nucleotides 1 to 1146 and approximately 99% identical to the BoNT/A3 subtype (A254) through nucleotides 1147 to 3450. This suggests that the BoNT/A2 subtype is a result of a recombination event between BoNT/A1 and BoNT/A3 lineages of gene sequences.

relationship of the different genetic backgrounds in the clostridia that contain these neurotoxin genes.

Sequencing and analysis of BoNT genes. To understand how conserved the sequences of the different BoNT genes are within *C. botulinum* strains of a given serotype, the full-length coding sequence of each BoNT/A, B, and E gene was amplified in overlapping segments by PCR and then sequenced. Comparisons of the neurotoxin sequences generated from the 60 BoNT/A genes sequenced here, as well as six previously published BoNT/A gene sequences, show that at least four distinct groups of BoNT/A sequences exist (Fig. 3). Ninety percent of the BoNT/A-producing strains in this study (54/60 strains) show little sequence variation in the BoNT/A gene and are of the previously reported BoNT/A1 subtype (9, 50). Within this subtype, 37 of the strains share identical sequences and differ from 16 of the remaining 17 strains in this subtype by two nucleotides. These 16 strains are A1(B) strains that contain a silent BoNT/B gene. Sequences were generated from six of the silent BoNT/B genes in these A1(B) strains and compared. All six of the silent BoNT/B sequences were similar to those reported under GenBank accession number AF300467 (26), which generate a truncated protein from a stop codon at amino acid 128. Four of the sequences (A148, A397, A404, and A406) were identical to each other but differed from that reported under accession number AF300467 by two single nucleotide polymorphisms. The other two sequences (A408 and A411) were identical to each other but different from the other four silent BoNT/B sequences by a single nucleotide polymorphism. The identification of these different silent BoNT/B gene sequences shows that there are more differences in clostridial strains than revealed by AFLP analysis and 16S rRNA and BoNT/A gene sequence analysis.

Besides the frequently occurring BoNT/A1 gene, three additional BoNT/A genes were identified (Fig. 3). Four strains produced the previously reported BoNT/A2 gene (28, 49). However, two previously unreported BoNT/A genes were also identified. One of these, termed BoNT/A3, was produced by a single strain (A254, also known as Loch Maree), which was isolated from a 1922 botulism outbreak in Scotland (29). An additional BoNT/A gene, BoNT/A4, was sequenced from the bivalent strain Ba207. Nucleotide and amino acid comparisons of the toxin genes of the BoNT/A1 to BoNTA4 subtypes are shown in Table 3. Nucleotide differences range from 3% to 8%, with amino acid differences ranging from 7% to 16% and with the most disparate toxin from BoNT/A1 being BoNT/A3. Given the level of amino acid differences, these two new BoNT/A genes almost certainly represent new BoNT/A subtypes.

Recombination analysis indicates that the A2 lineage, represented in Fig. 4 by isolate BoNT/A2 (Kyoto-F) (GenBank accession number X73423), is a relatively recent recombinant between the BoNT/A1 (A142) and BoNT/A3 (A254) lineages of BoNT/A gene sequences. The BoNT/A2 (strain Kyoto-F) sequence is close to 99% identical to BoNT/A1 strain A142 (this paper) over positions 1 through 1146 and close to 99% identical to BoNT/A3 strain A254 (this paper) over positions 1147 through 3450. The BoNT/A3 lineage (strain A254) has a region of the BoNT gene between bases 745 and 973 that is 74.7% identical to BoNT/A2 (strain Kyoto-F) and also 74.7% identical to BoNT/A1 (strain A142) (Fig. 4). This region, encoding a portion of the toxin light chain, is highly divergent. When this disparate region of BoNT/A3 (strain 254) was compared to sequences in the GenBank database using BLAST, the closest matches were all *Clostridium botulinum* type A toxin sequences, indicating that this sequence is not the result of recombination with any other known BoNT sequence and almost certainly did not result by recombination with another known serotype. The fact that the sequences with less than 99% identity to the query do not form parallel lines but rather have lines that intersect one another is suggestive of more ancient recombination events.

