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

ALISON K. EAST,* MANJU BHANDARI, JOANNA M. STACEY,
KATHRYN D. CAMPBELL, AND MATTHEW D. COLLINS

Department of Microbiology, Institute of Food Research, Reading,
Berkshire RG6 6BZ, United Kingdom

The cluster of genes encoding components of the botulinum neurotoxin (BoNT) complex was mapped in proteolytic (group I) *Clostridium botulinum* strains encoding BoNT types A, B, and F. Two different arrangements of genes were found: type A strain 62A and type B strain NCTC 7273 have similar organizations of genes encoding BoNT, the nontoxic nonhemagglutinin component (NTNH), hemagglutinin components, and P-21; type F strain Langeland has genes encoding BoNT, NTNH, and P-21, and a previously unidentified open reading frame encoding a protein of 416 amino acids. A group of type A strains typified by infant strain Kyoto-F, which is unlike type A strain 62A, lacks genes for hemagglutinin components and exhibits an organization similar to that of type F. Sequencing and pairwise analysis revealed the presence of possible chimeric sequences in some NTNH genes of proteolytic *C. botulinum*. Discordance in genealogical trees derived from different regions of the NTNH genes was observed which could be symptomatic of recombination and which may indicate that the NTNH gene represents a hot spot for such events within the cluster of genes encoding the BoNT complex. It is also evident that the phylogenetics of the NTNH gene, which is linked to the gene encoding BoNT, does not mirror the evolutionary history of the BoNT, upon which the *C. botulinum* species complex is defined and subdivided.

Strains of the species *Clostridium botulinum* produce the extremely toxic botulinum neurotoxin (BoNT). BoNT is the causative agent of the severe neuroparalytic illness in humans and animals referred to as botulism. BoNTs are classified into seven different types (A through to G) by reaction with specific antisera. *C. botulinum* strains are traditionally divided into four physiological groups, which are designated I to IV (18). Proteolytic group I *C. botulinum* strains produce a single toxin or, occasionally, dual toxins of type A, B, or F. Nonproteolytic group II *C. botulinum* strains invariably produce a single toxin of type B, E, or F. Group III strains produce BoNT type C or BoNT/D and are responsible for animal botulism, while group IV strains produce BoNT/G. Strains of groups I and II, especially those of types A and B, are important agents of both food-borne and infant botulism (18). In addition, some strains of the non-*C. botulinum* species are known to synthesize BoNT (viz., *Clostridium barati* [type F] and *Clostridium butyricum* [type E]). Recently, the evolutionary interrelationships of *C. botulinum* types A to G, *C. barati* type F, and *C. butyricum* type E and their BoNT genes have been systematically investigated. 16S rRNA sequencing revealed the presence of four distinct phylogenetic lines within *C. botulinum* which correlate with the four physiological groups (I to IV) (22). The same studies revealed that neurotoxicogenic *C. barati* and *C. butyricum* strains are authentic members of their respective species and are genealogically distinct from each other and the four *C. botuli-*

num lines (22). However, major differences in the pattern of relationships inferred from 16S rRNA and BoNT gene sequences were observed that were indicative of BoNT gene transfer between *C. botulinum* rRNA lines and the non-*C. botulinum* species (3, 10, 20).

BoNT is produced by *C. botulinum* as a noncovalently bound complex of two or more protein components. The smallest complex, the M complex (ca. 300 kDa), which is found in all strains (except type G) which produce neurotoxin, consists of BoNT with a component that has a similar size (ca. 150 kDa), the nontoxic nonhemagglutinin component (NTNH) (33). The size of the complex formed by *C. botulinum* shows some correlation with the toxin type encoded; strains encoding types E and F (26) and some type A strains (30) produce only the M complex, while strains encoding types A, B, C, and D, in addition, produce L complexes (ca. 500 kDa), which have associated hemagglutinin (HA) activity (28). Type G strains produce only the L complex (25). Many type A strains produce not only the M and L complexes, but also a larger LL complex (ca. 900 kDa). The purported function of the nontoxic components of the complex is to increase the stability of BoNT in the gastrointestinal tract (29). The gene encoding NTNH is located immediately upstream of that of BoNT in all toxin types (6), and genes for other components of the complexes have been shown clustered upstream of the NTNH gene in strains encoding types A and B (8) and type C (15, 19, 32, 33).

Recently, we showed that two distinct populations of *C. botulinum* type A strains exist, encoding distinct BoNT/A genes (designated BoNT/A1 and BoNT/A2) which can be distinguished from each other by DNA probes and by restriction fragment length polymorphism analysis (4). Type A1 and A2 strains encode BoNTs which show ca. 89% sequence identity

* Corresponding author. Mailing address: Department of Microbiology, Institute of Food Research, Earley Gate, Whiteknights Road, Reading, Berkshire RG6 6BZ, United Kingdom. Phone: 44 1734 357225. Fax: 44 1734 267917. Electronic mail address: alison.east@bbsrc.ac.uk.

