Genomic Characterization of Strains From a Cluster of Infant Botulism Type A in a Small Town in Colorado, United States

Lori Gladney*, Jessica L. Halpin and Carolina Lúquez

National Botulism Laboratory, Enteric Diseases Laboratory Branch, Division of Foodborne, Waterborne and Environmental Diseases, National Center for Emerging Zoonotic and Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States

Three cases of infant botulism were reported in a small Colorado town between 1981 and 1984. The first two cases occurred in 1981, 6 months apart, and the third case occurred in 1984. *Clostridium botulinum* type A was isolated from stool of all three case patients and from environmental samples of the patient's homes. An epidemiological investigation and follow-up study were conducted from 1981 to 1986 and concluded the cases were likely related. In this study, we sought to determine whether the *C. botulinum* type A clinical isolates were related to each other and to isolates obtained from environmental samples. We performed whole genome sequencing (WGS) for 17 isolates associated with this potential cluster of infant botulism. Fifteen isolates were confirmed to be *C. botulinum* type A(B) and contained botulinum toxin gene subtypes A1 and B5 by WGS; these strains formed a monophyletic cluster in a phylogeny and were considered closely related to each other (0–18 high-quality single-nucleotide polymorphisms), but distinct from other *C. botulinum* type A(B) in Colorado and elsewhere in the United States. Results of our study suggest that the three infant botulism cases could have represented a cluster due to a *C. botulinum* type A(B) strain present in the environment.

Keywords: *Clostridium botulinum*, botulism, infant botulism cluster, single-nucleotide polymorphism, high-quality SNP typing

INTRODUCTION

Clostridium botulinum are spore-forming Gram-positive bacteria that produce a potent neurotoxin responsible for the severe paralytic disease botulism. In the United States, there are on average 153 cases of botulism each year, and the majority (110 cases per year) are attributed to infant botulism.[1] *C. botulinum* strains produce seven well-characterized botulinum toxin types (A-G), although human illness in the United States is primarily caused by toxin types A and B, followed by E and F. Some *C. botulinum* strains, denominated bivalent, can produce more than one toxin type (Sobel, 2005; Raphael et al., 2014).

---

[1]https://www.cdc.gov/botulism/surveillance.html

C. botulinum is a diverse species and has historically been characterized into four metabolic groups designated Groups I, II, III, and IV, based on biochemical and microbiological traits (Raphael et al., 2014; Smith et al., 2015; Williamson et al., 2016). Group I consists of proteolytic strains that produce botulinum toxin types A, B, and F (Smith et al., 2015; Williamson et al., 2016). Group I strains are most often associated with human clinical cases and are the most common cause of botulism cases in the United States. Group I also includes bivalent strains (Ab, Af, Ba, and Bf), and strains designated type A(B), which carry types A and B botulinum neurotoxin (*bont*) genes, but only produce type A toxin (Franciosa et al., 1994; Hill et al., 2007). These *C. botulinum* A(B) strains have been discovered to be common among the US botulism cases (Raphael et al., 2014; Halpin et al., 2017).

The disease botulism can be classified into four natural-occurring forms: foodborne botulism (through ingestion of foods contaminated with botulinum toxin), wound botulism (through spores that germinate leading to growth in wounds, and toxin production *in situ*), infant botulism (through intestinal colonization in infants less than 1 year old), and adult intestinal colonization (through intestinal colonization of patients older than 1 year; Sobel, 2005). Infant botulism is the most common type of botulism in the United States, and it occurs in persons under 1 year old, through ingestion of *C. botulinum* spores that germinate in the intestines, leading to growth and production of botulinum toxin (Sobel, 2005). Spores of *C. botulinum* are commonly found in soil, and therefore, it has been proposed as the most probable source of spores in infant botulism cases (Koepke et al., 2008).

Here, we present a follow-up study, to characterize *C. botulinum* strains isolated from clinical and environmental samples associated with the investigation conducted by Istre et al. (1986), these cases were of interest since transmission of botulism is not considered communicable (does transmit from person to person) and the incidence of infant botulism is extremely low. The first two cases occurred in 1981, 6 months apart, while the third case occurred in September of 1984. All three infants lived in the same neighborhood, within 800 m of each other. The three cases were confirmed by detection of botulinum neurotoxin type A in patient stools using the mouse bioassay (Istre et al., 1986). *C. botulinum* type A was isolated from all three of the patient's stools. *C. botulinum* type A was also present in cultures of environmental samples taken in 1982 and 1985, following the second and third cases. Specifically, *C. botulinum* type A was found in soil and vacuum dust samples taken from the homes of the second and third cases and also from the crib of the second case patient (Istre et al., 1986). *C. botulinum* type A was also found in soil of the first case patient's home after the fact, but no house dust samples were obtained as the family moved prior to the onset of the second case (Istre et al., 1986). This investigation prompted a study in 1986, conducted by CDC, to test additional environmental samples from homes of infants in the community who were considered healthy control infants, as they were presumed

