Molecular Characterization of a Novel Botulinum Neurotoxin Type H Gene

Nir Dover,1 Jason R. Barash,1 Karen K. Hill,2 Gary Xie,2 and Stephen S. Arnon1

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

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

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

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

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

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

METHODS

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

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

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

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

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

RESULTS

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

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

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

A sequence identity comparison of BoNT/H to all known
BoNT toxin types with 1 representative from each further dis-
tinguished the novel type H sequence (Figure 2). The 7 known
BoNT types had nucleotide and amino acid sequence identities
of 53.9%–75.1% and 32.6%–64.1%, respectively (Figure 2).
Table 1. Comparison of Nucleotide and Amino Acid Pairwise Sequence Identities of Botulinum Neurotoxin (BoNT) Type H to Known BoNT Subtypes

<table>
<thead>
<tr>
<th>BoNT Subtype</th>
<th>Nucleotides</th>
<th>Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>66.0</td>
<td>50.9</td>
</tr>
<tr>
<td>A2</td>
<td>65.2</td>
<td>49.9</td>
</tr>
<tr>
<td>A3</td>
<td>66.7</td>
<td>49.1</td>
</tr>
<tr>
<td>A4</td>
<td>66.0</td>
<td>49.7</td>
</tr>
<tr>
<td>A5</td>
<td>65.5</td>
<td>50.3</td>
</tr>
<tr>
<td>B1</td>
<td>59.9</td>
<td>40.6</td>
</tr>
<tr>
<td>B2</td>
<td>60.5</td>
<td>41.3</td>
</tr>
<tr>
<td>B3</td>
<td>60.2</td>
<td>41.3</td>
</tr>
<tr>
<td>B4</td>
<td>59.4</td>
<td>41.7</td>
</tr>
<tr>
<td>B5</td>
<td>60.4</td>
<td>41.4</td>
</tr>
<tr>
<td>B6</td>
<td>60.2</td>
<td>41.2</td>
</tr>
<tr>
<td>B7</td>
<td>60.6</td>
<td>41.5</td>
</tr>
<tr>
<td>C</td>
<td>53.7</td>
<td>32.8</td>
</tr>
<tr>
<td>C/D</td>
<td>56.1</td>
<td>32.3</td>
</tr>
<tr>
<td>D</td>
<td>56.9</td>
<td>33.8</td>
</tr>
<tr>
<td>D/C</td>
<td>54.6</td>
<td>33.3</td>
</tr>
<tr>
<td>E1</td>
<td>65.6</td>
<td>48.7</td>
</tr>
<tr>
<td>E2</td>
<td>65.8</td>
<td>48.7</td>
</tr>
<tr>
<td>E3</td>
<td>65.5</td>
<td>48.5</td>
</tr>
<tr>
<td>E4</td>
<td>65.5</td>
<td>48.4</td>
</tr>
<tr>
<td>E5</td>
<td>65.6</td>
<td>48.4</td>
</tr>
<tr>
<td>E6</td>
<td>65.6</td>
<td>48.8</td>
</tr>
<tr>
<td>E7</td>
<td>65.3</td>
<td>48.7</td>
</tr>
<tr>
<td>E8</td>
<td>65.6</td>
<td>48.7</td>
</tr>
<tr>
<td>E9</td>
<td>66.2</td>
<td>48.9</td>
</tr>
<tr>
<td>F1</td>
<td>69.0</td>
<td>52.8</td>
</tr>
<tr>
<td>F2</td>
<td>67.1</td>
<td>50.5</td>
</tr>
<tr>
<td>F3</td>
<td>67.7</td>
<td>50.3</td>
</tr>
<tr>
<td>F4</td>
<td>67.9</td>
<td>51.7</td>
</tr>
<tr>
<td>F5</td>
<td>76.4</td>
<td>63.5</td>
</tr>
<tr>
<td>F6</td>
<td>67.7</td>
<td>50.8</td>
</tr>
<tr>
<td>F7</td>
<td>67.6</td>
<td>52.5</td>
</tr>
<tr>
<td>G</td>
<td>59.8</td>
<td>40.6</td>
</tr>
</tbody>
</table>