(1, 31, 35). These strains also differ in the composition of the progenitor toxin complex produced and the organization of genes in the toxin cluster (4). It has been shown that approximately 50% of *C. botulinum* type A strains contain unexpressed BoNT/B gene sequences (12). In the *C. botulinum* strains thus far examined, those which carry additional BoNT/B genes are invariably of type A1 (4). To date, knowledge of the gene organization and evolutionary interrelationships of the various progenitor toxin components has not been systematically examined. Since genes encoding these complex components are physically adjacent to genes responsible for BoNT production, it is not unreasonable to expect that their genealogical relationships would mirror that of the neurotoxin.

In this paper, we report elucidation of the organization of the cluster of genes encoding components of the BoNT complex in group I *C. botulinum* type A, B, and F strains. We report the finding of two distinct gene organizations: one for types A1 and B and another for types A2 and F. The complete sequence of the NTN gene is presented for strains encoding BoNT types A1, A2, and B, and genealogical analyses provide evidence for chimera-like sequences in some genes. Discordance in trees derived from different regions of the NTN genes, which is symptomatic of recombination, was observed, indicating that the NTN gene may be a hot spot for such events within the cluster of genes coding for the BoNT complex.

MATERIALS AND METHODS

Bacterial strains and plasmids. Genomic DNA was prepared from the clostridial strains listed in Table 1 according to the method described by Lawson et al. (23). *Escherichia coli* One Shot obtained from Invitrogen (R&D Systems, Cowley, United Kingdom) was used as the recipient for cloning.

PCR and cloning. PCR was carried out with Perkin-Elmer Ampli-Taq DNA polymerase (Applied Biosystems, Warrington, United Kingdom) as described previously (6) according to the scheme outlined in Fig. 1. The sequences of primers used are given in Table 2. PCR products were cloned with a TA cloning kit (Invitrogen) as described previously (6).

Hybridization analysis. Hybridizations were carried out with an ECL kit (Amersham International, Amersham, United Kingdom) according to the manufacturer's instructions. Slot blots were performed essentially as described by Campbell et al. (2), with hybridization carried out at 42°C and washing carried out at 50°C. The oligonucleotide hybridization probes used are listed in Table 2.

DNA sequencing. Double-stranded DNA was sequenced as described previously (2) with a U.S. Biochemicals Sequenase kit (Amersham) according to the manufacturer's instructions. Clones obtained from two separate PCRs were sequenced, and, in the case of discrepancy, a third independent clone was analyzed.

Analysis of data. Analysis was performed with the molecular biology software of University of Wisconsin Genetics Computer Group package (5). The neighbor-joining method described by Saitou and Nei (27) was used for construction of the phylogenetic trees. The stability of trees was assessed by bootstrap analysis (11). Bootstrap values were calculated from 250 replicates.

Nucleotide sequence accession numbers. The sequences presented here have been given EMBL/GenBank accession numbers as follows: NTN gene sequences from strains 62A, Kyoto-F, and NCTC 7273, respectively, X92973, X87974, and X78230; P-21-NTN gene sequences from strains Kyoto-F and Langeland, respectively, X96493 and X96494; P-21 partial gene sequences from strains Chiba-H and NCTC 9837, respectively, X96491 and X96492.

RESULTS

Analysis of the NTN genes. The regions upstream of the BoNT genes in type A strains 62A and Kyoto-F, type B strain NCTC 7273, and type F strain Langeland were amplified by PCR, and fragments were cloned as outlined in Fig. 1. The gene encoding NTN was found immediately upstream of BoNT, and its sequences in types A and B were determined. In type A strain 62A, the C-terminal 103 amino acids correspond to translation of the region upstream of BoNT/A reported by Binz et al. (1), and for type B strain NCTC 7273, the C-terminal 8 amino acids correspond to translation of the region upstream of BoNT/B (34). For type A strain Kyoto-F, the

TABLE 1. Details of strains used and results of PCR and probing experiments

Strain ^a	PCR with H1-H2	Probe HA33/A	PCR with P1-P2	PCR with I+-NN2	PCR with I--NN2
<i>C. botulinum</i> group I					
NCTC 887 (A1)	+	+	+	+	—
599 (A1)	+	+	+	+	—
603 (A1)	+	+	+	+	—
2775 (A1)	+	+	+	+	—
62A (A1)	ND ^b		+	+	—
547 (A1)	+	+	+	+	—
726 (A1)	+	+	+	+	—
1690 (A1)	+	+	+	ND	ND
NCTC 7272 (A1)	ND		+	ND	ND
NCTC 2916 [A1(B)]	+	+	+	+	+
NCTC 11199 [A1(B)]	+	+	+	+	+
2119 [A1(B)]	+	+	+	+	+
5311 [A1(B)]	+	+	+	+	+
7826 [A1(B)]	+	+	+	+	+
MDa10 [A1(B)]	+	+	+	+	+
588 [A1(B)]	+	+	+	+	+
13280 [A1(B)]	+	+	+	+	+
667 [A1(B)]	+	+	+	+	+
657 (A1/B)	+	+	ND	ND	ND
3281(Bf)	+	+	ND	+	+
Kyoto-F (A2)	ND		+	—	+
Chiba-H (A2)	—		+	—	+
NCTC 9837 (A2)	—		+	—	+
Y-8036 (A2)	—		—	—	+
7105-H (A2)	—		—	—	+
7103-H (A2)	—		—	—	+
KZ1828 (A2)	—		—	—	+
NCTC 2012 (A2)	—		—	—	+
Langeland (F)	—		+	—	+
<i>C. botulinum</i> group II					
Eklund 2B (B)	+	—	+	+	—
Scott 2129B (B)	+	—	+	+	—
Colworth 151 (B)	+	—	+	+	—
Hobbs FT50 (B)	+	—	+	+	—
Hazen 36208 (E)	—		—	—	+
Eklund 202F (F)	—		—	—	+
VPI 2093-1 (NT)	—		—	—	—
<i>C. botulinum</i> group IV					
ATCC 27322 (G)	—		+	—	—
<i>C. barati</i>					
ATCC 43756 (F)	—		—	—	—
<i>C. butyricum</i>					
ATCC 43755 (E)	—		—	—	+

