LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-222662

Genetic Diversity Among Botulinum Neurotoxin Producing Clostridial Strains

K. K. Hill, T. J. Smith, C. H. Helma, L. O. Ticknor, B. T. Foley, R. T. Svennson, J. L. Brown, E. A. Johnson, L. A. Smith, R. T. Okinaka, P. J. Jackson, J. D. Marks

July 7, 2006

Journal of Bacteriology

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Updated October 14, 2003

p-[00000000

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]

[1]Bioscience, [3]Decision Applications and [4]Theoretical Divisions, Los Alamos National Laboratory, Los Alamos, NM 87545;

[2]Integrated Toxicology Division, United States Army Medical Institute of Infectious Diseases (USAMRIID), Fort Detrick, MD 21702

[5]Department of Food Microbiology and Toxicology, Food Research Institute (FRI), University of Wisconsin, Madison, WI 53706

[6]Defense Biology Division, Lawrence Livermore National Laboratory, Livermore, CA 94551

[7]Department of Anesthesia and Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, Rm 3C-38, San Francisco General Hospital, 1001 Potrero Ave, San Francisco, CA 94110

*Corresponding author: Bioscience Division, MS M888, Los Alamos National Laboratory, Los Alamos, NM 87545, Phone: (505) 667-1309, FAX (505) 665-3024, email: khill@lanl.gov

This work was partially supported by DoD contract DAMD17-03-C-0076 and by NIAID cooperative agreement U01 AI056493

Keywords: *Clostridium botulinum* neurotoxin, sequence diversity, BoNT, AFLP, 16s rRNA

Running title: Genetic diversity of *Clostridium botulinum* strains

Abbreviations: AFLP: amplified fragment length polymorphism, BoNT: botulinum neurotoxin, rRNA: ribosomal ribonucleic acid, SNP: single nucleotide polymorphism, PCR: polymerase chain reaction

Abstract

Clostridium botulinum is a taxonomic designation for many diverse anaerobic spore-forming rod-shaped bacteria which have the common property of producing botulinum neurotoxins (BoNTs). The BoNTs are exoneurotoxins that can cause severe paralysis and even death in humans and various other animal species. A collection of 174 *C. botulinum* strains were examined by amplified fragment length polymorphism (AFLP) analysis and by sequencing of the 16S rRNA gene and BoNT genes to examine genetic diversity within this species. This collection contained representatives of each of the seven different serotypes of botulinum neurotoxins (BoNT A-G). Analysis of the16S rRNA sequences confirmed earlier reports of at least four distinct genomic backgrounds (Groups I-IV) each of which has independently acquired one or more BoNT serotypes through horizontal gene transfer. AFLP analysis provided higher resolution, and can be used to further subdivide the four groups into sub-groups. Sequencing of the BoNT genes from serotypes A, B and E in multiple strains 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 serotypes of BoNT were compared and show varying degrees of inter-relatedness and recombination as has been previously noted for the NTNH gene which is linked to BoNT. 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 treatment of botulism.

Introduction

Clostridium botulinum is a taxonomic collection of several distinct species of anaerobic gram positive spore-forming bacteria which produce the most poisonous substance known, botulinum neurotoxin (BoNT) (7). These organisms, along with related neurotoxin-producing species which for a variety of reasons were not included under the *C. botulinum* taxon, pose global health problems that affect both infants and adult humans, and can also affect wildlife, waterfowl and domestic animals. They cause intoxication through ingestion of the neurotoxin in contaminated foods and also in some cases toxicoinfections occur after contact with bacteria or bacterial spores (8, 18). These pathogens are ubiquitous and can be found in soils and sediments in both fresh water and marine environments (44).

BoNTs are classified by the Centers for Disease Control (CDC) as one of the six highest-risk threat agents for bioterrorism (the "Category A agents"), due to their extreme potency and lethality, ease of production and transport, and need for prolonged intensive care (1). Both Iraq and the former Soviet Union produced BoNT for use as weapons (5, 9) and the Japanese cult Aum Shinrikyo attempted to use BoNT for bioterrorism (1). Since the terrorist events of September 11, 2001 and the subsequent intentional terror release of anthrax spores, development of environmental toxin sensors, diagnostic tests for botulism, and specific countermeasures for 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 they produce.

C. botulinum strains are usually described as belonging within one of four different

Groups (I, II, III, and IV) based on physiologic characteristics (19). The toxins produced by them are categorized into seven serologically distinct groups (A through G), based on recognition by polyclonal serum (18). The BoNT is encoded by an approximately 3.8 kb gene, preceded by a non-toxic-non-hemagglutinin (NTNH) gene and several other genes encoding toxin-associated proteins (HA-17, HA-33, HA-70, p21 and/or p47) (3, 7, 12, 13, 34). The BoNT gene for serotype A, B, E and F strains 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 with the toxin operon (34). Strains producing inter-serotype recombinant toxins have been noted, primarily the C/D and D/C phage-encoded serotypes (32, 33). Additionally, several strains produce multiple toxins. Bivalent strains of *C. botulinum*, each producing two toxins of serotypes Ab, Ba, Af, and Bf have been noted (16, 36).

The genomic background housing these BoNT genes within *C. botulinum* has been identified as very diverse, and in addition other species have been known to harbor BoNT genes, such as *C. butyricum* (BoNT E) (2, 31), *C.baratii* (BoNT F) (17) and *C. argentinense* (BoNT G) (41). Previous16S rRNA analysis of many different *Clostridium* species has shown the *C. botulinum* form four distinct clusters, with each cluster representing one of the four different physiological Groups (I-IV) (7, 23). Previous 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 1 and Group II clusters (26). Like the 16S rRNA 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 (26). 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 house the very different BoNT genes.

Recently it has become appreciated that there is significant sequence diversity (subtypes) within the BoNT genes and toxins of at least six of the seven serotypes (37). The relationship between toxin gene diversity and clostridial genomic diversity is unknown. Such subtypes can differ by 2.6% to 31.6% percent at the amino acid level, and these differences can affect the binding and neutralization by monoclonal and polyclonal antibodies (14, 29, 37). Since 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 (37). Defining the extent of such toxin diversity would be a first step in the development of detection systems and countermeasures for prevention and treatment of botulism. 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 house these toxins.

In order to understand the extent of toxin gene diversity and the relationship between genomic diversity between *C. botulinum* serotypes and subtypes and other toxin-producing species of *Clostridium,* 174 toxin-producing strains were analyzed from a collection that included representatives of all neurotoxin serotypes (BoNT A-G). Several methods were used to examine the strains including sequencing of the 16S rRNA gene, analysis of the genome by amplified fragment length polymorphism (AFLP), and sequencing of BoNT/A, B and E neurotoxin genes. Nucleotide sequences of the 16S rRNA and BoNT gene 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 are presented, as well as the extent of toxin gene diversity and the relationship between toxin types, subtypes, and genomic differences.

Materials and Methods

Strains

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 (VPI) Anaerobe Laboratory collection. Strains were serotyped using antibody capture ELISA with serotype-specific monoclonal antibodies. In some cases, serotypes were confirmed using mouse neutralization (20). Silent (not expressed) BoNT/B genes were detected using real time PCR (6).

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 hours at 35°C then harvested by low speed centrifugation. The pellets were resuspended in 8.5ml TE buffer (10mM Tris, 1mM EDTA, pH 8.0) then quickly frozen in a dry ice/ethanol bath and stored at –70°C until further processing. Upon removal, the resuspended pellets underwent three successive freeze/thaw cycles of freezing in a dry ice/ethanol bath followed by melting at 65°C. Sodium dodecylsulfate (SDS) (450 ul) and 45ul of Proteinase K (10mg/ml) were added, mixed, and incubated at 42°C for 1 hour. After incubation, 1.5 ml of 5M NaCl solution and 1.4ml of a 10% CTAB (cetyl trimethyl ammonium bromide) solution were added, mixed thoroughly and incubated at 65°C for 10 minutes. Following this incubation, three organic extractions of the mixture were performed. The initial extraction involved the addition of an

equal volume of ~~a~~ chloroform/isoamyl alcohol (24:1) with incubation while rocking for 10 minutes. 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). 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 then resuspended in 2.0ml TE buffer. The DNA preparation was quantified with a spectrophotometer, diluted to 25 ug/ml, and analyzed on an agarose gel to determine quality.

