APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 2007, p. 2963–2975 Vol. 73, No. 9
0099-2240/07/$08.00+0 doi:10.1128/AEM.02623-06

Flagellin Diversity in *Clostridium botulinum* Groups I and II: a New Strategy for Strain Identification▽

Catherine J. Paul,[1,2] Susan M. Twine,[2] Kevin J. Tam,[1] James A. Mullen,[2] John F. Kelly,[2] John W. Austin,[1]* and Susan M. Logan[2]

Bureau of Microbial Hazards, HFPB, Health Canada, Sir Frederick G. Banting Research Centre, PL2204A2, Ottawa, Ontario, Canada K1A 0L2,[1] and Institute for Biological Sciences, National Research Council, 100 Sussex Dr., Ottawa, Ontario, Canada K1A 0R6[2]

Received 9 November 2006/Accepted 26 February 2007

Strains of *Clostridium botulinum* are traditionally identified by botulinum neurotoxin type; however, identification of an additional target for typing would improve differentiation. Isolation of flagellar filaments and analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) showed that *C. botulinum* produced multiple flagellin proteins. Nano-liquid chromatography-tandem mass spectrometry (nLC-MS/MS) analysis of in-gel tryptic digests identified peptides in all flagellin bands that matched two homologous tandem flagellin genes identified in the *C. botulinum* Hall A genome. Designated *flaA1* and *flaA2*, these open reading frames encode the major structural flagellins of *C. botulinum*. Colony PCR and sequencing of *flaA1*/*A2* variable regions classified 80 environmental and clinical strains into group I or group II and clustered isolates into 12 flagellar types. Flagellar type was distinct from neurotoxin type, and epidemiologically related isolates clustered together. Sequencing a larger PCR product, obtained during amplification of *flaA1*/*A2* from type E strain Bennett identified a second flagellin gene, *flaB*. LC-MS analysis confirmed that *flaB* encoded a large type E-specific flagellin protein, and the predicted molecular mass for FlaB matched that observed by SDS-PAGE. In contrast, the molecular mass of FlaA was 2 to 12 kDa larger than the mass predicted by the *flaA1*/*A2* sequence of a given strain, suggesting that FlaA is posttranslationally modified. While identification of FlaB, and the observation by SDS-PAGE of different masses of the FlaA proteins, showed the flagellin proteins of *C. botulinum* to be diverse, the presence of the *flaA1*/*A2* gene in all strains examined facilitates single locus sequence typing of *C. botulinum* using the flagellin variable region.

Clostridium botulinum is a gram-positive, spore-forming anaerobic bacterium that produces a potent neurotoxin. Botulinum neurotoxin (BoNT) causes botulism, a neuroparalytic disease which occurs after ingestion of food contaminated with BoNT or by other means of toxin exposure. *C. botulinum* can also colonize wounds and, rarely, the adult gastrointestinal tract as well as the infant gastrointestinal tract, where growth of the bacterium and subsequent production of BoNT results in wound botulism, adult colonization botulism, or infant botulism, respectively (4).

Strains of *C. botulinum* are divided into four distinct groups based on physiological characteristics. Only groups I and II contain strains that have been associated with human illness. Group I strains are proteolytic with an optimal growth temperature of 37°C, while group II strains are nonproteolytic and grow optimally at 30°C. Strains of *C. botulinum* are differentiated into serotypes A through G based on seroneutralization of BoNT. Group I strains can produce serotype A, B, or F BoNT, and group II strains can produce serotype B, E, or F BoNT. Group III strains produce either serotype C or D and are associated with botulism in animals, and group IV contains only strains producing BoNT serotype G (8, 28). In addition to categorizing strains based on physiological characteristics or BoNT serotype, other methods have been used for epidemiological and molecular typing of *C. botulinum*. These methods include the following: randomly amplified polymorphic DNA analysis; repetitive element sequence-based PCR; fatty acid analysis; amplified rRNA gene restriction analysis; pulsed-field gel electrophoresis (PFGE); and most recently, amplified fragment length polymorphism and gene sequencing (16S, neurotoxin) (18, 26, 33).

The flagella of gram-negative bacteria are the most widely characterized of the bacterial motility structures (5, 30). In general, flagella are composed of over 20 distinct structural proteins that assemble to form the flagellar basal body, hook, and filament, with the filament comprising around 20,000 subunits of the flagellin protein. Due to commonalities in folding and filament assembly, all bacterial flagellin proteins can be divided into three distinct subdomains: the N and C termini, which are conserved within a given species, and a middle variable region (31, 41). The flagellin monomers are arranged surrounding the central channel of the filament with the N and C termini of each protein buried in the interior nearest the filament channel. The flagellin variable domain protrudes outwards and, stacked together with other flagellin monomers, forms the external surface of the filament (30). The surface-exposed variable domain of the flagellin is antigenically diverse and forms the basis of a wide range of typing methods exploiting diversity at the DNA sequence level or structural differences at the protein level for strain identification (12, 49). Flagella are often involved in pathogenesis, with roles in mo-

* Corresponding author. Mailing address: Bureau of Microbial Hazards, Health Canada, 251 Sir Frederick Banting Driveway, Tunney's Pasture PL2204A2, Ottawa, Ontario, Canada K1A 0L2. Phone: (613) 957-0902. Fax: (613) 941-0280. E-mail: john_austin@hc-sc.gc.ca.

▽ Published ahead of print on 9 March 2007.

tility adhesion and, in some cases, the secretion of virulence factors (7, 19).

While the genetics, regulation, assembly, and physical structure of the gram-negative bacterial flagellum have been extensively investigated, less is known about the flagella of gram-positive bacteria, and in particular clostridial flagella. Several studies document preliminary investigations of flagella in the pathogens *Clostridium chauvoei*, *Clostridium haemolyticum*, *Clostridium novyi*, and *Clostridium septicum* as well as *Clostridium tyrobutyricum*, a clostridial species of relevance to the cheese industry (3, 22, 23, 39). The flagella of the human pathogen *Clostridium difficile* divides the species into 10 serogroups, has a role in the adherence of *C. difficile* to the mucous layer of the intestine, and is considered a potential vaccine candidate (36, 43, 45, 46). There are, however, no published investigations addressing the flagella of *C. botulinum*, a bacterium that has largely been studied only in the context of the neurotoxin.

This study examines flagellins from a range of isolates from different geographical, clinical, and environmental origins. We explored the diversity of flagellin structural proteins and gene sequences to determine the potential for using flagellin gene variable region sequences to improve genotyping of group I and group II *C. botulinum* strains.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The origin, source, and date of isolation for each of the *C. botulinum* strains used in this study are described in Table 1.

C. botulinum strains were grown for flagellar isolation in SPGY broth containing 5% (wt/vol) special peptone (Oxoid Inc., Basingstoke, United Kingdom), 0.5% (wt/vol) peptone (Difco, Tucker, GA), 2% (wt/vol) yeast extract (Difco), 0.4% (wt/vol) glucose (Difco) and 0.1% (wt/vol) sodium thioglycolate (Sigma Aldrich, St. Louis, MO) adjusted to a pH of 7.2 using HCl. For colony PCR, bacteria were grown on McClung-Toabe 1.5% agar (Difco) with 5% egg yolk and 0.5% yeast extract (MTEYE). Cells were grown for 24 to 48 h in an atmosphere of 10% H_2, 10% CO_2, and 80% N_2 at either 35°C (group I strains) or room temperature (group II strains). Cells were occasionally grown in 3.5-liter anaerobe jars (Oxoid Inc.) using AnaeroGen (Oxoid Inc.) for atmosphere production, at either 35°C or room temperature. *C. botulinum* strains were stored and archived at the Botulism Reference Service for Canada at −86°C on Microbank beads (Pro-Lab Diagnostics, Richmond Hill, Canada).

Escherichia coli strain DH10B (Invitrogen, Carlsbad, CA), grown in LB broth (Difco) at 37°C with shaking, was used for cloning manipulations.

Electron microscopy. Copper grids (Electron Microscopy Sciences, Fort Washington, PA) were covered with Formvar film and coated with carbon. For sample preparation, grids were floated on a drop of bacterial cells or purified flagella in 0.1 M Tris-HCl, pH 7.0, for 5 min and negatively stained with 1% (wt/vol) ammonium molybdate. Images were taken with a Zeiss EM902 transmission electron microscope operated at an accelerating voltage of 80 kV and recorded on Kodak electron image film SO-163 (Kodak, Rochester, NY).

Isolation and preliminary analysis of flagellar proteins. Cultures for flagellar isolation were grown overnight at the required temperature in SPGY broth. Cells were harvested and resuspended in 1/20 of the original volume, in 0.1 M Tris-HCl, pH 7.0. Flagella were sheared from the cell surface by 10 repetitions in a 50-ml tissue homogenizer. Flagella were isolated from the homogenate by removing whole cells by two low-speed centrifugations (5,000 × *g*; 15 min) before a high-speed centrifugation for 1 h at 130,000 × *g* (70 Ti rotor; Beckman Coulter Canada Inc., Mississauga, Ontario, Canada). Pellets containing flagella were washed once in ultrapure water (Gibco, Grand Island, NY), followed by a second centrifugation at 130,000 × *g* for 1 h and resuspension in ultrapure water. Flagella preparations were analyzed by standard sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on 12.5% acrylamide gels and stained with standard Coomassie blue, as described previously (38). Molecular weights of protein bands were determined using Quantity One software (Bio-Rad, Hercules, CA) with comparison to molecular mass standards. Protein bands for tryptic digest and mass spectrometry were excised from 12.5% acrylamide gels following SDS-PAGE and staining with Bio-Safe Coomassie (Bio-Rad). A ProQ Emerald 300 glycoprotein detection kit (Invitrogen, Carlsbad, CA) or a Dig glycan differentiation kit (Roche Diagnostics, Indianapolis, IN) was used following the manufacturer's instructions to examine posttranslational modification of flagellin proteins.