Comparison of the 53 BoNT/B genes sequenced for this

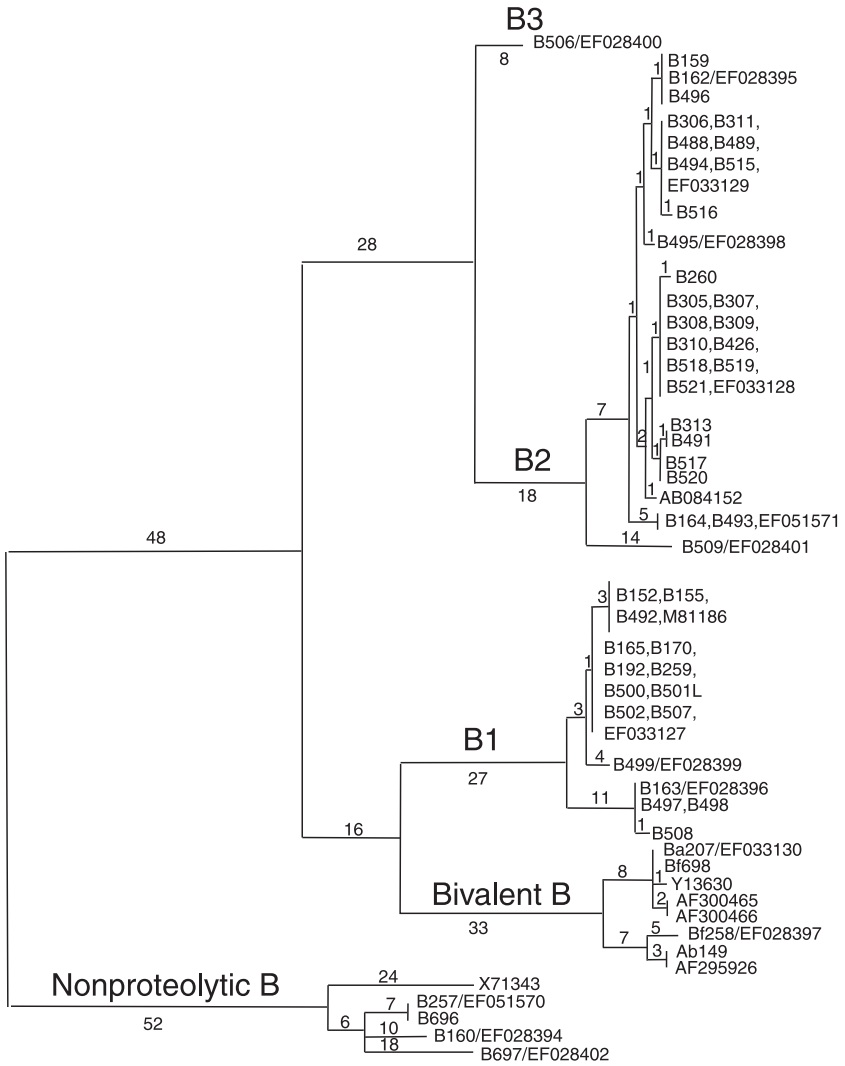


FIG. 5. Comparison of BoNT/B gene sequences. The full-length coding regions of the BoNT/B gene in 53 strains and seven GenBank sequences were aligned. Four distinct clusters that include the BoNT/B1 and BoNT/B2 and bivalent (Ab149, Ba207, Bf258, and Bf698) and nonproteolytic BoNT/B subtypes are apparent. Most strains are of the BoNT/B2 subtype, with 16 strains being of the BoNT/B1 subtype. Strain B506 is separate from the other BoNT/B2 strains and represents a newly identified variation in this serotype.

work and an additional 7 previously reported BoNT/B genes also demonstrated the existence of four, or possibly five, distinct groups. These groups represent the four previously described BoNT/B subtypes, BoNT/B1, BoNT/B2, bivalent BoNT/B, and nonproteolytic BoNT/B (Fig. 5) (20, 23, 27, 37, 48). Compared to BoNT/A, each subtype had more members, with BoNT/B2 being produced by the largest number of strains in our collection. There was also more nucleotide variation within members of each cluster compared to BoNT/A. Nucleotide and amino acid comparisons of the four BoNT/B subtypes are shown in Table 3. Nucleotide differences range from 2% to 4%, with amino acid differences ranging from 4% to 6%. A single BoNT/B isolate (B506) that differed from the closest BoNT/B2 strain by 33 nucleotides, which represents a 2% difference at the amino acid level, was sequenced.

A comparison of the sequences of the 21 BoNT/E genes reported in this work and an additional 15 previously published BoNT/E gene sequences revealed five distinct groups (Fig. 6). The predominant group contains 17 strains, producing a previously described BoNT/E gene that we have termed BoNT/E1 (35). Two groups contain sequences from only *C. butyricum* BoNT/E strains. One of these groups (E Ch. butyr.) contains 11 identical sequences from *C. butyricum* strains collected in China from soil and several food-borne cases of botulism (46). The other group (E It. butyr.) contains the *C. butyricum* BoNT/E sequences from an infant botulism case in Italy (35). Toxins in these three groups differ by 3 to 5% at the amino acid level and likely represent distinct subtypes (Table 3). Four of the strains (E185, E540, E545, and E549) are within one group that produces a previously unidentified BoNT/E that we have termed BoNT/E3. Another two strains (E544 and E546) also appear to represent a unique type of BoNT/E that we have termed BoNT/E2. Amino acid differences among BoNT/E1, E2, and E3 range from 1 to 3%.

