Implications of Genome-Based Discrimination between *Clostridium botulinum* Group I and *Clostridium sporogenes* Strains for Bacterial Taxonomy

Michael R. Weigand,^a Angela Pena-Gonzalez,^b Timothy B. Shirey,^c Robin G. Broeker,^c Maliha K. Ishaq,^c Konstantinos T. Konstantinidis,^{a,b} Brian H. Raphael^c

School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA^a; School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA^b; Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA^c

Taxonomic classification of *Clostridium botulinum* is based on the production of botulinum neurotoxin (BoNT), while closely related, nontoxic organisms are classified as *Clostridium sporogenes*. However, this taxonomic organization does not accurately mirror phylogenetic relationships between these species. A phylogenetic reconstruction using 2,016 orthologous genes shared among strains of *C. botulinum* group I and *C. sporogenes* clearly separated these two species into discrete clades which showed ~93% average nucleotide identity (ANI) between them. Clustering of strains based on the presence of variable orthologs revealed 143 *C. sporogenes* clade-specific genetic signatures, a subset of which were further evaluated for their ability to correctly classify a panel of presumptive *C. sporogenes* strains by PCR. Genome sequencing of several *C. sporogenes* strains lacking these signatures confirmed that they clustered with *C. botulinum* strains in a core genome phylogenetic tree. Our analysis also identified *C. botulinum* strains that contained *C. sporogenes* clade-specific signatures and phylogenetically clustered with *C. sporogenes* strains. The genome sequences of two *bont/B2*-containing strains belonging to the *C. sporogenes* clade contained regions with similarity to a *bont*-bearing plasmid (pCLD), while two different strains belonging to the *C. botulinum* clade carried *bont/B2* on the chromosome. These results indicate that *bont/B2* was likely acquired by *C. sporogenes* strains through horizontal gene transfer. The genome-based classification of these species used to identify candidate genes for the development of rapid assays for molecular identification may be applicable to additional bacterial species that are challenging with respect to their classification.

Botulinum neurotoxins (BoNT) cause neuromuscular paralysis associated with botulism and are produced by various clostridia, most notably *Clostridium botulinum*. *C. botulinum* is a Gram-positive, anaerobic, endospore-forming bacillus that can be classified on the basis of metabolic properties into four separate groups (groups I to IV). Group I *C. botulinum* strains are proteolytic, saccharolytic, and capable of producing BoNT types A, B, and F. Group II *C. botulinum* strains are nonproteolytic and can produce BoNT types B, E, and F, while group III strains produce BoNT types C and D. Group IV strains, which are also identified as *Clostridium argentinense*, produce BoNT type G.

The nontoxic species *Clostridium sporogenes* shares nearly identical metabolic properties with group I *C. botulinum*, including the formation of lipase-positive colonies when grown on egg yolk agar. Because of these similarities and the comparable heat resistance of its spores, *C. sporogenes* has been used as a surrogate organism for *C. botulinum* in the study of thermal processing for foods (1). Previous studies have shown that some strains of *C. sporogenes* and group I *C. botulinum* can be differentiated phenotypically by soluble protein expression, measured by polyacrylamide gel electrophoresis (2) and gas-liquid chromatography of trimethylsilyl derivatives of whole-cell hydrolysates (3).

Genetic studies such as DNA-DNA hybridization and 16S rRNA gene sequence analysis have shown that these species are closely related (4–7). While acknowledging a genetic relationship at the species level, Olsen et al. (8) proposed that the species name *C. botulinum* be reserved for toxigenic strains and that *C. sporogenes* be conserved for nontoxigenic strains in order to prevent confusion between these organisms, and this proposal has met with wide acceptance.

As genome sequencing becomes more commonly used for microbial identification, it is important to evaluate this taxonomic organization in a phylogenomic context. Bacterial species are often determined on the basis of a limited number of diagnostic phenotypic traits, while the ability of bacteria to form actual discrete units (i.e., species) is a topic of intense research (9–12). Moreover, horizontal gene transfer (HGT) of genes encoding diagnostic traits may act to further reduce the boundaries between bacterial species (13). In this study, we used genome sequence analysis to resolve *C. sporogenes* and group I *C. botulinum* strains on the basis of genetic relatedness between, as well as within, these named species. In addition, this analysis demonstrated a previously unrecognized role for HGT of some BoNT/B-encoding genes between these organisms.

Received 7 April 2015 Accepted 26 May 2015

Accepted manuscript posted online 5 June 2015

Citation Weigand MR, Pena-Gonzalez A, Shirey TB, Broeker RG, Ishaq MK, Konstantinidis KT, Raphael BH. 2015. Implications of genome-based discrimination between *Clostridium botulinum* group I and *Clostridium sporogenes* strains for bacterial taxonomy. *Appl Environ Microbiol* 81:5420–5429. doi:10.1128/AEM.01159-15.

Editor: D. W. Schaffner

Address correspondence to Brian H. Raphael, BRaphael@cdc.gov.

M.R.W. and A.P.-G. contributed equally to the work.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.01159-15>.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.