Amplified Fragment Length Polymorphism (AFLP) analysis of DNA samples

Each sample consisted of 100 ng of DNA that was digested with the restriction endonucleases *Eco*RI and *Mse*I and the resulting fragments were ligated to double-stranded adapters. The digested and ligated DNA was then amplified by PCR using *Eco*RI and *Mse*I +0/+0 primers (5'-GTAGACTGCGTACCAATTC-3' and 5'-GACGATGAGTCCTGAGTAA-3' respectively). Five μl of each product was used as template in subsequent selective amplifications using the +1/ +1 primer combination of 6-carboxyfluorescein-labeled *Eco*RI-T (5'-GTAGACTGCGTACCAATTCT-3') and *Mse*I-T (5'-GACGATGAGTCCTGAGTAAT-3'). 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 minute heat denaturation at 95°C, the reactions were loaded on an ABI 3100 automated fluorescent sequencer. Each set of AFLP reactions also contained a control DNA as a template. Inclusion of such a reaction in each run and analysis set allowed a comparison of

results from earlier archived analysis sets 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 by Ticknor *et al*. (42). 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 was used which compares the presence and absence of fragments of a given length. 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 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 (UPGMA) agglomerative hierarchical clustering method (25). All statistical data manipulations were done using codes developed in 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 sequencing of *C. botulinum* samples.

Representatives of the various BoNT producing *Clostridium* strains were selected to sequence the 16S rRNA. The primers 1492R (5'-GGTTACCTTGTTACGACTT-3') and 27F (5'-AGAGTTTGATCMTGGCTCAG-3') were used to PCR amplify about 1400 bases of the 1.5kb gene. The purified PCR template was then sequenced using these primers and internal primers 533Fb (5'-GCCAGCAGCNGCGGTAA-3'), 940Fb (5'-CGGGGGYCCGCACAAGC-3') and 910Rb (5'-GCCCCCGTCAATTYHTTTGAG-3'). The 16S rRNA phylogenetic tree was created from an alignment of 16S rRNA gene sequences, some of which were new to this study and the others of which were 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 1329 bases for phylogenetic analysis. The phylogenetic tree was calculated with PHYLIP dnadist and neighbor programs, using the F84 model of evolution and four sequences from Alkaliphilus genus (GenBank entries 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 removed to make the final figure.

BoNT gene Polymerase Chain Reaction (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 and sequencing primers for each of the neurotoxin gene fragments are listed in Table 2. Initial PCR 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 dNTP, 20 pmol of each primer, 2.5U of Amplitaq DNA polymerase (PerkinElmer, Inc., Boston MA) and approximately 1 ng template DNA in 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 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 hour at 80V. Gels were stained for 20 min with a solution containing 1 μg of ethidium bromide/ml, de-stained in distilled water, then visualized and photographed under UV light. PCR amplicons were purified using a Qiagen PCR Purification Kit (Qiagen Inc., Valencia, CA) then sequenced using ABI dye terminator 3.1 chemistry on 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/), CLUSTAL-W (http://www.ebi.ac.uk/clustalw/) and hand editing with BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) software and were

gapstripped 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 trees were rendered with treetool (http://packages.debian.org/unstable/science/TreeTool/). Intra- and inter- serotype BoNT gene recombination was explored with SimPlot (http://sray.med.som.jhmi.edu/SCRoftware/simplot/), and BioEdit. The Simplot analysis in Figure 4 is a comparison of different BoNT sequences and the BoNT/A2 sequence from GenBank accession number X73423. It was generated with a sliding window of 200 bp and the percent similarity between the two sequences 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 various researchers over many years. Table 1 lists the strains used in this study and describes the strain identification number from the original researcher or institution if known. The collection of the 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 and included five bivalent strains (Af695, Bf698, Bf258, Ba207, Ab149). For the purposes of the above classification, bivalent strains were classified based on the predominant toxin produced. All strains were tested to confirm that they produced toxin, either by ELISA and/or by 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 Analysis

Comparative analysis of the nucleotide sequences from the conserved 16S rRNA gene using a subset of 109 strains representing the various 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 species within the *Clostridium* genus. These results confirm and extend the earlier work of many others (7, 22, 23, 40). The majority of the strains in this collection are tightly clustered with16S 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 house all of the A, and most of the B and F neurotoxin genes

(23). The remaining strains form phylogenetically distant species clusters that define the remaining physiological Groups II, III and IV (22). The Group II strains include all of the neurotoxin E strains, nonproteolytic B strains, and nonproteolytic F strains, and are most closely related to species *C. beijerinckii* and *C. butyricum*. Group III (highly related to *C. novyi*) encode BoNT/C, D, 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 *C. subterminale* and *C. argentinense* and belong to BoNT serotype G that is encoded by a plasmid.

AFLP Analysis

The tree generated by AFLP analysis of the *Clostridium* DNAs is shown in Figure 2. The large branching patterns within this tree further resolve the tree obtained with the 16S rRNA gene. The AFLP tree clearly agrees with all other methods for classifying *C. botulinum* strains, including both genetic sequence analysis and the characterization of physiological differences used in classical bacteriology. These physiological differences have historically been used to categorize the *C. botulinum* into Groups I-IV. The AFLP tree shows a large separation between the proteolytic (Group 1) and nonproteolytic strains (Groups II, III, and IV) and forms distinct branches representing Groups I-IV separated by distances greater than 0.75. In general, the AFLP tree 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 strains representing the different BoNT/A and B subtypes show more diversity and are not clearly differentiated using AFLP analysis. Interestingly, four of the five bivalent strains included in this study (Bf698, Bf258, Ba207, and Ab149) also appear as a cluster, 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) (36).

The Group I BoNT/A strains are found within four AFLP clusters which generally contain just one toxin serotype or a single combination of two serotypes. One branch contains most of the BoNT/A producing strains examined (37/59). These strains produce BoNT/A1 and include the ATCC type strain 25763 (A146) and the BoNT/A Hall 174 strain (A143) sequenced by the Sanger Institute (http://www.sanger.ac.uk/projects/C_botulinum/). Of the 37 BoNT/A1 strains, 29 form a monophyletic cluster by analysis of both the BoNT gene sequence and AFLP data and suggest 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 to the other BoNT/A producing strains. These three distantly related clusters include: 1) a group of three of the four BoNT/A2 strains (Af695, A693 and A694); 2) a separate cluster containing the bivalent Ab strain (Ab149), a BoNT/A1 producing strain (A384) and three BoNT/B producing strains; and 3) the A254 strain, also known as Loch Maree. This strain, from a 1922 botulism incident in Scotland (30)**,** 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 1 proteolytic BoNT/B strains into smaller clusters which include the serologically distinct BoNT/B1 and BoNT/B2 producing strains (28),

and four bivalent BoNT/B producing strains (Ba207, A149, 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 US origin and associated with food borne cases due to improperly processed vegetables, where the BoNT/B2 strains are mostly from Europe and associated with animal cases or meat. It is interesting that the original BoNT/B2 strain was isolated from a case of infant botulism in Japan, and two recent published BoNT/B sequences from soil from Korea (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 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 tree where three nonproteolytic BoNT/B (B160, B257, B697) producing strains cluster and are most closely related to a nonproteolytic BoNT/F producing strain (F550).