Mass spectrometry analysis of flagellin proteins. Protein bands excised from Bio-Safe Coomassie blue (Bio-Rad)-stained SDS-PAGE gels were cut into 1-mm cubes. In microcentrifuge tubes, gel pieces were covered with 100 mM ammonium bicarbonate and 30% (vol/vol) acetonitrile for 5 min and washed three times with water or until the blue coloration disappeared. The predicted flagellin protein did not contain cysteine residues; therefore, samples were not subjected to reduction/alkylation. Gel pieces were treated with acetonitrile and then rehydrated with a trypsin-containing solution as described previously (47). Peptides were extracted from gel pieces by sonication for 20 min., then analyzed by nano-liquid chromatography–tandem mass spectrometry (nLC-MS/MS) using a 'CapLC' capillary chromatography system (Waters, Milford, MA) coupled to a 'QTOF Ultima' hybrid quadrupole time-of-flight mass spectrometer (Waters, Milford, MA) (47). MS/MS spectra obtained were searched against the Hall A genome sequence (http://www.sanger.ac.uk/Projects/C_botulinum) using MASCOT 2.0.1 (Matrix Science, United Kingdom). The following parameters for mass spectral identification were used: peptide tolerance of 1.5 Da, MS/MS tolerance of 0.8 Da, and possible 1 missed cleavage site. Peptide identifications were accepted if they met all of the following criteria: MASCOT peptide score >20, rank = 1, and mass accuracy <100 ppm. MASCOT scores evaluate the match between an MS/MS profile of a tryptic peptide and a protein sequence database and indicate the probability that the observed match is a random event. In addition, all MS/MS spectra were manually assessed for data quality and high-confidence identification, requiring a clear series of high mass y-ions and correct charge state assignment for all fragments.

Cloning, colony PCR, and sequencing of flagellin genes. Flagellin genes were amplified from chromosomal DNA preparations or from single colonies resuspended in 50 μl Tris-EDTA buffer. Chromosomal DNA was isolated as described previously with slight modifications (21). Primers cbotflaF (5′-CGCGG GGATCCATGATAATTAATCACAATTTAAATG-3′) and cbotflaR (5′-CGC GGGGATCCCTTAATAATTGAAGAACTCCTTGTG-3′) were designed to correspond to the N and C termini of a Hall A flagellin gene (*C. botulinum* genome sequence from the *C. botulinum* Sequencing Group at the Wellcome Trust Sanger Institute [ftp://ftp.sanger.ac.uk/pub/pathogens/cb]) and to include BamHI sites. PCR products were amplified using *Pwo* (Roche Diagnostics) or High Fidelity Triple Master (Eppendorf, Hamburg, Germany) polymerase, following the manufacturer's instructions, in an Applied Biosystems Gene Amp PCR system 9700 thermocycler (cycling parameters: 3 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 55/50°C, and 1 min at 72°C; and 2 min at 72°C). For colony PCR, 5 μl of bacterial cell/Tris-EDTA suspension was used instead of chromosomal DNA with a step of 10 min at 94°C added at the beginning of the PCR for cell lysis. Annealing temperatures were lowered from an initial temperature of 55°C until the product of the expected size could be visualized by standard agarose gel electrophoresis. All strains produced *flaA1*/*A2* amplicons at 50°C. PCR amplicons were purified either by excision of the desired product from preparative 1.0% agarose gels and Qiaquick gel extraction (QIAGEN, Hilden, Germany) or by using a Qiaquick nucleotide removal kit (QIAGEN). Initially, PCR products were ligated into an EcoRV site in vector pCR2.1 (Invitrogen), screened for insert following miniprep (QIAGEN) with BamHI digestion, and sequenced using M13F and M13R primers to vector sequence. Colony PCR products were purified using a Qiaquick nucleotide removal kit (QIAGEN) for direct amplicon sequencing using a Big Dye Terminator 3.1 kit (Applied Biosystems, Foster City, CA), cbotflaF and cbotflaR as sequencing primers, and an ABI3130 genetic analyzer (Applied Biosystems). All restriction enzymes were obtained from New England Biolabs (Ipswich, MA).

Analysis of DNA and protein sequences. Candidate Hall A flagellin genes were identified by tBlastN against the Hall A genome sequence (www.sanger.ac.uk/cgi-bin/blast/submitblast/c_botulinum) using flagellin protein sequences from *C. tyrobutyricum*, *C. difficile*, *C. chauvoei*, and *Bacillus subtilis* (GenBank accession numbers CAB44444.1, AAD46086.1, BAB13814.1, and NP_391416.1, respectively). Hall A flagellin open reading frames (ORFs) were identified by start and stop codons flanking DNA sequence identified by tBlastN as candidate flagellin genes and by comparison to the ORF assignments in homologous *Clostridium tetani* flagellin sequences (GenBank accession numbers AAO36250.1, AAO36226.1, and AAO36215.1). DNA sequences were compiled using Pregap and Gap4 from the Staden package of software for cloned flagellins (42) or Sequencher 4.6 software for colony PCR products (Gene Codes Corporation, Ann Arbor, MI).

TABLE 1. *C. botulinum* strains used in this study

Strain	Origin	Location[j]	Yr of isolation	Source or reference[a]
Group I				
Type A				
Hall A/ATCC 3502				11/IFR[b]
17A			1963	
62A				1/NCA[c]
A6				R. Woods/T. Midura
16037	Canned tomato	United States	1974	Leatherhead/IFR
CK2-A	Feces	Canada	1974	BRS[d]
F9604-A	Feces	Calgary, AB, Canada*	1996	BRS
F9801-A	Feces	Laval, QC, Canada*	1998	BRS
FE9909ACS Alberta	Feces	Loughheed, AB, Canada*	1999	BRS
FE0101AJO	Feces	Desbiens, QC, Canada	2001	BRS
GA0101AJO	Gastric liquid	Desbiens, QC, Canada	2001	BRS
PC0101AJO	Pork	St-Fulgence, QC, Canada	2001	BRS
NG0107ASA	Gastric liquid	Sanikiluaq, NN, Canada	2001	BRS
MUL0109ASA	Mullet fish	Gulf of Kuwait, Kuwait	2001	BRS
FE0205A1AK	Feces	Calgary, AB, Canada*	2002	BRS
FE0207AMB	Feces	Meaford, ON, Canada	2002	BRS
FE0303A1YO	Feces	Toronto, ON, Canada*	2003	BRS
SO300A1	Soil	Kuujuaraapiik, QC, Canada	2003	BRS
INWB2202A2	Seal intestine	Kangiqsualujjuaq, QC, Canada	2004	BRS
Type AB				
NCTC 2916	Canned corn	Colorado	1929	M. Wictome/IFR
FE9504ACG	Feces	Baie Comeau, QC, Canada*	1995	BRS
Type B				
IB1-B	Feces	Peterborough, ON, Canada*	1979	BRS
MRB	Mushrooms	Montreal, QC, Canada	1973	BRS
13983B	Asparagus	Unknown	1962	2/NCA
427-2-76	Honey	Modesto, CA	1976	D. Arvant
426B	Honey	Modesto, CA	1976	R. Wood/T. Midura
1366-1-77	Honey	San Francisco, CA*	1977	T. Midura
1344-1-77	Feces	San Francisco, CA*	1977	T. Midura
920A276	Feces	Modesto, CA*	1976	T. Midura
368B	Feces	Modesto, CA*	1976	T. Midura
BL 143	Fish	United Kingdom	1986	IFR
FE9508BRB	Feces	Sherbrooke, QC, Canada	1995	BRS
FE9508BPD	Feces	Sherbrooke, QC, Canada	1995	BRS
PA9508B	Pate	Sherbrooke, QC, Canada	1995	BRS
FE9904BMT	Feces	Toronto, ON, Canada	1999	BRS
GA0108BEC	Gastric liquid	Rivière-du-Loup, QC, Canada	2001	BRS
FE0507BLP	Feces	Ottawa, ON, Canada*	2005	BRS
EN0509BLP	Dust	Ottawa, ON, Canada*	2005	BRS
Type F				
Langeland	Liver paste	Langeland, Denmark	1960	V. Moller/Catherwood[e]
H461297F	Honey	Wisconsin	1998	BRS
Group II				
Type B				
2B	Marine sediment	Pacific coast, United States	1960s	M. W. Eklund/H. Solomon[f]
17B	Marine sediment	Pacific coast, United States	1960s	M. W. Eklund/H. Solomon
DB-2	Marine sediment	Pacific coast, United States	1968	M. W. Eklund/H. Solomon
Prevot 59				IFR
KAP-B-3	Kapchunka	California	1981	H. Solomon
KAP-B-8	Kapchunka	California	1981	H. Solomon
II60-15B	Feces	British Columbia, Canada	1982	BRS
Type E				
Eruss	Sturgeon	Russia	1935	24/NR[g]
Gordon	Clinical	Kuujjuaq, QC, Canada	1975	L. Gauvreau[h]
Bennett	Gastric liquid	Happy Valley, NL, Canada	1976	BRS
FE9507EEA	Feces	Kangiqsualujjuaq, QC, Canada	1995	BRS
MI9507E	Misiraq	Kangiqsualujjuaq, QC, Canada	1995	BRS
F9508EMA	Feces	Kuujjuaq, QC, Canada	1995	BRS
VI9508E	Seal meat	Kuujjuaq, QC, Canada	1995	BRS
FE9508EPB	Feces	Tasiujaq, QC, Canada	1995	BRS
S9510E	Seal meat	Kuujjuaq, QC, Canada	1995	BRS
GA9709EHS	Gastric liquid	Kangiqsualujjuaq, QC, Canada	1998	BRS

Continued on following page

TABLE 1—*Continued*

Strain	Origin	Location[j]	Yr of isolation	Source or reference[a]
GA9709ENS	Gastric liquid	Kangiqsualujjuaq, QC, Canada	1998	BRS
GA9709EJA	Gastric liquid	Kangiqsualujjuaq, QC, Canada	1998	BRS
FE9709ELB	Feces	Kangiqsualujjuaq, QC, Canada	1998	BRS
FE9709EBB	Feces	Kangiqsualujjuaq, QC, Canada	1998	BRS
FE9909ERG	Feces	Inuvik, NU, Canada	1999	BRS
FE0005EJT	Feces	Inuvik, NU, Canada	2000	BRS
MU0005EJT	Muktuk	Inuvik, NU, Canada	2000	BRS
SW280E	Saltwater	Koksoak River, QC, Canada	2001	BRS
SO309E2	Shoreline soil	Hudson Bay, QC, Canada	2001	BRS
SO325E1	Shoreline soil	Ungava Bay, QC, Canada	2001	BRS
SO326E1	Shoreline soil	Ungava Bay, QC, Canada	2001	BRS
SO329E1	Shoreline soil	Ungava Bay, QC, Canada	2001	BRS
SP455456E2	Coastal rock	Ungava Bay, QC, Canada	2001	BRS
SOKR-23E1	Marine sediment	Koksoak River, QC, Canada	2002	BRS
SOKR-29E1	Shoreline soil	Ungava Bay, QC, Canada	2002	BRS
SOKR-37E2	Freshwater sediment	Koksoak River, QC, Canada	2002	BRS
HNB0804E	Honey	New Brunswick, Canada	2004	BRS
Type F				
70F	Marine sediment	California coast	1960s	M. W. Eklund/H. Solomon
190F	Marine sediment	California coast	1960s	M. W. Eklund/H. Solomon
202F	Marine sediment	California coast	1965	M. W. Eklund/H. Solomon
205F	Marine sediment	California coast	1960s	M. W. Eklund/H. Solomon
610F	Salmon	Columbia River, OR	1966	Craig/H. Solomon
19501F	Marine sediment	Oregon coast	1960s	M. W. Eklund/J. S. Crowther[i]

[a] First name is the initial source and the second is the donor source.
[b] M. Peck, Institute for Food Research, Norwich, United Kingdom.
[c] National Canners Association.
[d] Botulism Reference Service for Canada, Health Canada, Ottawa, Ontario, Canada.
[e] Canadian Food Inspection Agency.
[f] H. Solomon, U.S. Food and Drug Administration, Washington, DC.
[g] NR, no records of the donor source found.
[h] L. Gauvreau, Laboratoire de microbiologie, Centre hospitalier de l'Université Laval, Ste-Foy, QC, Canada.
[i] J. S. Crowther, Unilever Research, Sharnbrook, Bedford, England.
[j] An asterisk indicates a strain associated with a case of infant botulism.

