frontiers | Research Topics

New insights on botulism, botulinum neurotoxins, and botulinum toxin-producing clostridia

Edited by
Fabrizio Anniballi, John W. Austin and Theresa Smith

Published in
Frontiers in Microbiology
Frontiers in Public Health

ISSN 1664-8714
ISBN 978-2-83251-555-6
DOI 10.3389/978-2-83251-555-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

New insights on botulism, botulinum neurotoxins, and botulinum toxin-producing clostridia

Topic editors

Fabrizio Anniballi — Department of Food Safety, Nutrition and Veterinary Public Health, National Institute of Health (ISS), Italy

John W. Austin — Health Canada, Canada

Theresa Smith — Northern Arizona University, United States

Citation

Anniballi, F., Austin, J. W., Smith, T., eds. (2023). *New insights on botulism, botulinum neurotoxins, and botulinum toxin-producing clostridia*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-555-6

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 18 July 2022
DOI 10.3389/fmicb.2022.979653

OPEN ACCESS

EDITED AND REVIEWED BY
Giovanna Suzzi,
University of Teramo, Italy

*CORRESPONDENCE
Fabrizio Anniballi
fabrizio.anniballi@iss.it

SPECIALTY SECTION
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 27 June 2022
ACCEPTED 04 July 2022
PUBLISHED 18 July 2022

CITATION
Anniballi F, Smith TJ and Austin JW (2022) Editorial: New insights on botulism, botulinum neurotoxins, and botulinum toxin-producing clostridia. *Front. Microbiol.* 13:979653. doi: 10.3389/fmicb.2022.979653

Editorial: New insights on botulism, botulinum neurotoxins, and botulinum toxin-producing clostridia

Fabrizio Anniballi[1]*, Theresa J. Smith[2] and John W. Austin[3]

[1]Microbiological Food Safety and Food-Borne Diseases Unit, Department of Food Safety, Nutrition and Veterinary Public Health, National Reference Centre for Botulism, Istituto Superiore di Sanità, Rome, Italy, [2]Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, United States, [3]Botulism Reference Service for Canada, Microbiology Research Division, Bureau of Microbial Hazards, Food Directorate, Health Products and Food Branch, Ottawa, ON, Canada

KEYWORDS

botulinum neurotoxins, botulinum toxin-producing clostridia, wound botulinum, animal botulism, CRISPR-Cas systems

Editorial on the Research Topic
New Insights on Botulism, Botulinum Neurotoxins, and Botulinum Toxin-producing Clostridia

Botulism is a life-threatening disease affecting humans, and many warm-blooded animals and fishes. Botulism is caused by botulinum neurotoxins (BoNTs) produced by anaerobic, spore-forming bacteria belonging to the genus *Clostridium*. Although the disease in humans is rare, each botulism case constitutes a public health emergency. A prompt clinical diagnosis and laboratory confirmation are essential for the correct management of patients and activating proper public health measures addressed to avoid the further cases. Conversely, botulism is more common in animals and may result in a high mortality rate, eliciting environmental, and economic concerns.

Both human and animal botulism have been classified into different forms based on route of exposure to BoNTs. Foodborne and iatrogenic botulism are based on ingestion/injection of toxin, while infant botulism and wound botulism are toxicoinfections that result after the growth and expression of bacteria within the body. On a worldwide basis, food-borne botulism, due to the consumption of foods contaminated with preformed toxins, is the most widely recognized form. Home-prepared foods are mainly responsible for sporadic cases, and small outbreaks occur primarily at the familiar level. Commercially produced foods may cause large outbreaks affecting multiple regions or countries. Since botulism can result in respiratory failure requiring hospitalization in intensive care units for weeks or months, these latter outbreaks pose a challenge for local and regional health systems. As highlighted by Lúquez et al. in their manuscript on food-borne botulism outbreaks that occurred in the US from 2001 to 2017, modern hospitals can admit to intensive care units only a limited number of patients. Thus, a local outbreak of foodborne botulism could easily overwhelm

the hospital system with worried asymptomatic or paucisymptomatic people, delaying ventilatory assistance and botulinum antitoxin administration to those who need it most. Local health systems should consider this when implementing and revising their emergency preparedness plans. For optimal management of a botulism case/outbreak, it is fundamental that early clinical diagnosis and rapid activation of the epidemiological surveillance system is aimed at identifying the implicated food vehicle and avoiding new cases.