In addition, four bivalent strains, Ab (Ab149), Ba (Ba207), and Bf (Bf698 and Bf258) included in this study cluster together at the 0.2 level in this area of the AFLP tree. The genetic backgrounds of these four strains cannot be differentiated by AFLP, and their 16S rRNA genes were found to be greater than 99.93% identical to one another. By comparison, A150 and Bf258 separated by AFLP clustering, were 99.78% identical to each other in 16S rRNA sequence comparisons. 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 appears to

indicate very recent horizontal transfer of these toxin genes into the same bacterial lineage. Interestingly, they all appear to be from infant cases of botulism from different geographic locations: Sweden, Texas, New Mexico, and Utah.

Group II *C. botulinum* BoNT/E producing strains, which are usually associated with fish and other marine mammals, appear within their own branch of the AFLP tree. The 21 BoNT/E producing strains include samples from salmon, whale and soil from the Olympian forest. The placement of these Group II BoNT/E strains in a distinct branch of the AFLP tree 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 to these other 20 *C. botulinum* strains and is from an infant botulism case in Italy (31). 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 received from two different research collections (USAMRIID and VPI) of the "beluga" strain were intentionally included in these experiments. These beluga isolates are indistinguishable and add confidence to the results obtained using strains collected by different researchers over many years.

Most of the Group III BoNT/C(17/19) and BoNT/D (6/6) serotypes form a distinct branch of the AFLP tree. 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, C659), seven of which are from Western Europe. These strains are linked to disease in mammals, and

they are either proven to be or thought to be type C strains. Another cluster of five strains, shown to be C/D strains, can be mapped to marine or fresh water sediments. Three of the strains (C525, C526, C527) are from marine sediments in the US. Strain C209, which differs slightly, is from Japan. Other Group III strains are from the US (C529), Japan (D701), South America (C700) and Africa (C524, C699, D535). Two of these isolates, C523 and C650, were identical strains that were blindly 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) that shows these strains to be most similar to the BoNT/G serotypes.

The last cluster includes the seven BoNT/G strains in the AFLP tree. This plasmid encoded toxin gene was first identified from soil in Argentina (16). It is not clear if one of the strains in this study is from the Argentinian soil, however five strains are from human autopsy specimens from Switzerland (38, 39). Four of the five autopsy specimens cluster together, however the fifth (G193) is closer to G190 of unknown origin than to the others.

The results of the AFLP analysis supports previous AFLP analysis that showed that this technique could differentiate Group I and II *C. botulinum* strains (26). The current work extends those findings to include strains representative of Group III and IV. This analysis illustrates the relationship of the different genetic backgrounds in the *Clostridia* that house 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 then sequenced. Comparisons of the neurotoxin sequences generated from the 60 BoNT/A genes sequenced here, as well as six previously published BoNT/A genes, shows that four distinct groups of BoNT/A sequences were identified (Figure 3). Most of the BoNT/A producing strains in this study (54/60) show little sequence variation in the BoNT/A gene and are of the previously reported BoNT/A1 subtype (10, 47). Within this subtype, 37 of the strains share the identical sequence 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 GenBank accession AF300467 (27) that generate a truncated protein from a stop codon at amino acid 128. Four of the sequences (A148, A397, A404, A406) were identical to each other but differed from AF300467 by two single nucleotide polymorphisms (SNPs). The other two sequences (A408, A411) were identical to each other but different from the other four silent BoNT/B sequences by a SNP. The identification of these different silent BoNT/B gene sequences shows that there are more differences in clostridial strains than revealed by AFLP analysis, 16S rRNA and BoNT/A sequence analysis. There is some geographic correlation among these silent BoNT/B strains. Ten of the 16 silent BoNT/B strains were from a US collection and seven specifically map to a region stretching from Scottsbluff, Nebraska south through eastern Colorado to northeastern New Mexico. In contrast, none of the 18 BoNT/A1 strains from a European collection harbored silent BoNT/B genes (personal communication).

Besides the frequently occurring BoNT/A1 gene, three additional BoNT/A genes were

identified (Figure 3). Four strains produced the previously reported BoNT/A2 (29, 46). However, two previously unreported BoNT/A genes were also identified. One of these, termed BoNT/A3, was produced by a single strain (A253, also known as Loch Maree) which was isolated from a 1922 botulism outbreak in Scotland (30). 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-A4 subtypes are shown in Table 3. Nucleotide differences range from 3% to 8%, with amino acid differences ranging from 7% to 16%, 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 Figure 4 by isolate BoNT/A2 (Kyoto-F, GenBank accession 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 to 1146, and close to 99% identical to BoNT/A3 strain A254 (this paper) over positions 1147 to 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) (Figure 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 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 not 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 which intersect one another, is suggestive of more ancient recombination events.

Comparison of the 53 BoNT/B genes sequenced for this work and an additional seven previously published BoNT/B genes also showed four, possibly five, distinct groups. These represent the four previously described BoNT/B subtypes, BoNT/B1, BoNT/B2, bivalent BoNT/B, and nonproteolytic BoNT/B (Figure 5) (21, 24, 28, 36, 45). Compared to BoNT/A, each subtype had more members, with BoNT/B2 being produced by the largest number of strains. 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%-4%, with amino acid differences ranging from 4%-6%. A single BoNT/B (B506) was sequenced that differed from the closest BoNT/B2 strain by 33 nucleotides, which represents a 2% difference at the amino acid level. It is unknown whether this represents a new BoNT/B subtype (BoNT/B3), as the subtype delineation indicates a difference in immune recognition by monoclonal antibodies.

Comparison of the 21 BoNT/E genes sequenced for this work and an additional 15 published BoNT/E genes showed five distinct groups (Figure 6). The predominant group contains 17 strains, producing a previously described BoNT/E which we have termed BoNT/E1 (35). Two groups contain sequences from only *C. butyricum* BoNT/E strains. One of these groups contains 11 identical sequences from *C. butyricum* strains in China isolated from soil and several food borne cases of botulism (43). The other group contains the *C. butyricum* BoNT/E sequences from an infant botulism case in Italy (35). Toxins in these three groups differ by 3-5%

at the amino acid level and most likely represent distinct subtypes (Table 3). Four of the strains (E185, E540, E545, E549) are within one group which produces a previously unidentified BoNT/E we have termed BoNT/E3. Another two strains (E544, E546) also appear to represent a unique group of BoNT/E which we have termed BoNT/E2. Amino acid differences between BoNT/E1, E2, and E3 range from 1-3%. It in unknown whether this level of difference would represent new BoNT/E subtypes, as the subtype delineation indicates a difference in immune recognition by monoclonal antibodies.

A comparison of the BoNT nucleotide sequences from these strains and from available GenBank sequences representing all of the serotypes is shown in Figure 7. The tree 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 comprised of the C and D toxins. This relationship among the *C. botulinum* neurotoxin genes is different than the results obtained from the 16S rRNA gene and AFLP analysis (compare to Figures 1 and 2). The relationships among the group designations (I-IV) is also not maintained (compare to Figure 2), and this suggests that the toxin gene has 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 16S rRNA conserved gene sequences illustrate how the different serotypes are closely related to other *Clostridia* species, AFLP analysis is consistent with the 16s rRNA data but adds significant resolution to the genomic background that houses the different neurotoxin genes, and finally the diversity within and between the seven BoNT serotypes reveal a completely different phylogeny within this species that suggests intra and inter species 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. Since *C. butyricum* and *C baratii* strains have been identified that contain BoNT genes (11, 35), the taxonomy of the toxin producing *Clostridia* has become more complex. The tree generated from 16S rRNA gene sequence data suggests that the various botulinum neurotoxins that define the species, *Clostridium botulinum,* are actually housed in genomes from four different clostridial species. The 16S rRNA tree shows that BoNT/A and BoNT/B and BoNT/F producing strains are closely related to each other and evolved from a common ancestor. However, the genomes for the BoNT/C, D, E and G producing strains have 16S profiles that closely align to distant clostridial relatives *C. novyi*/*C. haemolyticum, C. baratii,* and *C. suberminale*/*C. argentinense,* respectively (Figure 1). These results should not change the basic nomenclature for *C. botulinum* to avoid confusion and because these taxonomic designations are based on strong phenotypic as well as genotypic characteristics. But the presence of related toxin genes in distant relatives of the

clostridia, serves as a reminder that horizontal gene transfer has played a significant role in the evolution of *Clostridium botulinum.*