Flagellin gene sequences were obtained from this study with the exception of Hall A CBO0242, CBO2666, CBO2695, and *flaA1/2* sequences, which were extracted from the Hall A genome sequence (*C. botulinum* genome sequence from the *C. botulinum* Sequencing Group at the Wellcome Trust Sanger Institute ftp://ftp.sanger.ac.uk/pub/pathogens/cb). Flagellin protein sequences used for dendrogram construction were translated from gene sequences obtained in this study, the Hall A genome sequence (see above), or GenBank (see figure legends for accession numbers). In silico translation of *flaA* and *flaB* was performed using ExPASy proteomic tools (http://ca.expasy.org/tools/). To avoid analyzing multiple flagellin variants from a population survey, a representative single flagellin protein sequence was selected for a single species. Multiple flagellins present in one contiguous genome or flagellins that originated from only one species were also included. Dendrograms were generated using Bionumerics 4.5 (Applied Maths, Kortrijk, Belgium). The dendrogram comparing flagellin gene sequences used the unweighted-pair group method with arithmetic averages (UPGMA) with default settings and the neighbor-joining method incorporating the Kimura 2P parameter. The dendrogram comparing flagellin proteins was generated using UPGMA followed by the neighbor-joining method, but no correction was applied. To accommodate alignment of flagellins from multiple species and differing in molecular mass by as much as 25 kDa, the maximum number of gaps allowed was increased from default to 98 and the gap penalty was reduced to 50%. All dendrograms were bootstrapped using 2,000 simulations.

Variable and hypervariable regions in flagellin sequences were assigned by inspection of sequence alignments. Assignment of *flaB* as a flagellin gene was determined by comparison of the predicted amino acid sequence to the NCBI database using BlastP and tBlastX.

Nucleotide sequence accession numbers. All full and partial sequences of *C. botulinum* flagellin genes were deposited in GenBank with the following accessions numbers: *flaB*, DQ658239; and *flaA*, DQ844946 to DQ845031.

RESULTS

TEM shows variation in the degree of flagellation among *C. botulinum* strains. Fifteen strains of *C. botulinum* were examined for flagellation by transmission electron microscopy (TEM). *C. botulinum* strain Hall A was not flagellated (data not shown); however, almost all other group I strains produced multiple peritrichous flagella when grown in SPGY broth (Fig. 1A and B). Group I type F strain Langeland produced only one or two filaments per cell. All group II strains examined produced multiple filaments (Fig. 1C and D). The flagella from group II type F strain 610F appeared short and fragile, with some cells observed not producing filaments (data not shown).

To resolve the protein components making up the filaments, flagella were sheared from the surface of *C. botulinum* cells. Examination by TEM of initial flagellar preparations from select group I and group II strains showed that the shearing technique isolated large amounts of long intact flagellar filaments. Several filaments demonstrated the characteristic single-ring structure of a gram-positive flagellar basal body (Fig. 1C, inset).

Group I and group II strains produce both common and distinct flagellins. Flagellar filaments sheared from the surface of *C. botulinum* strains were analyzed by SDS-PAGE, and the resultant protein profiles for group I strains (Fig. 2A) differed significantly from those of group II (Fig. 2B). Molecular

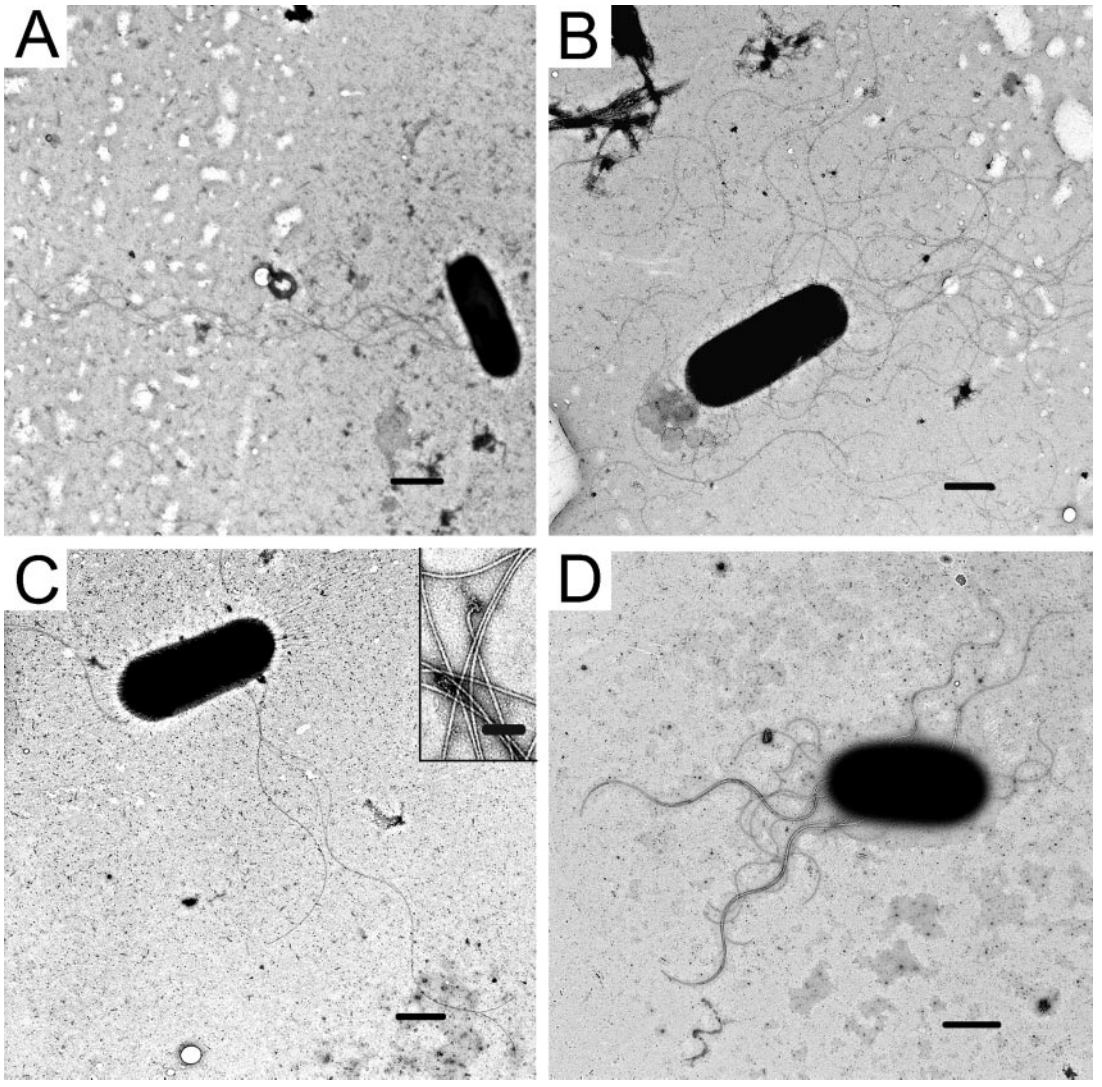


FIG. 1. TEMs of negatively stained *C. botulinum* showing the flagellation of group I and group II strains. Although several strains did not produce flagellar filaments, the majority of strains produced several to multiple peritrichous flagella. (A) Type A strain FE9909ACS Alberta (scale bar = 2.5 μm). (B) Group I type B strain FE0507BLP (scale bar = 1.1 μm). (C) Group II type B strain KapB3 has several filaments (scale bar = 1.1 μm); inset panel, group II flagellar filament from strain GA9709EHS with basal body and hook structure attached (scale bar = 0.2 μm). (D) Type E strain Bennett (scale bar = 2.5 μm).

weights of the putative flagellin protein bands were assessed by comparison to molecular mass standards and are listed in Table 2. Sheared flagellar filaments, regardless of group or serotype, contained one to three flagellin subunits with molecular masses ranging from 29 to 32 kDa (Fig. 2A and B; Table 2). Although not apparent in Fig. 2, type E strains also produced these smaller flagellins; however, they were only visible when large amounts of sample were loaded (data not shown). Protein samples prepared from strain Hall A did not contain any predominant protein, confirming the observation by TEM that this strain does not produce flagella (Fig. 2A, lane 1). In contrast to group I, flagellar proteins isolated from group II strains produced multiple flagellins with much larger differences in apparent molecular mass than those observed for group I flagellins (Fig. 2, compare panels A and B). In addition to the 29- to 33-kDa flagellins common to both groups, SDS-PAGE of group II flagellar preparations showed protein bands with approximate apparent molecular masses of 38 to 42 kDa and/or 48 to 53 kDa, suggesting the existence of additional distinct putative flagellins. Within group II, the 48- to 53-kDa flagellins were only observed in type E strains (Fig. 2B, lanes 3 to 5). To confirm that the difference in flagellin proteins between groups was not due to growth at different temperatures for the two groups, flagella were isolated from group I strains grown at room temperature and group II strains grown at 35°C and flagellin proteins were analyzed by SDS-PAGE. No changes were detected in either the amounts or observed molecular masses of the flagellin proteins (data not shown).