doi:10.1128/AEM.01159-15

TABLE 1 Characteristics of strains examined in this study^a

Strain ^b	Toxin type	Species	Location	Year	Source
CDC41370	Ab	<i>C. botulinum</i>	Mexico	1996	Food
CDC41370NT	NT	<i>C. botulinum</i>	NA	NA	Spontaneous mutant of CDC41370
CDC68016	B1	<i>C. botulinum</i>	West Virginia	2013	Stool
CDC68016NT	NT	<i>C. botulinum</i>	NA	NA	Spontaneous mutant of CDC68016
ATCC 7949	B2	<i>C. botulinum</i>	Unknown	Unknown	Unknown
Prevot25 NCASE	B2	<i>C. botulinum</i>	Unknown	Unknown	Unknown
Prevot1662	B2	<i>C. botulinum</i>	Unknown	Unknown	Unknown
ATCC 51387	B2	<i>C. botulinum</i>	Unknown	Unknown	Unknown
CDC66221	B6	<i>C. botulinum</i>	Colorado	2011	Stool
11579	NT	<i>C. sporogenes</i>	Unknown	1972	Unknown
85-3852	NT	<i>C. sporogenes</i>	Unknown	Unknown	Unknown
87-0535	NT	<i>C. sporogenes</i>	Unknown	Unknown	Unknown
88-0163	NT	<i>C. sporogenes</i>	Unknown	1988	Blood
CDC22719	NT	<i>C. sporogenes</i>	Colorado	1977	Chili sauce
CDC22720	NT	<i>C. sporogenes</i>	Colorado	1977	Cheese
CDC21678	NT	<i>C. sporogenes</i>	Hawaii	1976	Abdominal drainage
CDC21494	NT	<i>C. sporogenes</i>	Germany	1976	Canned ham
CDC22679	NT	<i>C. sporogenes</i>	Puerto Rico	1977	Canned meat
CDC22680	NT	<i>C. sporogenes</i>	Puerto Rico	1977	Canned meat
CDC22753	NT	<i>C. sporogenes</i>	South Carolina	1977	Wound (foot)
CDC22798	NT	<i>C. sporogenes</i>	South Carolina	1977	Blood
CDC23091	NT	<i>C. sporogenes</i>	Oklahoma	1978	Gastric fluid
CDC23284*	NT	<i>C. sporogenes</i>	Maryland	1978	Abdominal fluid
CDC23285*	NT	<i>C. sporogenes</i>	Maryland	1978	Peritoneal fluid
CDC24474	NT	<i>C. sporogenes</i>	Maryland	1979	Abdominal fluid
CDC24545 ^c	NT	<i>C. sporogenes</i>	New Zealand	1979	Dotterel (bird)
CDC24442 ^c	NT	<i>C. sporogenes</i>	New Zealand	1979	Dotterel (bird)
CDC24533	NT	<i>C. sporogenes</i>	Arizona	1979	Wound (leg)
CDC24968	NT	<i>C. sporogenes</i>	Missouri	1979	Stool
CDC24726	NT	<i>C. sporogenes</i>	North Carolina	1979	Liver paste
NCTC 534	NT	<i>C. sporogenes</i>	Unknown	~1920	Human
CDC24614	NT	<i>C. sporogenes</i>	Florida	1979	Wound
CDC35120 [†]	NT	<i>C. sporogenes</i>	California	1980	Blood
CDC35121 [†]	NT	<i>C. sporogenes</i>	California	1980	Lung
CDC35566	NT	<i>C. sporogenes</i>	Missouri	1980	Unknown
CDC35197 [‡]	NT	<i>C. sporogenes</i>	Alaska	1980	Wound drainage
CDC35196 [‡]	NT	<i>C. sporogenes</i>	Alaska	1980	Wound drainage

^a NT, nontoxic; NA, not applicable.^b Isolates from same individual share the same symbol.^c Isolates from identical cultures.

MATERIALS AND METHODS

Initial characterization of bacterial strains. All strains examined in this study were grown in Trypticase-peptone-glucose-yeast extract (TPGY) medium (Remel, Lenexa, KS) at 35°C under anaerobic conditions. Single colonies were isolated from egg yolk agar (EYA). Strains presumptively identified as *C. sporogenes* (Table 1) based on microbiological characteristics and lack of toxicity using the mouse bioassay (14) were reexamined using the botulinum toxin gene real-time PCR assay (distributed by the Laboratory Response Network, CDC) to demonstrate absence of botulinum neurotoxin genes (A to G) and by botulinum toxin enzyme-linked immunosorbent assay (ELISA) (15) to demonstrate lack of toxin (types A, B, E, and F) production. Additional toxic and nontoxic *C. botulinum* and *C. sporogenes* strains examined in this study are also described in Table 1.

DNA extraction and draft genome sequence assembly. Genomic DNA from all isolates was extracted from TPGY cultures as previously described (16) and purified using the DNA Clean & Concentrator-5 kit (Zymo Research, Irvine, CA) prior to library construction. Genome sequencing was performed using either the Illumina MiSeq or Life Technologies Ion Torrent PGM instrument (Table 2). Illumina reads were filtered

and trimmed at both the 5' and 3' ends based on a Phred score threshold (Q) of 20 using SolexaQA (17). For each isolate, quality-trimmed reads were assembled first in parallel runs of Velvet v1.0.13 (18) and SOAPdenovo2 (19) with a range of *k*-mer values, as described previously (20). The resulting precontigs were pooled and then assembled using the Genome Sequencer (GS) *de novo* assembler software v2.0.01.14 (Roche, Branford, CT). Ion Torrent reads were assembled using the GS *de novo* assembler. For both platforms, assembled contigs were ordered using Mauve Contig Mover (21) with *C. botulinum* strain Loch Maree as a reference, and assembly validation was performed with Mauve Assembly Metrics (22). The resulting draft genome sequences were annotated with the Rapid Annotation using Subsystems Technology (RAST) server (23). Despite a large number of contigs in some assemblies, more than 95% of the core genes were fully recovered, suggesting that no major sequencing gaps remained.

The nucleotide sequences of *bont*/B2 and *bont*/B6 genes were determined by sequence read mapping using CLC bio Genomics Workbench version 7.5. For *bont*/B2-containing genomes, the ATCC 7929 *bont*/B sequence (GenBank accession number EF028395) was selected as the reference for read mapping, while the Okayama 2011 *bont*/B sequence

TABLE 2 Statistics of the genome sequences generated in this study

Strain	Platform ^a	No. of assembled reads	No. of contigs	N ₅₀ ^b	Estimated genome size (Mb)	Coverage	GenBank accession no.
CDC41370	PGM	536,228	980	7,449	4.1	26×	LAGI01000000
CDC41370NT	PGM	695,428	254	30,380	3.8	37×	LAGJ01000000
CDC68016	PGM	654,420	872	9,380	4.2	33×	LAGK01000000
CDC68016NT	PGM	392,523	1,714	3,800	4.3	18×	LAGL01000000
ATCC 7949	PGM	1,012,546	206	50,992	3.8	57×	LAGE01000000
Prevot25NCASE	PGM	1,088,680	192	42,026	4.1	56×	LAGN01000000
Prevot1662	PGM	775,204	593	19,719	4.9	30×	LAGM01000000
ATCC 51387	PGM	1,178,142	302	31,395	4.3	56×	LAGD01000000
11579	PGM	577,378	317	27,263	4.4	26×	JZJN01000000
85-3852	PGM	634,904	305	30,089	4.1	32×	JZJO01000000
87-0535	MiSeq	1,078,369	226	46,656	3.8	37×	JZJP01000000
88-0163	MiSeq	1,596,670	247	40,700	3.8	53×	JZJQ01000000
CDC66221	PGM	422,086	818	9,229	4.1	17×	LAGO01000000
CDC23284	PGM	580,480	506	17,646	4.1	28×	LAGF01000000
CDC24442	PGM	476,812	368	14,470	3.1	27×	LAGG01000000
CDC24533	PGM	793,767	344	24,303	4.1	41×	LAGH01000000

^a PGM, Life Technologies Ion Torrent personal genome machine. MiSeq was from Illumina.
^b N₅₀ is the contig length at which half of the total assembly is present in contigs of that length or longer.