The AFLP analysis of these strains illustrates clustering by group designation and by toxin serotype. The tree divides the strains into clusters which follow the Group I-IV designations, which are based on physiological characteristics. AFLP clearly separates the proteolytic and nonproteolytic types 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 tree 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 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 various 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 the Ba207 houses a completely new BoNT/A subtype 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 serotypes of BoNT (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 tendency has been reported by others (4). Of the ten strains isolated from infant, in this study, four are bivalent. These ten bivalent strains are located in different branches of the AFLP tree and include a *C. butyricum* BoNT/E producing strain (E543) from Italy. Examination of the sequences of the BoNT/A, B and E genes from these infant strains shows that the toxin gene frequently represents a unique cluster within the serotype. Within both the BoNT/A and the BoNT/B gene trees, 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, though, that the infant strains 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 (A-G) suggests

the BoNT gene has evolved separately in different genomic backgrounds. The tree 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 comprised of the C and D toxins. These relationships are different than the Group I-IV designations supported by the 16S rRNA and AFLP analysis. This discordant phylogeny suggests gene transfer among different clostridial species and *C. botulinum* has contributed to the movement of the BoNT gene into various genetic backgrounds. The nucleotide differences within these neurotoxin genes are a result of both natural variation and selection pressure. Recombination events, similar to that illustrated in Figure 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 which have previously been reported (32). Several recombination events within the NTNH genes of A1, B and F strains have also been noted (12).

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 ref. (37)). 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, which we have termed BoNT/A3 and BoNT/A4. Subtypes have historically been defined by the differential binding of monoclonal antibodies (14, 29, 37), and the 15% and 11% amino acid difference between BoNT/A1, A3, and A4 would certainly result in differential binding of some BoNT/A monoclonal antibodies (37). 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 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 would await 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 would be 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 trees or overall pairwise sequence distances. For example, BoNT/A2 represents a recombination of the 5' end of the BoNT/A1 heavy chain gene with the 3' end of the BoNT/A3 gene. This analysis permits identification of regions of the 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 would identify regions for generation of antibodies that would cross react with chimeric BoNT/C-D.

In conclusion, the toxins of *C. 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. Historically the expression of these toxins has been used to taxonomically identify these *Clostridia* as *C. botulinum*. The presence of these toxins in different genetic backgrounds has been observed and suggests their movement both within the species and between other species.

Most of these bacteria are found throughout the world and yet there is no 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 diversity observed in one of the most lethal neurotoxins known. The BoNTs produced by these clostridial species show dramatic sequence differences both within and between serotypes. Identifying the extent of these differences is the crucial first step in the development of detection devices, vaccines, and therapeutics.

Acknowledgements

Funding for this project was provided by the Department of Homeland Security, National Bioforensics Analysis Center and NIAID Cooperative Agreement U01 AI056493. We would also like to thank the DOE Joint Genome Institute (JGI) at Los Alamos National Laboratory for their support by providing technical assistance and facilities for DNA sequencing.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

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. 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:**207-11.
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:**773-82.
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:**1713-5.
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.
6. **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:**141-5.

7. **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.

8. **Control, C. f. D.** 1998. Botulism in the United States, 1899-1998. handbook for epidemiologists, clinicians,and laboratory workers. Atlanta, Georgia **U.S. Department of Health and Human Services, Public Health Service:**downloadable at http://www.bt.cdc.gov/agent/botulism/index,asp.

9. **Council, U. N. S.** 1995. Tenth report of the executive committee of the special commision 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 Commision. United Nations Security Council.

10. **Dineen, S. S., M. Bradshaw, and E. A. Johnson.** 2003. Neurotoxin gene clusters in Clostridium botulinum type A strains: sequence somparison and evolutionary implications. Curr Microbiol **46:**345-52.

11. **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:**262-8.

12. **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:**1105-12.

13. **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.

14. **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:**285-91.

15. **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.

16. **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.

17. **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:**654-5.

18. **Hatheway, C. L.** 1995. Botulism: the present status of the disease. Curr Top Microbiol Immunol **195:**55-75.

19. **Hatheway, C. L.** 1990. Toxigenic clostridia. Clin Microbiol Rev **3:**66-98.

20. **Hatheway, C. L., and J. L. Ferreira.** 1996. Detection and identification of Clostridium botulinum neurotoxins. Adv Exp Med Biol **391:**481-98.

21. **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:**101-10.

22. **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:**103-10.

23. **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 Van Leeuwenhoek **64:**273-83.

24. **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.

25. **Kaufman, L. and P. J. Rousseeuw.** 1990. Finding groups in data: an introduction to cluster analysis. John Wiley and Sons, Inc. New York, N.Y.

26. **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:**1148-54.

27. **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:**159-64.

28. **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:**4811-6.

29. **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:**767-74.

30. **Leighton, G.** 1923. Botulism and food preservation (The Loch Maree Tragedy). Collins & Sons, Glasgow.

31. **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:**201-2.

32. **Moriishi, K., M. Koura, N. Abe, N. Fujii, Y. Fujinaga, K. Inoue, and K. Ogumad.** 1996. Mosaic structures of neurotoxins produced from Clostridium botulinum types C and D organisms. Biochim Biophys Acta **1307:**123-126.

33. **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:**662-7.

34. **Popoff, M., and J.-C. Marvaud.** 1999. Structural and genomic features of clostridial neurotoxins, p. 202-228. *In* A. JE and F. JH (ed.), The Comprehensive Sourcebook of Bacterial Protein Toxins. Academic Press, London.

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:**107-13.

36. **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:**312-8.

37. **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.

38. **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:**22-7.

39. **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 **1:**237-41.

40. **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:**253-8.

41. **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 Clsotridium hastiforme. Int. J. Syst. bacteriol. **38:**375-381.

42. **Ticknor, L. O., A. B. Kolsto, 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:**4863-73.

43. **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:**4992-7.

44. **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:**964-5.

45. **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:**2345-54.

46. **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:**547-56.

47. **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.

Figure Legends

Figure 1. Phylogenetic tree of *Clostidium* species based on 16S rRNA gene.

Neighbor-joining tree of 54 sequences from GenBank accessions and 36 sequences representative of the strains from this collection, that illustrates the genetic diversity within the *Clostiridia*. *C. botulinum* strains cluster into four distinct groups that follow the Group I-IV designation historically based on physiological characteristics. These groups are interspersed among the 27 other *Clostridia* species in the tree. The tree was constructed using an alignment of 16S rRNA sequences which contained 1329 bases after removal of columns containing more than 80% gap characters, and includes sequences from bivalent, nonproteolytic and proteolytic toxin producing strains.

Figure 2. AFLP-based tree of 174 *C. botulinum* strains.

DNA fragments generated from the restriction enzyme digestion of each of the strain DNAs were ligated to linkers and selectively amplified. Forty DNA fragments generated by AFLP experiments were used as a fingerprint to represent each of the strains. If forty 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 1) and nonproteolytic strains (Groups II, III, and IV) and distinct branches representing Groups I-IV. The AFLP groups are also generally distinct toxin serotypes. The distance measure or genetic distance is the proportion of fragments that two samples do not have in common.

Figure 3. Comparison of BoNT/A gene sequences.

The full length coding region of the BoNT/A gene in 60 strains and 6 GenBank sequences were aligned. Four distinct subtypes are visible. Most strains (54) are of the BoNT/A1 subtype with four 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-which show significant sequence variation compared to BoNT/A1 and A2 subtypes.