not to be experiencing symptoms of botulism. Stool samples from these healthy infants were also collected. Ten stool samples, 22 soil samples, and 13 vacuum dust samples were collected: 91% (20/22) of soil samples from the homes of control infants yielded *C. botulinum* type A, and 77% (10/13) of vacuum dust samples from the homes of control infants also yielded *C. botulinum* type A. *C. botulinum* type A was not identified in the stool of any control infants; however, *C. botulinum* type B was found in the stools of two control infants (unpublished data). In this study, we employed a whole genome sequencing (WGS) approach to further characterize the clinical and environmental isolates associated with this investigation and to determine (1) whether the three cases represented a cluster of infant botulism cases and (2) whether the *C. botulinum* type A clinical isolates were related to the isolates found in environmental samples.

MATERIALS AND METHODS

Strains

Seventeen *C. botulinum* isolates were selected from the CDC strain collection for further characterization (refer to **Table 1**): four type A isolates from patient stool samples associated with cases one and three (unfortunately, the isolate from case 2 was not available in CDC's strain collection), two type B isolates from stools of two healthy infants, and 11 type A isolates from environmental samples (from cases two and three and the homes of six healthy infants). Eight of the isolates were from the original investigation by Istre et al., 1986, and nine isolates were from the unpublished study in 1986. Additional isolates for geographical perspective were also included in the study, based on availability of WGS data (refer to **Supplementary Table 1**). Strains were grown at 35°C under anaerobic conditions in Trypticase Peptone Yeast Extract (Remel, Lenexa, KS). All strains were coded prior to the study to comply with a human subjects research protocol. DNA was extracted using a modified MasterPure DNA extraction protocol (Epicenter, Madison, WI; Halpin et al., 2019).

Sequencing

Isolates were sequenced on either the Ion Torrent S5 or Illumina MiniSeq. DNA libraries were prepared for WGS using the Nextera DNA Flex kit (Illumina, San Diego, CA) and sequenced on the MiniSeq (Illumina, San Diego, CA) with 2 × 150 bp chemistry. DNA libraries were also prepared using the Ion Torrent Chef (Thermo Fisher Scientific, Waltham, MA) and sequenced on the Ion Torrent S5 (Thermo Fisher Scientific, Waltham, MA) using previously established methods (Halpin et al., 2019).

Genome Assembly and Quality Filtering

Genome sequences were downloaded from NCBI or assembled from raw reads using SPAdes. SPAdes version 3.14.0 and the option --careful were used for Illumina reads, while

TABLE 1 | *C. botulinum* strains included in this study.

CDC strain No./ NCBI accession	Year	Botulism type	Source type	Serotype *mouse bioassay*[c]	Toxin subtype *WGS*	Case information
CDC36910 SAMN17597327	1981	Infant	Stool	A	A1,B5	Case 1 clinical sample
CDC36911 SAMN17597328	1981	Infant	Stool	A	A1,B5	Case 1 clinical sample
CDC37208 SAMN17597330	1982	N/A	Vacuum dust[a]	A	A1,B5	Case 2 environmental sample
CDC37220 SAMN17597329	1982	N/A	Crib springs	A	A1,B5	Case 2 environmental sample
CDC37369 SAMN17597331	1982	N/A	Soil	A	A1,B5	Case 2 environmental sample
CDC37370 SAMN17597332	1982	N/A	Soil	A	A1,B5	Case 2 environmental sample
CDC39239 SAMN17597333	1984	Infant	Stool	A	A1,B5	Case 3 clinical sample
CDC39242 SAMN17597334	1984	Infant	Stool	A	A1,B5	Case 3 clinical sample
CDC31846 SAMN17597318	1986	N/A	Stool	B	B1	Healthy infant 1
CDC31956 SAMN17597320	1986	N/A	Soil	A	A1,B5	Associated with healthy infant 1
CDC31965 SAMN17597321	1986	N/A	Vacuum dust	A	A1,B5	Associated with healthy infant 2
CDC31966 SAMN17597322	1986	N/A	Soil	A	A1,B5	Associated with healthy infant 3
CDC31969 SAMN17597323	1986	N/A	Soil	A	A1,B5	Associated with healthy infant 4
CDC31973 SAMN17597324	1986	N/A	Vacuum dust	A	A1,B5	Associated with healthy infant 4
CDC31851 SAMN17597319	1986	N/A	Stool	B	Non-toxigenic[b]	Healthy infant 5
CDC31976 SAMN17597326	1986	N/A	Vacuum dust	A	A1,B5	Associated with healthy infant 5
CDC31974 SAMN17597325	1986	N/A	Vacuum dust	A	A1,B5	Associated with healthy infant 6

[a]*Vacuum cleaner dust was collected from the grandmother's home.*
[b]*Isolate does not harbor a bont/B gene – strain may have lost a plasmid with the toxin gene.*
[c]*Data obtained from previous laboratory results in the 1980s.*

SPAdes version 3.13.0 and the options -sc, --iontorrent, and --careful were used for Ion Torrent reads (Bankevich et al., 2012). Short contigs (<500 bp) were removed using CG-Pipeline script run_assembly_filterContigs.pl (Kislyuk et al., 2010). Sequence accession numbers are associated with NCBI Bioproject PRJNA428620; these can be found in **Table 1**.