(GenBank accession number [AB665558](#)) was used for *bont*/B6-containing genomes.

Additional comparative analysis included alignment of selected genome sequences with pCLD (from *C. botulinum* strain Okra) using the BLAST Ring Image Generator (BRIG) (24) and alignment of the chromosomally located toxin gene cluster region of specific genomes containing *bont*/B2 using the R package genoPlotR (25). The locations of specific genes within these toxin gene clusters were predicted using GeneMarkS (26).

Core genome phylogenetic reconstruction. All predicted protein-coding genes from each of the sequenced genomes and available reference sequences were compared using an all-versus-all BLAST search as described previously (27). This analysis identified shared reciprocal best matches in all pairwise genome comparisons (core orthologs) of *C. sporogenes* and either all *C. botulinum* strains or only group I *C. botulinum* strains. These core orthologs were individually aligned using MUSCLE (28). The resulting alignments were concatenated to create a whole-genome alignment, and the phylogeny of genomes was reconstructed by computing a maximum-likelihood distance matrix with RAxML (29, 30).

Genetic signature identification. The collection of variable genes in each genome, defined as genes absent in 1 or more genomes, was identified in the all-versus-all BLAST search described above. The presence or absence of these variable genes was used to hierarchically cluster genomes using complete linkage across a centered Pearson correlation similarity matrix using Cluster 3.0 (31). Variable genes were considered *C. sporogenes* specific if present in at least 7 out of 8 genomes and absent in all group I *C. botulinum* genomes. Conversely, genes present in 11 out of 13 group I *C. botulinum* genomes but absent in all *C. sporogenes* genomes were classified as *C. botulinum* specific. These ratios correspond to at least 80% of the population of strains examined for each species (representing values >1 standard deviation from the mean of the population) and are therefore expected to be robust to spurious matches.

Functional annotation of predicted proteins encoded by species-specific genes was bioinformatically inferred using a combination of three independent approaches. First, sequences were searched against the UniProtKB/Swiss-Prot database (32) using DELTA-BLAST (33) with a bit score cutoff of 200. Second, conserved domains were predicted by RPS-BLAST against the position-specific score matrices of the NCBI Conserved Domain Database (34) using an E-value threshold of 0.01. Finally, a BLASTX search against the NCBI nonredundant (nr) database was performed with a bit score cutoff of 200 and identity greater than 96%. The results of these independent searches were manually inspected, and the

corresponding genes were functionally annotated based on the consensus of the three.

Genetic signature PCR. A total of 143 putative *C. sporogenes*-specific orthologs were identified, from which 11 candidates were selected for primer design for PCR amplification. The annotated functions and corresponding NCBI locus ID for the candidate signatures are shown in Table S1 in the supplemental material. PCR amplification was performed on genomic DNA extracted from 24 *C. sporogenes* isolates using primers targeting each of the 11 biomarkers (shown in Table 3). PCR thermocycling conditions consisted of an initial denaturation at 95°C for 3 min followed by 35 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 90 s with a final extension of 72°C for 10 min. The amplified products were examined on

TABLE 3 *C. sporogenes* clade-specific gene signatures

Gene or locus	Annotated function	PCR primers (5'→3') ^a
<i>hsdMSR</i>	Type I restriction modification	TGAATCGGAAACCGATGGAC TGCCACTTGGCTTCATTCT
<i>ssuABC</i>	Alkane sulfonate transport	TAATCC CTGGGCTTTACCT CCAGCTGTTATCTCATGCCA
<i>isdAFD</i>	Heme transport	CCAGGAAAGGGCAAAACCA TGGAGACCAAGACACTTTA
<i>thiHG</i>	Thiazole biosynthesis	RCGCTTGTGTCCAKCTAAAT GCCGGGTATGGAGYTAATTG
<i>lipAS</i>	Lipoate protein ligase and synthase	TGTGACGAAGCTAATTGTCCT CTTTTCCCCAAGACCTACCA
<i>ltaS</i>	Lipoteichoic acid synthase	ATGGGAGAGGCCAGAAAGTA TGCCTCTCAGATACCATCC A
<i>gadA</i>	Glutamate decarboxylase	AGATATTGGAGCGCTTGAGA CTTGTCTCCCCAATCTGGTT
<i>dacB</i>	D-Alanyl-D-alanine carboxypeptidase	TCCAGAAGCCTGCTCTATCC AAAGTGCTGGTGCTGTAT
<i>nifS</i>	Cysteine desulfurase	ATAGACTCTGCTCAAACCGC GTGTAATCC GCATCTGGTCA
<i>dapL</i>	N-Acetyl-L,L-diaminopimelate deacetylase	ACGCCTTCTGTAGCATCAAA GCCTGGACTTTAGGTACTGC
<i>spoVB</i>	Stage V sporulation protein B	GTGTTTACGTCCTGCTTTTCA TGCGCTCTAGAGGAAGATACT

^a The pairs of primer sequences represent forward (top) and reverse (bottom) primers.

FIG 1 Core gene phylogenetic trees of *C. sporogenes* and *C. botulinum* genomes. An initial tree was based on the maximum-likelihood distance calculated from the alignment of 179 shared orthologous genes of *C. sporogenes* and *C. botulinum* groups I, II, and III. The tree in the inset was based on the maximum-likelihood distance calculated from the alignment of 1,451 shared orthologous genes restricted to *C. sporogenes* and group I *C. botulinum*. Specific clades discussed in the text are labeled. The BoNT produced by each strain is indicated in brackets (NT, nontoxic). Genomes sequenced in this study are in red. Additional genome sequences were retrieved from GenBank.

1% ethidium bromide-stained agarose gels and visualized under UV transillumination.

16S rRNA sequencing. A ~1.3-kb region of the 16S rRNA gene from presumptively identified *C. sporogenes* strains was amplified and sequenced using primers that were previously described (35). Sanger sequencing was performed using the GenomeLab dye termination cycle sequencing kit and a CEQ8000 genetic analysis system (Beckman Coulter, Brea, CA). Sequences were assembled and edited with Sequencher v 4.8 (Gene Codes, Ann Arbor, MI).