A comparison of the BoNT nucleotide sequences from these strains and from available GenBank sequences representing all

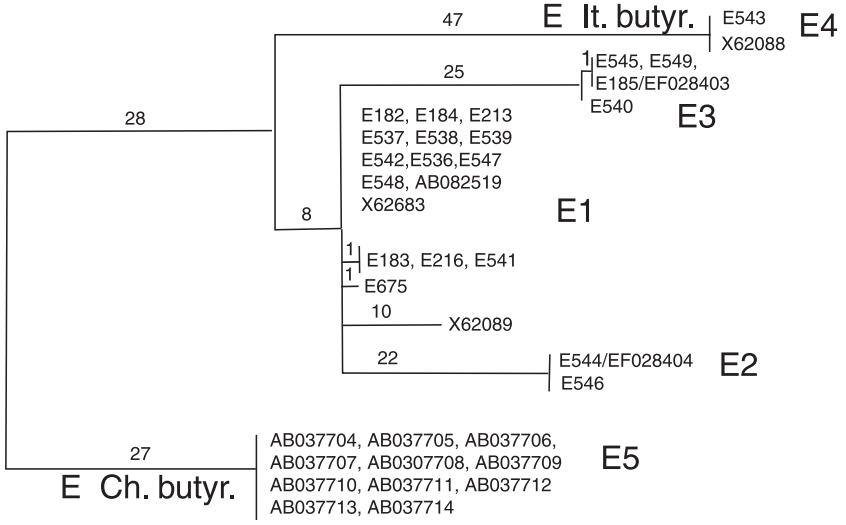


FIG. 6. Comparison of BoNT/E gene sequences. The full-length coding regions of the BoNT/E gene in 21 strains and 15 GenBank sequences were aligned, resulting in five clusters labeled E1 to E5. Two clusters contain sequences from *C. butyricum* BoNT/E strains collected in Italy (E It. butyr.) or China (E Ch. butyr.). The other subtypes include BoNT/E1 and E2 and a newly identified subtype, labeled BoNT/E3, containing four members (E185, E540, E545, and E549).

of the serotypes is shown in Fig. 7. The dendrogram indicates that the seven BoNT genes form three distinct clusters: a large cluster containing the A, E, and F neurotoxins; a second cluster comprised of the B and G toxins; and a third cluster comprised of the C and D toxins. This relationship among the *C. botulinum* neurotoxin genes is different from the results based on the 16S rRNA gene sequence and AFLP analysis (compare Fig. 7 to Fig. 1 and 2). The relationships among the group designations (groups I to IV) are also not maintained (compare Fig. 7 to Fig. 2), suggesting that the toxin gene evolved separately in different genomic backgrounds.

DISCUSSION

More than 170 strains of BoNT-producing clostridial strains were analyzed by different molecular methods to evaluate the genetic diversity and understand the evolutionary history within this species. The conserved 16S rRNA gene sequences illustrate how the different serotypes are closely related to other clostridial species, the AFLP analyses are consistent with the 16S rRNA gene data but add significant resolution to the genomic background that contains the different neurotoxin genes, and finally, the diversity within and between the seven BoNT serotypes reveals a completely different phylogeny within this species that suggests intra- and interspecies transfer of these genes.

The taxonomy of the *C. botulinum* species has historically been based on the identification and/or expression of botulinum toxin genes (38). Since *C. butyricum* and *C baratii* strains that contain BoNT genes have been identified (2, 10, 16, 35), the taxonomy of the toxin-producing clostridia has become more complex. The dendrogram generated using 16S rRNA gene sequence data suggests that the different botulinum neurotoxins that define the species *Clostridium botulinum* are actually contained in genomes from four different clostridial species. The 16S rRNA gene dendrogram demonstrates that BoNT/A-, BoNT/B-, and BoNT/F-producing strains are closely related to each other and to *Clostridium sporogenes* and probably evolved from a common ancestor. However, the genomes for the BoNT/C-, D-, E-, and G-producing strains have 16S rRNA gene sequence profiles that closely align to distant clostridial relatives including *C. novyi*/*C. haemolyticum*, *C. baratii*, and *C. subterminale* (Fig. 1). The results reported here should not change the basic nomenclature for *C. botulinum* in order to avoid confusion and because these taxonomic designations have been based on strong phenotypic as well as genotypic characteristics. However, the presence of related toxin genes in distantly related clostridia serves as a reminder that horizontal gene transfer has played a significant role in the evolution of *Clostridium botulinum.*

AFLP analysis of these strains illustrates clustering by group designation and by toxin serotype. The AFLP-based dendrogram divides the strains into clusters that follow the group I to group IV designations, which are based on physiological characteristics. AFLP analysis clearly separates the proteolytic and nonproteolytic groups and shows the relationship of the genomic backgrounds among strains that are usually defined by the expression of a single 3.8-kb BoNT gene into one of seven different serotypes. This AFLP analysis shows a close relationship of BoNT/A1 subtypes to the A1(B) strains that are distant from the BoNT/A2 and BoNT/A3 subtypes that lie within the BoNT/B1 and BoNT/B2 subtypes. AFLP also shows relationships among the proteolytic BoNT/B and BoNT/F isolates that are mirrored in the nonproteolytic BoNT/B and BoNT/F branches. AFLP analysis supports the group III clustering of BoNT/C and BoNT/D serotypes and the clustering of the group IV BoNT/G strains as distinct from the other serotypes.