Toxin Gene Location and Type Determination

bont gene subtypes were determined with CLC Genomics Workbench v10 Map to Reference feature (CLC bio, Aarhus, Denmark). Location of *bont* genes was determined using a command-line nucleotide blast with Blast+ version 2.9.0 (Camacho et al., 2009), and their location on either the chromosome or a plasmid was predicted using nucleotide blast of the *C. botulinum* database on NCBI. Seven-gene Multi-Locus Sequence Typing (MLST) types were determined by querying the *C. botulinum* PubMLST database[3] (Jacobson et al., 2008).

WGS Analysis

A single-nucleotide polymorphism (SNP) analysis was performed using only the highest quality SNPs. An appropriate reference genome for a high-quality SNP analysis was identified by first querying our in-house genome sequence repository. These in-house sequences were then compared to additional genome sequences on NCBI, using Mashtree (Katz et al., 2019). An appropriate reference was selected based on two criteria: closeness to the study strains in the Mashtree and completeness of the reference assembly. Plasmid sequences were identified in assemblies using PLSDB (Galata et al., 2019) and removed from the reference sequence to avoid erroneous SNP counts due to plasticity of some strains. Read pairs were preprocessed with Lyve-SET version 1.1.4f scripts prior to performing SNP analyses; read pairs were first interleaved with shuffleSplitReads.pl. and then cleaned with run_assembly_trimClean.pl. with the options --min_avg_quality 24 and --nosingletons (Katz et al., 2017). Lyve-SET version 1.1.4f was used for the SNP analysis with the following settings and the cleaned reads: --allowedFlanking 5 --min_alt_frac

0.75 --min_coverage 10 --mask-phages --mask-cliffs --mapper smalt --numcpus 10 --numnodes 50 (Katz et al., 2017). Lyve-SET quality filters for SNP coverage, percent consensus, and flanking distance ensure that only the highest quality SNPs are used in the analysis and that low-quality regions susceptible to recombination and horizontal exchange are not included (Katz et al., 2017). External reference genome sequences were also placed in the Lyve-SET asm folder for WGS read simulation as needed to root the tree. Phylogenetic trees were visualized using Mega7 (Kumar et al., 2016), and SNP distance matrices were visualized using a text editor or Microsoft Excel (Microsoft, Albuquerque, NM).

RESULTS

Toxin Gene Location and Type Determination

All 15 *C. botulinum* type A clinical and environmental isolates were confirmed to be A(B), *bont* gene subtypes A1 and B5 by WGS, and MLST type ST6. The *bont* genes were predicted to be on the chromosome for all *C. botulinum* type A(B) isolates associated with this investigation. Interestingly, the type A and type B toxin genes were located on a single 311,258 bp contig in one of the strains (CDC31973). The presence of both type A and type B toxin genes on the chromosome has been seldomly reported in *C. botulinum* type A(B) strains (Franciosa et al., 2009; Gonzalez-Escalona and Sharma, 2020). One *C. botulinum* type B isolate (CDC31846) from healthy infant stool was confirmed to harbor *bont*/B1 gene and was predicted to be on a plasmid. CDC31851, originally classified as *C. botulinum* type B by mouse bioassay in 1986, did not carry a *bont* gene and was determined as non-toxigenic in this study. The absence of the toxin gene in this isolate CDC31851 could have been due to loss of a plasmid. Both CDC31846 and CDC31851 were confirmed to be MLST type ST96.

WGS Analysis

The *C. botulinum* type A(B) study strains in **Table 1** clustered together in the Mashtree and also with other unrelated *C. botulinum* type A(B) from the United States (**Figure 1; Supplementary Table 1**). However, they did not cluster with *C. botulinum* type A(B) isolated from other infant and foodborne botulism cases in Colorado (CDC61017, CDC64216, CDC64218, and CDC64223; **Figure 1; Supplementary Table 1**).