Clinical diagnosis might be difficult because of the rarity of botulism and because, at the onset, symptoms and clinical signs may be non-specific. The difficulty increases if the patient is a drug user affected by wound botulism because some neurologic signs may be confused with the drug effects. In wound botulism cases, the physician's awareness and timely diagnosis are crucial to adopting suitable treatments, such as botulinum antitoxin administration and debridement of the wound. In their manuscript, Middaugh et al. confirmed that early clinical diagnosis and prompt administration of antitoxin prevents respiratory failure and reduces hospital stay. In addition, they highlighted the importance of wound botulism surveillance among close contacts with persons who injected drugs, as well as the importance of health alerts to raise clinicians' awareness of wound botulism. Molecular surveillance of microbial isolates could help identify the source of contamination in injectable drugs. Since *Clostridium botulinum* is an environmental microorganism, spores may contaminate the drug at different stages of production and processing, such as where the raw materials were grown and during cutting and diluting. Contamination at the site of use also cannot be ruled out. To trace the contamination source of drug consumed by an injection drug user affected by wound botulism, Halpin et al. studied the phylogenetic relatedness of *C. botulinum* isolated in Hawaii from wound and infant botulism cases. *C. botulinum* strains isolated from infant botulism cases were chosen because they have been considered representative of the environment in which cases occurred. As the main finding, Halpin et al. found a high homology among isolates, supporting the hypothesis that the organism implicated in wound botulism in this case contaminates the drug paraphernalia or the wound itself locally rather than at the production site or during transport.

Molecular sub-typing of isolated strains is crucial for source attribution. In this respect, several molecular biology techniques have been developed. The Multiple Locus Variable-number of tandem repeat Analysis (MLVA) was successfully adopted by Souillard et al. during their investigations of an outbreak of cattle botulism. Souillard et al. identified a poultry house as the source of contamination in a massive type D/C cattle botulism outbreak on a mixed dairy and broiler farm in France. This was accomplished by monitoring carriage in the broilers, the ventilation system of the poultry house, and contamination of equipment from the hatchery used for delivering the chicks. As a second-generation sub-typing technique, MLVA is less discriminating than Whole Genome Sequencing (WGS); however, as a PCR-based technique, it can be performed using only a tiny amount of DNA template. If a species-specific scheme is used (as Souillard et al. have done), MLVA avoids the strain isolation and purification step that is time-consuming and particularly challenging for the microbial agents of animal botulism.

For genomic characterization of strains isolated from a cluster of infant botulism type A cases, Gladney et al. used WGS single-nucleotide polymorphism (SNP) analysis. Using Lyve-SET high-quality SNP analysis, they were able to differentiate strains isolated in their cluster from other *C. botulinum* type A(B), demonstrating the high-resolution level of this technique. Another molecular approach adopted for bacterial sub-typing uses Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) analysis. CRISPR and CRISPR-associated proteins (Cas) are a prokaryotic adaptative immunity and regulatory system that interferes with invading phages and plasmids and can be used to gain direct insight into horizontal gene transfer events. The analysis of CRISPR-Cas systems was carried out by Le Gradiet et al. testing 58 *Clostridium novyi sensu lato* genospecies (responsible for animal botulism), while Wentz et al. profiled endogenous CRISPR-Cas systems from 241 Group I *C. botulinum* (proteolytic strains) and *Clostridium sporogenes* genomes. Le Gradiet et al. conducted their studies to explore CRISPR-Cas systems in the *Clostridium novyi sensu lato* strains to evaluate their presence, determine their characteristics, and explore the protospacer origins to gain insight into the mobile genetic elements (MGE) interacting with this taxon. They found that CRISPR-Cas systems are numerous in *Clostridium novyi sensu lato* strains and may present in the bacterial chromosome or MGE. The components carrying out CRISPR-Cas systems seem to have been recruited as anti-MGE systems and for inter-MGE conflicts to protect mainly against a restricted number of MGEs. In their investigations, Wentz et al. found that the *bont* gene cluster was not directly targeted by endogenous CRISPR-Cas systems in Group I *C. botulinum* and *C. sporogenes*. An extensive study on the evolution of *bont* gene clusters harbored by Group I *C. botulinum* was performed by Smith et al. They found that strains isolated in the northern hemisphere harbor *bont* gene clusters containing *ha* genes, whilst those isolated in the southern hemisphere primarily harbor *bont* gene clusters containing *orfX* genes. In addition, they found that the movement of *bont* gene clusters into non-neurotoxigenic clostridia is a one-way process that occurs *via* introduction within extrachromosomal plasmids followed by chromosomal integration.

Among CRISPR-Cas systems, CRISPR-Cas9 is widely used as a versatile tool to perform gene editing. In this respect, Mertaoja et al. published an elegant paper which exemplified CRISPR-Cas9 bookmark technology to construct mutants suitable to investigate the sporulation

mechanism in Group II *C. botulinum* (non-proteolytic) strains. Moreover, they demonstrated the potential of the bookmark technology in functional studies in which a chromosomal gene is replaced, in a two-step process, with a derivatized copy.