^a Toxin types are given in parentheses after strain designations. A1(B), production of BoNT/A but with silent BoNT/B present on the genome; A1/B, BoNT/A and BoNT/B (dual toxin) production; Bf, BoNT/B and BoNT/F production, but with BoNT/F produced only in minor amounts; NT, nontoxic.

^b ND, not determined.

C-terminal 16 amino acids of NTN correspond to the sequence upstream of BoNT/A reported by Willems et al. (35). The proteins encoded by strain 62A (type A1), Kyoto-F (type A2), and NCTC 7273 (type B) sequences consisted of 1,193, 1,159 and 1,197 amino acids, respectively, and showed homology to published NTN sequences of toxins of other types (6, 13, 14, 33). While this work was in progress, Fujita et al. (16) reported the NTN sequence of type A strain A-NIH, which is identical to that reported here for type A strain 62A. The overall homology shown between these NTNs (>60% iden-

FIG. 1. Strategy for PCR amplification and cloning of fragments encoding components of the BoNT complex from proteolytic *C. botulinum* type A, B, and F strains. The strains used and the approximate positions of binding of PCR primers are shown. Primer sequences are given in Table 2. The positions of genes encoding the components of the complex and direction of transcription are shown. The sizes of the intergenic regions are indicated. Dotted lines, sequences previously reported (type A strains 62A [1] and Kyoto-F [35], type B strains NCTC 7273 [8] and Danish [34], and type F strain Langeland [10]).

tity) is higher than that shown between the BoNTs, which can be as low as ca. 34% identity (3, 10). From an alignment of the NTN1 amino acid sequences (data not shown), it is clear that although the proteins show a high degree of homology, several insertions and/or deletions of one or more amino acids are

evident. The largest and most striking of these is at position 115, where there are an extra 33 amino acids in types A1 and B compared with A2. The extra 33 residues are present in the NTN1 sequence of *C. botulinum* type C (33) but are absent from those of *C. botulinum* type E (14), *C. butyricum* (13), and nonproteolytic *C. botulinum* type F (6). Preliminary sequence analysis of the region in proteolytic *C. botulinum* type F strain Langeland shows that it too lacks these 33 amino acids (9). Since all of the strains examined which possess this insert (types A1, B, and C) have HA components in their progenitor toxin complex, it is interesting to speculate that this part of the NTN1 may somehow be involved in interaction with HAs. It has recently been shown that the NTN1 component of type A, C, and D strains of the M complex has been proteolytically cleaved at a position within the extra 33 amino acids, while the NTN1 component of the L complex is unprocessed (16). Since the L complex contains components which have HA activity, while the M complex lacks them, processing within this region may prevent (directly or indirectly) the HA components from interacting to form the larger complex.

To examine the occurrence of this insert in the NTN1 genes of other strains, PCR primer I+, which binds to a region within the extra 33 residues (positions 122 to 127 [Table 2]), and primer I-, which spans the site at which the additional residues are present in some strains (residues 110 to 117 for NTN1/A2 [Table 2]), were used in combination with conserved primer NN2 (positions 692 to 699 [Table 2]). PCR products were obtained with primer combination I+-NN2 with all strains encoding type A1 [including A1(B)] and B tested but not with any of those encoding type A2 (Table 1). With primer combination I--NN2, positive results were obtained for all strains encoding types A2, E, F, and, unexpectedly, all type A strains containing silent BoNT/B sequences [i.e., A1(B)]. Since products were obtained with both pairs of primers for type A1(B) strains, this indicates that NTN1 genes encoding proteins with and without the additional 33 amino acids are