Nucleotide sequence accession numbers. Newly determined sequence data have been deposited in GenBank under the accession numbers shown in Table 2.

RESULTS

Core genome phylogenetic reconstruction. An all-versus-all BLAST comparison indicated that the pangenome for strains of *C. sporogenes* and *C. botulinum* (including groups I, II, and III) was comprised of 16,229 genes. Of these, 179 genes were shared by all strains (core orthologs), and a phylogenetic reconstruction based on their concatenated alignment revealed four distinct clades (Fig. 1). As expected, strains of *C. botulinum* groups I, II, and III were

well separated. The *C. sporogenes* strains were most closely related to a clade of group I *C. botulinum* strains, consistent with previous genomic studies (36, 37), but formed their own discrete clade. The *C. sporogenes* clade contained *C. sporogenes* strains 88-0163, ATCC 15579, 87-0535, PA3679, 85-3852, and 11579 and two *C. botulinum* type B strains, Osaka05 and Okayama2011. Furthermore, the average pairwise genome aggregate average nucleotide identity (ANI) between genomes within the group I *C. botulinum* clade and the *C. sporogenes* clade was 93.4%, below the 95% ANI cutoff frequently used for species demarcation (38), supporting the classification of these strains into separate bacterial species. Further comparison of strains in the *C. sporogenes* and group I *C. botulinum* clades revealed a pangenome (for both clades combined) comprised of 9,306 orthologs, of which 2,016 were core and 7,290 were variable among these closely related genomes.

Clade-specific gene identification. Hierarchical clustering based on presence or absence of the variable genes (absent in one or more genomes) among the *C. sporogenes* and *C. botulinum* strains examined in this study again separated *C. sporogenes*

FIG 2 Hierarchical clustering based on the presence or absence of variable orthologs. (A) The presence (yellow) or absence (black) of variable orthologs ($n = 16,412$) in *C. sporogenes* and *C. botulinum* group I genomes was used to perform hierarchical clustering using a complete linkage across a centered Pearson correlation. (B) Identification of 143 orthologs genes that are present in the genomes of 7 out of 8 *C. sporogenes* strains and 91 genes that are present in 11 out of 13 group I *C. botulinum* strains.

strains from *C. botulinum* group I, group II, and group III mirroring the phylogenetic relationships inferred from the core genome phylogeny (data not shown). When the analysis was restricted to the 16,412 variable genes present in the *C. sporogenes* and group I *C. botulinum* genomes specifically, the *C. sporogenes* strains continued to form a discrete cluster, suggesting that there exist clade-specific gene signatures in *C. sporogenes* genomes (Fig. 2A).

Selective filtering of variable orthologs identified 143 clade-specific genes that were present in at least 7 out of 8 available *C. sporogenes*-clade genomes (strains 11579, 85-3852, 88-0163, ATCC 15579, 87-0535, PA3679, Osaka05, and Okayama2011) and absent in the genomes associated with the group I *C. botulinum* clade (Fig. 2B). Of these, 65 were annotated as encoding hypothetical proteins. Annotation of the remaining genes included various functions related to nutrient uptake (e.g., heme transport and alkane sulfonate uptake) and bacterial cell defense against invasion of foreign DNA (e.g., type 1 restriction modification system). Correspondingly, 91 genes were observed in at least 11 out of the 13 genomes associated with the group I *C. botulinum* clade that were absent from all *C. sporogenes* strains examined. From these, 44 genes were annotated as hypothetical proteins, while the remaining genes were functionally associated with metabolic pathways (e.g., V-type ATP synthase) and cell wall modifications.

Genetic signatures for clade discrimination. Eleven signatures from the set of *C. sporogenes* clade-specific genomes, each representing a different operon or function, were further evaluated and annotated for potential use in PCR assays. The candidate signatures were selected from throughout the *C. sporogenes* genome in order to evenly sample the genome backbone (see Fig. S1 in the supplemental material). Manual curation of bioinformatic evidence was used to infer the putative function of each gene, and these annotation assignments are summarized in Table 3.

A total of 24 nontoxic lipase-positive strains presumptively

identified as *C. sporogenes* were selected from the CDC culture collection to test the efficacy of *C. sporogenes* clade-specific PCR assays targeting the identified gene signatures. While PCR results were variable for some genes, a subset of five loci (*hsdMSR*, *thiHG*, *lipAS*, *ltaS*, and *dapL*) showed highly consistent results for the nontoxic strains examined (Fig. 3). These five conserved signatures distinguished the 24 nontoxic strains into two groups; those positive for the presence of all five signatures (referred to as “signature positive”; 18 strains) and those negative for all five signatures (“signature negative”; 6 strains).

Three putative *C. sporogenes* strains were selected for genome sequencing to determine whether the genetic signatures detected by PCR were predictive of the clade assigned by core genome phylogeny. As expected, strain CDC23284, which was signature positive, clustered with the *C. sporogenes* clade accordingly (Fig. 1). Both CDC24533 and CDC24442 were signature negative, but only the CDC24533 clustered with the group I *C. botulinum* clade, while strain CDC24442 formed a distinct cluster (Fig. 1). Notably, the genome sequences of these three strains were also consistent with the 16S rRNA gene sequence analysis of all 24 strains (i.e., strains lacking the *C. sporogenes* clade-specific signatures based on PCR analysis were phylogenetically distinct from strains containing the signatures [see Fig. S2 in the supplemental material]). The 16S rRNA gene sequences of strains CDC23284 and CDC24533 differed by 0.3% compared to each other and by ~1% compared to strain CDC24442. Strain CDC24442 shares 16S rRNA gene sequence identity with *C. sporogenes* subsp. *tusciae* biovar *pennavorans* (39) and clearly represents a distinct species, consistent with its distinct position in the core genome phylogeny.