In summary, the articles included in this Research Topic provide an overview of current research activities in the botulism field and illustrate the growing use of genomic techniques in epidemiologic investigations as well as basic research applications. SNP analysis of whole-genome sequences was used to investigate the relatedness of wound botulism isolates in Hawaii and infant botulism isolates in Colorado, and MLVA studies were used to trace the source of contamination in an outbreak involving cattle. Genomic analysis of a range of clostridial strains having *orfX*+ *bont* gene clusters provided insights as to the evolutionary origins and spread of these strains. The realization of the importance of CRISPR-Cas systems in manipulating mobile genetic elements within bacteria has inspired articles investigating the links between these systems and mobile genetic elements, including exploitation of these systems to better understand the factors contributing to spore formation in botulinum neurotoxin-producing clostridia. These articles highlight the main challenges and propose solutions to improve management and control of the disease, stressing the power of the newer techniques as suitable tools to enhance botulism prevention measures and address future trends in botulism research.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors would like to acknowledge all the 53 colleagues who submitted their high-quality manuscripts and the reviewers that dedicated their time in the peer-reviewing process.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 January 2021
doi: 10.3389/fmicb.2021.617269

CRISPR-Cas9-Based Toolkit for *Clostridium botulinum* Group II Spore and Sporulation Research

Anna Mertaoja[1†], Maria B. Nowakowska[1†], Gerald Mascher[1], Viivi Heljanko[1], Daphne Groothuis[2], Nigel P. Minton[2] and Miia Lindström[1]*

[1]*Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland,* [2]*Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), Biodiscovery Institute, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom*

OPEN ACCESS

Edited by:
John W. Austin, Health Canada, Canada

Reviewed by:
Yi Wang, Auburn University, United States
Bo Segerman, National Veterinary Institute, Sweden
Francois Wasels, IFP Energies nouvelles, France

***Correspondence:**
Miia Lindström
miia.lindstrom@helsinki.fi

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *14 October 2020*
Accepted: *05 January 2021*
Published: *27 January 2021*

Citation:
Mertaoja A, Nowakowska MB, Mascher G, Heljanko V, Groothuis D, Minton NP and Lindström M (2021) CRISPR-Cas9-Based Toolkit for Clostridium botulinum Group II Spore and Sporulation Research. Front. Microbiol. 12:617269. doi: 10.3389/fmicb.2021.617269

The spores of *Clostridium botulinum* Group II strains pose a significant threat to the safety of modern packaged foods due to the risk of their survival in pasteurization and their ability to germinate into neurotoxigenic cultures at refrigeration temperatures. Moreover, spores are the infectious agents in wound botulism, infant botulism, and intestinal toxemia in adults. The identification of factors that contribute to spore formation is, therefore, essential to the development of strategies to control related health risks. Accordingly, development of a straightforward and versatile gene manipulation tool and an efficient sporulation-promoting medium is pivotal. Our strategy was to employ CRISPR-Cas9 and homology-directed repair (HDR) to replace targeted genes with mutant alleles incorporating a unique 24-nt "bookmark" sequence that could act as a single guide RNA (sgRNA) target for Cas9. Following the generation of the sporulation mutant, the presence of the bookmark allowed rapid generation of a complemented strain, in which the mutant allele was replaced with a functional copy of the deleted gene using CRISPR-Cas9 and the requisite sgRNA. Then, we selected the most appropriate medium for sporulation studies in *C. botulinum* Group II strains by measuring the efficiency of spore formation in seven different media. The most effective medium was exploited to confirm the involvement of a candidate gene in the sporulation process. Using the devised sporulation medium, subsequent comparisons of the sporulation efficiency of the wild type (WT), mutant and "bookmark"-complemented strain allowed the assignment of any defective sporulation phenotype to the mutation made. As a strain generated by complementation with the WT gene in the original locus would be indistinguishable from the parental strain, the gene utilized in complementation studies was altered to contain a unique "watermark" through the introduction of silent nucleotide changes. The mutagenesis system and the devised sporulation medium provide a solid basis for gaining a deeper understanding of spore formation in *C. botulinum*, a prerequisite for the development of novel strategies for spore control and related food safety and public health risk management.

Keywords: *Clostridium botulinum* Group II, CRISPR-Cas9, sporulation medium, *spo0A*, spore

INTRODUCTION

Clostridium botulinum is a Gram-positive, strictly anaerobic, spore-forming bacterium. *C. botulinum* strains are traditionally classified into four genetically and physiologically diverse groups designated I–IV (Hatheway, 1993). The critical characteristic of all four groups is their ability to produce extremely potent neurotoxin which is the causative agent of botulism, a rare but deadly disease affecting humans and animals (Sobel, 2005). *C. botulinum* strains of Group I and Group II are usually associated with human botulism, which classically manifests as a food poisoning. This form of botulism results from consumption of foods where *C. botulinum* spores germinated and outgrew into a neurotoxigenic culture. Their toxinogenic potential at refrigeration temperatures makes *C. botulinum* Group II strains a particular food safety concern (Lindström et al., 2006). Moreover, *C. botulinum* spores can colonize wounds or the digestive tract of infants and susceptible adults with compromised gut microbiota, causing the toxicoinfectious forms of wound and intestinal botulism, respectively (Harris, 2015).