TABLE 2. Oligonucleotide primers used in PCR and probing experiments

PCR primer or probe	Sequence 5'-3'	Target amino acids ^a	Target positions ^b
PCR primers			
BT3	TGCTTCTGGTGGTGGATTAAATCTCC	GDLNPPPEA	57-65 (BoNT/A)
BT4	CTCCCGTACCTCTCGCAAATGG	PFARGTG	25-31 (BoNT/B)
H1	AAATTGTTACCACTCTCTGTAAGGC	IVTISCKA	16-23 (HA-33)
H2	AATTAATACTTGAATAGCAGTTCGGT	GTAIVGFN	275-282 (HA-33)
HAII-1	AT(A/T)TATAATTACCATAGG	PNGNYNI	9-15 (HA-17)
I+	ATATGATTACTTTTGGATCAGCAC	MITFGSA	122-127 (NTNH/A1)
I-	CTTATGAATATAAACCTGGAGATTAT	YEYKPGDY	110-117 (NTNH/A2)
NN1	ATA(A/G)ATTC(A/T)CCAGTGGG(A/T/C)AA	I(N/D)SPVDN	9-15 (NTNH)
NN2	ATTTTCTTGTGCTAATATTGA	SILAQEND	692-699 (NTNH)
NN3	TTAGTTTCTTAGATCAATGGTGG	SFLDQWW	669-675 (NTNH)
NN4	AGAAAGGTATTATGGAGACA	ERYGYE	44-49 (NTNH)
NN5	TTAGAAATCATAAATTCC(T/A)CC(G/A)TC	DGGIYDS	60-66 (NTNH)
NN6	TGATTATCTTTAGATACT	VSKDN	157-161 (NTNH)
NN7	TAAGGGGGCGTGTATAGTA	Noncoding	Upstream of NTN1
NTNH-1	CTTGCTCTAACTATTACAAC	VVIVRAR	18-24 (NTNH)
P1	GATAT(A/T)(C/G/T)TG(A/T)(A/G)(T/C)CAT(C/T)T(A/T)TGG	DI(L/V)xH(F/L)W	44-50 (P-21)
P2	T(T/C)TTATATA(C/T)AG(A/C)(T/C)TGACGACT	SRQ(S/A)(V/I)YK	167-173 (P-21)
Probes			
H33/A	CTTCTAATGATCAAAATA	SNDQN	231-235 (HA-33)
H33/Bp	TCCCGGTAACGGTAACGT	PGNGNV	33-38 (HA-33)
A33/Bnp	GCTATGACAGATCTAAG	AMTDL	174-179 (HA-33)

^a Amino acids for oligonucleotide target positions. x, any amino acid.

^b Binding positions are given with reference to the following: BoNT/A (1), BoNT/B (34), HA-33 (8), HA-17 (19), NTN1 (6), and P-21 (Fig. 6).

FIG. 2. Pairwise comparison of NTN sequences from proteolytic *C. botulinum* type A strains 62A (A1) and Kyoto-F (A2), type B strain NCTC 7273 (Bp), and type A(B) strain 667. 667'A', A cluster; 667'B', B cluster (21). Comparisons were performed on blocks of 100 amino acids (shown at the top). Shading indicates the degrees of homology.

present. Hutson et al. (21) have sequenced the two BoNT gene clusters encoded by *C. botulinum* type A1(B) strain 667, which revealed that the NTN gene of the A cluster (i.e., upstream of the expressed BoNT/A) lacks the extra 33 amino acids while the NTN encoded by the B cluster possesses them.

Pairwise analyses were performed on the NTN gene sequences of proteolytic *C. botulinum*. It was evident from these analyses that for some pairs of sequences there was considerable variation in the degree of homology along their lengths. The variability in the pattern of relationship between different positions in the NTN sequences of representative combinations of *C. botulinum* strains is illustrated in Fig. 2. For example, the NTNs of proteolytic *C. botulinum* types A and B show very high sequence relatedness for approximately the first 600 residues (>99% identity). In contrast, the amino acids from position ca. 600 to the C terminus have a much-reduced relatedness, ca. 66% identity. The reverse is observed for the NTNs of *C. botulinum* types A1 and A2, with the C-terminal end showing high sequence identity (ca. 93% for position 670 to the end) compared with that for the N-terminal region (ca. 67% for positions 1 to 669). Matrices of pairwise sequence homology data from different regions of the NTN (residues 1 to 550, 551 to 1020, and 1021 to the C terminus) were used to construct comparative genealogical trees (Fig. 3). Major differences in the three trees reinforce the incongruence in the pattern of relationships along the length of the NTNs. Such marked discordance could be symptomatic of recombination and suggests that the NTN gene is a hot spot for such events within the BoNT complex cluster of genes. Hutson et al. (21) reported that the NTN gene adjacent to the silent BoNT/B gene (B cluster) in *C. botulinum* type A(B) strain 667 was chimeric, with the 5' and 3' regions of the gene exhibiting high homology with the corresponding regions of the type B NTN gene (Fig. 3), while a 417-amino-acid sequence in the central region was identical to that in the NTN whose gene was located adjacent to the expressed BoNT/A gene (A cluster) (Fig. 2). The results of the present study demonstrate that such chimera-like NTN genes may be commonplace among proteolytic *C. botulinum*.

A comparison of the homology values along the length of the cluster of genes encoding the BoNT complex for type A strains 62A and Kyoto-F and type B strain NCTC 7273 is shown in Fig. 4. In addition to the variation in the degree of homology in the coding regions, further weight is given to the idea of

recombination by observation of the intergenic regions. In these strains, the intergenic regions are of identical size and (almost) identical sequence between HA-17 and HA-33, HA-33 and P-21, and P-21 and NTN in type A strain 62A and type B strain NCTC 7273, while the corresponding region in type A strain Kyoto-F is noncomparable (Fig. 4). In contrast, the intergenic region between the NTN and BoNT genes is identical in size and sequence in the two type A strains, while that of type B strain NCTC 7273 is different (Fig. 4).