Identification of putative flagellin genes in the Hall A genome sequence. tBlastN searches using protein sequences of known flagellins from *Bacillus* and *Clostridium* species (GenBank accession numbers: *C. tyrobutyricum* flagellin,

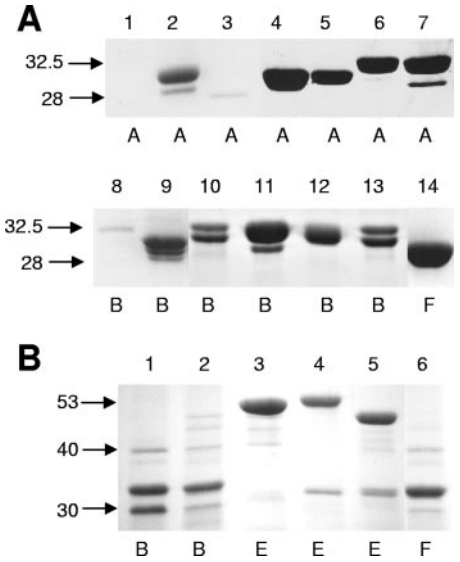


FIG. 2. SDS-PAGE profiles of flagellar proteins. (A) Group I flagellins. Lane 1, Hall A; lane 2, FE9504ACG; lane 3, 17A; lane 4, FE9909ACS Alberta; lane 5, FE0303A1YO; lane 6, MUL0109ASA; lane 7, PC0101AJO; lane 8, MRB; lane 9, FE0507BLP; lane 10, PA9508B; lane 11, 920A276; lane 12, FE9904BMT; lane 13, FE9508BPD; and lane 14, Langeland. (B) Group II flagellins. Lane 1, 17B; lane 2, KapB3; lane 3, ERuss; lane 4, Bennett; lane 5, GA9709EHS; and lane 6, 610F. Molecular masses in kilodaltons are indicated by arrows. Serotypes of strains are indicated by a letter below each lane. All predominant bands were identified as flagellin by mass spectrometry, with the exception of the 30-kDa molecular mass band in panel B, lanes 1, 2, and 6 (see Table 2 and the text). The minor band at 38 kDa in lanes 1, 2, and 6 was also identified as flagellin.

CAB44444.1; *C. difficile*, AAD46086.1; *C. chauvoei*, BAB13814.1; and *B. subtilis*, BAB58984.1) identified five ORFs within the Hall A genome sequence as putative flagellin genes (CBO0242, CBO2666, CBO2695, CBO2730, and CBO2731; Fig. 3A). Four of the five putative Hall A flagellins had homologs within the *C. tetani* E88 genome sequence, with the genetic organization of the four ORFs also conserved. CBO0242, CBO2666, CBO2695, CBO2730, and CBO2731 all encoded putative flagellin proteins with predicted molecular masses of 29.8 to 31.2 kDa. Compared via BlastP to the NCBI GenBank database, all five ORFs matched numerous bacterial flagellins and contained protein subdomains that corresponded to conserved flagellin N and C termini domains (Fig. 3; also data not shown). ORFs CBO2730 and CBO2731, located in tandem orientation in the Hall A genome (base pairs 2882498 to 2883325 and 2883539 to 2884366, respectively), were 99% identical and were similar to the structural subunit of the well-characterized *B. subtilis* flagellin, Hag. CBO0242, CBO2666, and CBO2695 were all homologous to other known flagellins but represented unique and individual putative Hall A flagellin genes (Fig. 3B).

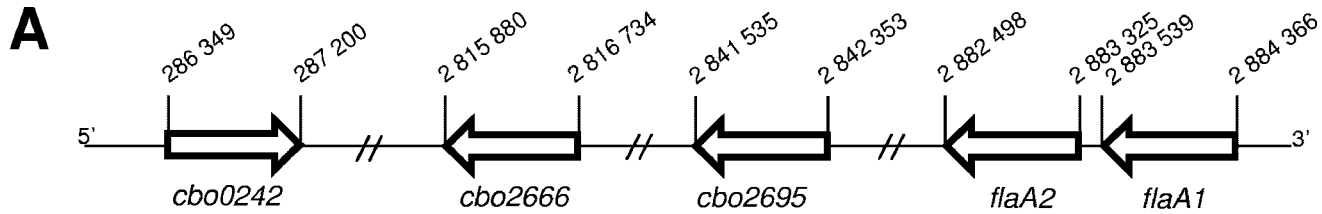
The alternative RNA polymerase sigma factors σ^D, in gram-positive bacteria, or σ^F, in gram-negative bacteria, transcribe flagellin structural genes and recognize highly conserved −35 or −10 promoter elements (16). We were unable to identify conserved promoter signatures for σ^D, σ^{54}, or σ^{70} upstream of any of the putative flagellin ORFs (13, 16, 32). One explanation for a lack of conserved promoter elements is the low GC content of the *C. botulinum* upstream sequences (22%); however, sequences similar to low GC flagellar promoter regions also could not be identified upstream of any of the *C. botulinum* flagellins (14, 35). The only consensus regulatory sequence identified was the ribosome binding site (CAGGG AGGAA) of the *B. subtilis* flagellin structural gene, *hag* (32). This motif was identified upstream of ORFs CBO2730 and CBO2731.

Characterization of the flagellin structural proteins, FlaA1 and FlaA2. To determine if any of the five Hall A flagellin ORFs were present as flagellar filament structural subunits, peptide sequences from putative group I and group II flagellins previously identified by SDS-PAGE were obtained by nLC-MS/MS (Table 2 and Fig. 4). Peptide MS/MS spectra from most protein bands matched the predicted amino acid sequences for ORFs CBO2730 (predicted molecular mass, 29.816 kDa) and CBO2731 (predicted molecular mass, 29.845 kDa) (Fig. 4A). We have designated CBO2731 as *flaA1* and CBO2730 as *flaA2*, as our data indicate that these homologous genes encode the major structural proteins of the *C. botulinum* flagellar filament. No peptides matched any of three additional putative flagellins identified in this study or the putative flagellin identified by Dineen et al. (10), indicating that these flagellins are either not present in the purified filaments or are present as minor components. A 30-kDa band in group II type

TABLE 2. Identification of flagellin proteins by tryptic digest and nLC-MS/MS analysis

Strain	Mol mass (kDa)		MASCOT score (no. of peptides)
	Predicted	Observed	
Group I			
Type A			
FE9909ACS Alberta	29.4	30.3	450 (8)
17A	30	28.0	308 (6)
Type B			
MRB	29.5	32.1	253 (5)
FE0507BLP	29.6	28.9	213 (3)
		30.7	174 (2)
		32.7	206 (3)
Type AB			
FE9504ACG	29.7	28.6	363 (6)
		30.5	265 (5)
Type F			
Langeland	29.4	29.3	133 (3)
Group II			
Type B			
17B	29.1	32.6	132 (2)
		38.3	79 (1)
		43.6	86 (1)
Kapchunka B3	29.1	33.0	81 (1)
Type E			
Bennett	28.9	30.2[a]	81 (1)
		31.8	188 (3)
		53.0	84 (1)
GA9709EHS	28.9	30.2[a]	437 (7)
		31.5	176 (3)
		47.7	102 (1)
Type F			
610F	29.3	33.2	82 (1)
		38.8	70 (1)
		41.9	79 (1)

[a] Bands were excised from overloaded SDS-PAGE gels and are not visible in Fig. 2.



Gene size (bp)	851	854	818	827	827
Locus	CBO0242	CBO2666	CBO2695	CBO2730	CBO2731
Predicted molecular mass (kDa)	30.376	31.281	30.364	29.845	29.816
Highest BlastP hit (Locus; Genbank Accession number)	Flagellin *Geobacillus kaustophilus* HTA426 (GK3131; BAD77416)	Flagellin *Clostridium tetani* E88 (CTC01679; NP_782278)	Flagellin *Clostridium tetani* E88 (CTC01691; NP_782289)	Flagellin *Clostridium tetani* E88 (CTC01715; NP_782313)	Flagellin *Clostridium tetani* E88 (CTC01715; NP_782313)
% identical residues (sequence coverage)	50% (152/301)	44% (125/284)	47% (126/267)	78% (218/277)	78% (218/277)
Expect value	$2e^{-68}$	$3e^{-65}$	$7e^{-63}$	$3e^{-105}$	$2e^{-105}$

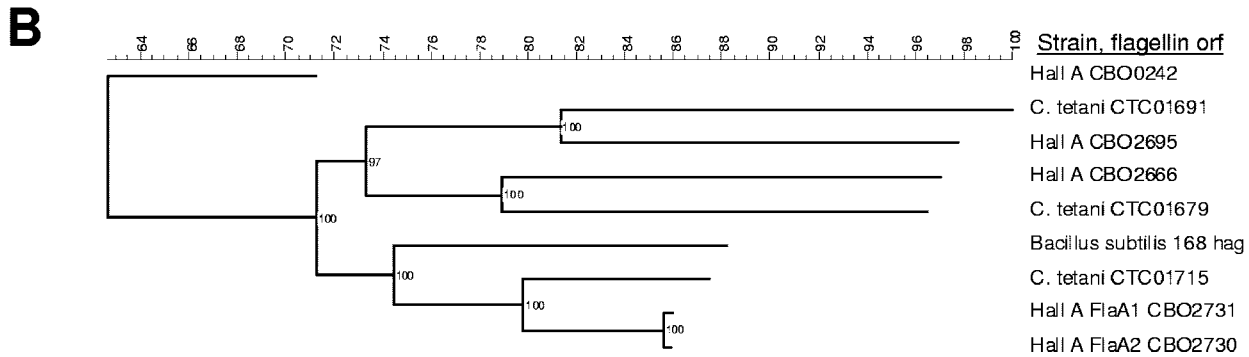


FIG. 3. Identification of putative flagellin genes in the Hall A genome sequence. (A) Arrows indicate the ORF identified as a putative flagellin with numbers indicating the location in base pairs of the ORF in the Hall A genome DNA sequence. (B) Comparison of the protein sequences of all Hall A putative flagellins and flagellin homologs in *C. tetani* and *B. subtilis.* Of the putative Hall A flagellins, CBO2730 and CBO2731 are most similar in protein sequence to the known structural subunit Hag of *B. subtilis* flagellin. The scale above the dendrogram is proportional to the similarity between sequences. Bootstrap values from 2,000 simulations are included adjacent to each branch.

B and F strains and a 50-kDa band in a group II type B strain (Fig. 2, lane 2) contained peptide sequences matching subunits of an electron-transferring flavoprotein (*Clostridium beijerinckii*), an NAD-dependent beta-hydroxybutyryl coenzyme A (CoA) dehydrogenase (*Clostridium saccharobutylicum*), 3-hydroxybutyryl-CoA dehydrogenase (*C. tetani* E88, locus tag CTC02423), and related proteins. An enzyme complex comprising propionyl-CoA dehydrogenase and electron-transferring flavoprotein has been characterized in *Clostridium propionicum*, where it comprises an estimated one-third of the soluble protein in a cell extract, suggesting that in this study it may be a contaminant of the flagellar preparations (17). A 30-kDa band was also observed in type E strains, and this band did contain peptides matching FlaA1/A2. This flagellin protein, however, is only a minor component of type E flagellar filaments, as it was observed only when SDS-PAGE gels were overloaded fivefold and is not visible in Fig. 2B (lanes 3 to 5).