Any interventions for controlling the risks posed by *C. botulinum* spores rely on the fundamental understanding of the spore composition and resistance properties, as well as the processes of spore formation, spore germination, and neurotoxin production. High-quality mechanistic research is reliant on precise deletions, genome manipulations, and efficient screening assays for the targeted phenotypes. Some progress has been made in recent times as a consequence of the development of appropriate genome editing tools (Heap et al., 2007; Ng et al., 2013; Minton et al., 2016; Cañadas et al., 2019; Ingle et al., 2019; Kuehne et al., 2019). These include the extensively utilized ClosTron together with various allelic exchange methodologies which have been exploited to successfully construct insertion and deletion mutants of sporulation and toxin-related genes in *C. botulinum* (Dahlsten et al., 2013; Kirk et al., 2014; Mascher et al., 2014, 2017; Zhang et al., 2014; Clauwers et al., 2016, 2017). The ClosTron, however, like any insertional mutagen is compromised by the fact that mutants may have altered phenotypes due to polar effects on genes located downstream of the intron insertion (Heap et al., 2007). Although polar effects may be overcome using allelic exchange to make in-frame, markerless deletion mutants (reviewed in Minton et al., 2016), the strategy utilized so far in *C. botulinum* Group II strains has resulted in the insertion of an antibiotic resistance gene (Clauwers et al., 2016, 2017). To make markerless deletions, counterselection markers are necessary, such as *pyrE* and *codA* (Minton et al., 2016), but their deployment requires the use of minimal media. Finding a minimal medium that is applicable to metabolically diverse *C. botulinum* strains is, however, challenging (Whitmer and Johnson, 1988).

Recently, the ease with which allelic exchange may be deployed in clostridia, including *C. botulinum* Group I strains, has been considerably improved through the exploitation of clustered regularly interspaced short palindromic repeats-CRISPR-associated protein 9 (CRISPR-Cas9; Cañadas et al., 2019; Ingle et al., 2019; McAllister and Sorg, 2019). Using such an approach, the allelic-exchange driven gene deletion makes the resulting mutant cells immune to Cas9-cleavage. This enables direct selection of mutant colonies, as opposed to having to be passaged onto specialized media where counterselection markers are employed. Consequently, the reduced number of steps leading to the mutant isolation decreases the likelihood of accumulating undesired ancillary mutations [single nucleotide polymorphisms (SNPs) and/or insertion/deletions (InDels)] that may affect phenotype. Thus, CRISPR-Cas9-based methods offer distinct advantages for generating mutants in *C. botulinum* Group II strains.

Another critical component in the fundamental spore research is a medium that efficiently promotes sporulation and ensures sufficient spore yields for downstream applications. While *Bacillus* spp. sporulation is mainly induced by starvation (Grossman and Losick, 1988), the factors triggering clostridial sporulation appear more complex (Perkins and Tsuji, 1962; Tsuji and Perkins, 1962; Roberts, 1965, 1967; Strasdine and Melville, 1968; Hawirko et al., 1979). Furthermore, the four different *C. botulinum* groups have distinct physiology and metabolism and therefore they respond differently to environmental cues (Woods and Jones, 1987) and assumed triggers of sporulation (Roberts, 1967; Solomon and Kautter, 1979). The routine growth media used for *C. botulinum* Group II strains yield insufficient numbers of spores for downstream applications [e.g., tryptone-peptone-glucose-yeast extract broth (TPGY) 10^5 spores/ml (Mascher et al., 2017)]. Several studies have shown evidence that biphasic media consisting of a solid agar phase and a liquid water phase can support efficient sporulation of Group II strains. However, the solid phase is prone to crumbling and thus makes sampling for many downstream applications difficult (Bruch et al., 1968; Peck et al., 1992). In order to facilitate spore biology research in *C. botulinum* Group II, a medium triggering efficient sporulation and enabling easy sampling is urgently needed.

Here we took the opportunity to exemplify CRISPR-Cas9 "bookmark" technology, a recently proposed concept (Seys et al., 2020) for gold standard complementation studies, which exploits CRISPR-Cas9 to replace the mutant allele *in situ* with a wild type (WT) allele. Moreover, we established a novel medium for the characterization of spore formation in *C. botulinum* Group II strains and used it to determine the suitability of a previously described CRISPR-Cas9 system (Ingle et al., 2019) for spore mutant generation in these strains. To exemplify these systems, we targeted the *spo0A* gene of *C. botulinum* Group II strain Beluga as its inactivation has been shown to result in an asporogenous phenotype (Al-Hinai et al., 2015; Mascher et al., 2017). The exemplified procedures form a solid basis for future spore-related studies in *C. botulinum* Group II strains.

MATERIALS AND METHODS

Experimental Design

The study consisted of four parts: (1) sporulation mutants of *C. botulinum* Group II type E strain Beluga were constructed using a novel CRISPR-Cas9 bookmark approach (Seys et al., 2020).

(2) Growth and sporulation of WT Beluga was tested in seven different biphasic media. The spore and viable cell counts were enumerated directly after inoculating the medium, and subsequently at 1 and 6 days after inoculation. (3) Growth and sporulation of three different Group II strains (Beluga, Eklund 17B, FT10F) were tested in cooked meat medium-TPGY (CMM-TPGY), which was the medium selected as best meeting the study criteria in (2), and in two control media: TPGY and Duncan-Strong (DS), the latter supporting the sporulation of *Clostridium perfringens* (Duncan and Strong, 1968). Spore and viable cell counts were enumerated directly after inoculating the medium, and subsequently at 1 day, 1 week, and 2 weeks after inoculation. Phase-contrast microscopy of the cultures was performed 2 weeks after inoculation. (4) The mutant strains constructed in (1) were characterized and validated in CMM-TPGY. Spore and viable cell count assays and phase-contrast microscopy were performed as in (3).