Horizontal transfer of botulinum toxin genes. Since neurotoxin gene loss should not affect core genome phylogenetic reconstruction, we also sequenced the genomes of two additional pairs of strains where a toxic and spontaneous nontoxic

FIG 3 Gene signature PCR. The results of gene signature PCR assays for each locus are shown in the columns, which are labeled across the top. The presence of a PCR product for each assay among the strains examined (indicated to the left of the figure) is indicated in red. The dotted box indicates the loci that demonstrated invariant PCR results among the strains examined.

isolate were available. *C. botulinum* type B strain CDC68016 was isolated from the stool of an infant with botulism, and a nontoxic isolate (CDC68016NT) was also obtained during its laboratory cultivation. Both strains clustered with the *C. sporogenes* clade in the core genome phylogeny (Fig. 1). While strain CDC68016 was initially classified as *C. botulinum* based primarily on the production of botulinum neurotoxin, this strain and its nontoxic progeny (CDC68016NT) are more appropriately viewed as belonging to the *C. sporogenes* clade based on their phylogenetic relationships. These findings also suggest that the type B neurotoxin gene present in strain CDC68016 was horizontally acquired, presumably from a *C. botulinum* donor. Conversely, the *C. botulinum* type Ab strain CDC41370 (isolated from a food source) and its corresponding nontoxic derivative (CDC41370NT) clustered within the group I *C. botulinum* clade.

In order to examine whether strains containing the same toxin gene variant would belong to the same genomic clade, we sequenced strain CDC66221. This strain contains the *bont/B6* variant (Fig. 4), similar to strains Osaka05 and Okayama2011, which belong to the *C. sporogenes* clade (Fig. 1). However, core genome phylogeny placed strain CDC66221 within the *C. botulinum* clade (Fig. 1). Hence, strain CDC66221 can be considered a member of *C. botulinum* according to a genome-based taxonomy, suggesting that strains with the same neurotoxin gene variant do not always group within the same clade.

To better appreciate the frequency of horizontal exchange of neurotoxin genes between strains belonging to either the *C. botulinum* or *C. sporogenes* clade, we analyzed four additional *C. botulinum* type B strains (Prevot1662, ATCC 51387, Prevot25 NCASE, and ATCC 7949) containing the neurotoxin gene variant *bont/B2* (Fig. 4). Two strains (ATCC 51387 and Prevot 1662) were positive and two strains (ATCC 7949 and Prevot 25 NCASE) were negative

for the *C. sporogenes* clade-specific locus *hsdS* by PCR (data not shown). Not surprisingly, the two *hsdS*-negative strains clustered with the *C. botulinum* clade and the two *hsdS*-positive strains clustered with the *C. sporogenes* clade in the core genome phylogeny (Fig. 1). Taken together, these data reveal that the transfer of type B neurotoxin genes between strains belonging to clades representing *C. botulinum* and *C. sporogenes* has occurred in multiple genomes and involved different *bont/B* sequence variants (e.g., strains ATCC 51387 and Prevot 1662, carrying *bont/B2*, and Osaka05 and Okayama2011, carrying *bont/B6*; all four strains group within the *C. sporogenes* clade).

Diversity of type B neurotoxin gene-bearing plasmids. Comparison of the *bont/B2*-containing genomes of *C. sporogenes* clade strains ATCC 51387 and Prevot 1662 with the *bont/B1*-bearing plasmid (pCLD) of strain Okra (Fig. 5A) revealed a high level of nucleotide sequence similarity. These findings are consistent with the presence of a PCR product for the previously described PL-6 plasmid marker (40) in strains ATCC 51387 and Prevot 1662 (data not shown). In contrast to these strains, which may have acquired *bont/B2* via the transfer of plasmid DNA, the *C. botulinum* clade strains ATCC 7949 and Prevot25 NCASE appear to contain *bont/B2* within their chromosomes, colocated within the *oppA-brnQ* operon (Fig. 5B), similar to the location of the unexpressed *bont/B* found in *C. botulinum* type A(B) strain NCTC2916 (36).

Similarly, the genomes of *bont/B6*-containing strains Okayama2011 and CDC66221 also shared a high level of similarity with pCLD. The sequence of a 185-kb extrachromosomal element in strain Osaka05 (GenBank accession number NZ_BA000059) also contains the *bont/B6* toxin gene cluster but shares less similarity with pCLD (Fig. 5A). While all three of these *bont/B6*-containing strains appear to carry the toxin gene on a plasmid, they differ with respect to their core genome phylogeny, suggesting

FIG 4 Neighbor-joining tree of *bont/B* sequences. *bont/B2* sequences are in blue, and *bont/B6* sequences are in red. Representative sequences of other *bont/B* variants were retrieved from GenBank as follows: B1, strain Okra (NC_010379); B3, strain CDC795 (EF028400); B4, strain 17B (EF051570); B5, strain CDC4013 (GU271943); B7, strain NCTC3807 (JN120760); B8, strain Chaiyaphum_2014 (KM067395).

that such a plasmid may be able to exist in strains of either the *C. botulinum* or *C. sporogenes* clade.

DISCUSSION

Previous studies have highlighted the challenging taxonomy of *C. botulinum* and *C. sporogenes* (7, 37). In one study, the 16S rRNA genes of sequences belonging to 110 *Clostridium* species were compared to identify molecular markers capable of differentiating various *Clostridium* species (7). However, *C. botulinum* and *C. sporogenes* were found to cluster together in a single clade of the resulting phylogenetic tree. Additionally, comparative analysis of *in silico* restriction enzyme digestion showed that both species have exact profiles for the restriction enzymes AluI, BfaI, HaeIII, RsaI, Tru9I, and SmaI. The digestion profile for DpnII demonstrated that *C. sporogenes* was segregated into two separate populations, one of which showed exact homology with *C. botulinum* while the other was distinct from all other *Clostridium* species. These data, along with variations in DNA-DNA hybridization among different strains of *C. sporogenes* compared to *C. botulinum* (4, 5), suggest that *C. sporogenes* may be polyphyletic.

More recently, Bradbury et al. (37) reported the genome se-

quencing and annotation of *Clostridium sporogenes* strain PA3679, which has been used as a surrogate for group I *C. botulinum* in thermal processing of foods for over 70 years (41, 42). The percentage of shared *k*-mers ($k = 25$ bp) between the *C. sporogenes* PA3679 genome and several genomes of *C. botulinum* revealed that more than 85% of the *C. sporogenes* PA3679 genome matches group I *C. botulinum* strains (containing *bont/A1*). In addition, alignment of 16S rRNA gene sequences indicated a 99 to 100% nucleotide similarity between PA 3679 and several proteolytic *C. botulinum* strains, as well as other *C. sporogenes* strains.