Identification of the flaVR. While MS/MS of tryptic peptides identified the expressed flagellin proteins as FlaA1/A2, the majority of flagellins observed by SDS-PAGE showed apparent molecular weights that differed from, and in some cases were significantly larger than, those predicted for Hall A FlaA1/A2. The *flaA1/A2* genes were amplified from each strain for which the flagellin proteins had been examined by MS/MS, to determine the predicted molecular weight for each FlaA1/A2 protein (Table 2). By using primers cbotflaF and cbotflaR, designed to the conserved regions of Hall A *flaA1/A2*, an amplicon corresponding to the approximate expected size of 827 bp was obtained from each strain. Cloning and sequencing of each product gave an ORF that when trans-

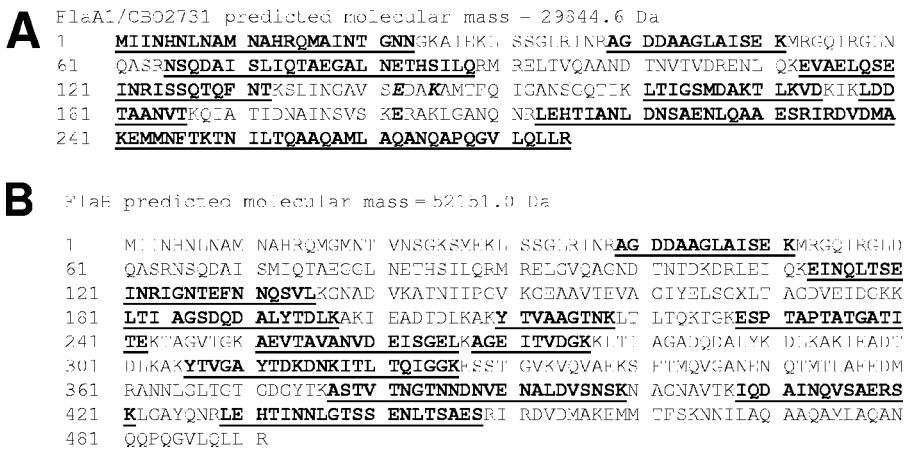


FIG. 4. nLC-MS/MS analysis matched peptides from the tryptic digests of the expressed flagellin proteins to FlaA1 (Hall A) or FlaB (Bennett). (A) Protein sequence for Hall A FlaA1/CBO2731. Peptides in bold and underlined were identified following MS/MS of various flagellin proteins from different strains of group I and group II *C. botulinum.* At least one peptide matching the Hall A FlaA1 sequence was identified in all flagellin proteins examined. Amino acids 142 and 145 are indicated in bold with italics: these residues differ between FlaA1 and FlaA2 protein sequences ($E_{142} \rightarrow K_{142}$ and $K_{145} \rightarrow D_{145}$, respectively). (B) FlaB from type E strain Bennett. Peptides identified by MS/MS of FlaB are in bold and underlined. While some peptides (i.e., AGDDAAGLAISEK) were present in both FlaA1/A2 and FlaB, a number of peptides matched only the sequence predicted for FlaB. For MASCOT scores and the number of peptides from each putative flagellin protein matching Hall A FlaA1 or FlaB (when applicable), see Table 2.

lated was identified as *flaA1/A2* by extensive sequence identity to the conserved regions of Hall A *flaA1/A2*, as expected. Alignment of DNA sequences identified a flagellin variable region (flaVR) corresponding to base pairs 277 to 672 of the Hall A *flaA1/A2* (data not shown). Sequencing of independent *E. coli* clones originating from the same PCR amplicon in some cases allowed the resolution of the separate *flaA1* and *flaA2* allelic ORFs. Alignment of *flaA1* and *flaA2* sequence pairs from FE9504ACG, GA9709EHS, and the Hall A genome showed that, while the majority of the flaVR sequence was conserved between the *flaA1* and *flaA2* of a given strain, single base pairs and, in two cases, small clusters of base pairs were not conserved. Base pairs differing between *flaA* alleles were designated as a distinct allelic sequence (DAS). In some cases, the location of the DAS differed between group I and group II *flaA1* and *flaA2* pairs: the two small clusters of DAS were located at base pairs 420 to 438 and base pair 486 in group I strains and base pairs 541 to 553 and base pair 569 in group II strains (all base pair locations reference Hall A *flaA1*).

Group I and group II strains posttranslationally modify FlaA1/A2. The predicted molecular masses for FlaA1/A2 for several strains were determined from *flaA1/A2* sequences for the corresponding strain (Table 2). Comparison of the observed molecular masses of FlaA1/A2 proteins, following SDS-PAGE, to that predicted from the translated *flaA1/A2* sequences showed that most FlaA1/A2 proteins were larger than expected. Several species of clostridia are known to posttranslationally modify flagellin proteins. Several bacterial species, including *Campylobacter jejuni*, *Pseudomonas aeruginosa*, and *Listeria monocytogenes*, glycosylate their flagellin proteins (27). The putative modifications could not be immediately identified as carbohydrates, however, as both a periodate oxidation-based assay to detect glycan modification and screening with a panel of digoxigenin-labeled lectins failed to react with any of the *C. botulinum* flagellins (data not shown).

Identification of *flaB*, a novel type E flagellin. During the amplification of *flaA1/A2* from type E strains, a second amplicon of 1.3 kb was detected in addition to the expected PCR product for *flaA1/A2*. Cloning and sequencing of the 1.3-kb PCR product from type E strain Bennett identified a putative flagellin ORF with similar conserved regions to that of Hall A *flaA1/A2* but with a substantially larger and distinct variable region, comprising 1,000 bp in contrast to the roughly 400-bp region assigned for the *flaA* genes. nLC-MS/MS analysis of tryptic digests of a 53-kDa band present in Bennett flagellar preparations matched the peptide sequences predicted from this larger flagellin gene (Fig. 4B). Due to the presence of multiple flagellins as well as the similarity in the N- and C-terminal protein sequence to that of the smaller FlaA1/A2, we have designated the larger flagellin gene as *flaB* (GenBank accession number DQ658239). FlaB differs from FlaA in that it does not appear to be posttranslationally modified: translation of *flaB* gave a predicted molecular mass of 52.151 kDa, closely correlating to the 53-kDa observed molecular mass of the large flagellin from strain Bennett (Fig. 2B, lane 4). As a large flagellin protein was only observed in flagellar filaments isolated from type E *C. botulinum*, and a large *flaB* ORF did not exist in the Hall A genome sequence, FlaB appears to be a type E-specific flagellin.

BlastP analysis suggested that FlaB is one of several large clostridial flagellins, with 46% identity to FliC of *C. chauvoei*, 38% identity to FliC of *C. haemolyticum*, and 37% identity to FliC of *C. septicum*. Dendrogram comparison of FlaB to other known and putative *Clostridium* flagellin sequences reflected this sequence conservation, clustering FlaB with the flagellin from *C. chauvoei* and *C. septicum* while the FlaA1/A2 sequences from type E strain Bennett clustered with the Hall A FlaA1/A2 proteins and a flagellin from *C. tetani* (Fig. 5). The FlaA1/A2 proteins clustered together with the well-characterized *B. subtilis* Hag flagellin subunit. FlaB clustered in a second