Investigation of the HA-33 genes of types A and B. We recently showed that type A strain NCTC 7272 and type B strains NCTC 7273 and 17B encode genes for HA-17, HA-33, and P-21 (linked in this order) upstream of those for NTN and BoNT (Fig. 1) (8). The region encoding HA-33 and P-21 was cloned from type A strain 62A (Fig. 1) and was found to be identical to that from type A strain NCTC 7272 (8). To determine if the HA-33 gene is present in other strains, PCR primers based on conserved amino acids were developed (primers H1 and H2 [Table 2]). Positive results (PCR product of ca. 1.2

FIG. 3. Trees revealing the discordance in relationships along the lengths of NTNs, residues 1 to 550 (i), residues 551 to 1020 (ii), and residue 1021 to C terminus (iii). NTN sequences from the following were compared: proteolytic *C. botulinum* type A strains 62A (A1) and Kyoto-F (A2), type B strain NCTC 7273 (Bp), type A(B) strain 667, A cluster (667'A') and B cluster (667'B') (21), and nonproteolytic *C. botulinum* type F strain Eklund 202F (Fnp) (6). The numbers on the tree indicate bootstrap values for branch points; only values of >90 are shown.

FIG. 4. Homology along the length of the BoNT gene cluster for *C. botulinum* type A and B strains. Closed arrows, amino acid identity values for proteins encoded by the genes, open arrows, nucleic acid identity values for intergenic regions. Values for NTNH refer to residues 1 to 600 and 601 to the C terminus.

kb) were obtained for *C. botulinum* strains encoding type A1 [including A1(B)] and B but not for those encoding type A2, E, F, and G or toxigenic strains of *C. barati* and *C. butyricum* (Table 1). Although the lack of a PCR product may be due to the mismatch of primers with the target gene, these results correlate with the presence or absence of HA activity in the progenitor complexes; strains encoding types A and B, which produce complexes with HA activity, give positive results, while strains which lack HA activity (types A2, E, and F and strains of *C. barati* and *C. butyricum*) give no PCR product. Although type G strains produce complexes which contain components with HA activity, purification of the components shows that none correspond to HA-33 of type A, B, or C (25).

Hybridization of the PCR products with probe H33/A (Fig. 5A) shows a positive result for all type A1, A1(B), A1/B, and Bf strains. In contrast, the probe designed to detect HA-33 of proteolytic B (H33/Bp) gave a signal only with the positive control and not with any other strains, including the type A1(B) strains (Fig. 5C). The probe to detect HA-33 of non-proteolytic type B (H33/Bnp) hybridized only with the PCR product from the nonproteolytic B strains (Fig. 5B). Interestingly, all of the A1(B) strains examined hybridized only with probe H33/A, indicating the presence of only type A HA-33. Sequence analysis of *C. botulinum* type A1(B), strain 667, showed only one HA-33 gene to be present; although it was located in the B cluster, it showed the highest homology with

FIG. 5. (A to C) Slot blots with probes designed to detect the genes encoding HA-33 type A, HA-33 type B from nonproteolytic *C. botulinum*, and HA-33 type B from proteolytic *C. botulinum*, respectively; (D) key with strain type, showing hybridization with the following strains: 547 (1a), 588 (1b), 599 (1c), 603 (1d), 726 (1e), NCTC 887 (1f), 5311 (2a), 2119 (2b), NCTC 2916 (2c), 2775 (2d), 7826 (2e), 1690 (2f), NCTCC 11199 (3a), 13280 (3b), MDa10 (3c), Scott 2129B (3d), Colworth 151 (3e), Hobbs FT50 (3f), Eklund 2B (4a), pAET70 (cloned HA-33 gene from type A strain NCTC 7272) (8) (4b), pAET59 (cloned HA-33 gene from proteolytic type B strain NCTC 7273) (8) (4c), pAET55 (cloned HA-33 gene from nonproteolytic type B strain Eklund 17B) (8) (4d), 657 (4e), 3281 (4f), and 667 (5a).

FIG. 6. Comparison of derived amino acid sequences for P-21s from *C. botulinum* type A strains 62A (A1), Kyoto-F (A2K), Chiba-H (A2C), and NCTC 9837 (A29); type B strains NCTC 7273 (Bp), and Eklund 17B (Bnp) (8); type C strain 468 (19); type F strain Langeland; and UviA from *C. perfringens* (17). Black and shaded areas, identical and similar amino acids in five or more sequences, respectively; dots, absent residues.

HA-33/A (ca. 99% identity) and possessed the hybridization site of the H33/A probe (21). The single Bf strain examined (strain 3281) hybridized only with the HA-33 type A probe (Fig. 5). Further experiments are under way to investigate this result.