In this study, we compared draft genome sequences to resolve *C. sporogenes* and *C. botulinum* group I strains by both core genome phylogeny and variable gene content analysis. Several *bont*-containing *C. botulinum* group I strains, such as Osaka05 and Okayama2011, were more closely related to *C. sporogenes* strains. Recent work by Kenri et al. (43) using whole-genome SNP analysis demonstrated that these two strains were distantly related to other *C. botulinum* group I strains examined. These results emphasize the conflicting nature of taxonomic designations based on the presence or absence of the botulinum neurotoxin gene, which does not reflect true phylogenetic relationships between strains.

FIG 5 BLAST analysis of draft genome sequences. (A) Genomes were compared using BRIG with the *bont/B*-bearing plasmid (pCLD) in the *C. botulinum* type B strain Okra (GenBank accession number [NC_010379](#)). Genomes shown include (from innermost to outermost ring) ATCC 7949, Prevot 25 NCASE, ATCC 51387, Prevot 1662, Osaka05 (extrachromosomal element 2), Okayama2011, CDC66221, CDC41370NT, CDC41370, CDC68016NT, and CDC68016. Regions with >50% nucleotide identity compared to pCLD are shaded. Shading color corresponds to the *bont/B* sequence variant, as follows: blue, *bont/B*2; red, *bont/B*6; green, *bont/B*5; and purple, *bont/B*1. The positions of toxin gene cluster genes are indicated by black arrows. (B) The nucleotide sequence similarities of contigs containing the *bont/B* neurotoxin gene complex from assembled draft genome sequences ATCC 7949 and Prevot25 NCASE were compared. As a reference, the *oppA-brnQ* operon of *C. botulinum* type A(B) strain NCTC2916 (GenBank accession number [NZ_ABDO02000001](#)) was also compared to ATCC 7949. Regions sharing high nucleotide similarity are shaded.

Kalia et al. (7) suggested, “It may not be too inappropriate to suggest that *C. sporogenes* is perhaps a sub-species of *C. botulinum* or it may find its appropriate place if *C. botulinum* can be reclassified as 4 different sub-species.” Our results support the latter, namely, that *C. sporogenes* is a distinct species of the genus, closely related to group I *C. botulinum*, because *C. sporogenes* strains are monophyletic on the core genome tree, show less than 95% ANI to their group I *C. botulinum* relatives but more than 95% ANI among themselves, and are characterized by specific genes and pathways that are rarely found in their relatives (Table 3). These genes may underlie important phenotypic differences between *C. sporogenes* and group I *C. botulinum* that remain to be elucidated. Identifying the associated phenotypes should be the subject of future work toward developing fast diagnostic tests for each of these two important bacterial taxa. While it is possible that some of the gene signatures for *C. sporogenes* may be found in other clostridial (or even nonclostridial) species, phylogenetic analysis would likely distinguish *C. sporogenes* from other species.

Among a panel of 24 nontoxic presumptive *C. sporogenes* isolated primarily from clinical samples, 18 were positive for *C. sporogenes* clade signatures. While the majority of BoNT-producing isolates examined in this study belonged to the *C. botulinum* group I clade, we observed some type B toxin producing isolates within the *C. sporogenes* clade suggesting that some toxin gene variants may be highly mobile among these two groups of organisms. Consistent with these findings, Carter et al. (44) identified only one named *C. botulinum* type B strain that grouped together

with various *C. sporogenes* strains examined based on hybridization of a larger collection of strains against a ATCC 3502-specific (a group I *C. botulinum* type A strain) DNA microarray.

In a study of 63 *C. botulinum* type B strains, Franciosa et al. (45) found that more than half contained *bont/B* on a plasmid. Interestingly, neurotoxin gene-bearing plasmids were found among multiple *bont/B* sequence variants. At least some of these plasmids are likely to be mobile by conjugation, as shown by Marshall et al. (46). These findings are consistent with the results reported here that at least five *C. sporogenes*-like strains (CDC68016, ATCC 51387, Prevot 1662, Osaka05, and Okayama2011) appear to have acquired the neurotoxin gene via the horizontal transfer of plasmid DNA. Similar to another report (47), we also observed type B toxin gene loss in strains CDC41370 and CDC68016. The nontoxic strain CDC24533 belonging to the *C. botulinum* clade, specifically within a subclade containing strains that carry their toxin genes on plasmids (36, 48), presumably underwent plasmid loss. This strain was isolated from necrotic tissue of a leg wound and not associated with a case of wound botulism, indicating that loss of toxicity by this strain probably occurred prior to infection of the wound rather than during subsequent laboratory cultivation. Collectively, these results further support the conclusion that the current practice of classifying strains as *C. botulinum* or *C. sporogenes* based on the presence or absence of the botulinum toxin genes, respectively, cannot be highly reliable from a phylogenetic perspective, because toxin gene loss can result in confounding classification of the resulting nontoxic progeny.

In this study, two strains containing *bont/B2* on a plasmid were associated with the *C. sporogenes* clade, while strains carrying this gene on the chromosome were located within the *C. botulinum* clade. Smith et al. (49) recently reported that the genome sequence of another *bont/B2*-containing strain (Prevot 594) with its neurotoxin gene on a plasmid was more related to strain Osaka05 than other toxigenic strains examined, suggesting that it was also a member of the *C. sporogenes* clade. Similarly, *in silico* multilocus sequencing typing of the draft genome sequence of *C. botulinum* strain 450 (isolated from a case of wound botulism and found to contain *bont/B2*) demonstrated a close phylogenetic relationship with *C. sporogenes* strain ATCC 15579 (50). Future work should include the study of a larger population of *bont/B2*-containing strains to determine the frequency at which *bont/B2* is plasmid borne among strains belonging to the *C. sporogenes* clade. The results presented here, which are based on a limited number of strains, indicate that the transfer of neurotoxin genes among these clades may not be a rare event.

Distinguishing *C. sporogenes* strains from *C. botulinum* isolates that have lost toxicity has important implications. Since nontoxic isolates from distinct phylogenetic clades may have different phenotypic properties, it may be useful to consider a broader range of nontoxic organisms for *C. botulinum* surrogates (especially in food protection studies) than simply those identified as *C. sporogenes*.