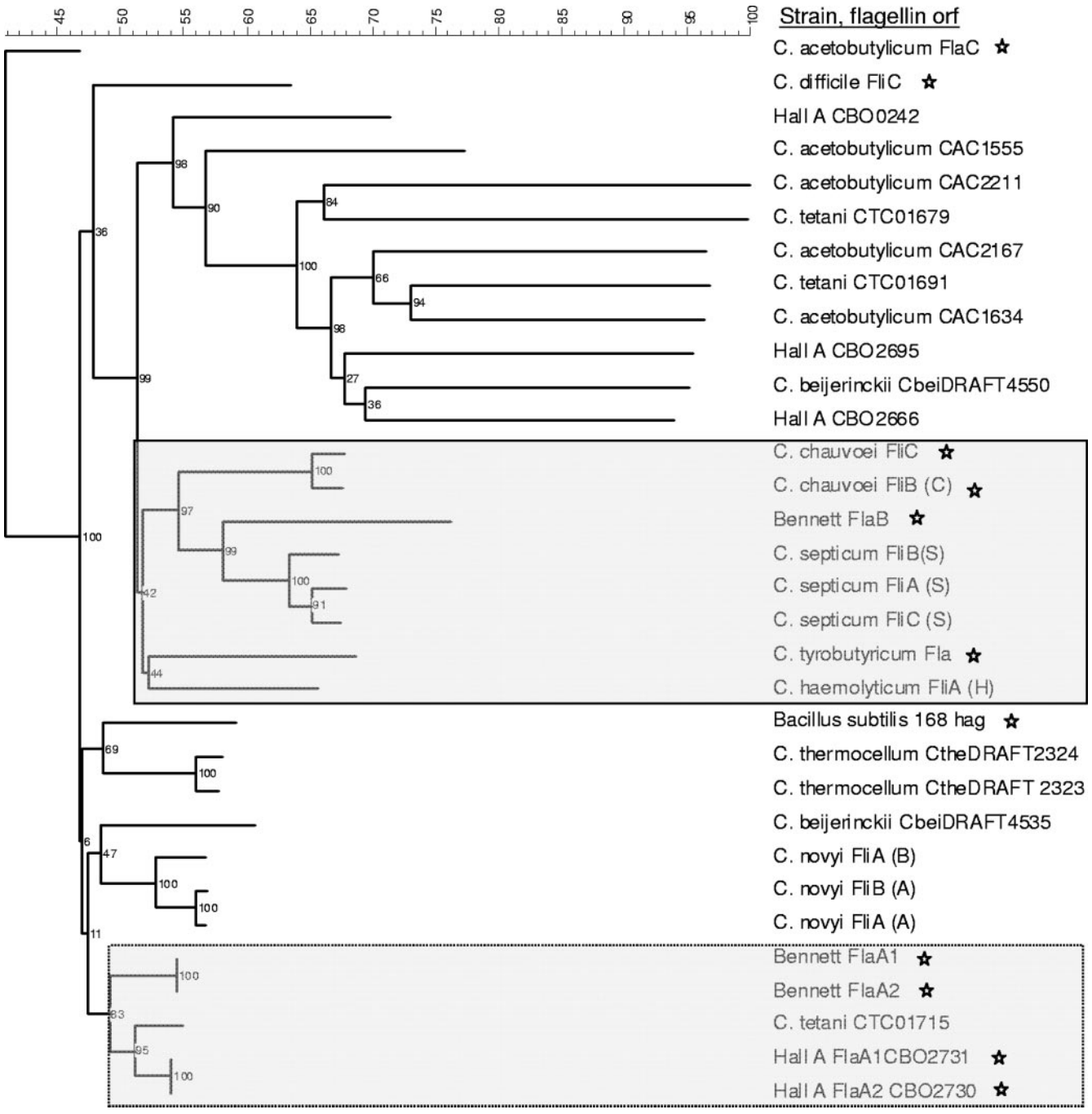
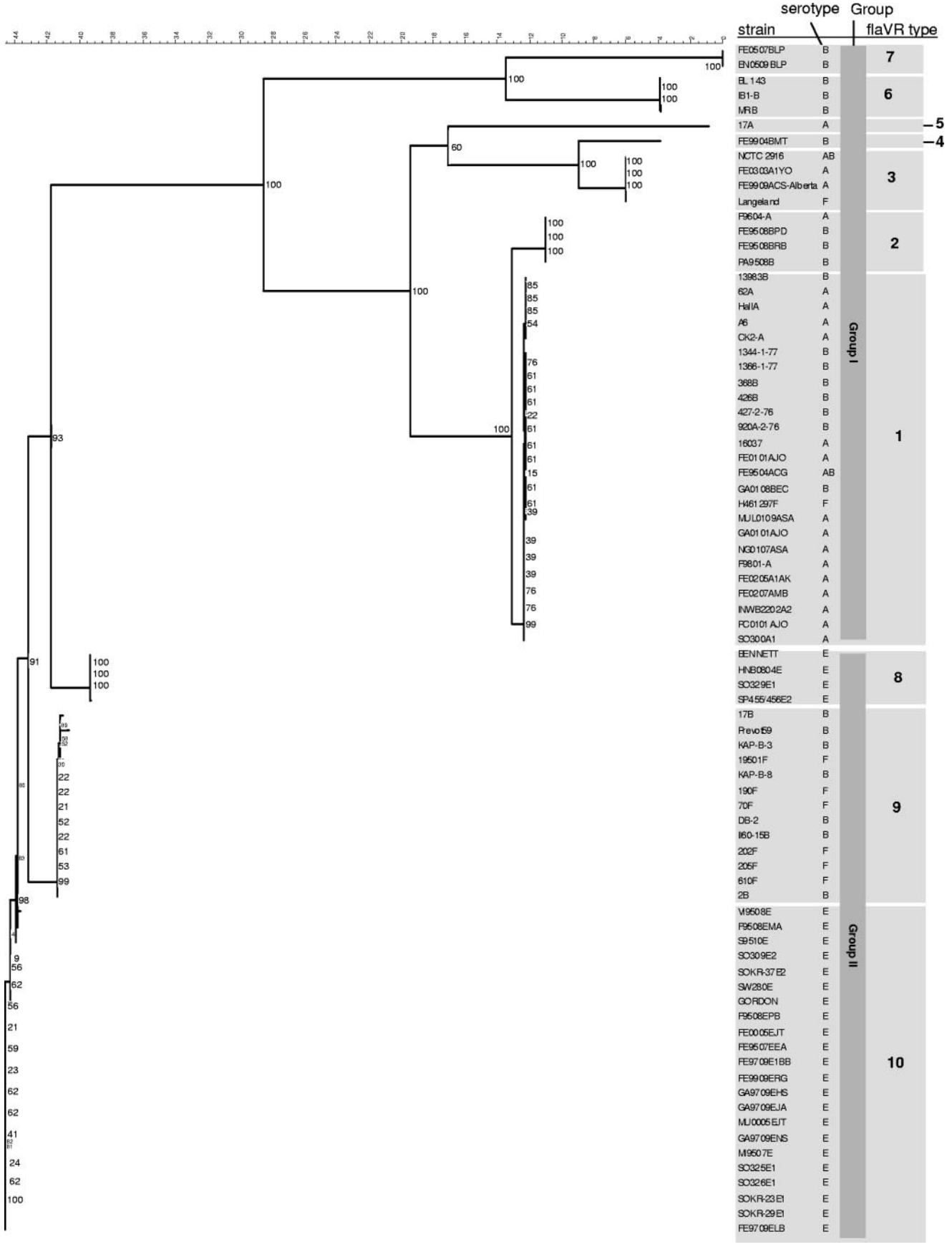


FIG. 5. Dendrogram analysis of protein sequences for known flagellins from different members of the genus *Clostridium.* Protein sequences for *C. botulinum* FlaA1/A2 from strain Hall A and strain Bennett and FlaB from strain Bennett were compared by generating a dendrogram using the UPGMA and nearest-neighbor-joining methods in Bionumerics 4.5 with a large maximum gap size of 98 to accommodate flagellins with large differences in molecular mass. Flagellins indicated with an asterisk have been documented in the literature as expressed proteins. In the case of ORF pairs, where it was not specified which was expressed, both ORFs were given an asterisk. Gray shaded boxes show subclusters containing *C. botulinum* flagellins. *C. botulinum* Hall A FlaA1 and FlaA2 sequences were obtained from the Hall A genome project at www.sanger.ac.uk/Projects/C_botulinum. For all other flagellin sequences, GenBank accession numbers are as follows: *C. botulinum* Bennett FlaA1 (DQ845000), FlaA2 (DQ845001), and FlaB (DQ658239.1); *C. acetobutylicum* ATCC 824 (genome strain) FlaC (AAC16553.1), CAC1634 (AAK79601.1), CAC1555 (AAK79522.1), CAC2211 (AAK80168.1), and CAC2167 (AAK80125.1); *C. beijerinckii* NCIMB 8052 CbeiDRAFT4550 (ZP_00907552.1) and CbeiDRAFT4535 (ZP_00907537.1); *C. chauvoei* FliC (BAB13814.1) and FliB (C) (BAB87728.1); *C. difficile* FliC (AAF09167.1); *C. haemolyticum* FliA (H) (BAB87738.1); *C. novyi* FliA (B) (BAB87737.1), FliB (A) (BAB87735.1), and FliA (A) (BAB87734.1); *C. septicum* FliA (S) (BAB87729.1), FliC (S) (BAB87732.1), and FliB (S) (BAB87730.1); *C. tetani* E88 (genome strain) CTC01679 (AAO36215.1), CTC01691 (AAO36226.1), and CTC01715 (AAO36250.1); *C. thermocellum* CtheDRAFT 2324 (ZP_00504695.1) and CtheDRAFT 2323 (ZP_00504694.1); *C. tyrobutyricum* Fla (CAB44444.1); and *B. subtilis* 168 genome strain Hag (NP_391416.1). The scale above the dendrogram is proportional to the similarities between sequences. Bootstrap values from 2,000 simulations are included adjacent to each branch.


strain
serotype
Group
flaVR type
FE0507BLP B
EN0509BLP B
BL143 B
IB1-B B
MRB B
17A A
FE9904BMT B
NCTC2916 AB
FE0303A1YO A
FE9909ACS-Alberta A
Langeland F
F9604-A A
FE9508BPD B
FE9508BRB B
PA9508B B
13983B B
62A A
HallA A
A6 A
CK2-A A
1344-1-77 B
1366-1-77 B
368B B
426B B
427-2-76 B
920A-2-76 B
16037 A
FE0101AJO A
FE9504ACG AB
GA0108BEC B
H461297F F
MJL0109ASA A
GA0101AJO A
NG0107ASA A
F9801-A A
FE0205A1AK A
FE0207AMB A
INWB2202A2 A
FC0101AJO A
SO300A1 A
BENNETT E
HNB0804E E
SO329E1 E
SP455/456E2 E
17B B
Prevot59 B
KAP-B-3 B
19501F F
KAP-B-8 B
190F F
70F F
DB-2 B
I60-15B B
202F F
205F F
610F F
2B B
M9508E E
F9508EMA E
S9510E E
SO309E2 E
SOKR-37E2 E
SW280E E
GORDON E
F9508EPB E
FE0005EJT E
FE9507EEA E
FE9709E1BB E
FE9909ERG E
GA9709EHS E
GA9709EJA E
MJ0005EJT E
GA9709ENS E
M9507E E
SO325E1 E
SO326E1 E
SOKR-23E1 E
SOKR-29E1 E
FE9709ELB E
Group I
Group II
1
2
3
4
5
6
7
8
9
10

major branch that divided into two separate clusters, with one containing FlaB and its homologs while the other contained the putative Hall A flagellins, CBO0242, CBO2666, and CBO2695. Interestingly, in addition to *C. botulinum*, *C. novyi*, *Clostridium thermocellum*, *C. septicum*, and *C. chauvoei* also contained multiple copies of similar flagellins.

While the *flaA1/A2* and *flaB* genes account for the 29- to 33-kDa and 48- to 53-kDa flagellin proteins, respectively, no gene was identified that would encode a 38- to 42-kDa flagellin protein. Peptides from several of the 38- to 42-kDa flagellin proteins matched the conserved N terminus of Hall A FlaA1/A2, suggesting that they may be FlaA1/A2 proteins with extensive posttranslational modification.

Sequencing of the *flaA1/A2* flagellin variable region from multiple strains of *C. botulinum*. The use of the flaVR sequence to replace traditional flagellar H antigen serotyping for identification of strains by single locus sequence typing has been effective in typing a variety of motile gram-negative bacteria (49). Identification of the *flaA1/A2* flaVR in both groups I and II suggested that a PCR-based single locus sequence typing method for *C. botulinum* would facilitate strain typing that would not be required to accommodate the physiological differences between the two groups.

To expedite the discrimination of *C. botulinum* isolates, colony PCR was used to amplify *flaA1/A2* sequences from an additional 67 strains. Direct sequencing of colony PCR amplicons with primers cbotflaF and cbotflaR gave sequences that completely spanned the flaVR region. A total data set of 80 *flaA1/A2* flaVR sequences was analyzed using Bionumerics 4.5. Only sequences corresponding to the flaVR were selected for dendrogram analysis: sequences corresponding to conserved N and C termini were excluded to maximize differences between strains; the DAS clusters (see above) were excluded to limit sequence variation being examined to that between strains and to exclude *flaA1* and *flaA2* allelic variation. The flaVR sequences clustered group I and group II strains into 10 flaVR types, with group I isolates falling into flaVR types 1 to 7 and group II strains into flaVR types 8 to 10 (Fig. 6). The largest group I cluster contained the sequence for the genome strain, Hall A, and was designated flaVR1.