The *C. botulinum* type A(B) study strains formed a monophyletic cluster in a Lyve-SET phylogeny and are considered closely related to each other (2–18 high-quality SNPs), but distinct from other closely related *C. botulinum* type A(B) from the United States (CDC43305, CDC43306, CDC69094, and CDC64144; **Figures 2A,B**). Interestingly, the clinical isolates (CDC36910, CDC36911, CDC39239, and CDC39242) had fewer
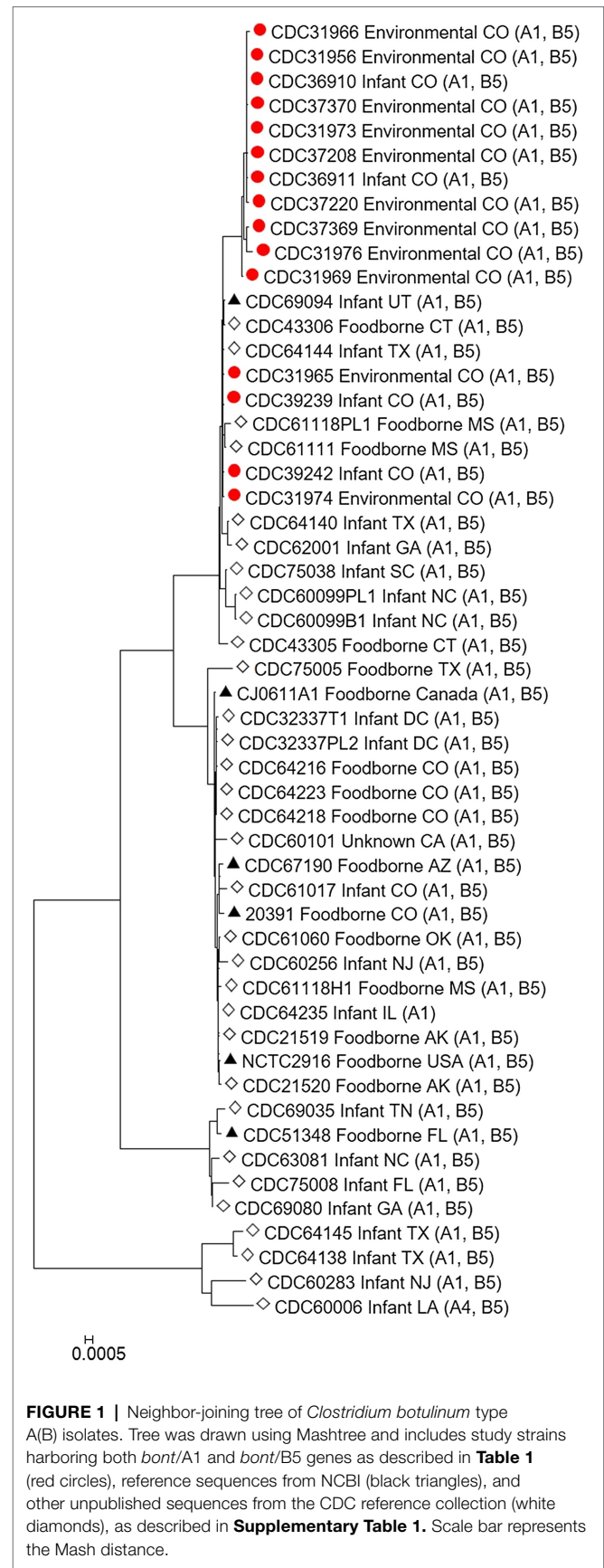
FIGURE 1 | Neighbor-joining tree of *Clostridium botulinum* type A(B) isolates. Tree was drawn using Mashtree and includes study strains harboring both *bont*/A1 and *bont*/B5 genes as described in **Table 1** (red circles), reference sequences from NCBI (black triangles), and other unpublished sequences from the CDC reference collection (white diamonds), as described in **Supplementary Table 1.** Scale bar represents the Mash distance.
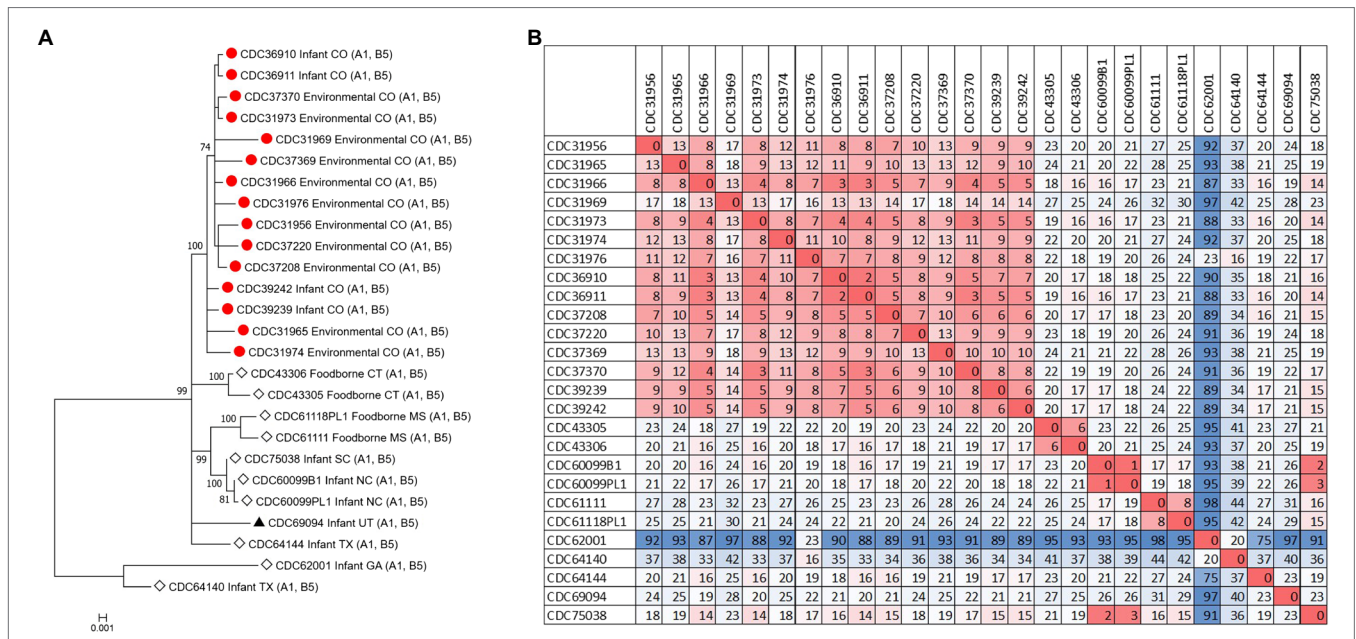
FIGURE 2 — (A) Maximum-likelihood whole genome high-quality SNP tree (panel A) with tip labels:

- CDC36910 Infant CO (A1, B5)
- CDC36911 Infant CO (A1, B5)
- CDC37370 Environmental CO (A1, B5)
- CDC31973 Environmental CO (A1, B5)
- CDC31969 Environmental CO (A1, B5)
- CDC37369 Environmental CO (A1, B5)
- CDC31966 Environmental CO (A1, B5)
- CDC31976 Environmental CO (A1, B5)
- CDC31956 Environmental CO (A1, B5)
- CDC37220 Environmental CO (A1, B5)
- CDC37208 Environmental CO (A1, B5)
- CDC39242 Infant CO (A1, B5)
- CDC39239 Infant CO (A1, B5)
- CDC31965 Environmental CO (A1, B5)
- CDC31974 Environmental CO (A1, B5)
- CDC43306 Foodborne CT (A1, B5)
- CDC43305 Foodborne CT (A1, B5)
- CDC61118PL1 Foodborne MS (A1, B5)
- CDC61111 Foodborne MS (A1, B5)
- CDC75038 Infant SC (A1, B5)
- CDC60099B1 Infant NC (A1, B5)
- CDC60099PL1 Infant NC (A1, B5)
- CDC69094 Infant UT (A1, B5)
- CDC64144 Infant TX (A1, B5)
- CDC62001 Infant GA (A1, B5)
- CDC64140 Infant TX (A1, B5)

Bootstrap values shown: 74, 100, 99, 100, 100, 99, 100, 81. Scale bar: 0.001.

(B) Distance matrix of SNP counts from Lyve-SET:

	CDC31956	CDC31965	CDC31966	CDC31969	CDC31973	CDC31974	CDC31976	CDC36910	CDC36911	CDC37208	CDC37220	CDC37369	CDC37370	CDC39239	CDC39242	CDC43305	CDC43306	CDC60099B1	CDC60099PL1	CDC61111	CDC61118PL1	CDC62001	CDC64140	CDC64144	CDC69094	CDC75038
CDC31956	0	13	8	17	8	12	11	8	8	7	10	13	9	9	9	23	20	20	21	27	25	92	37	20	24	18
CDC31965	13	0	8	18	9	13	12	11	9	10	13	13	12	9	10	24	21	20	22	28	25	93	38	21	25	19
CDC31966	8	8	0	13	4	8	7	3	5	5	7	4	4	5	5	18	16	16	17	23	21	87	33	16	19	14
CDC31969	17	18	13	0	13	17	16	13	13	14	17	18	14	14	14	27	25	24	26	32	30	97	42	25	28	23
CDC31973	8	9	4	13	0	8	7	4	4	5	8	9	3	5	5	19	16	16	17	23	23	88	33	16	20	14
CDC31974	12	13	8	17	8	0	11	10	8	9	12	13	11	9	9	22	20	20	21	27	24	92	37	20	25	18
CDC31976	11	12	7	16	7	11	0	7	7	8	9	12	8	8	8	22	18	19	20	26	24	23	16	19	22	17
CDC36910	8	11	3	13	4	10	7	0	2	5	8	9	5	7	7	20	17	18	18	25	24	90	35	18	21	16
CDC36911	8	9	5	13	4	8	7	2	0	5	8	9	3	5	5	19	16	16	17	23	21	88	33	16	20	14
CDC37208	7	10	5	14	5	9	8	5	5	0	7	10	6	6	6	20	17	17	18	23	20	89	34	16	21	15
CDC37220	10	13	7	17	8	12	9	8	8	7	0	13	9	9	9	23	18	19	20	26	24	91	36	19	24	18
CDC37369	13	13	4	18	9	13	12	9	9	10	13	0	10	10	10	24	21	21	22	28	26	93	38	21	25	19
CDC37370	9	12	4	14	3	11	8	5	3	6	9	10	0	8	8	22	19	19	20	26	24	91	36	19	22	17
CDC39239	9	9	5	14	5	9	8	7	5	6	9	10	8	0	6	20	17	17	18	24	22	89	34	17	21	15
CDC39242	9	10	5	14	5	9	8	7	5	6	9	10	8	6	0	20	17	17	18	24	22	89	34	17	21	15
CDC43305	23	24	18	27	19	22	22	20	19	20	23	24	22	20	20	0	6	23	22	26	25	95	41	23	27	21
CDC43306	20	21	16	25	16	20	18	17	16	17	18	21	19	17	17	6	0	20	21	25	24	93	37	20	25	19
CDC60099B1	20	20	16	24	16	20	19	18	16	17	19	21	19	17	18	23	20	0	1	17	17	93	38	21	26	2
CDC60099PL1	21	22	17	26	17	21	19	18	17	18	20	22	20	18	22	22	21	1	0	19	18	95	39	22	26	3
CDC61111	27	28	23	32	23	27	26	25	23	23	26	28	26	24	26	26	25	17	19	0	8	98	44	27	31	16
CDC61118PL1	25	25	21	30	21	24	20	24	26	24	25	25	24	24	24	25	24	17	18	8	0	95	42	24	29	15
CDC62001	92	93	87	97	88	92	23	90	88	89	91	93	91	89	89	95	93	93	95	98	95	0	20	75	97	91
CDC64140	37	38	33	42	33	37	16	35	33	34	36	38	36	34	34	41	37	38	39	44	42	20	0	37	40	36
CDC64144	20	21	16	25	16	20	18	17	16	17	18	21	19	17	17	23	22	27	24	27	24	75	37	0	23	19
CDC69094	24	25	19	28	20	25	22	21	20	21	24	25	22	21	21	27	25	26	26	31	29	97	40	23	0	23
CDC75038	18	19	14	23	14	18	17	16	14	15	18	19	17	15	15	21	19	2	3	16	15	91	36	19	23	0

FIGURE 2 | (A) Maximum-likelihood whole genome high-quality SNP tree of *C. botulinum* type A(B) isolates harboring both *bont*/A1 and *bont*/B5 genes. Tree was drawn using Lyve-SET and includes study strains as described in **Table 1** (red circles), reference sequences from NCBI (black triangles), and other unpublished sequences from *C. botulinum* type A(B) strains from the CDC reference collection (white diamonds). Scale bar represents the number of changes per 100 sites. Bootstrap values greater than 70 are shown. **(B)** Distance matrix of SNP counts from Lyve-SET. The matrix was drawn in Excel using conditional formatting.

SNP differences between isolates (5–7 high-quality SNPs), highlighting that the additional diversity (up to 18 SNP differences) in the *C. botulinum* type A(B) cluster is due to the environmental isolates, which were taken at different time points (see **Table 1; Figures 2A,B**). The two isolates recovered from healthy infant stools (*C. botulinum* type B CDC31846 and the non-toxigenic isolate CDC31851) were 28 SNPs different from each other and were considered unrelated (data not shown).

DISCUSSION

In this study, we sought to characterize *C. botulinum* type A isolated from clinical specimens from infant botulism cases reported in Colorado between 1981 and 1984 (Istre et al., 1986), and other *C. botulinum* type A isolated from environmental samples. Unfortunately, the clinical isolate associated with the second case was not available in CDC's strain collection and could not be included as part of this study; while this is a limitation of the study, we did include environmental isolates associated with the second case as part of this study. In a subsequent study done by CDC in 1986, stool samples were collected from 10 healthy control infants as well as soil and dust samples from their homes. The control stools were negative for *C. botulinum* type A, but two infants had *C. botulinum* type B in their stool (unpublished data). It is unclear how the two presumed healthy infants had *C. botulinum* type B in their stool, and to our knowledge, there are no reports of healthy infants harboring *C. botulinum* (Dowell et al., 1977). Additionally, *C. botulinum* type A was found in the environmental samples collected from homes of the healthy control infants. Our findings confirm this and add more to the investigation – the study strains also had a silent *bont*/B5 gene (Franciosa et al., 1994).