Figure 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 subtype of Kyoto-F GenBank accession number X73423. This plot illustrates that the BoNT/A2 subtype is close to 99% identical to the BoNT/A1 subtype (A142) over nucleotide positions 1 to 1146, and close to 99% identical to the BoNT/A3 subtype (A254) over positions 1147 to 3450 showing the BoNT/A2 subtype is a result to a recombination event between BoNT/A1 and BoNT/A3 lineages of gene sequences.

Figure 5. Comparison of BoNT/B gene sequences.

The full length coding regions of the BoNT/B gene in 53 strains and 7 GenBank sequences were aligned. Four distinct clusters that include the BoNT/B1 and BoNT/B2, bivalent (Ab149, Ba207,Bf258, Bf698) and nonproteolytic BoNT/B subtypes are visible. Most strains are of the BoNT/B2 subtype with 16 of the BoNT/B1 subtype. Strain B506 is separate from the other BoNT/B2 strains and represents newly identified variation in this serotype.

Figure 6. Comparison of BoNT/E gene sequences.

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

Figure 7. Comparison of the seven different serotype BoNT gene sequences.

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 seen by 16S rRNA or AFLP analysis. Nonproteolytic and

bivalent strains (Ba207 and Ab149) as well as representatives of the various subtypes are included.

Table 1. *C. botulinum* strains analyzed.

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

F658	F	Langeland	G193	G	2738	G196	G	2740
			G194	G	1353	G197	G	2741
G190	G	5/18/78	G195	G	2739	G198	G	2742

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

Primer ID	Type	Sequence	Location*
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

*Location of primer sequence of BoNT A within GenBank accession X73423
Location of primer sequence of BoNT B within GenBank accession X71343
Location of primer sequence of BoNT E within GenBank accession X62683

Table 3. Nucleotide and amino acid identities in strains representing BoNT/A, B and E subtypes[1]. The coding sequence for a representative of each different subtype of BoN/A, B and E was compared to determine overall homology at both the nucleic acid and amino acid level; npB (nonproteolytic B), bvB (bivalent B), It butyr (Italian *C. butyricum* strain BL5262), Ch butyr (Chinese *C. butyricum* strain LCL155).

BoNT/A	A1 (A150 Hall)	A2 (Ab149 CDC 1436)	A3 (A254 Loch Maree)	A4 (Ba207 strain 657)	
A1	-----	95 %/90 %	92 %/85 %	94 %/89 %	
A2		-----	97 %/93 %	94 %/88 %	
A3			-----	92 %/84 %	
A4				-----	
BoNT/B	B1 (B155 okra)	B2 (B162 213B)	B3 (B506 CDC 795)	NpB[2] (B257 Eklund 17B)	BvB[3] (Ba207 strain 657)
B1	-----	98 %/96 %	98 %/96 %	96 %/93 %	98 %/96 %
B2		-----	99 %/98 %	96 %/94 %	97 %/95 %
B3			-----	96 %/94 %	98 %/96 %
npB[2]				-----	96 %/93 %
bvB[3]					-----
BoNT/E	E1 (E213Beluga)	E2 (E544 CDC5247)	E3 (E185 AlaskaE43)	E It butyr[4] (E543 BL5262)	E Ch butyr[5] (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 %
E Ch butyr					-----

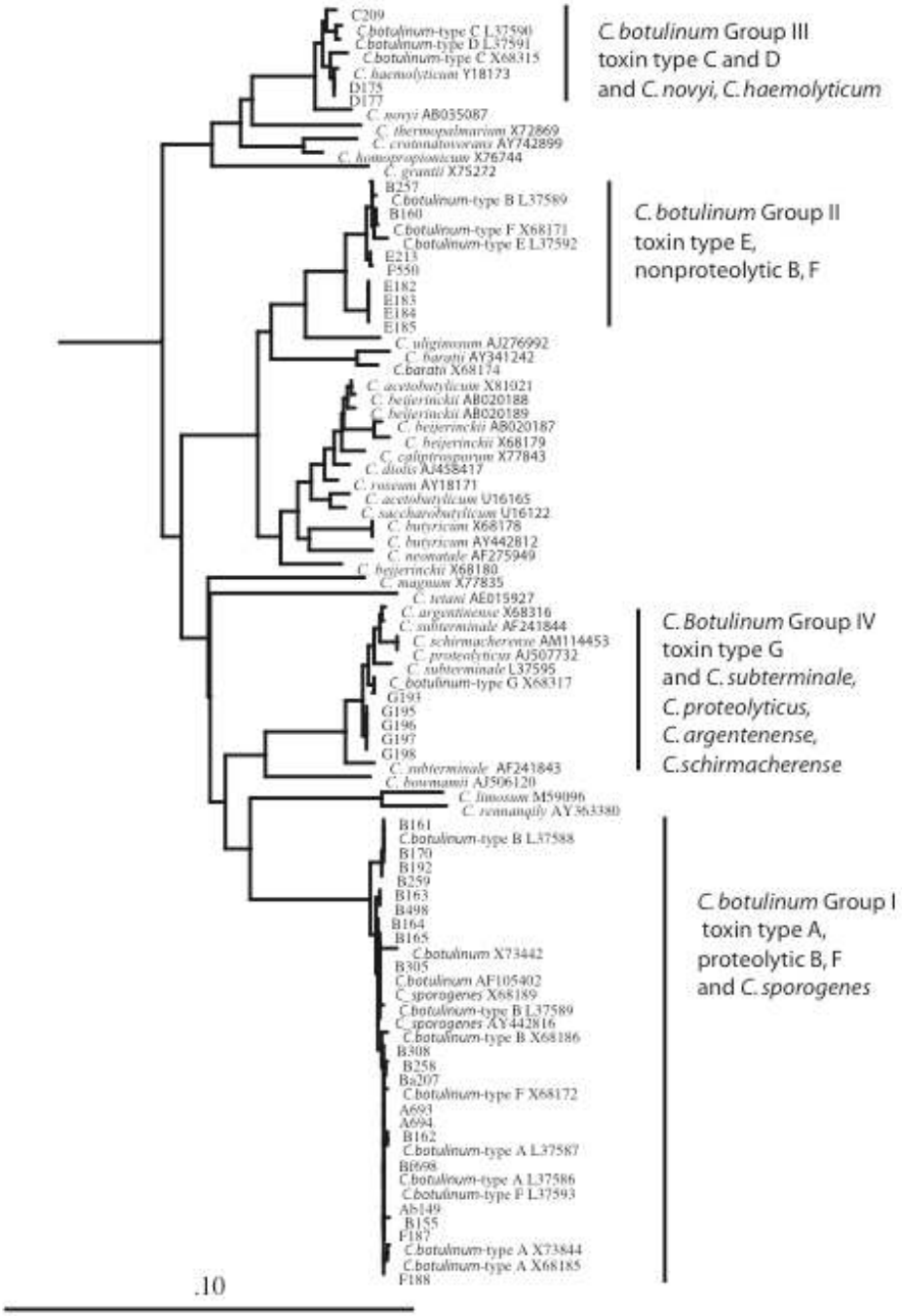
[1]Identity at DNA level / identity at amino acid level