Within group II, flaVR sequences were less variable (maximum 8% difference; Bennett and MU0005EJT) than those of group I strains (maximum 39% difference; FE0705BLP/EN0705BLP and 17A), with flaVR sequence clearly able to differentiate between group I and group II strains. No obvious correlation to geographic location, chronological time, food source, or type of botulism (i.e., food-borne, infant) could be resolved. Two flaVR types were defined by epidemiologically related strains: flaVR2 contained the epidemiologically related isolates PA9508B, FE9508BRB, and FE9508BPD and the unrelated strain F9604-A; and flaVR7 contained only FE0507BLP and EN0507BLP.

DISCUSSION

The flagella produced by group I and group II *C. botulinum* possess several standard characteristics of other well-characterized flagella. Electron microscopy showed a reduced number of basal body rings, consistent with the less-complex gram-positive membrane structure, a hook structure, and multiple, unsheathed, peritrichous filaments. Isolation of flagellar filaments and SDS-PAGE showed multiple flagellin monomer protein bands, suggesting that the filaments are complex in nature and not comprised of a single flagellin structural protein. This is in distinct contrast to the closely related species *Clostridium acetobutylicum*, where a single flagellin structural protein, FlaC, has been identified (29, 34, 35). Analysis of the *C. botulinum* Hall A genome sequence revealed five potential ORFs encoding flagellin proteins; however, nLC-MS/MS analysis identified peptides from purified flagellar filaments that matched only two of the putative ORFs. Arranged in tandem in the Hall A genome, these ORFs were designated *flaA1* and *flaA2* and, with the exception of type E strains, encode the major flagellin structural subunit of group I and group II *C. botulinum* strains.

Present in both group I and group II strains, the 29- to 33-kDa FlaA1/A2 flagellins had apparent molecular masses larger than those predicted from the corresponding *flaA1/A2* gene sequences, suggesting that these flagellins are posttranslationally modified. In some strains, e.g., FE0507BLP, two or more protein bands were resolved following SDS-PAGE of flagellin filaments, with all proteins identified as FlaA1/A2. As at the predicted molecular mass FlaA1 and FlaA2 would be indistinguishable by SDS-PAGE, the observed multiple banding patterns may be a result of different levels of FlaA1/A2 posttranslational modification occurring within an individual strain. Posttranslational modification of surface proteins is well characterized in gram-positive bacteria, and modification of flagellin proteins has been documented (40). While the precise modification is not known, *C. difficile* posttranslationally modifies flagellin, and both *C. tyrobutyricum* and *C. acetobutylicum* modify flagellin proteins with glycans (6, 29, 44). A number of gram-negative organisms have been shown to produce glycosylated flagellins (27).

The diversity in the molecular masses of flagellin proteins was observed to be greater among group II isolates than those of group I, with flagellin proteins of 38 to 42 kDa in type B and F strains and the additional 49- to 52-kDa FlaB in type E strains. While the *flaB* gene was identified, an ORF encoding a 38- to 42-kDa flagellin was not, and the only peptide se-

FIG. 6. Dendrogram analysis of the DNA sequence of the flaVR. A dendrogram was constructed for the flaVR DNA sequences from 80 *C. botulinum* strains using the UPGMA with default settings followed by a global alignment and the nearest-neighbor-joining method incorporating the Kimura 2P parameter. The designation and BoNT serotype for each strain are listed to the right of the tree. Light gray bars flank flaVR sequences clustering into group I or group II. Dark gray bars flank sequences of assigned flaVR types within group I or group II. The scale above the dendrogram is proportional to evolutionary time, due to incorporation of the Kimura 2P parameter for nucleotide sequences. Bootstrap values from 2,000 simulations are included adjacent to each branch.

quences obtained from the 38- to 42-kDa flagellins matched FlaA1/A2. A distinct genetic locus, that is not represented in the Hall A genome and cannot be amplified with the *flaA1/A2* primers, could be present in group II strains and contain sequence encoding conserved peptides common to both this ORF and the FlaA1/A2 proteins. Conversely, the approximately 12-kDa difference in mass could be a unique and larger form of posttranslational modification of the FlaA1/A2 proteins, distinct from that of the 28- to 32-kDa FlaA1/A2. *C. difficile* and *C. acetobutylicum* flagellins are significantly larger than their predicted molecular weights, with posttranslational modifications of approximately 8 kDa and 12.5 kDa, respectively (29, 44).

Within group II, type E strains were distinct in that their flagellar filaments contained primarily the large and apparently unmodified flagellin, FlaB. A recent study identified large (65-kDa) and small (33-kDa) flagellins that comprised the thick and thin filaments, respectively, of *Bradyrhizobium japonicum* flagella (20). TEM micrographs of several type E strains did not consistently show morphologically distinct filaments, suggesting that FlaA1/A2 and FlaB may be found in the same filament. Type E strains possess other unique characteristics, including the mechanism of regulation of the type E BoNT (9).

The majority of existing tools for genetic characterization of *C. botulinum* are able to discriminate to the strain level for group II but are limited for group I strains (26). Typing strains by sequencing of the neurotoxin gene still clusters large numbers of strains together, with maximum differences between subtypes not exceeding 8% for group I strains (18). Sequencing of type A neurotoxin genes showed that 90% of serotype A strains are of subtype A1, with divisions in this subcluster requiring resolution of a single base pair change in approximately 3.8 kb of gene sequence (18). 16S sequences within group I are also highly identical (8). The maximum diversity in group I flaVR sequence was 39% over a region of 400 bp, providing a level of strain-to-strain discrimination not previously possible within this group. This large region of diversity in the flaVR nucleotide sequence is advantageous for establishing epidemiological relationships as the likelihood that two strains will have 100% identical flaVR sequences is low, single base pair errors related to a DNA polymerase error during amplification or gene sequencing will not have a significant impact, and lastly, all sequence required for typing can be obtained in a single sequencing reaction. It must, however, be established to what overall extent *flaA* varies among *C. botulinum* strains, as the ultimate usefulness of this typing will be influenced by the total amount of variation at the *flaA* locus. For example, as strains FE0507BLP and EN0507BLP are the only two strains that possess flagellin genes in a deep-branching cluster, they are likely related. However, with many strains in the *flaA* type I cluster (HallA, 62A, and FE0303IYO), strains within this flaVR group are obviously not all from the same epidemiological source. PCR-based detection and typing based on the nucleotide sequence of the flagellin gene has allowed differentiation between closely related strains for *C. jejuni*, *Borrelia* spp., *P. aeruginosa*, and members of the *Burkholderia cepacia* complex (48, 49).

In group I, flaVR typing placed strains of serotypes A, B, and in some cases F in the same cluster. This is not surprising given that horizontal gene transfer of BoNT and associated genes has likely scattered these genes into various genetic backgrounds, including other clostridia species (10, 26, 37). With the mobility and extensive sequence conservation of the BoNT genes, inferring phylogeny based on the *C. botulinum* flagellin gene sequence could be preferable, with nucleotide changes in the variable region representing a passage of time over which the mutations have been accumulated. As tandem flagellin genes are subject to recombination, the stability of the *flaA* sequence would need to be established, although in contrast to the flagellins in *Campylobacter*, conservation of DNA sequence between *flaA1* and *flaA2* suggests that recombination might not be noticeable (15). Since identification of strains by either flagellin or BoNT sequencing appears uncorrelated, sequencing of BoNT genes and *flaA1/A2* would rapidly differentiate isolates of the same BoNT serotype. This approach would be analogous to *Escherichia coli* O and H antigen typing.

Within group II, flaVR sequences were less diverse than those within group I, with three group II flaVR types loosely associated with BoNT serotype, and flaVR sequence differing only by a maximum of 8% at the nucleotide level. PFGE analysis of the group II strains used in this study resolved significant genetic differences while still clustering group II serotype B and F strains separately (25), showing that for group II strains, other methods may be preferred for phylogenetic analysis. Sequencing of the flaVR region did clearly separate group I strains from those of group II, suggesting that rapid methods such as restriction fragment length polymorphism, based on group-specific sequences, could first identify the group of a strain and thus determine the most suitable downstream method (flaVR sequencing, PFGE) for genotypic analysis.

While gene sequence analysis based on BoNT requires multiple PCR conditions or previous knowledge of a strain's serotype, the *flaA* gene, by virtue of DNA sequence conserved across groups I and II, can be amplified in a single PCR from an uncharacterized strain of *C. botulinum*. The ability of the *flaA1/A2* primers to detect group III and group IV *C. botulinum* strains and other clostridia species is currently being investigated. Analysis of flagellin proteins has shown that this diversity is further amplified at the protein level, where each strain produces flagellins in different amounts and with different apparent molecular masses. It thus appears that while the flagellin genes can be used to differentiate among strains of *C. botulinum*, the potential to further discriminate among isolates based on posttranslational modification of flagellin proteins remains to be investigated.

ACKNOWLEDGMENTS

We thank Greg Sanders for assistance with electron microscopy and Nadia Mykytczuk for preliminary flagellin peptide identification.

This work was supported by project CRTI-02-0091TA of the Defense Research and Development Canada CBRN Research and Technology Initiative (J.W.A. and S.M.L.).

REFERENCES

1. **Andersen, A. A.** 1951. A rapid plate method of counting spores of *Clostridium botulinum*. J. Bacteriol. **62:**425–432.
2. **Anellis, A., and R. B. Koch.** 1962. Comparative resistance of strains of *Clostridium botulinum* to gamma rays. Appl. Microbiol. **10:**326–330.
3. **Arnold, F., L. Bedouet, P. Batina, G. Robreau, F. Talbot, P. Lecher, and R. Malcoste.** 1998. Biochemical and immunological analyses of the flagellin of *Clostridium tyrobutyricum* ATCC 25755. Microbiol. Immunol. **42:**23–31.