Four out of the five bivalent strains in this study cluster together within a branch of the AFLP-based dendrogram that also contains strain A254, which produces BoNT/A3. This branch is also related to a branch containing three BoNT/A2-producing strains, including the remaining bivalent strain, Af695. These bivalent strains contain BoNT/A, B, and F genes

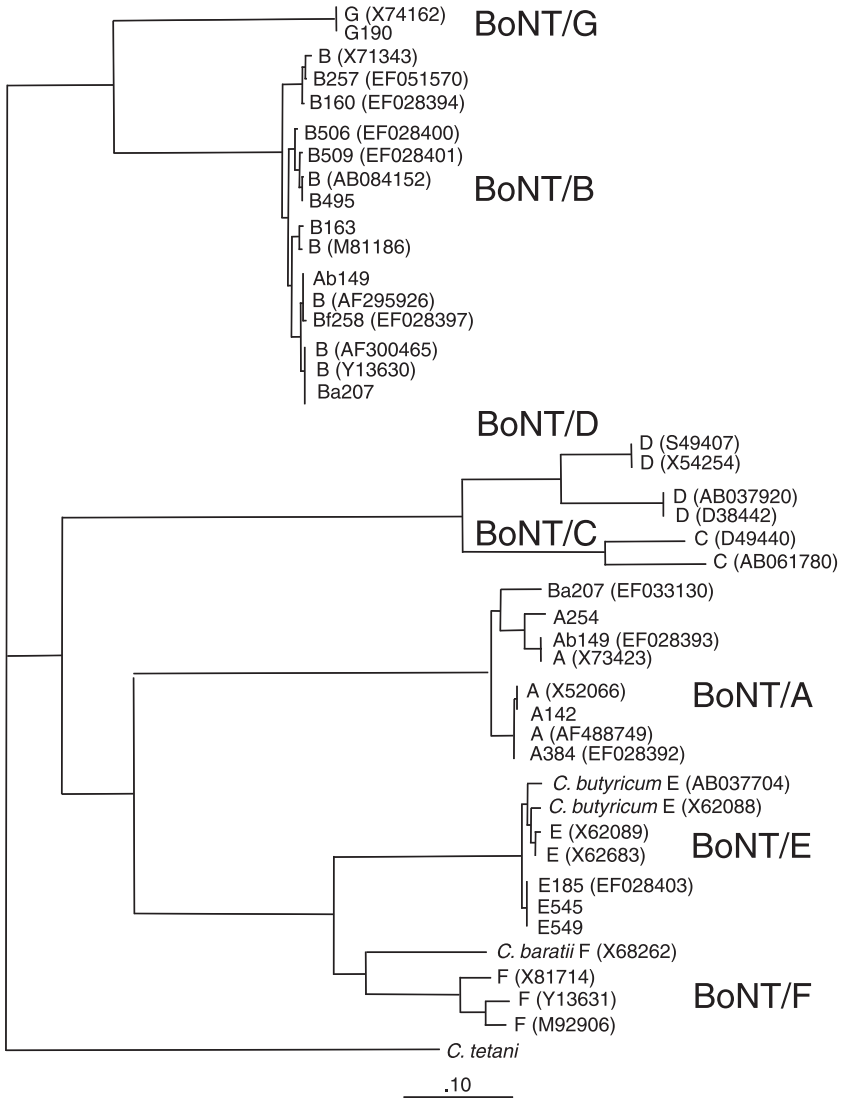


FIG. 7. Comparison of the seven different serotypes of BoNT gene sequences. Shown is a neighbor-joining alignment of the nucleotide coding regions of the seven BoNT genes (A through G) including the tetanus toxin. The comparison of the BoNT genes shows a different relationship of the serotypes than what is found based on 16S rRNA genes or AFLP analysis. Nonproteolytic and bivalent strains (Ba207 and Ab149) and representatives of the different subtypes are included.

and were all isolated from infant botulism cases in different geographic locations. The genomes of these isolates cannot be distinguished by AFLP, yet these strains contain different combinations of neurotoxin genes. The sequences of the individual neurotoxin genes show that the BoNT/B gene sequence in all of these strains is of the same subtype (not identical sequences) but that the BoNT/A genes differ, representing different subtypes. Ab149 contains a BoNT/A2 subtype sequence, but Ba207 contains a completely new BoNT/A subtype that we have termed BoNT/A4. These results suggest either that two lineages of a single strain already carrying the BoNT/B gene acquired the BoNT/A2 and BoNT/A4 genes horizontally or that two strains carrying BoNT/A2 and BoNT/A4 genes both acquired the same BoNT/B gene horizontally. Southern blotting or genome analysis of toxin gene integration sites would be necessary to distinguish between these possibilities.