Strains, Media, and Growth Conditions

All the *C. botulinum* and *Escherichia coli* strains used are listed in **Supplementary Table S1**. *C. botulinum* cultures were routinely grown at 30°C in an anaerobic workstation (MG1000 anaerobic workstation; Don Whitley Scientific Ltd., Shipley, United Kingdom) with an atmosphere of 85% N_2, 10% CO_2 and 5% H_2. Spore or glycerol stocks of *C. botulinum* strains were revived in TPGY broth composed of 5% (w/v) tryptone, 0.5% peptone, 2% yeast extract (Difco, BD Diagnostic Systems, Sparks, MD), 0.4% glucose (VWR Chemicals, Leuven, Belgium), and 0.1% sodium thioglycolate (Merck, Darmstadt, Germany) by inoculating 5 ml of the medium with either 5 µl of spore stock or 50 µl of glycerol stock and incubating until growth reached stationary phase. An overnight culture was prepared by inoculating 5 ml of TPGY with 50 µl of revived culture. All the media used for *C. botulinum* growth experiments were deoxygenated prior to use, either by boiling for 20 min or by exposure to anaerobic conditions for at least 24 h. All tests were run in triplicate.

In part 1 of the study, transformed strains were grown on TPGY plates containing 1.5% (w/v) bacteriological agar supplemented with 250 µg ml^{-1} cycloserine (Merck) and 15 µg ml^{-1} thiamphenicol (Merck). *E. coli* strains used for cloning and conjugation procedures were grown in Luria-Bertani (LB; Invitrogen, Carlsbad, CA) broth or on LB agar plates incubated aerobically at 37°C. For molecular cloning, LB was supplemented with 25 µg ml^{-1} chloramphenicol (Merck) and for conjugation with 25 µg ml^{-1} chloramphenicol and 30 µg ml^{-1} kanamycin (Merck).

In part 2 of the study, *C. botulinum* strain Beluga was grown in seven different biphasic media: (i) egg-yolk-agar (EYA)-H_2O consisting of a solid EYA phase [2% peptone, 0.5% yeast extract, 0.5% tryptone, 0.5% NaCl, 10% (v/v) egg yolk emulsion (Oxoid Microbiology Products, Basingstoke, United Kingdom), and 1.5% agar] and a liquid phase of sterile water; (ii) TPGY-H_2O consisting of a solid TPGY agar phase (TPGY broth with 1.5% bacteriological agar (Amresco, Ohio, United States) and a liquid phase of sterile water; (iii) Reinforced Clostridial Medium (RCM)-H_2O consisting of a solid RCM phase [Reinforced Clostridial Medium (Lab M limited, Lancashire, United Kingdom) prepared according to manufacturer's instructions, and 1.5% bacteriological agar] and a liquid phase of sterile water; (iv) CMM)-H_2O consisting of a solid phase of 10% (w/v) cooked meat medium (Oxoid) with 0.1% glucose and 1.5% agar, and sterile water; (v) CMM-TPGY consisting of a solid CMM phase as in (iv) and a liquid phase of TPGY broth; (vi) agar-TPGY consisting of a solid phase of 1.5% agar and a liquid phase of TPGY broth, and (vii) agar-H_2O consisting of a solid phase of 1.5% agar and a liquid phase of sterile water. All biphasic media used in part 2 had a 150-ml solid phase and a 20-ml liquid phase. The liquid phase was inoculated with 5 ml of overnight culture in TPGY to a total volume of 25 ml.

In part 3 of the study, Group II strains Beluga, Eklund 17B, and FT10F were grown in CMM-TPGY containing a solid phase of 75 ml and a 50 ml liquid phase. TPGY broth and DS medium [composed of Modified Duncan-Strong Medium (HiMedia Laboratories Pvt. Ltd., Mumbai, India] prepared according to the manufacturer's instructions, volume 75 ml) were used as controls. The media were inoculated with 75 µl of overnight culture in TPGY.

In part 4 of the study, Beluga sporulation mutant strains, a plasmid control strain, and the WT strain were grown in CMM-TPGY supplemented with 15 µg ml^{-1} thiamphenicol when appropriate.

Plasmid Design and Mutant Construction

In part 1 of the study, we constructed sporulation mutants of *C. botulinum* Beluga using a novel CRISPR-Cas9 bookmark approach (Seys et al., 2020). All PCR reactions for the cloning procedures were performed using the KOD Hot-Start high fidelity DNA polymerase (Merck). For DNA purification, the GeneJET PCR Purification Kit (Thermo Fisher Scientific, Waltham, MA) or Monarch DNA Gel Extraction Kit (New England Biolabs, Ipswich, MA) was used. Ligation reactions were performed using T4 DNA ligase (Promega Corporation, Madison, WI). For the restriction digestion of DNA, NEB restriction endonucleases were used (New England Biolabs). Plasmid extractions were carried out using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Colony PCR screening was performed using the Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific). The DNA concentration was measured using the NanoDrop 1,000 Spectrophotometer (Thermo Fisher Scientific). For DNA size reference in gel electrophoresis, 1 kb Plus DNA Ladder was used (New England Biolabs).