4. **Austin, J. W.** 2001. *Clostridium botulinum*, p. 329–349. *In* M. P. Doyle, R. L. Beuchat, and T. J. Montville (ed.), Food microbiology: fundamentals and frontiers. ASM Press, Washington, DC.
5. **Bardy, S. L., S. Y. M. Ng, and K. F. Jarrell.** 2003. Prokaryotic motility structures. Microbiology **149:**295–304.
6. **Bedouet, L., F. Arnold, G. Robreau, P. Batina, F. Talbot, and A. Binet.** 1998. Evidence for an heterogeneous glycosylation of the *Clostridium tyrobutyricum* ATCC 25755 flagellin. Microbios **94:**183–192.
7. **Blocker, A., K. Komoriya, and S.-I. Aizawa.** 2003. Type III secretion systems and bacterial flagella: insights into their function from structural similarities. Proc. Natl. Acad. Sci. USA **100:**3027–3030.
8. **Collins, M. D., and A. K. East.** 1998. Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. J Appl. Microbiol. **84:**5–17.
9. **Couesnon, A., S. Raffestin, and M. R. Popoff.** 2006. Expression of botulinum neurotoxins A and E, and associated non-toxin genes, during the transition phase and stability at high temperature: analysis by quantitative reverse transcription-PCR. Microbiology **152:**759–770.
10. **Dineen, S. S., M. Bradshaw, and E. A. Johnson.** 2003. Neurotoxin gene clusters in *Clostridium botulinum* type A strains: sequence comparison and evolutionary implications. Curr. Microbiol. **46:**345–352.
11. **Duff, J. T., G. G. Wright, J. Klerer, D. E. Morre, and R. H. Bibler.** 1957. Studies on immunity to toxins of *Clostridium botulinum*. J. Bacteriol. **73:**42–47.
12. **Fields, P. I., K. Blom, H. J. Hughes, L. O. Helsel, P. Feng, and B. Swaminathan.** 1997. Molecular characterization of the gene encoding H antigen in *Escherichia coli* and development of a PCR-restriction fragment length polymorphism test for identification of *E. coli* O157:H7 and O157:NM. J. Clin. Microbiol. **35:**1066–1070.
13. **Fredrick, K., T. Caramori, Y. Chen, A. Galizzi, and J. D. Helmann.** 1995. Promoter architecture in the flagellar regulon of *Bacillus subtilis*: high-level expression of flagellin by the sigmaD RNA polymerase requires an upstream promoter element. Proc. Natl. Acad. Sci. USA **92:**2582–2586.
14. **Ge, Y., I. G. Old, I. Saint Girons, and N. W. Charon.** 1997. Molecular characterization of a large *Borrelia burgdorferi* motility operon which is initiated by a consensus σ^{70} promoter. J. Bacteriol. **179:**2289–2299.
15. **Harrington, C. S., F. M. Thomson-Carter, and P. E. Carter.** 1997. Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: implications for the flagellin gene typing scheme. J. Clin. Microbiol. **35:**2386–2392.
16. **Helmann, J. D.** 1991. Alternative sigma factors and the regulation of flagellar gene expression. Mol. Microbiol. **5:**2875–2882.
17. **Hetzel, M., M. Brock, T. Selmer, A. J. Pierik, B. T. Golding, and W. Buckel.** 2003. Acryloyl-CoA reductase from *Clostridium propionicum.* An enzyme complex of propionyl-CoA dehydrogenase and electron-transferring flavoprotein. Eur. J. Biochem. **270:**902–910.
18. **Hill, K. K., T. J. Smith, C. H. Helma, L. O. Ticknor, B. T. Foley, R. T. Svensson, J. L. Brown, E. A. Johnson, L. A. Smith, R. T. Okinaka, P. J. Jackson, and J. D. Marks.** 2007. Genetic diversity among botulinum neurotoxin-producing clostridial strains. J. Bacteriol. **189:**818–832.
19. **Josenhans, C., and S. Suerbaum.** 2002. The role of motility as a virulence factor in bacteria. Int. J. Med. Microbiol. **291:**605–614.
20. **Kanbe, M., J. Yagasaki, S. Zehner, M. Gottfert, and S. Aizawa.** 2007. Characterization of two sets of subpolar flagella in *Bradyrhizobium japonicum*. J. Bacteriol. **189:**1083–1089.
21. **Keto-Timonen, R., M. Nevas, and H. Korkeala.** 2005. Efficient DNA fingerprinting of *Clostridium botulinum* types A, B, E, and F by amplified fragment length polymorphism analysis. Appl. Environ. Microbiol. **71:**1148–1154.
22. **Kojima, A., K. Amimoto, T. Ohgitani, and Y. Tamura.** 1999. Characterization of flagellin from *Clostridium chauvoei*. Vet. Microbiol. **67:**231–237.
23. **Kojima, A., I. Uchida, T. Sekizaki, Y. Sasaki, Y. Ogikubo, M. Kijima, and Y. Tamura.** 2000. Cloning and expression of a gene encoding the flagellin of *Clostridium chauvoei*. Vet. Microbiol. **76:**359–372.
24. **Kushnir, E. D., T. M. Breen, and S. S. Paikina.** 1937. Sources of infection of sturgeons (red fish) with *Bacillus botulinum*. Zh. Mikrobiol. Epidemiol. Immunobiol. **19:**80–85.
25. **Leclair, D., F. Pagotto, J. M. Farber, B. Cadieux, and J. W. Austin.** 2006. Comparison of DNA fingerprinting methods for use in investigation of type E botulism outbreaks in the Canadian Arctic. J. Clin. Microbiol. **44:**1635–1644.
26. **Lindstrom, M., and H. Korkeala.** 2006. Laboratory diagnostics of botulism. Clin. Microbiol. Rev. **19:**298–314.
27. **Logan, S. M.** 2006. Flagellar glycosylation—a new component of the motility repertoire? Microbiology **152:**1249–1262.
28. **Lund, B. M., and M. W. Peck.** 2000. *Clostridium botulinum*, p. 1057–1109. *In* B. M. Lund, T. C. Baird-Parker, and G. W. Gould (ed.), The microbiological safety and quality of food. Aspen Publishers Inc., Gaithersburg, MD.
29. **Lyristis, M., Z. L. Boynton, D. Petersen, Z. Kan, G. N. Bennett, and F. B. Rudolph.** 2000. Cloning, sequencing, and characterization of the gene encoding flagellin, flaC, and the post-translational modification of flagellin, FlaC, from *Clostridium acetobutylicum* ATCC824. Anaerobe **6:**69–79.
30. **Macnab, R. M.** 2003. How bacteria assemble flagella. Annu. Rev. Microbiol. **57:**77–100.
31. **Macnab, R. M.** 2004. Type III flagellar protein export and flagellar assembly: protein export/secretion in bacteria. Biochim. Biophys. Acta **1694:**207–217.
32. **Mirel, D. B., and M. J. Chamberlin.** 1989. The *Bacillus subtilis* flagellin gene (*hag*) is transcribed by the σ^{28} form of RNA polymerase. J. Bacteriol. **171:**3095–3101.
33. **Nevas, M., M. Lindstrom, S. Hielm, K. J. Bjorkroth, M. W. Peck, and H. Korkeala.** 2005. Diversity of proteolytic *Clostridium botulinum* strains, determined by a pulsed-field gel electrophoresis approach. Appl. Environ Microbiol. **71:**1311–1317.
34. **Paredes, C. J., K. V. Alsaker, and E. T. Papoutsakis.** 2005. A comparative genomic view of clostridial sporulation and physiology. Nat. Rev. Microbiol. **3:**969–978.
35. **Paredes, C. J., I. Rigoutsos, and E. T. Papoutsakis.** 2004. Transcriptional organization of the *Clostridium acetobutylicum* genome. Nucleic Acids Res. **32:**1973–1981.
36. **Pituch, H., P. Obuch-Woszczatynski, N. van den Braak, A. van Belkum, M. Kujawa, M. Luczak, and F. Meisel-Mikolajczyk.** 2002. Variable flagella expression among clonal toxin A−/B+ *Clostridium difficile* strains with highly homogeneous flagellin genes. Clin. Microbiol. Infect. **8:**187–188.
37. **Pourshaban, M., G. Franciosa, L. Fenicia, and P. Aureli.** 2002. Taxonomic identity of type E botulinum toxin-producing *Clostridium butyricum* strains by sequencing of a short 16S rDNA region. FEMS Microbiol. Lett. **214:**119–125.
38. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
39. **Sasaki, Y., A. Kojima, H. Aoki, Y. Ogikubo, N. Takikawa, and Y. Tamura.** 2002. Phylogenetic analysis of PCR detection of *Clostridium chauvoei*, *Clostridium haemolyticum*, *Clostridium novyi* types A and B, and *Clostridium septicum* based on the flagellin gene. Vet. Microbiol. **86:**257–267.
40. **Schaffer, C., and P. Messner.** 2004. Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. Glycobiology **14:**31R–42R.
41. **Schoenhals, G., and C. Whitfield.** 1993. Comparative analysis of flagellin sequences from *Escherichia coli* strains possessing serologically distinct flagellar filaments with a shared complex surface pattern. J. Bacteriol. **175:**5395–5402.
42. **Staden, R., K. F. Beal, and J. K. Bonfield.** 2000. The Staden package, 1998. Methods Mol. Biol. **132:**115–130.
43. **Tasteyre, A., M. C. Barc, A. Collignon, H. Boureau, and T. Karjalainen.** 2001. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. Infect. Immun. **69:**7937–7940.
44. **Tasteyre, A., M. C. Barc, T. Karjalainen, P. Dodson, S. Hyde, P. Bourlioux, and P. Borriello.** 2000. A *Clostridium difficile* gene encoding flagellin. Microbiology **146:**957–966.
45. **Tasteyre, A., T. Karjalainen, V. Avesani, M. Delmee, A. Collignon, P. Bourlioux, and M. C. Barc.** 2000. Phenotypic and genotypic diversity of the flagellin gene (*fliC*) among *Clostridium difficile* isolates from different serogroups. J. Clin. Microbiol. **38:**3179–3186.
46. **Tasteyre, A., T. Karjalainen, V. Avesani, M. Delmée, A. Collignon, P. Bourlioux, and M. C. Barc.** 2001. Molecular characterization of *fliD* gene encoding flagellar cap and its expression among *Clostridium difficile* isolates from different serogroups. J. Clin. Microbiol. **39:**1178–1183.
47. **Twine, S. M., N. C. S. Mykytczuk, M. Petit, T.-L. Tremblay, P. Lanthier, J. W. Conlan, and J. F. Kelly.** 2005. *Francisella tularensis* proteome: low levels of ASB-14 facilitate the visualization of membrane proteins in total protein extracts. J. Proteome Res. **4:**1848–1854.
48. **Winstanley, C.** 2003. Improved flagellin genotyping in the *Burkholderia cepacia* complex. FEMS Microbiol. Lett. **229:**9–14.
49. **Winstanley, C., and J. A. Morgan.** 1997. The bacterial flagellin gene as a biomarker for detection, population genetics and epidemiological analysis. Microbiology **143:**3071–3084.