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

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

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

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

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

The BoNT/H predicted heavy chain amino acid sequence was further divided into its translocation and binding domains and compared to the homologue domains of BoNT types A and F (Figure 5B1 and 5B2, respectively). The translocation domain of BoNT/H had a novel sequence that was more closely
Figure 1. Phylogenetic analysis of nucleotide sequences representing all 7 bont serotypes and their subtypes. One representative of each toxin subtype is shown. The phylogenetic tree was generated using the neighbor-joining method and the Kimura 2-parameter model. Genetic distance and bootstrap values are shown. Note that the novel bont/H creates a distinct lineage separated from the 7 bont/A-G types. Botulinum toxin subtypes, and strains and GenBank accession numbers used in the analysis, are as follows: A1 = NCTC2916, X52086; A2 = Kyoto, CP001581; A3 = Loch Maree, CP000963; A4 = 657Ba, CP001081; A5 = IBCA94-0216, FJ959094; B1 = Okra, CP000940; B2 = Prevot 25 NCASE, EF033129; B3 = CDC795, EF028400; B4 = Eklund 17B, CP001057; B5 = 657Ba, CP001081; B6 = Osaka05, AB302852; B7 = NCTC3807, JN120760; C = Strain 573, AB200359; C/D = Strain S19, FN436022; D = BVD/−3, X54254; D/C = Strain OFD17, AB461921; E1 = NCTC 11219, X62683; E2 = CDC5257, EF028404; E3 = Alaska E43, CP001078; E4 = ATCC 43755, X62088; E5 = LCL 095, AB037714; E6 = K35, AM695752; E7 = IBCA87-0192, JN695729; E8 = E134, JN695730; E9 = CCDC56177, JX424534; F1 = Langeland, CP000728; F2 = CDC5028, JY13209; F3 = CDC5086, GU213218; F4 = CDC5086, GU213218; F5 = CDC5086, GU213218; F6 = IBCA66-5436, HJ441176; F7 = CDC5086, GU213218; and G = NCFS 3012, X74162.
related to the translocation domain of BoNT/F than to the translocation domain of BoNT/A (Figure 5B1). In contrast, the binding domain of BoNT/H was more closely related to the binding domain of BoNT/A than to the binding domain of BoNT/F (Figure 5B2).

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

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

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

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

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

The \( \text{bont/H} \) Toxin Gene Cluster and Its Genomic Location

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

Supplementary Table 1 compares the pairwise identities of nucleotide and predicted amino acid sequences of the genes of the \( \text{bont/H} \) orfX toxin gene cluster to those of the types A and F orfX gene clusters available in GenBank. The \( \text{ntnh} \) of the \( \text{bont/H} \) toxin gene cluster was most similar to the \( \text{ntnh} \) of the \( \text{bont/F1} \) toxin gene cluster (95.8% and 92.8% nucleotide and amino acid identity, respectively), while the \( \text{p47} \), \( \text{botR} \), \( \text{orfX1} \), \( \text{orfX2} \), and \( \text{orfX3} \) genes were most similar to their homologue genes in the type A1 toxin gene cluster in strain NCTC2916 (Supplementary Table 1).
BLAST search upstream of the bont/H toxin gene cluster (ie, upstream of orfX3) identified overall synteny and approximately 93% sequence homology to the upstream sequence of the bont/A4 toxin gene cluster in plasmid pCLJ that resides in C. botulinum strain 657 type Ba (GenBank accession no. CP001081). This homology extended approximately 31 kb upstream, where a putative gene coding for an IS110 transposase was located. Interestingly, a second copy of the same putative transposase-coding gene (99.8% nucleic acid identity) was located 1570 bp downstream of bont/H (Figure 6B). Alignment of the sequences upstream and downstream of these 2 IS110 transposases with the genome sequence of strain ATCC 3502 revealed that the bont/H toxin gene cluster is part of an insert of 54 464 bp (flanked by the 2 IS110 transposases) located in the chromosome between nucleotides 2 308 161 and 2 307 116 (strain ATCC 3502 numbering), in place of deleted gene cbo2164 (Figure 6B).