Four of the five strains producing two BoNT serotypes (bivalent strains) were isolated from infants with botulism. This high proportion of bivalent strains found in infants might reflect sample bias within this collection, but this has been reported previously by others (4). Of the 10 strains isolated from infants, 4 were found to be bivalent in this study. These 10 strains that affected infants are located in different branches of the AFLP dendrogram and include a *C. butyricum* BoNT/E-producing strain (E543) from Italy. An examination of the sequences of the BoNT/A, B, and E genes from these strains from infants shows that the toxin gene frequently represents a unique cluster within the serotype. Within both the BoNT/A and the BoNT/B gene sequence-based dendrograms, three of the nine clusters in the trees contain BoNT produced by strains from infant cases; all of the bivalent strains producing BoNT/B form a unique cluster, as does the single strain (E543) producing

BoNT/E. It must be noted, however, that the strains obtained from infants in this collection were deliberately chosen for their unusual characteristics and that a large collection of infant isolates may show higher percentages of the more common BoNT/A1 and BoNT/B1 subtypes.

The neurotoxin gene sequence comparisons of all of the toxin serotypes (serotypes A to G) suggest that the BoNT gene has evolved separately in different genomic backgrounds. The dendrogram indicates that the seven BoNT genes form three distinct clusters: a large cluster consisting of the A, E, and F neurotoxins; a second cluster comprised of the B and G toxins; and a third cluster comprised of the C and D toxins. These relationships are different from the group I to group IV designations supported by the 16S rRNA gene sequences and AFLP analysis. This discordant phylogeny suggests gene transfer among different clostridial species and suggests that *C. botulinum* has contributed to the movement of the BoNT gene into various genetic backgrounds. The nucleotide differences within these neurotoxin genes are probably a result of both natural variation and selection pressure. Recombination events, similar to that illustrated in Fig. 4, where an A1/A3 recombination created BoNT/A2, can also be found within other toxin gene lineages, including many C/D and D/C interserotype recombination events that have previously been reported (31). Several recombination events within the nontoxic nonhemagglutinin genes of A1, B, and F strains have also been described (11).

The current analysis of 134 BoNT/A, B, and E toxin genes significantly increases our understanding of the extent of subtype variability within these three serotypes. The neurotoxin sequences demonstrate that there is more diversity within these toxin serotypes than previously known (summarized in reference 39). Two new BoNT/A genes, one new BoNT/B gene, and two new BoNT/E genes were identified. The two new BoNT/A genes clearly represent new BoNT/A subtypes that we have termed BoNT/A3 and BoNT/A4. Subtypes have historically been defined by the differential binding of monoclonal antibodies (13, 28, 39), and the 15% and 11% amino acid differences between BoNT/A1, A3, and A4 would certainly result in differential binding of some BoNT/A monoclonal antibodies (39). The toxins encoded by the new BoNT/B gene (BoNT/B3) and the new BoNT/E genes (BoNT/E2 and E3) differ from BoNT/B1 and BoNT/E1 by 4%, 1%, and 2% at the amino acid level, respectively. It is not clear whether these new toxins represent new toxin subtypes using the historical standard of monoclonal antibody binding. While single amino acid changes can cause a loss of antibody binding, whether the amino acid differences in these toxins are large enough to result in differential monoclonal antibody binding is unknown and must await the completion of studies using panels of monoclonal antibodies. However, lacking monoclonal antibody studies, subtypes could also be defined based on nucleotide or, more appropriately, amino acid differences, especially where multiple members are identified from different strains. This is the case for the BoNT/E2 and E3 genes.

Accurate analyses and understanding of the recombinations between toxin genes of different serotypes and subtypes may be more helpful for identifying potential vaccines and therapeutic antibodies than relying on phylogenetic dendrograms or overall pairwise sequence distances. For example, BoNT/A2 represents a recombination of the 5′ end of the BoNT/A1 light-chain gene with the 3′ end of the BoNT/A3 gene. This analysis permits the identification of regions of BoNT that could be used to generate antibodies that can cross-react with all three subtypes. Similarly, knowledge of the recombination site between BoNT/C and D will allow the identification of targeted regions for the generation of antibodies that cross-react with chimeric BoNT/C and D.