The source of *C. botulinum* spores in infant botulism cases is rarely determined, but exposure is thought to be through the environment, by ingestion of microscopic dust particles, or by consumption of honey (Dabritz et al., 2014). Furthermore, our results suggest that soil or dust may have been the probable source of spores in these three infant botulism cases, since the *C. botulinum* type A(B) strains isolated from environmental samples were closely related to the *C. botulinum* type A(B) strains isolated from stool samples. Some diversity (up to 18 SNPs) was observed among the environmental isolates, while very little diversity was observed in the clinical isolates (5 to 7 SNPs). This finding is interesting because the number of SNP differences is similar to what we would expect to see between unrelated clusters of *C. botulinum* A(B) in point source outbreaks (e.g., foodborne outbreaks). Raphael et al. reported SNP ranges of 3–8 between epidemiologically linked *C. botulinum* A(B) isolates of clinical and food pairs in six foodborne outbreaks using a reference-free SNP analysis (Raphael et al., 2014). Halpin et al. reported a range of 2–6 SNPs among *C. botulinum* clinical isolates and an epidemiologically linked contaminated powder infant formula isolate harboring the *bont*/B7 gene using a reference approach in Lyve-SET (Halpin et al., 2019). The differences observed among the

environmental isolates in this study may be a result of the diversity of *C. botulinum* spores in the environment where they are collected.

Our study demonstrates the utility of WGS analysis and its enhanced resolution for resolving clusters of *C. botulinum*, in particular of *C. botulinum* type A(B) strains, which have been historically challenging to differentiate (Hill et al., 2007; Raphael et al., 2014; Smith et al., 2015; Halpin et al., 2017). By using Lyve-SET high-quality SNP analysis, we were able to differentiate the study isolates from other closely related *C. botulinum* type A(B) isolates from Colorado and the United States that were not a part of this cluster, but otherwise appear related in a simple distance-based Mashtree; for example, CDC69094, CDC43306, CDC64144, CDC61118PL1, and CDC61111 which cluster in the Mashtree with study strains were unrelated to the study strains in the Lyve-SET phylogeny (see **Figures 1, 2A**). Katz et al. previously showed the usefulness of Lyve-SET for investigating closely related strains within outbreaks of *Listeria monocytogenes, Escherichia coli, Salmonella enterica*, and *Campylobacter jejuni* (Katz et al., 2017). At the time of the original investigation, the mouse bioassay was used for confirming the infant botulism cases, and for identifying the botulinum toxin serotype. The mouse bioassay does not, and it is not intended to provide resolution to resolve clusters of closely related strains nor identify an unexpressed botulinum toxin gene. Additional subtyping methods, such as pulsed-field gel electrophoresis, seven-gene MLST, multi-loci variable number of tandem repeat analysis, and amplified fragment length polymorphism analysis, have been utilized to subtype *C. botulinum* isolates in the last 25 years, but none of these subtyping tools can provide the level of resolution needed for resolving clusters of *C. botulinum* type A(B) strains (Raphael et al., 2014; Smith et al., 2015; Halpin et al., 2017). For instance, our study showed that all of the type A(B) isolates shared seven-gene MLST type ST6 with other type A(B) isolates that were not part of the cluster of infant botulism cases (data not shown). Our study builds on previous work to show the utility of reference-based whole genome SNP typing to successfully resolve clusters of *C. botulinum* (Gonzalez-Escalona et al., 2014; Raphael et al., 2014; Williamson et al., 2016; Gonzalez-Escalona and Sharma, 2020). In conclusion, results of this study suggest that these *C. botulinum* type A(B) isolates obtained from infant botulism cases are genetically related to each other and to the environmental isolates obtained from soil and dust samples.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found online at the NCBI Sequence Read Archive (SRA): https://www.ncbi.nlm.nih.gov/sra/PRJNA428620.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at https://www.frontiersin.org/articles/10.3389/fmicb.2021.688240/full#supplementary-material

REFERENCES

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10:421. doi: 10.1186/1471-2105-10-421

Dabritz, H. A., Hill, K. K., Barash, J. R., Ticknor, L. O., Helma, C. H., Dover, N., et al. (2014). Molecular epidemiology of infant botulism in California and

elsewhere, 1976–2010. *J. Infect. Dis.* 210, 1711–1722. doi: 10.1093/infdis/jiu331

Dowell, V. R. Jr., McCroskey, L. M., Hatheway, C. L., Lombard, G. L., Hughes, J. M., and Merson, M. H. (1977). Coproexamination for botulinal toxin and *Clostridium botulinum*. A new procedure for laboratory diagnosis of botulism. *JAMA* 238, 1829–1832. doi: 10.1001/jama.1977.03280180033021

Franciosa, G., Ferreira, J. L., and Hatheway, C. L. (1994). Detection of type A, B, and E botulism neurotoxin genes in *Clostridium botulinum* and other clostridium species by PCR: evidence of unexpressed type B toxin genes in type A toxigenic organisms. *J. Clin. Microbiol.* 32, 1911–1917. doi: 10.1128/JCM.32.8.1911-1917.1994