### The bont/B Neurotoxin Gene Cluster

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

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

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

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

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

---

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

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>---</td>
<td>90.9</td>
<td>91.2</td>
<td>96.2</td>
<td>80.2</td>
<td>93.2</td>
<td>82.0</td>
<td>69.0</td>
</tr>
<tr>
<td>F2</td>
<td>83.4</td>
<td>---</td>
<td>98.9</td>
<td>91.8</td>
<td>82.9</td>
<td>95.2</td>
<td>79.7</td>
<td>67.1</td>
</tr>
<tr>
<td>F3</td>
<td>83.9</td>
<td>97.1</td>
<td>---</td>
<td>91.8</td>
<td>83.0</td>
<td>95.1</td>
<td>79.4</td>
<td>67.7</td>
</tr>
<tr>
<td>F4</td>
<td>92.5</td>
<td>83.5</td>
<td>83.8</td>
<td>---</td>
<td>81.2</td>
<td>93.7</td>
<td>81.1</td>
<td>67.9</td>
</tr>
<tr>
<td>F5</td>
<td>69.8</td>
<td>74.4</td>
<td>74.0</td>
<td>69.7</td>
<td>---</td>
<td>83.1</td>
<td>74.5</td>
<td>76.4</td>
</tr>
<tr>
<td>F6</td>
<td>87.7</td>
<td>90.2</td>
<td>90.1</td>
<td>87.1</td>
<td>74.1</td>
<td>---</td>
<td>80.0</td>
<td>67.7</td>
</tr>
<tr>
<td>F7</td>
<td>73.7</td>
<td>68.9</td>
<td>69.3</td>
<td>72.2</td>
<td>63.9</td>
<td>69.9</td>
<td>---</td>
<td>67.6</td>
</tr>
<tr>
<td>H</td>
<td>52.8</td>
<td>50.5</td>
<td>50.3</td>
<td>51.7</td>
<td>63.5</td>
<td>50.8</td>
<td>52.5</td>
<td>---</td>
</tr>
</tbody>
</table>
gene sequence identified it as belonging to *C. botulinum* group I. The variable region of flaA (flaVR) may serve for further genotyping of *C. botulinum* [28, 29]. The IBCA10-7060 flaVR sequences were identical to type 1 flaVR, a group type that contains multiple type A, B, and AB strains [29].

**DISCUSSION**

*C. botulinum* strain IBCA10-7060 was found to produce BoNT type B and the novel BoNT type H [20]. We used a combination of Sanger and next-generation sequencing methods to
identify and characterize both toxin gene clusters and found that one was bont/B2. The second toxin gene cluster contained an overlap nucleotide sequence that did not match any of the 7 known bont types (ie, A–G). Thus, our sequence characterization provides a molecular explanation for the inability of either a botulinum antitoxins ABF mixture or a US Army heptavalent F(ab′)2 botulinum antitoxin A–G to neutralize IBCA10-7060 culture filtrate [20].

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

The possible origins of bont/H are uncertain. bont/H did not result from a simple recombination event between bont/F5 and bont/A because the approximately middle third of bont/H is <78% identical to the homologous section of any known bont/A–G nucleotide sequence (Figure 4). The 3′ one-third terminus of bont/H is closely related to the various toxin type A subtypes, with the percentage of sequence identity highest (96.0%–97.4%) for the 764 nucleotides at the 3′ terminus itself (Figure 4 and 5B2). We speculate that bont/H resulted from either: (1) multiple recombination events between a bont/F5-like gene, an as
yet-undiscovered second bont gene, and a third bont/A gene, each contributing approximately one-third of its gene sequence; or (2) a primordial bont gene that also contributed its light chain coding sequence to bont/F5 and its 3′ one-third terminus (HC) to bont/A.

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

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

The bont/H toxin gene cluster had a unique chromosomal location different from those used by other chromosomally
located toxin gene clusters [3, 10, 19]. This novel chromosomal location was between the genes cbo2163 and cbo2165, where an insertion event appears to have deleted gene cbo2164 and replaced it with the bont/H toxin gene cluster and other genetic material between nucleotides 2,308,161 and 2,307,116 (strain ATCC 3502; Figure 6B). cbo2164 codes for an AraC family transcriptional regulator, and its sequence is conserved among group I C. botulinum strains. We do not know how the lack of cbo2164 affects the strain IBCA10-7060 cells. In contrast, all other chromosomally located botulinum toxin gene clusters (including the bont/B2 toxin gene cluster of strain IBCA10-7060) in C. botulinum group I strains are found within the arsC, oppA/brnQ, and pulE loci and in group II strains within the rara locus [3, 10, 19].

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

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

### Supplementary Data

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

### Notes

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

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

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

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

### References