Genomic and metagenomic approaches for microbial identification will likely lead to fewer laboratories attempting to detect botulinum toxin in culture. The results presented here clearly suggest that genomic sequencing technologies are highly reliable and robust in correctly classifying new strains and assessing their toxin production potential (i.e., identification of neurotoxin genes). While it is difficult to determine based on traditional techniques whether a nontoxic strain with microbiological properties resembling proteolytic *C. botulinum* or *C. sporogenes* may simply be the result of toxin gene loss, the genetic signatures developed in this study can help in determining the genomic background of these nontoxic and toxic strains and also correctly interpreting previous strain classifications. The presence of *C. sporogenes* clade-specific genes would provide investigators with a high-confidence genomic signature for “true” *C. sporogenes* strains. Nontoxic strains lacking such signatures may well be nontoxic isolates of *C. botulinum* and, in the context of a botulism case investigation, may necessitate screening larger numbers of isolates for toxicity. Collectively, the results presented here show that genome-derived data can offer robust resolution and classification of *C. botulinum* or *C. sporogenes* strains, even in the presence of HGT. These approaches provide a more phylogenetically accurate classification *C. botulinum* and *C. sporogenes* strains and promote a coherent bacterial species concept for these groups. Our results also underscore the limitations of the current bacterial classification system, which is frequently based on laboratory-assessed phenotypes that can be easily transferable between genetically distinct species, while providing an example of how to advance bacterial classification toward a genome-based taxonomy. This work should be applicable to additional bacterial taxa that are challenging with respect to classification.

ACKNOWLEDGMENTS

We thank Susan Maslanka and Carolina Lúquez for critical review of the manuscript. Genome sequencing of some strains using the MiSeq platform was performed by members of the PulseNet USA Team (CDC) and National Enteric Reference Laboratory (CDC).

This work was supported by funds made available from the Centers for Disease Control and Prevention, Office of Public Health Preparedness and Response and in part by the U.S. National Science Foundation under award no. 1241046 (to K.T.K.). A.P.-G. was supported by Colciencias-Colombian Administrative Department for Science, Technology and Innovation through a doctoral fellowship. T.B.S. and M.K.I. were supported by fellowships with the Oak Ridge Institute for Science and Education.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

- Brown JL, Tran-Dinh N, Chapman B. 2012. *Clostridium sporogenes* PA 3679 and its uses in the derivation of thermal processing schedules for low-acid shelf-stable foods and as a research model for proteolytic *Clostridium botulinum*. J Food Prot 75:779–792. <http://dx.doi.org/10.4315/0362-028X.JFP-11-391>.
- Cato EP, Hash DE, Holdeman LV, Moore WEC. 1982. Electrophoretic study of *Clostridium* species. J Clin Microbiol 15:688–702.
- Farshy DC, Moss CW. 1970. Characterization of clostridia by gas chromatography: differentiation of species by trimethylsilyl derivatives of whole-cell hydrolysates. Appl Environ Microbiol 20:78–84.
- Lee WH, Riemann H. 1970. The genetic relatedness of proteolytic *Clostridium botulinum* strains. J Gen Microbiol 64:85–90. <http://dx.doi.org/10.1099/00221287-64-1-85>.
- Nakamura S, Okado I, Nakashio S, Nishida S. 1977. *Clostridium sporogenes* isolates and their relationships to *C. botulinum* on deoxyribonucleic acid reassociation. J Gen Microbiol 100:395–401. <http://dx.doi.org/10.1099/00221287-100-2-395>.
- Hutson RA, Thompson DE, Lawson PA, Schocken-Iturino RP, Böttger EC, Collins MD. 1993. Genetic interrelationships of proteolytic *Clostridium botulinum* types A, B, and F and other members of the *Clostridium botulinum* complex as revealed by small-subunit rRNA gene sequences. Antonie Van Leeuwenhoek 64:273–283.
- Kalia VC, Mukherjee T, Bhushan A, Joshi J, Shankar P, Huma N. 2011. Analysis of the unexplored features of *rrs* (16S rDNA) of the genus *Clostridium* (2011). BMC Genomics 12:18. <http://dx.doi.org/10.1186/1471-2164-12-18>.
- Olsen I, Johnson JL, Moore LVH, Moore WEC. 1995. Rejection of *Clostridium putrificum* and conservation of *Clostridium botulinum* and *Clostridium sporogenes*. Int J Syst Bacteriol 45:414. <http://dx.doi.org/10.1099/00207713-45-2-414>.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. Science 323:741–746. <http://dx.doi.org/10.1126/science.1159388>.
- Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J. 2005. Opinion: re-evaluating prokaryotic species. Nat Rev Microbiol 3:733–739. <http://dx.doi.org/10.1038/nrmicro1236>.
- Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A 102:2567–2572. <http://dx.doi.org/10.1073/pnas.0409727102>.
- Konstantinidis KT, Ramette A, Tiedje JM. 2006. The bacterial species definition in the genomic era. Philos Trans R Soc Lond B Biol Sci 361: 1929–1940. <http://dx.doi.org/10.1098/rstb.2006.1920>.
- Lawrence JG. 2002. Gene transfer in bacteria: speciation without species? Theor Popul Biol 61:449–460. <http://dx.doi.org/10.1006/tpbi.2002.1587>.
- Maslanka SE, Solomon HM, Sharma S, Johnson EA. 2013. *Clostridium botulinum* and its toxins. In Doores S, Salfinger Y, Tortorello ML, Wilke BW (ed), Compendium of methods for the microbiological examination of foods, 5th ed. American Public Health Association, Washington, DC.
- Maslanka SE, Lúquez C, Raphael BH, Dykes JK, Joseph LA. 2011. Utility of botulinum toxin ELISA A, B, E, F kits for clinical laboratory investigations of human botulism. Botulinum J 2:72–92. <http://dx.doi.org/10.1504/TBJ.2011.041817>.
- Raphael BH, Choudoir MJ, Lúquez C, Fernández R, Maslanka SE.