All the primers (Metabion International AG) used are listed in **Supplementary Table S2**. To construct the pMTL431511-Beluga Δ*spo0A*::bm plasmid for generating a Δ*spo0A*::bm genome alteration, we followed a two-step cloning procedure. Firstly, we constructed CRISPR-Cas9 plasmid pMTL431511-Beluga Δ*spo0A*, which served as a template for the final pMTL431511-Beluga Δ*spo0A*::bm plasmid, and carried a copy of a gene encoding truncated but functional Cas9 (Ingle et al., 2019). For the construction of pMTL431511-Beluga Δ*spo0A*,

we generated a knockout (KO) cassette consisting of fused, 1,000-bp long regions flanking *spo0A* in the genome, hereafter called the left and right homology arms (LHA and RHA, respectively). The KO cassette was designed to remove the genomic *spo0A* codons 3–270, inclusively, generating an in-frame deletion of *spo0A*. Fragments encoding LHA and RHA were amplified from the genomic DNA of *C. botulinum* Beluga WT strain using two primer pairs (i) F _LHA_*spo0A*-AsiSI with R_LHA_*spo0A* and (ii) F_RHA_*spo0A* with R_RHA_*spo0A*-AscI. Obtained PCR products were purified and fused in splicing by overhang extension-polymerase chain reaction (SOE-PCR) using a pair of primers F_LHA_*spo0A*-AsiSI and R_RHA_*spo0A*-AscI. The resulting DNA fragment contained a complete 2,000-bp *spo0A* KO cassette, flanked by AsiSI and AscI restriction sites. The 20-nt long DNA sequence encoding sgRNA which redirects Cas9 toward *spo0A*, was designed using the CRISPR-Cas9 guide design tool available online in the Benchling platform[1] with default settings. The pMTL431511-compatible DNA fragment containing template for sgRNA was generated in PCR reaction using the universal 95-nt primer R_sgRNA-AsiSI and customized 107-nt primer F_*spo0A*_sgRNA-SalI which harbored a previously designed template for the sgRNA. The utilized oligonucleotides carried 30-bp overhangs, which enabled their mutual hybridization and thus further nucleotide incorporation to the remaining ssDNA sequence in a PCR reaction. The resulting 172-bp product encoded *spo0A*-specific sgRNA sequence (5'-CATGCTATAGAAGTAGCGTG-3') fused to Cas9-recognized RNA scaffold flanked by restriction sites SalI and AsiSI compatible with the pMTL431511 modular shuttle vector. pMTL431511 linearized with AscI and SalI was ligated with the digested KO cassette and sgRNA-containing fragment in a ratio of 1:3:3, respectively. Chemically competent NEB 5-alpha *E. coli* cells were transformed with the ligation mixture, plated on selective solid LB medium, and incubated overnight at 37°C. Resulting antibiotic-resistant colonies were screened in colony PCR with the primer pair F_P*thl*_scr and 83XXX-LR, which flank the insert-containing plasmid region. Positive colonies were incubated overnight in 5 ml of selective LB broth at 37°C with shaking. The plasmid pMTL431511-Beluga Δ*spo0A* was extracted from *E. coli* cultures and sequenced using primer pair F_*spo0A*_seq and R_*spo0A*_seq in Sanger sequencing to confirm the correct insert sequence.

Vector pMTL431511-Beluga Δ*spo0A*::bm was constructed by inserting 24-bp bookmark sequence into the LHA and RHA junction localized on pMTL431511-Beluga Δ*spo0A*. The bookmark sequence was designed to contain an efficient 20-nt Cas9-targeting sequence, 3-nt Protospacer Adjacent Motif (PAM), and an additional adenine to maintain the gene alteration in-frame. Shuttle vector pMTL431511-Beluga Δ*spo0A* was spliced into two fragments in PCR amplification: the first fragment was generated using the F_*spo0A*_bm and R_*cas* primers, and the second with F_*cas* and R_*spo0A*_bm. Primers F_*spo0A*_bm and R_*spo0A*_bm contained complementary overhangs which encoded the bookmark sequence, and were used to generate the insertion into the LHA and RHA junction. The resulting fragments were fused in a NEBuilder reaction (New England Biolabs) following the manufacturer's recommendations. Positive *E. coli* clones were selected through colony PCR screening using the primer pair F_bm_scr and 83XXX-LR. The obtained plasmid pMTL431511-Beluga Δ*spo0A*::bm was sequenced with Sanger sequencing to verify the correct sequence of the bookmark insertion.