In conclusion, the neurotoxins produced by *Clostridium tetani*, *C. butyricum*, and *C. baratii* are as similar, or more similar, to *C. botulinum* neurotoxins as the various serotypes of BoNT are to each other (36). Historically, the expression of these neurotoxins has been used to taxonomically identify these clostridia as *C. botulinum* or *C. tetani*. The presence of these toxins in different genetic backgrounds suggests their movement both within the species and among other species. Most of these bacteria are distributed throughout the world, yet there is no known geographical relationship to the genetic diversity. Environmental niches, geographic distribution, and gene transfer mechanisms among these spore-forming clostridia must all interact to produce the sequence diversity observed in one of the most lethal neurotoxins known. The BoNTs produced by these clostridial species show sequence differences both within and between serotypes. Identifying the extent of these differences is the crucial first step in the development of improved diagnostics and therapeutics for the treatment of botulism.

ACKNOWLEDGMENTS

Funding for this project was provided by the Department of Homeland Security, National Bioforensics Analysis Center, DoD contract DAMD17-03-C-C0076, and NIAID cooperative agreement U01 AI056493. Some of this work was conducted under the auspices of the U.S. Department of Energy.

We thank the DOE Joint Genome Institute (JGI) at Los Alamos National Laboratory for their support by providing technical assistance and facilities for DNA sequencing. We thank Stephen Arnon for his thorough review of the manuscript.

Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army.

REFERENCES

1. **Arnon, S. A., R. Schecter, T. V. Inglesby, D. A. Henderson, J. G. Bartlett, M. S. Ascher, E. Eitzen, A. D. Fine, J. Hauer, M. Layton, S. Lillibridge, M. T. Osterholm, T. O'Toole, G. Parker, T. M. Perl, P. K. Russell, D. L. Swerdlow, and K. Tonat.** 2001. Botulinum toxin as a biological weapon: medical and public health management. JAMA **285:**1059–1070.
2. **Aureli, P., L. Fenicia, B. Pasolini, M. Gianfranceschi, L. M. McCroskey, and C. L. Hatheway.** 1986. Two cases of type E infant botulism caused by neurotoxigenic *Clostridium butyricum* in Italy. J. Infect. Dis. **154:**207–211.
3. **Balding, P., E. R. Gold, D. A. Boroff, and T. A. Roberts.** 1973. Observations on receptor specific proteins. II. Haemagglutination and haemagglutination-inhibition reactions of *Clostridium botulinum* types A, C, D and E haemagglutinins. Immunology **25:**773–782.
4. **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.
5. **Bozheyeva, G., Y. Kunakbayev, and D. Yeleukenov.** 1999. Former soviet biological weapons facilities in Kazakhstan: past, present, and future. Center for Nonproliferation Studies, Monterey Institute of International Studies, Monterey, CA.
6. **Centers for Disease Control and Prevention.** 1998. Botulism in the United States, 1899–1998. Handbook for epidemiologists, clinicians, and laboratory workers. Public Health Service, U.S. Department of Health and Human Services, Atlanta, GA. http://www.bt.cdc.gov/agent/botulism/index.asp.
7. **Christensen, D. R., L. J. Hartman, B. M. Loveless, M. S. Frye, M. A. Shipley, D. L. Bridge, M. J. Richards, R. S. Kaplan, J. Garrison, C. D. Baldwin, D. A. Kulesh, and D. A. Norwood.** 2006. Detection of biological