Elucidation of P-21 gene sequences in *C. botulinum* types A2 and F. A gene encoding a potential regulator of the BoNT complex was reported for *C. botulinum* types A and B (designated P-21) (8) and type C (designated *orf-22*) (19). This gene is clustered with those of the components of the BoNT complex, although it maps to different positions in strains of different toxin types. PCR primers based on conserved regions of P-21 (P1 and P2 [Table 2]) were designed and used to investigate the presence of this gene in other strains (Table 1). No PCR products were obtained for any of the strains of nonproteolytic *C. botulinum* type F, *C. botulinum* type E, *C. barati*, or *C. butyricum* examined. PCR amplification products with the expected size (ca. 400 bp) were obtained for all strains encoding types A1, A1(B), B, and G for the single proteolytic *C. botulinum* type F strain and three of the type A2 strains examined (Table 1). The PCR amplification products from the proteolytic *C. botulinum* type F strain (Langeland) and the three type A2 strains (Kyoto-F, Chiba-H, and NCTC 9837) were cloned, and their sequences were determined. Very high sequence homology was observed between the P-21 proteins from these type A2 strains and type F strain Langeland (ca. 91% sequence identity). The P-21 proteins from type A2 strains and type F strain Langeland show ca. 63% amino acid identity with the P-21 proteins from *C. botulinum* type A1 and type B strains (proteolytic and nonproteolytic), ca. 55% sequence identity with the *orf-22* gene product from *C. botulinum* type C (19), and ca. 26% identity with the *uviA* gene product from *Clostridium perfringens* (17). UviA is thought to be a DNA-binding protein with a possible regulatory function (24). Figure 6 shows a comparison of the sequences of the P-21

proteins of *C. botulinum* and UviA from *C. perfringens*. To determine the position of the P-21 gene in strains Langeland and Kyoto-F, PCR strategies were employed. The P-21 gene was found to map in an identical position ca. 2 kb upstream of the NTNH gene in both of these strains. The orientation of P-21 is such that it is transcribed away from the 5' end of the NTNH gene (Fig. 1).

Analysis of a new open reading frame. In an earlier study, we reported the 3' end of an unknown *orf* located upstream of NTNH gene in nonproteolytic *C. botulinum* type F strain Eklund 202F (6). Cloning and sequencing the region between the NTNH and P-21 genes revealed an *orf*, which was preceded by a putative Shine-Dalgarno sequence, encoding a protein of 416 amino acids in both proteolytic *C. botulinum* type F (Langeland) and type A2 (Kyoto-F) strains. This protein has been designated P-47, since it has relative molecular weight values of 47,476 in strain Kyoto-F and of 47,420 in strain Langeland. A comparison of the derived amino acid sequences of P-47 proteins for strains Kyoto-F and Langeland shows ca. 79% identity overall but ca. 96% identity in the first 200 amino acids. The translated 3' end of the *orf* in *C. botulinum* type F strain Eklund 202F shows homology to the C termini of P-47 proteins encoded in strains Langeland and Kyoto-F (ca. 66 to 69% identity over 271 amino acids [7]). Interestingly, the 3' end of this *orf* is also found in *C. botulinum* type A(B) strain 667 (21), and translation of sequence upstream of the NTNH gene in *C. botulinum* type E (14) and *C. butyricum* (13) reveals an *orf* encoding a protein with homology to P-47s of Kyoto-F and Langeland (ca. 54 to 80% identity). This gene maps in identical positions in these strains (Fig. 1). The gene encoding P-47 is, therefore, widely distributed among strains encoding BoNT and is linked or clustered with genes encoding neurotoxin complex components. It may be pertinent that this gene has been found only in toxin gene clusters which lack genes for HA components of the complex.

DISCUSSION

It has long been recognized that strains of *C. botulinum* are physiologically heterogeneous, and four distinct phenotypic groups (I to IV) are recognized (18). These four metabolically distinct groups do not, however, necessarily correlate with the serological specificities of the BoNT produced (18). Genealogical analysis based on 16S rRNA gene sequences has shown that *C. botulinum* consists of four highly divergent lines which correlate with the four phenotypic divisions (22). Recent sequencing studies of the genes encoding BoNT have unequivocally demonstrated the marked discordance between trees depicting natural relationships (i.e., 16S rRNA) of the organisms and those of the various BoNT types produced (e.g., see references 3, 10, and 20). In particular, the results of these comparative sequencing studies are strongly indicative of BoNT gene transfer between *C. botulinum* metabolic groups or rRNA lines (viz., BoNT/B in groups I and II, BoNT/E in group II and toxigenic *C. butyricum*, and BoNT/F in groups I and II and toxigenic *C. barati*). As outlined earlier, the gene encoding BoNT is clustered with several other genes encoding components of the progenitor complex. Since BoNT is responsible for botulism and BoNT is the primary trait for defining and subdividing the so-called species *C. botulinum*, a thorough understanding of the evolutionary interrelationships of the genes encoding this toxin and other components of the progenitor complex is essential. Knowledge of the organization of the cluster of genes encoding BoNT complex components is very limited. Information on the genealogy of the various progenitor proteins is also unknown, and since genes encoding these are proximal to that responsible for BoNT production, it is likely that their genealogical patterns would be similar to that of the neurotoxins. In this study, we have examined the arrangements of genes encoding components of the BoNT complex in proteolytic (group I) *C. botulinum* type A, B, and F strains. Two different arrangements were found. Type A strain 62A and type B strain NCTC 7273 possessed similar organizations, viz., *BoNT*, *NTNH*, *P-21*, and *HA* components; type F strain Langeland and infant type A strain Kyoto-F exhibited gene organizations similar to each other, viz., *BoNT*, *NTNH*, *P-47*, and *P-21*, and appeared to lack *HA* genes. In addition to this discordance in gene organization versus BoNT type, we have shown that the genealogy of the *NTNH* (the gene encoding *NTNH* is linked to the *BoNT* gene) does not mirror the evolutionary history of BoNT. Furthermore, we have found evidence of chimeric *NTNH* gene sequences, which could be symptomatic of recombination, and indicate that the *NTNH* gene may be a hot spot for such events within the BoNT complex gene cluster. There is clearly a strong case to extend the described studies to embrace further group I strains and *C. botulinum* groups (including other BoNT-producing species) to obtain a better understanding of population structure within this important pathogen and a complete picture of the evolutionary histories of its neurotoxin and associated proteins. Such genetic information is of great relevance to the taxonomy of *C. botulinum*, which is currently defined and subdivided on the basis of the type of neurotoxin produced.