Of note, the final CRISPR-Cas9 vector pMTL431511-Beluga Δ*spo0A*::bm was generated using a two-step procedure where we utilized a previously constructed pMTL431511-Beluga Δ*spo0A* vector available in our laboratory. If no pre-existing vector is available, we would recommend performing a regular one-step cloning.

For the construction of vector pMTL431511-Beluga::*spo0A*-wm, a 2,804-bp complementation cassette was assembled. The fragments encoding LHA and RHA were amplified from the genomic DNA of *C. botulinum* Beluga WT strain, using two primer pairs: (i) F_LHA_*spo0A*-AsiSI with R_*spo0A*_wm, and (ii) F_*spo0A*_wm with R_RHA_*spo0A*-AscI. The resulting DNA fragments were fused in SOE-PCR yielding a full complementation cassette. A sgRNA-encoding bookmark-targeting fragment (5'-GTACGACACCTCGATCACCA-3') was synthesized as described above, using primers F_bm_sgRNA-SalI and R_sgRNA-AsiSI. Ligation of DNA fragments and *E. coli* transformation were performed as described earlier. Positive colonies were screened with colony PCR using primers F_*spo0A*_seq and R_*spo0A*_seq. The constructed plasmid pMTL431511-Beluga *spo0A*-wm was sequenced to confirm the correct insert sequence.

The conjugation donor *E. coli* CA434 (Purdy et al., 2002) was transformed with an appropriate plasmid and grown overnight in liquid selective LB broth at 37°C with shaking. An aliquot of 1 ml of stationary-phase overnight culture was washed with 1 ml of anaerobic phosphate-buffered saline (PBS). The *E. coli* cell pellet was re-suspended with 200 µl of *C. botulinum* overnight culture in an anaerobic chamber. The cell suspension was spotted on an anaerobic TPGY agar plate and incubated to ensure conjugation for 8–10 h at 30°C in the anaerobic chamber. The bacterial growth was scraped with a sterile inoculation loop and re-suspended in 500 µl of anaerobic PBS which was subsequently spread onto three selective TPGY-agar plates. The plates were incubated at 30°C for 24 h in the chamber. The *C. botulinum* mutant colonies resulting from conjugation of *C. botulinum* WT strain with *E. coli* CA434 pMTL431511-Beluga Δ*spo0A*::bm were screened through colony PCR using primers F_*spo0A*_scr and R_*spo0A*_scr. Positive mutant colonies were plated on non-selective TPGY-agar plates anaerobically for 2 days at 30°C. Several resulting random colonies were picked and replica-plated on non-selective and selective TPGY agar plates. The clones growing only on non-selective TPGY plates were stored and confirmed by PCR as plasmid-cured. The plasmid loss frequency for *C. botulinum* Beluga strain is around 50–60% of screened colonies.

Primers F_*spo0A*_scr and R_*spo0A*_seq were used to screen for complemented *C. botulinum* colonies resulting from conjugation of *C. botulinum* Δ*spo0A*::bm with *E. coli* CA434 pMTL431511-Beluga::*spo0A*-wm. The PCR products obtained from colony PCR reactions were separated in an agarose

[1]https://www.benchling.com/

gel (1.5% w/v) to verify the amplicon size. After initial confirmation of the complementation, the PCR products derived from positive clones were purified and 600 ng of each was digested with PstI-HF (New England Biolabs) for 1 h at 37°C. An aliquot of 25 µl of the digests were separated in an agarose gel to reveal the resulting digestion pattern. Sanger sequencing confirmed the successful complementation and correct single-nucleotide modifications. CRISPR-Cas9 plasmid was cured from the positive mutant colonies as described above.

Spore and Viable Cell Count Assays

In parts 2–4 of the study, spore counts were determined using spore heating assays. A 200-µl aliquot of culture was incubated for 20 min at 60°C to eliminate vegetative cells. Heat-treated samples were serially diluted in fresh TPGY and, whenever applicable, supplemented with 15 µg ml^{-1} thiamphenicol, in 96-well plates and incubated anaerobically for 5 days at 30°C. The most probable number technique with three tubes was applied to enumerate heat-stable spores (Blodgett, 2000). Non-heated aliquots were similarly investigated to estimate viable cell counts.

Phase-Contrast Microscopy

Phase-contrast microscopy was used to monitor sporulation 14 days after inoculation in parts 3 and 4 of the study. The cultures were gently mixed and 200-µl culture aliquots were centrifuged for 2 min at 5000 × *g*, the supernatant was discarded, and cell pellet was re-suspended in 30–50 µl of anaerobic PBS. A small amount (2 µl) of resulting cell suspension was immobilized on a thin layer of 1.7% agarose (SeaKem LE agarose, Lonza) coated on the surface of a microscopy slide. Cells were visualized in DMi8 Leica inverted fluorescent microscope equipped with HC PL APO 100x/1.40 OIL PH3 objective and Hamamatsu Orca Flash V2 LT camera. Microscopy images were captured using the Metamorph Basic Acquisition for Microscope software and adjusted and analyzed with COREL PaintShop Pro x9 and Fiji ImageJ version 1.51, respectively.