threat agents by real-time PCR: comparison of assay performance on the R.A.P.I.D., the LightCycler, and the Smart Cycler platforms. Clin. Chem. **52:**141–145.
8. **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.
9. **Dineen, S. S., M. Bradshaw, and E. A. Johnson.** 2003. Neurotoxin gene clusters in *Clostridium botulinum* type A strains: sequence comparison and evolutionary implications. Curr. Microbiol. **46:**345–352.
10. **East, A. K., M. Bhandari, S. Hielm, and M. D. Collins.** 1998. Analysis of the botulinum neurotoxin type F gene clusters in proteolytic and nonproteolytic *Clostridium botulinum* and *Clostridium barati*. Curr. Microbiol. **37:**262–268.
11. **East, A. K., M. Bhandari, J. M. Stacey, K. D. Campbell, and M. D. Collins.** 1996. Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic *Clostridium botulinum* types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component. Int. J. Syst. Bacteriol. **46:**1105–1112.
12. **East, A. K., J. M. Stacey, and M. D. Collins.** 1994. Cloning and sequencing of a hemagglutinin component of the botulinum neurotoxin complex encoded by the *Clostridium botulinum* types A and B. Syst. Appl. Microbiol. **17:**306–312.
13. **Gibson, A. M., N. K. Modi, T. A. Roberts, P. Hambleton, and J. Melling.** 1988. Evaluation of a monoclonal antibody-based immunoassay for detecting type B *Clostridium botulinum* toxin produced in pure culture and an inoculated model cured meat system. J. Appl. Bacteriol. **64:**285–291.
14. **Gimenez, D. F., and A. S. Ciccarelli.** 1970. Another type of *Clostridium botulinum*. Zentralbl. Bakeriol. 1 Abt. Orig. A Med. Mikrobiol. Infecktionskr. Parasitl. **215:**221–224.
15. **Gimenez, D. F., and J. A. Gimenez.** 1993. Serological subtypes of botulinal neurotoxins, p. 421–431. *In* B. R. DasGupta (ed.), Botulinum and tetanus neurotoxins. Plenum Press, New York, NY.
16. **Hall, J. D., L. M. McCroskey, B. J. Pincomb, and C. L. Hatheway.** 1985. Isolation of an organism resembling *Clostridium barati* which produces type F botulinal toxin from an infant with botulism. J. Clin. Microbiol. **21:**654–655.
17. **Hatheway, C. L.** 1995. Botulism: the present status of the disease. Curr. Top. Microbiol. Immunol. **195:**55–75.
18. **Hatheway, C. L.** 1990. Toxigenic clostridia. Clin. Microbiol. Rev. **3:**66–98.
19. **Hatheway, C. L., and J. L. Ferreira.** 1996. Detection and identification of *Clostridium botulinum* neurotoxins. Adv. Exp. Med. Biol. **391:**481–498.
20. **Hutson, R. A., M. D. Collins, A. K. East, and D. E. Thompson.** 1994. Nucleotide sequence of the gene coding for non-proteolytic *Clostridium botulinum* type B neurotoxin: comparison with other clostridial neurotoxins. Curr. Microbiol. **28:**101–110.
21. **Hutson, R. A., D. E. Thompson, and M. D. Collins.** 1993. Genetic interrelationships of saccharolytic *Clostridium botulinum* types B, E and F and related clostridia as revealed by small-subunit rRNA gene sequences. FEMS Microbiol. Lett. **108:**103–110.
22. **Hutson, R. A., D. E. Thompson, P. A. Lawson, R. P. Schocken-Itturino, E. C. Bottger, and M. D. Collins.** 1993. Genetic interrelationships of proteolytic *Clostridium botulinum* types A, B, and F and other members of the *Clostridium botulinum* complex as revealed by small-subunit rRNA gene sequences. Antonie Leeuwenhoek **64:**273–283.
23. **Ihara, H., T. Kohda, F. Morimoto, K. Tsukamoto, T. Karasawa, S. Nakamura, M. Mukamoto, and S. Kozaki.** 2003. Sequence of the gene for *Clostridium botulinum* type B neurotoxin associated with infant botulism, expression of the C-terminal half of heavy chain and its binding activity. Biochim. Biophys. Acta **1625:**19–26.
24. **Kaufman, L., and P. J. Rousseeuw.** 1990. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, Inc., New York, NY.
25. **Keto-Timonen, R., M. Nevas, and H. Korkeala.** 2005. Efficient DNA fingerprinting of *Clostridium botulinum* types A, B, E, and F by amplified fragment length polymorphism analysis. Appl. Environ. Microbiol. **71:**1148–1154.
26. **Kirma, N., J. L. Ferreira, and B. R. Baumstark.** 2004. Characterization of six type A strains of *Clostridium botulinum* that contain type B toxin gene sequences. FEMS Microbiol. Lett. **231:**159–164.
27. **Kozaki, S., Y. Kamata, T. Nishiki, H. Kakinuma, H. Maruyama, H. Takahashi, T. Karasawa, K. Yamakawa, and S. Nakamura.** 1998. Characterization of *Clostridium botulinum* type B neurotoxin associated with infant botulism in Japan. Infect. Immun. **66:**4811–4816.
28. **Kozaki, S., S. Nakaue, and Y. Kamata.** 1995. Immunological characterization of the neurotoxin produced by *Clostridium botulinum* type A associated with infant botulism in Japan. Microbiol. Immunol. **39:**767–774.
29. **Leighton, G.** 1923. Botulism and food preservation (the Loch Maree tragedy). Collins & Sons, Glasgow, United Kingdom.