ACKNOWLEDGMENTS

We thank Charles Hatheway and Genji Sakaguchi for the generous gift of strains.

M.B. is the recipient of a BBSRC research studentship.

REFERENCES

1. Binz, T., H. Kurazono, M. Wille, J. Frevert, K. Wernars, and H. Niemann. 1990. The complete sequence of botulinum neurotoxin type A and comparison with other clostridial neurotoxins. *J. Biol. Chem.* **265**:9153–9158.

2. Campbell, K. D., M. D. Collins, and A. K. East. 1993. Gene probes for identification of the botulinum neurotoxin gene and specific identification of neurotoxin types B, E, and F. *J. Clin. Microbiol.* **31**:2255–2262.
3. Campbell, K. D., M. D. Collins, and A. K. East. 1993. Nucleotide sequence of the gene coding for *Clostridium botulinum* (*Clostridium argentinense*) type G neurotoxin: genealogical comparison with other clostridial neurotoxins. *Biochem. Biophys. Acta* **1216**:487–491.
4. Cordoba, J. J., M. D. Collins, and A. K. East. 1995. Studies on the genes encoding botulinum neurotoxin type A of *Clostridium botulinum* from a variety of sources. *Syst. Appl. Microbiol.* **18**:13–22.
5. Devereux, J., P. Haeberli, and D. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387–395.
6. East, A. K., and M. D. Collins. 1994. Conserved structure of genes encoding components of botulinum neurotoxin complex M and the sequence of the gene coding for the nontoxic component in non-proteolytic *Clostridium botulinum* type F. *Curr. Microbiol.* **29**:69–77.
7. East, A. K., and M. D. Collins. Unpublished data.
8. 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 *Clostridium botulinum* types A and B. *Syst. Appl. Microbiol.* **17**:306–312.
9. East, A. K., J. M. Stacey, S. Hiem, and M. D. Collins. Unpublished data.
10. Elmore, M. J., R. A. Hutson, M. D. Collins, N. J. Bodsworth, S. M. Whelan, and N. P. Minton. 1995. Nucleotide sequence of the gene coding for proteolytic (group I) *Clostridium botulinum* type F neurotoxin: genealogical comparison with other clostridial neurotoxins. *Syst. Appl. Microbiol.* **18**:23–31.
11. Felsenstein, J. 1989. PHYLIP-phylogeny inference package (version 3.2). *Cladistics* **5**:164–166.
12. Franciosa, G., J. L. Ferreira, and C. L. Hatheway. 1994. Detection of type A, B and E botulinum neurotoxin genes in *Clostridium botulinum* and other *Clostridium* species by PCR: evidence of unexpressed type B toxin genes in type A toxigenic organisms. *J. Clin. Microbiol.* **32**:1911–1917.
13. Fujii, N., K. Kimura, N. Yokosawa, K. Oguma, T. Yashiki, K. Takeshi, T. Ohyama, E. Isogai, and H. Isogai. 1993. Similarity in nucleotide sequence of the gene encoding nontoxic component of botulinum toxin produced by toxigenic *Clostridium butyricum* strain BL6340 and *Clostridium botulinum* type E strain Mashike. *Microbiol. Immunol.* **37**:395–398.
14. Fujii, N., K. Kimura, N. Yokosawa, T. Yashiki, K. Takeshi, K. Tsuzuki, and K. Oguma. 1993. The complete nucleotide sequence of the gene encoding the nontoxic component of *Clostridium botulinum* type E progenitor toxin. *J. Gen. Microbiol.* **139**:79–86.
15. Fujinaga, Y., K. Inoue, S. Shimazaki, K. Tomochika, K. Tsuzuki, N. Fujii, T. Watanabe, T. Ohyama, K. Takeshi, K. Inoue, and K. Oguma. 1994. Molecular construction of *Clostridium botulinum* type C progenitor toxin and its gene organization. *Biochem. Biophys. Res. Commun.* **205**:1291–1298.
16. Fujita, R., Y. Fujinaga, K. Inoue, H. Nakajima, H. Kumon, and K. Oguma. 1995. Molecular characterization of two forms of nontoxic-nonhemagglutinin components of *Clostridium botulinum* type A progenitor toxins. *FEBS Lett.* **376**:41–44.
17. Garnier, T., and S. T. Cole. 1988. Studies of UV-inducible promoters from *Clostridium perfringens* in vivo and in vitro. *Mol. Microbiol.* **2**:607–614.
18. Hatheway, C. L. 1992. *Clostridium botulinum* and other clostridia that produce botulinum neurotoxin, p. 3–20. In A. W. Hauschild and K. L. Dodds (ed.), *Clostridium botulinum—ecology and control in foods*. Marcel Dekker, New York.
19. Hauser, D., M. W. Eklund, P. Boquet, and M. R. Popoff. 1994. Organization of the botulinum neurotoxin C1 gene and its associated non-toxic protein genes in *Clostridium botulinum* C468. *Mol. Gen. Genet.* **243**:631–640.
20. Hutson, R. A., M. D. Collins, A. K. East, and D. E. Thompson. 1994. Nucleotide sequence of the gene coding for non-proteolytic *Clostridium botulinum* type B neurotoxin: comparison with other clostridial neurotoxins. *Curr. Microbiol.* **28**:101–110.
21. Hutson, R. A., M. D. Collins, Y. Zhou, J. Jiang, H. Sugiyama, E. A. Johnson, and C. L. Hatheway. 1996. Genetic characterization of *Clostridium botulinum* type A containing a silent type B neurotoxin gene sequences. *J. Biol. Chem.* **271**:10786–10792.
22. Hutson, R. A., D. E. Thompson, P. A. Lawson, R. P. Schocken-Iturrino, E. C. Botter, and M. D. Collins. 1993. Genetic interrelationships of proteolytic *Clostridium botulinum* type A, B, and F and other members of the *Clostridium botulinum* complex as revealed by 16S rRNA gene sequences. *Antonie van Leeuwenhoek* **64**:278–283.
23. Lawson, P. A., S. E. Gharbia, H. N. Shah, and D. J. Clark. 1989. Recognition of *Fusobacterium nucleatum* subgroups Fn-1, Fn-2 and Fn-3 by ribosomal RNA gene restriction patterns. *FEMS Microbiol. Lett.* **65**:41–46.
24. Lonetto, M., M. Gribskov, and C. A. Gross. 1992. The σ^{70} family: sequence conservation and evolutionary relationships. *J. Bacteriol.* **174**:3843–3849.
25. Nukina, M., Y. Mochida, S. Sakaguchi, and G. Sakaguchi. 1991. Difficulties of molecular dissociation of *Clostridium botulinum* type G progenitor toxin. *FEMS Microbiol. Lett.* **79**:165–170.
26. Ohishi, I., and G. Sakaguchi. 1975. Molecular construction of *Clostridium*