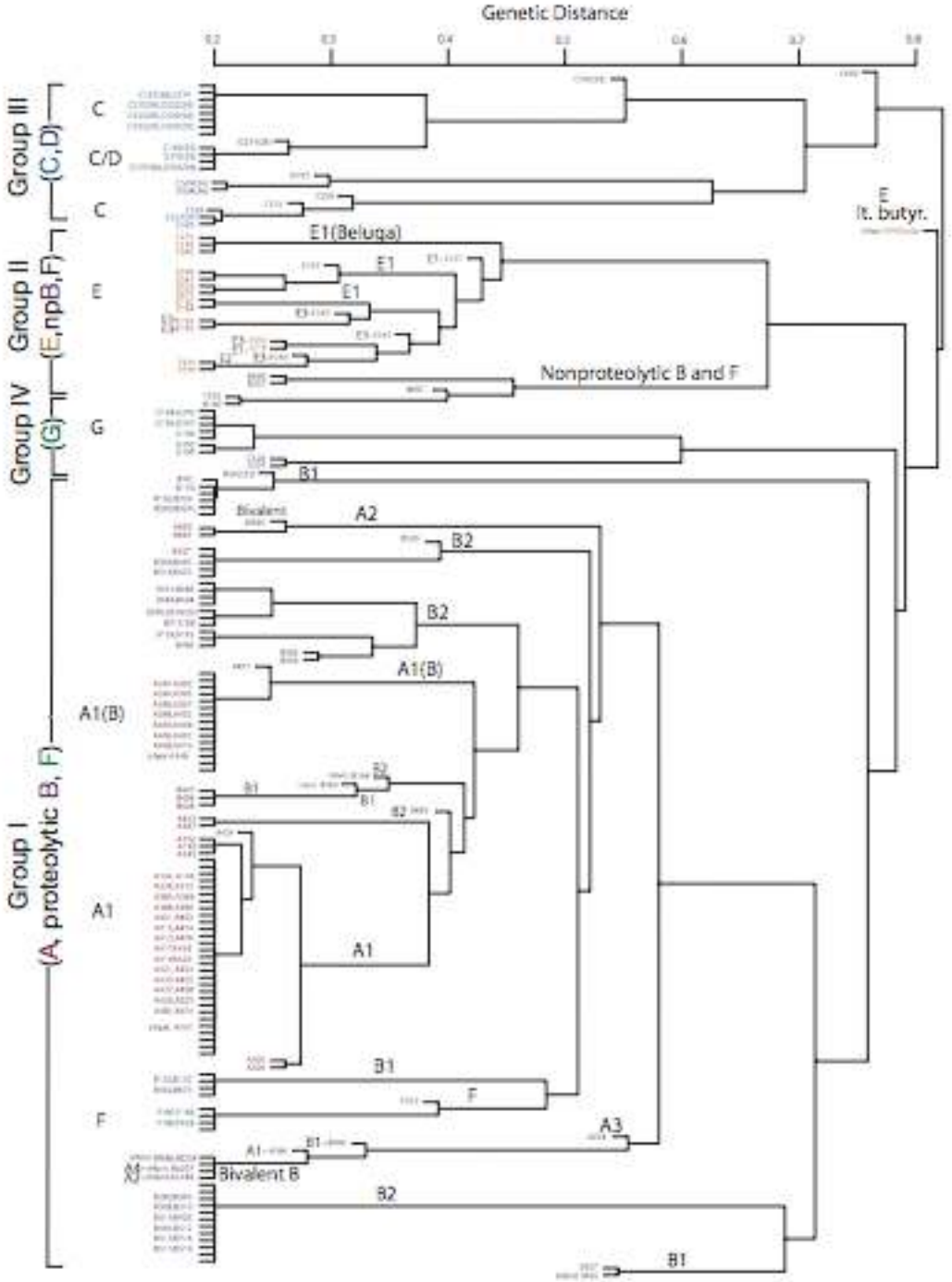
[2]npB (nonproteolytic B)

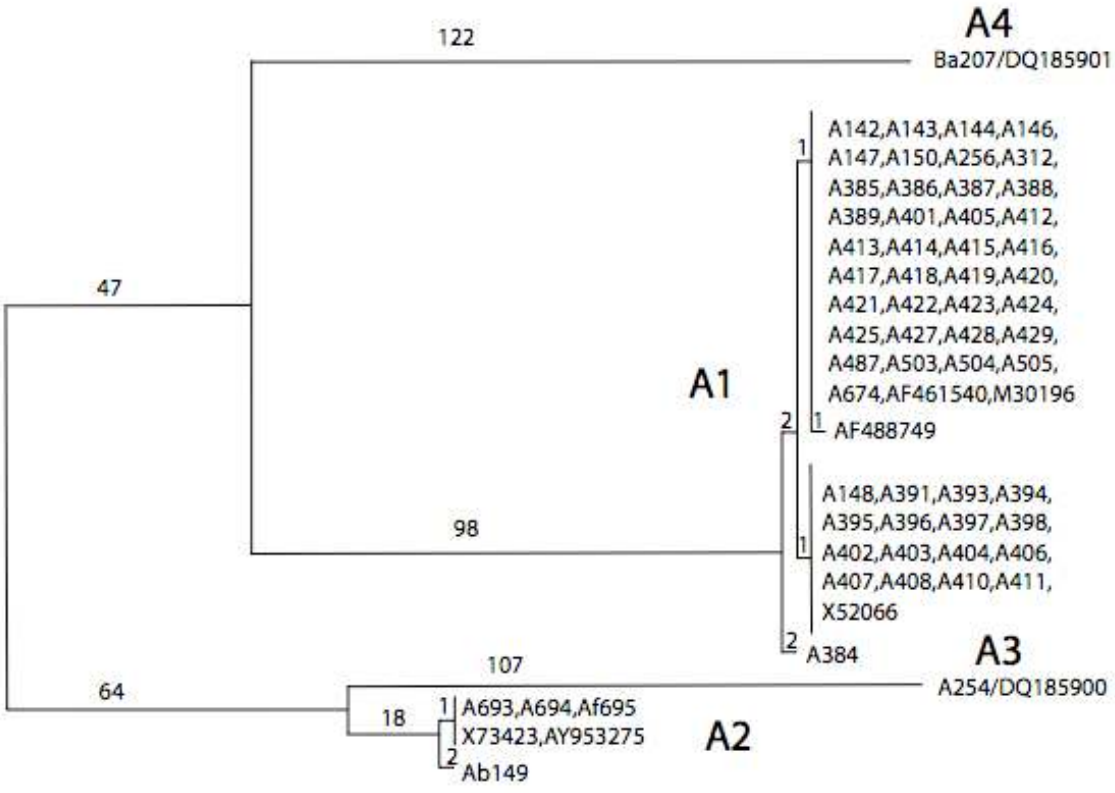
[3]bvB (bivalent B)

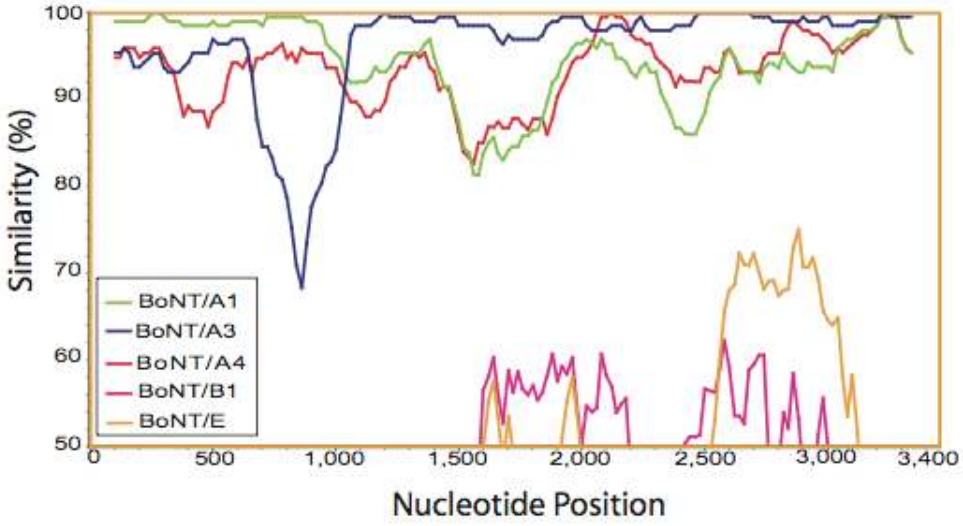
[4]It butyr (Italian *C. butyricum* strain BL5262)