Franciosa, G., Maugliani, A., Scalfaro, C., and Aureli, P. (2009). Evidence that plasmid-borne botulinum neurotoxin type B genes are widespread among *Clostridium botulinum* serotype B strains. *PLoS One* 4:e4829. doi: 10.1371/journal.pone.0004829

Galata, V., Fehlmann, T., Backes, C., and Keller, A. (2019). PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res.* 47, D195–D202. doi: 10.1093/nar/gky1050

Gonzalez-Escalona, N., and Sharma, S. K. (2020). Closing *Clostridium botulinum* group I genomes using a combination of short- and long-reads. *Front. Microbiol.* 11:239. doi: 10.3389/fmicb.2020.00239

Gonzalez-Escalona, N., Timme, R., Raphael, B. H., Zink, D., and Sharma, S. K. (2014). Whole-genome single-nucleotide-polymorphism analysis for discrimination of *Clostridium botulinum* group I strains. *Appl. Environ. Microbiol.* 80, 2125–2132. doi: 10.1128/AEM.03934-13

Halpin, J. L., Dykes, J. K., Katz, L., Centurioni, D. A., Perry, M. J., Egan, C. T., et al. (2019). Molecular characterization of *Clostridium botulinum* harboring the bont/B7 gene. *Foodborne Pathog. Dis.* 16, 428–433. doi: 10.1089/fpd.2018.2600

Halpin, J. L., Joseph, L., Dykes, J. K., McCroskey, L., Smith, E., Toney, D., et al. (2017). Pulsotype diversity of *Clostridium botulinum* strains containing serotypes A and/or B genes. *Foodborne Pathog. Dis.* 14, 494–501. doi: 10.1089/fpd.2017.2280

Hill, K. K., Smith, T. J., Helma, C. H., Ticknor, L. O., Foley, B. T., Svensson, R. T., et al. (2007). Genetic diversity among botulinum neurotoxin-producing clostridial strains. *J. Bacteriol.* 189, 818–832. doi: 10.1128/JB.01180-06

Istre, G. R., Compton, R., Novotny, T., Young, J. E., Hatheway, C. L., and Hopkins, R. S. (1986). Infant botulism. Three cases in a small town. *Am. J. Dis. Child.* 140, 1013–1014. doi: 10.1001/archpedi.1986.02140240059027

Jacobson, M. J., Lin, G., Whittam, T. S., and Johnson, E. A. (2008). Phylogenetic analysis of *Clostridium botulinum* type A by multi-locus sequence typing. *Microbiology* 154, 2408–2415. doi: 10.1099/mic.0.2008/016915-0

Katz, L. S., Griswold, T., Morrison, S. S., Caravas, J. A., Zhang, S., den Bakker, H. C., et al. (2019). Mashtree: a rapid comparison of whole genome sequence files. *J. Open Source Softw.* 4:1762. doi: 10.21105/joss.01762

Katz, L. S., Griswold, T., Williams-Newkirk, A. J., Wagner, D., Petkau, A., Sieffert, C., et al. (2017). A comparative analysis of the Lyve-SET phylogenomics pipeline for genomic epidemiology of foodborne pathogens. *Front. Microbiol.* 8:375. doi: 10.3389/fmicb.2017.00375

Kislyuk, A. O., Katz, L. S., Agrawal, S., Hagen, M. S., Conley, A. B., Jayaraman, P., et al. (2010). A computational genomics pipeline for prokaryotic sequencing projects. *Bioinformatics* 26, 1819–1826. doi: 10.1093/bioinformatics/btq284

Koepke, R., Sobel, J., and Arnon, S. S. (2008). Global occurrence of infant botulism, 1976–2006. *Pediatrics* 122, e73–e82. doi: 10.1542/peds.2007-1827

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Raphael, B. H., Shirey, T. B., Luquez, C., and Maslanka, S. E. (2014). Distinguishing highly-related outbreak-associated *Clostridium botulinum* type A(B) strains. *BMC Microbiol.* 14:192. doi: 10.1186/1471-2180-14-192

Smith, T. J., Hill, K. K., and Raphael, B. H. (2015). Historical and current perspectives on *Clostridium botulinum* diversity. *Res. Microbiol.* 166, 290–302. doi: 10.1016/j.resmic.2014.09.007

Sobel, J. (2005). Botulism. *Clin. Infect. Dis.* 41, 1167–1173. doi: 10.1086/444507

Williamson, C. H., Sahl, J. W., Smith, T. J., Xie, G., Foley, B. T., Smith, L. A., et al. (2016). Comparative genomic analyses reveal broad diversity in botulinum-toxin-producing clostridia. *BMC Genomics* 17:180. doi: 10.1186/s12864-016-2502-z