30. **McCroskey, L. M., C. L. Hatheway, L. Fenicia, B. Pasolini, and P. Aureli.** 1986. Characterization of an organism that produces type E botulinal toxin but which resembles *Clostridium butyricum* from the feces of an infant with type E botulism. J. Clin. Microbiol. **23:**201–202.
31. **Moriishi, K., M. Koura, N. Abe, N. Fujii, Y. Fujinaga, K. Inoue, and K. Oguma.** 1996. Mosaic structures of neurotoxins produced from *Clostridium botulinum* types C and D organisms. Biochim. Biophys. Acta **1307:**123–126.
32. **Moriishi, K., M. Koura, N. Fujii, Y. Fujinaga, K. Inoue, B. Syuto, and K. Oguma.** 1996. Molecular cloning of the gene encoding the mosaic neurotoxin, composed of parts of botulinum neurotoxin types C1 and D, and PCR detection of this gene from *Clostridium botulinum* type C organisms. Appl. Environ. Microbiol. **62:**662–667.
33. **Mullaney, B. P., M. G. Pallavincini, and J. D. Marks.** 2001. Epitope mapping of neutralizing botulinum neurotoxin A antibodies by phage display. Infect. Immun. **69:**6511–6514.
34. **Popoff, M., and J.-C. Marvaud.** 1999. Structural and genomic features of clostridial neurotoxins, p. 202–228. *In* J. E. Alouf and J. H. Freer (ed.), The comprehensive sourcebook of bacterial protein toxins. Academic Press, London, United Kingdom.
35. **Poulet, S., D. Hauser, M. Quanz, H. Niemann, and M. R. Popoff.** 1992. Sequences of the botulinal neurotoxin E derived from *Clostridium botulinum* type E (strain Beluga) and *Clostridium butyricum* (strains ATCC 43181 and ATCC 43755). Biochem. Biophys. Res. Commun. **183:**107–113.
36. **Raffestin, S., J. C. Marvaud, R. Cerrato, B. Dupuy, and M. R. Popoff.** 2004. Organization and regulation of the neurotoxin genes in *Clostridium botulinum* and *Clostridium tetani*. Anaerobe **10:**93–100.
37. **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.
38. **Smith, L. D. S.** 1977. Botulinum: the organism, its toxins, the disease. Charles C. Thomas, Springfield, IL.
39. **Smith, T. J., J. Lou, I. 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.
40. **Sonnabend, O., W. Sonnabend, R. Heinzle, T. Sigrist, R. Dirnhofer, and U. Krech.** 1981. Isolation of *Clostridium botulinum* type G and identification of type G botulinal toxin in humans: report of five sudden unexpected deaths. J. Infect. Dis. **143:**22–27.
41. **Sonnabend, O. A., W. F. Sonnabend, U. Krech, G. Molz, and T. Sigrist.** 1985. Continuous microbiological and pathological study of 70 sudden and unexpected infant deaths: toxigenic intestinal *Clostridium botulinum* infection in 9 cases of sudden infant death syndrome. Lancet **i:**237–241.
42. **Stackebrandt, E., I. Kramer, J. Swiderski, and H. Hippe.** 1999. Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*. FEMS Immunol. Med. Microbiol. **24:**253–258.
43. **Suen, J. C., C. L. Hatheway, A. G. Steigerwalt, and D. J. Brenner.** 1988. *Clostridium argentinense*, sp. nov.: a genetically homogeneous group composed of all strains of *Clostridium botulinum* toxin type G and some nontoxigenic strains previously identified as *Clostridium subterminale* or *Clostridium hastiforme*. Int. J. Syst. Bacteriol. **38:**375–381.
44. **Ticknor, L. O., A. B. Kolstø, K. K. Hill, P. Keim, M. T. Laker, M. Tonks, and P. J. Jackson.** 2001. Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. Appl. Environ. Microbiol. **67:**4863–4873.
45. **United Nations Security Council.** 1995. Tenth report of the Executive Committee of the Special Commission established by the Secretary-General pursuant to paragraph 9(b) (I) of Security Council resolution 687 (1991), and paragraph 3 of resolution 699 (1991) on the activities of the Special Commission. United Nations Security Council, New York, NY.
46. **Wang, X., T. Maegawa, T. Karasawa, S. Kozaki, K. Tsukamoto, Y. Gyobu, K. Yamakawa, K. Oguma, Y. Sakaguchi, and S. Nakamura.** 2000. Genetic analysis of type E botulinum toxin-producing *Clostridium butyricum* strains. Appl. Environ. Microbiol. **66:**4992–4997.
47. **Ward, B. Q., B. J. Carroll, E. S. Garrett, and G. B. Reese.** 1967. Survey of the U.S. Atlantic coast and estuaries from Key Largo to Staten Island for the presence of *Clostridium botulinum*. Appl. Microbiol. **15:**964–965.
48. **Whelan, S. M., M. J. Elmore, N. J. Bodsworth, J. K. Brehm, T. Atkinson, and N. P. Minton.** 1992. Molecular cloning of the *Clostridium botulinum* structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence. Appl. Environ Microbiol. **58:**2345–2354.
49. **Willems, A., A. K. East, P. A. Lawson, and M. D. Collins.** 1993. Sequence of the gene coding for the neurotoxin of *Clostridium botulinum* type A associated with infant botulism: comparison with other clostridial neurotoxins. Res. Microbiol. **144:**547–556.
50. **Zhang, L., W.-J. Lin, S. Li, and K. R. Aoki.** 2003. Complete DNA sequences of the botulinum neurotoxin complex of *Clostridium botulinum* type A-Hall (Allergan) strain. Gene **315:**21–32.