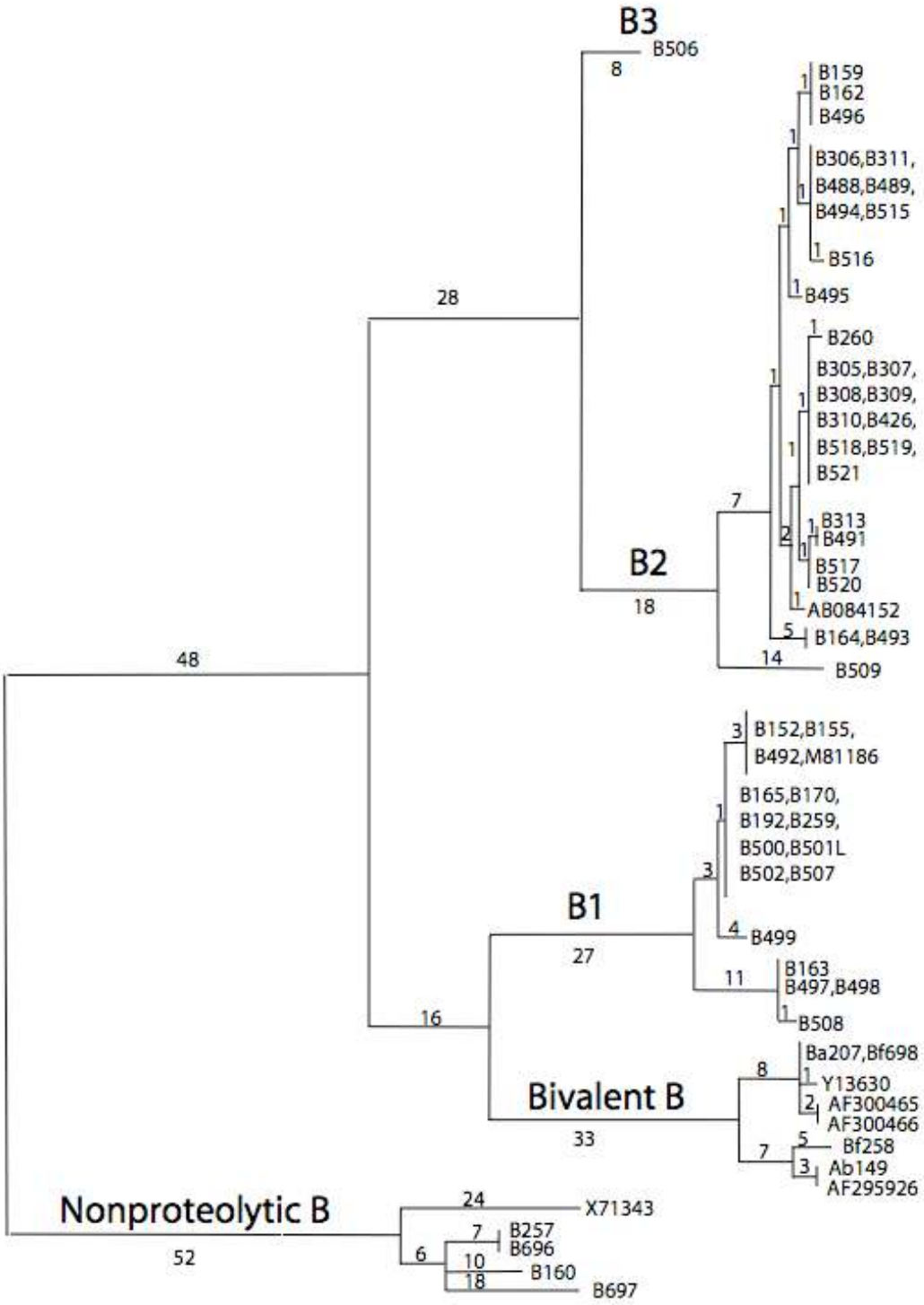
[5]Ch butyr (Chinese *C. butyricum* strain LCL155).

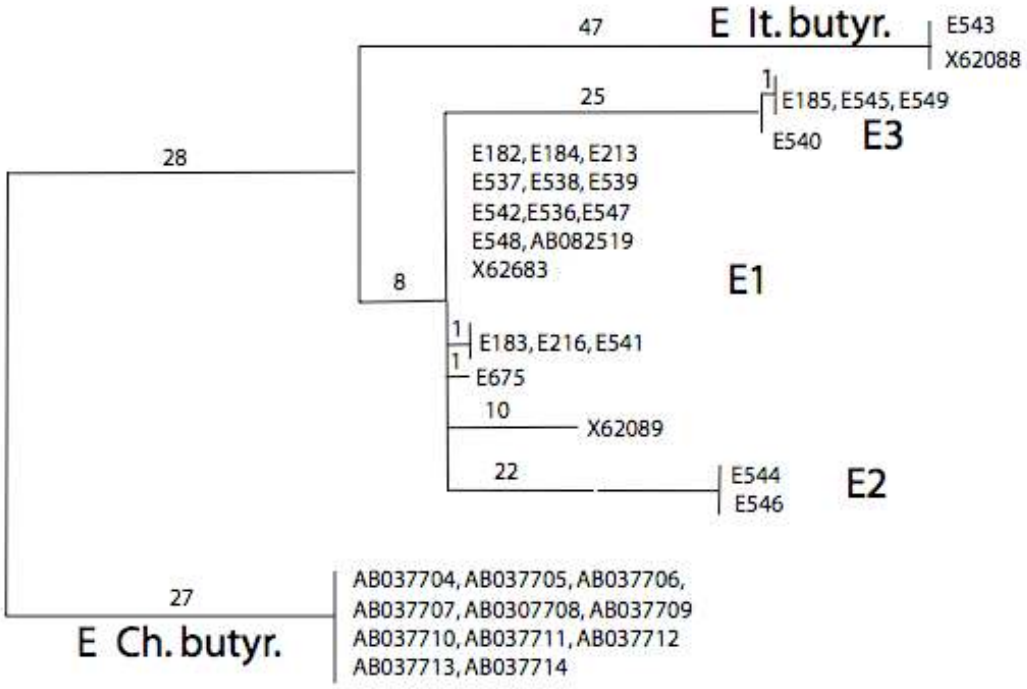

C209
C.botulinum-type C L37590
C.botulinum-type D L37591
C.botulinum-type C X68315
C. haemolyticum Y18173
D175
D177
C. novyi AB035087
C. thermopalmarium X72869
C. cratonatoverans AY742899
C. homopropionicum X76744
C. grantii X75272
B257
C.botulinum-type B L37589
B160
C.botulinum-type F X68171
C.botulinum-type E L37592
E213
F550
E182
E183
E184
E185
C. uligimosum AJ276992
C. baratii AY341242
C.baratii X68174
C. acetobutylicum X81021
C. beijerinckii AB020188
C. beijerinckii AB020189
C. beijerinckii AB020187
C. beijerinckii X68179
C. calniroseparum X77843
C. diolis AJ458417
C. rasetti AY18171
C. acetobutylicum U16165
C. saccharobutylicum U16122
C. butyricum X68178
C. butyricum AY442812
C. neonatale AF275949
C. beijerinckii X68180
C. magnum X77835
C. tetani AE015927
C. argentinense X68316
C. subterminale AF241844
C. schirmacherense AM114453
C. proteolyticus AJ507732
C. subterminale L37595
C.botulinum-type G X68317
G193
G195
G196
G197
G198
C. subterminale AF241843
C. bowmanii AJ506120
C. limosum M59096
C. sennangily AY363380
B161
C.botulinum-type B L37588
B170
B192
B259
B163
B498
B164
B165
C.botulinum X73442
B305
C.botulinum AF105402
C. sporogenes X68189
C.botulinum-type B L37589
C. sporogenes AY442816
C.botulinum-type B X68186
B308
B258
Ba207
C.botulinum-type F X68172
A693
A694
B162
C.botulinum-type A L37587
B698
C.botulinum-type A L37586
C.botulinum-type F L37593
Ab149
B155
F187
C.botulinum-type A X73844
C.botulinum-type A X68185
F188
.10
C. botulinum Group III
toxin type C and D
and C. novyi, C. haemolyticum
C. botulinum Group II
toxin type E,
nonproteolytic B, F
C. Botulinum Group IV
toxin type G
and C. subterminale,
C. proteolyticus,
C. argentenense,
C.schirmacherense
C. botulinum Group I
toxin type A,
proteolytic B, F
and C. sporogenes