- botulinum* type F progenitor toxin. Appl. Microbiol. **29**:444–447.
27. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for constructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
 28. Somers, E., and B. R. DasGupta. 1991. *Clostridium botulinum* types A, B, C1, and E produce proteins with or without hemagglutinating activity: do they share common amino acids sequences and genes. J. Prot. Chem. **10**:415–425.
 29. Sugii, S., I. Ohisi, and G. Sakaguchi. 1977. Correlation between oral toxicity and in vitro stability of *Clostridium botulinum* type A and B toxins of different molecular sizes. Infect. Immun. **16**:910–914.
 30. Tabita, K., S. Sakaguchi, and G. Sakaguchi. 1991. Distinction between *Clostridium botulinum* type A strains associated with food-borne botulism and those with infant botulism in Japan in intraintestinal toxin production in infant mice and some other properties. FEMS Microbiol. Lett. **79**:251–256.
 31. Thompson, D. E., J. K. Brehm, J. D. Oultram, T. J. Swinfield, C. C. Shone, T. Atkinson, J. Melling, and N. P. Minton. 1990. The complete amino acid sequence of the *Clostridium botulinum* type A neurotoxin deduced by nucleotide sequence analysis of the encoding gene. Eur. J. Biochem. **189**:73–81.
 32. Tsuzuki, K., K. Kimura, N. Fuji, N. Yokosawa, T. Indoh, T. Murakami, and K. Oguma. 1990. Cloning and complete nucleotide sequence of the gene for the main component of hemagglutinin produced by *Clostridium botulinum* type C. Infect. Immun. **58**:3173–3177.
 33. Tsuzuki, K., K. Kimura, N. Fuji, N. Yokosawa, and K. Oguma. 1992. The complete nucleotide sequence of the gene coding for the nontoxic-nonhemagglutinin component of *Clostridium botulinum* type C progenitor toxin. Biochem. Biophys. Res. Commun. **183**:1273–1279.
 34. Whelan, S. M., M. J. Elmore, N. J. Bodsworth, J. K. Brehm, T. Atkinson, and N. P. Minton. 1992. Molecular cloning of the *Clostridium botulinum* structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence. Appl. Environ. Microbiol. **58**:2345–2354.
 35. Willems, A., A. K. East, P. A. Lawson, and M. D. Collins. 1993. Sequence of the gene coding for the neurotoxin of *Clostridium botulinum* type A associated with infant botulism: comparison with other clostridial neurotoxins. Res. Microbiol. **144**:547–556.