2010. Sequence diversity of genes encoding botulinum neurotoxin type F. *Appl Environ Microbiol* 76:4805–4812. <http://dx.doi.org/10.1128/AEM.03109-09>.
17. Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11:485. <http://dx.doi.org/10.1186/1471-2105-11-485>.
18. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
19. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaogian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. <http://dx.doi.org/10.1186/2047-217X-1-18>.
20. Luo C, Tsementzi D, Kyrpides NC, Konstantinidis KT. 2012. Individual genome assembly from complex community short-read metagenomic datasets. *ISME J* 6:898–901. <http://dx.doi.org/10.1038/ismej.2011.147>.
21. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* 25:2071–2073. <http://dx.doi.org/10.1093/bioinformatics/btp356>.
22. Darling AE, Tritt A, Eisen JA, Facciotti MT. 2011. Mauve assembly metrics. *Bioinformatics* 27:2756–2757. <http://dx.doi.org/10.1093/bioinformatics/btr451>.
23. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNiel LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
24. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. <http://dx.doi.org/10.1186/1471-2164-12-402>.
25. Guy L, Kultima JR, Andersson SG. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26:2334–2335. <http://dx.doi.org/10.1093/bioinformatics/btq413>.
26. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618.
27. Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT. 2011. Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci U S A* 108:7200–7205. <http://dx.doi.org/10.1073/pnas.1015622108>.
28. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <http://dx.doi.org/10.1093/nar/gkh340>.
29. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267.
30. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <http://dx.doi.org/10.1093/bioinformatics/btl446>.
31. de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. *Bioinformatics* 20:1453–1454. <http://dx.doi.org/10.1093/bioinformatics/bth078>.
32. UniProt Consortium. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 42:D191–D198. <http://dx.doi.org/10.1093/nar/gkt1140>.
33. Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. *Biol Direct* 7:2. <http://dx.doi.org/10.1186/1745-6150-7-2>.
34. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH. 2005. CDD: a conserved domain database for protein classification. *Nucleic Acids Res* 33:D192–D196.
35. Hill KK, Smith TJ, Helma CH, Ticknor LO, Foley BT, Svensson RT, Brown JL, Johnson EA, Smith LA, Okinaka RT, Jackson PJ, Marks JD. 2007. Genetic diversity among botulinum neurotoxin-producing clostridial strains. *J Bacteriol* 189:818–832. <http://dx.doi.org/10.1128/JB.01180-06>.
36. Hill KK, Xie G, Foley BT, Smith TJ, Munk AC, Bruce D, Smith LA, Brettin TS, Detter JC. 2009. Recombination and insertion events involving the botulinum neurotoxin complex genes in *Clostridium botulinum* types A, B, E and F and *Clostridium butyricum* type E strains. *BMC Biol* 7:66. <http://dx.doi.org/10.1186/1741-7007-7-66>.
37. Bradbury M, Greenfield P, Midgley D, Li D, Tran-Dinh N, Vrieskoop F, Brown JL. 2012. Draft genome sequence of *Clostridium sporogenes* PA 3679, the common nontoxic surrogate for proteolytic *Clostridium botulinum*. *J Bacteriol* 194:1631–1632. <http://dx.doi.org/10.1128/JB.06765-11>.
38. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81–91. <http://dx.doi.org/10.1099/ijs.0.64483-0>.
39. Ionata E, Canganella F, Bianconi G, Benno Y, Sakamoto M, Capasso A, Rossi M, La Cara F. 2008. A novel keratinase from *Clostridium sporogenes* bv. pennavorans bv. nov., a thermotolerant organism isolated from solfatara muds. *Microbiol Res* 163:105–112. <http://dx.doi.org/10.1016/j.micres.2006.08.001>.
40. Raphael BH, Bradshaw M, Kalb SR, Joseph LA, Lúquez C, Barr JR, Johnson EA, Maslanka SE. 2014. *Clostridium botulinum* strains producing BoNT/F4 or BoNT/F5. *Appl Environ Microbiol* 80:3250–3257. <http://dx.doi.org/10.1128/AEM.00284-14>.
41. McClung LS. 1937. Studies on anaerobic bacteria. X. Heat stable and heat labile antigens in the botulinus and related groups of spore-bearing anaerobes. *J Infect Dis* 60:122–128.
42. Townsend CT, Esty JR, Baselt FC. 1938. Heat-resistance studies on spores of putrefactive anaerobes in relation to determination of safe processes for canned foods. *J Food Sci* 3:323–346. <http://dx.doi.org/10.1111/j.1365-2621.1938.tb17065.x>.
43. Kenri T, Sekizuka T, Yamamoto A, Iwaki M, Komiya T, Hatakeyama T, Nakajima H, Takahashi M, Kuroda M, Shibayama K. 2014. Genetic characterization and comparison of *Clostridium botulinum* isolates from botulism cases in Japan between 2006 and 2011. *Appl Environ Microbiol* 80:6954–6964. <http://dx.doi.org/10.1128/AEM.02134-14>.
44. Carter AT, Paul CJ, Mason DR, Twine SM, Alston MJ, Logan SM, Austin JW, Peck MW. 2009. Independent evolution of neurotoxin and flagellar genetic loci in proteolytic *Clostridium botulinum*. *BMC Genomics* 10:115. <http://dx.doi.org/10.1186/1471-2164-10-115>.
45. Franciosa G, Maugliani A, Scalfaro C, Aureli P. 2009. Evidence that plasmid-borne botulinum neurotoxin type B genes are widespread among *Clostridium botulinum* serotype B strains. *PLoS One* 4:e4829. <http://dx.doi.org/10.1371/journal.pone.0004829>.
46. Marshall KM, Bradshaw M, Johnson EA. 2010. Conjugative botulinum neurotoxin-encoding plasmids in *Clostridium botulinum*. *PLoS One* 11:e11087. <http://dx.doi.org/10.1371/journal.pone.0011087>.
47. Umeda K, Seto Y, Kohda T, Mukamoto M, Kozaki S. 2012. Stability of toxigenicity in proteolytic *Clostridium botulinum* type B upon serial passage. *Microbiol Immunol* 56:338–341. <http://dx.doi.org/10.1111/j.1348-0421.2012.00441.x>.
48. Marshall KM, Bradshaw M, Pellett S, Johnson EA. 2007. Plasmid encoded neurotoxin genes in *Clostridium botulinum* serotype A subtypes. *Biochem Biophys Res Commun* 14:49–54.
49. Smith TJ, Hill KK, Xie G, Foley BT, Williamson CH, Foster JT, Johnson SL, Chertkov O, Teshima H, Gibbons HS, Johnsky LA, Karavis MA, Smith LA. 2015. Genomic sequences of six botulinum neurotoxin-producing strains representing three clostridial species illustrate the mobility and diversity of botulinum neurotoxin genes. *Infect Genet Evol* 30:102–113. <http://dx.doi.org/10.1016/j.meegid.2014.12.002>.
50. Fillo S, Giordani F, Anselmo A, Fortunato A, Palozzi AM, De Santis R, Ciammarucconi A, Spagnolo F, Anniballi F, Fiore A, Auricchio B, De Medici D, Lista F. 2015. Draft genome sequence of *Clostridium botulinum* B2 450 strain from wound botulism in a drug user in Italy. *Genome Announc* 3:e00238–15. <http://dx.doi.org/10.1128/genomeA.00238-15>.