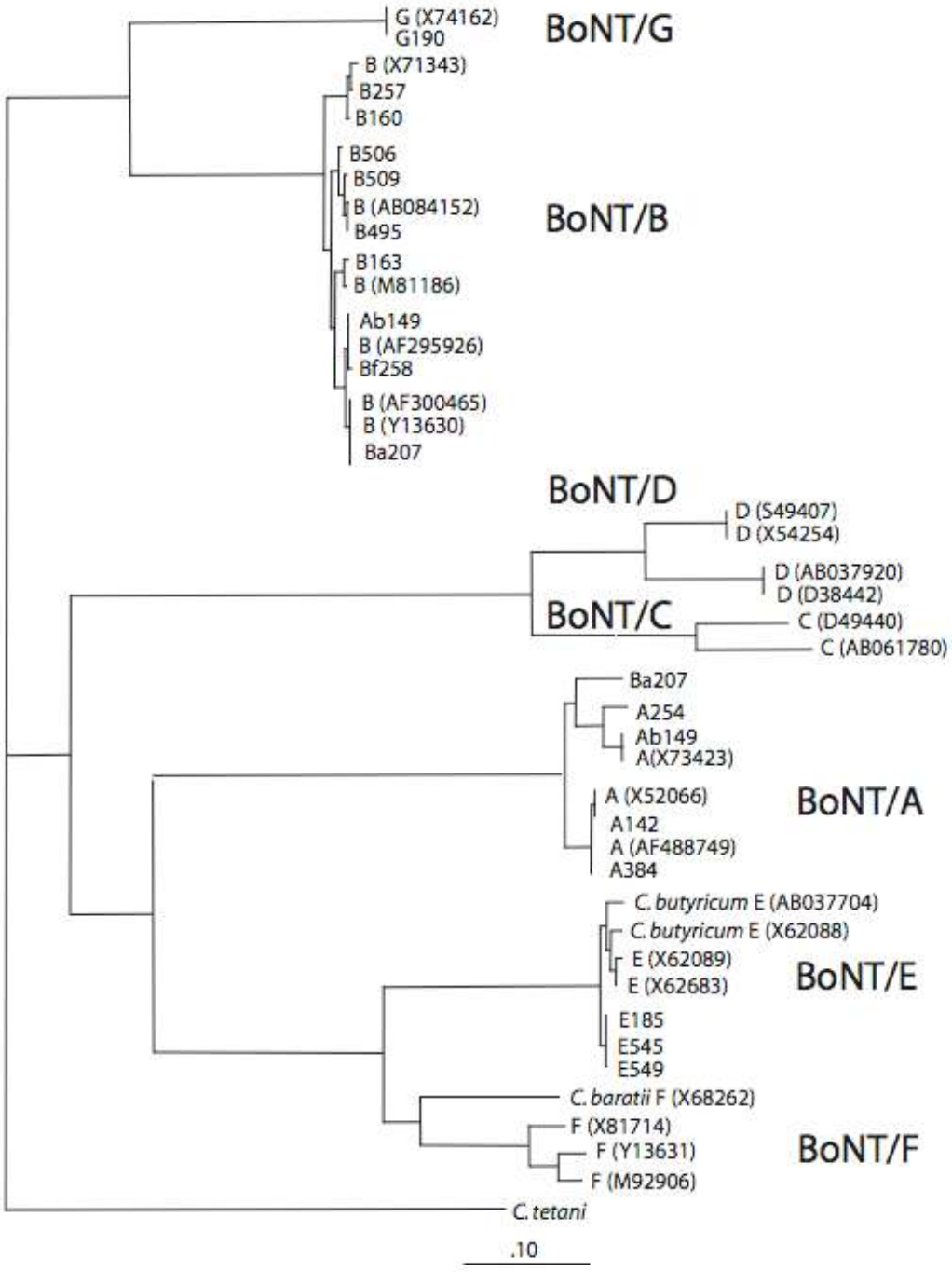

Genetic Distance
Group III
(C,D)
Group II
(E,npB,F)
Group IV
(G)
Group I
(A, proteolytic B, F)
C
C/D
C
E
G
A1(B)
A1
F
M
E1(Beluga)
E1
E1
Nonproteolytic B and F
B1
Bivalent
A2
B2
B2
A1(B)
B1
B1
B2
A1
B1
F
A3
A1
B1
Bivalent B
B2
B1
E
It. butyr.


A4
Ba207/DQ185901
122
47
A142,A143,A144,A146,
A147,A150,A256,A312,
A385,A386,A387,A388,
A389,A401,A405,A412,
A413,A414,A415,A416,
A417,A418,A419,A420,
A421,A422,A423,A424,
A425,A427,A428,A429,
A487,A503,A504,A505,
A674,AF461540,M30196
AF488749
A1
98
A148,A391,A393,A394,
A395,A396,A397,A398,
A402,A403,A404,A406,
A407,A408,A410,A411,
X52066
A384
A3
A254/DQ185900
107
64
18
A693,A694,Af695
X73423,AY953275
Ab149
A2


100
90
80
70
60
50
Similarity (%)
BoNT/A1
BoNT/A3
BoNT/A4
BoNT/B1
BoNT/E
0
500
1,000
1,500
2,000
2,500
3,000
3,400
Nucleotide Position


B3
B506
8
B159
B162
B496
B306,B311,
B488,B489,
B494,B515
B516
B495
28
B260
B305,B307,
B308,B309,
B310,B426,
B518,B519,
B521
7
B313
B491
B517
B520
B2
18
AB084152
B164,B493
5
14
B509
48
B152,B155,
B492,M81186
B165,B170,
B192,B259,
B500,B501L
B502,B507
B1
27
B499
B163
B497,B498
11
B508
16
Ba207,Bf698
Y13630
AF300465
AF300466
Bivalent B
33
Bf258
Ab149
AF295926
Nonproteolytic B
52
24
X71343
B257
B696
6
10
B160
18
B697


47
E It. butyr.
E543
X62088
1
25
E185, E545, E549
E540
E3
28
E182, E184, E213
E537, E538, E539
E542,E536,E547
E548, AB082519
X62683
E1
8
1
E183, E216, E541
1
E675
10
X62089
22
E544
E546
E2
27
E Ch. butyr.
AB037704, AB037705, AB037706,
AB037707, AB0307708, AB037709
AB037710, AB037711, AB037712
AB037713, AB037714


G (X74162)
G190
BoNT/G
B (X71343)
B257
B160
B506
B509
B (AB084152)
B495
BoNT/B
B163
B (M81186)
Ab149
B (AF295926)
Bf258
B (AF300465)
B (Y13630)
Ba207
BoNT/D
D (S49407)
D (X54254)
D (AB037920)
D (D38442)
BoNT/C
C (D49440)
C (AB061780)
Ba207
A254
Ab149
A(X73423)
A (X52066)
BoNT/A
A142
A (AF488749)
A384
C. butyricum E (AB037704)
C. butyricum E (X62088)
E (X62089)
BoNT/E
E (X62683)
E185
E545
E549
C. baratii F (X68262)
F (X81714)
F (Y13631)
BoNT/F
F (M92906)
C. tetani
.10