Statistical Analysis

One-way ANOVA was used to compare the total viable cell and spore counts of each strain in different media at different time points in part 3 of the study. Student's *t*-test was used to compare total viable cell counts or spore counts of mutant and complemented strains to the WT or respective plasmid control strain in part 4 of the study.

RESULTS AND DISCUSSION

Construction of *C. botulinum* Group II Mutants Using CRISPR-Cas9 Bookmark Approach

The assignment of function to a particular gene product requires that the phenotypic properties of a mutant culture in which the encoding gene has been deleted are compared to those of the progenitor, WT culture. However, the phenotype of the isolated mutant may be a consequence of the acquisition of ancillary mutations elsewhere in the genome during the mutagenesis procedure, which may be responsible for the changed phenotype. To assign the observed phenotype to the absence of the deleted gene, it is essential that a complementation test is performed in which a functional copy of the deleted gene is introduced back into the mutant cells and the phenotype of the complemented mutant shown to be restored to that of the progenitor, WT cell. As the complementation method itself can also cause unintended changes in phenotype, especially if the introduced gene is on an autonomous plasmid, we chose to experimentally validate a recently proposed concept of bookmark complementation (Seys et al., 2020). In essence, the procedure restores the mutant to the WT configuration by replacing the mutant allele in the chromosome with the WT gene. It relies on prior incorporation of a unique 24-nucleotide (nt) bookmark sequence into the mutant allele that acts as a single guide RNA target for its subsequent Cas9-mediated substitution with the WT allele.

As the first practical demonstration of the bookmark concept (Seys et al., 2020), we chose to target the *spo0A* of our model *C. botulinum* Group II strain Beluga. Accordingly, we used the previously described (Seys et al., 2020) KO system, based on the CRISPR-Cas9 of *Streptococcus pyogenes*, to delete *spo0A* and replace it with a mutant allele comprising a 24-nt bookmark sequence flanked by the first and last two codons of the structural gene. The bookmark (**Figure 1**) was designed by applying a GC-rich sequence comprised an efficiently cleaved 20-nt CRISPR-Cas9 target (efficiency score calculated using algorithm available on https://www.benchling.com/), 3-bp PAM and an additional base pair maintaining the resulting gene deletion in frame. An nBLAST analysis of the bookmark sequence against the order *Clostridiales* indicated no matches in the genome sequences currently available in public databases, suggesting it could find wide use in these bacteria. By applying this approach, we generated a *C. botulinum* Beluga Δ*spo0A*::bm deletion mutant. Following the transfer of the CRISPR-Cas9 deletion vector, we performed colony PCR screening of 12 randomly picked antibiotic-resistant colonies. Six out of the twelve colonies screened were deletion mutants, and thus the

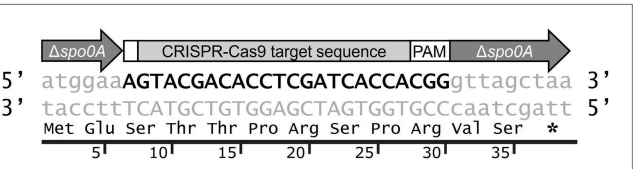


FIGURE 1 | Schematic design of Δ*spo0A*::bm mutation introduced into *Clostridium botulinum* Beluga genome using CRISPR-Cas9 genome modification tool. The 39-bp Δ*spo0A*::bm open reading frame encodes the two first and the two last amino acids of the wild type (WT) *spo0A* (gray arrows), and a bookmark sequence (bold and capitalized sequence). The bookmark sequence consists of an additional adenine (small white segment), a unique CRISPR-Cas9 targeting sequence, and a Protospacer Adjacent Motif (PAM) sequence.

mutant generation success was 50% (**Supplementary Figure S1**). CRISPR-Cas9 deletion vector was cured from the mutant strain and Sanger sequencing confirmed the expected sequence of the Δ*spo0A*::bm modification (**Supplementary Figure S2**).

To demonstrate the applicability of the gene bookmarking approach in *C. botulinum*, we restored a functional *spo0A* copy into the original chromosomal locus by redirecting the CRISPR-Cas9 machinery toward the introduced bookmark sequence in the Δ*spo0A*::bm mutant. The functional copy of *spo0A* was not the native gene as this would result in a complemented strain that would be indistinguishable from the progenitor WT cell. Consequently, contamination of a culture with the WT strain would be difficult to rule out (Seys et al., 2020). Instead, the gene used for complementation should contain a "watermark" sequence that would distinguish the complemented clones from the WT (Seys et al., 2020). Importantly, the nucleotide changes made should not cause a change to the encoded protein. The applied approach is to make changes that either create, or remove, a restriction enzyme recognition sequence. This allows the authenticity of the clone to be rapidly established through restriction digestion of an amplified PCR fragment encompassing the region in question (Ng et al., 2013). Here we chose to introduce a watermark sequence, comprising silent changes to five separate codons (267A > T, 270A > T, 273 T > A, 285 T > A, 288A > T) in the *spo0A* gene (**Supplementary Figure S5**). When designing the watermark sequence, we considered the codon usage of *C. botulinum* species by replacing the WT codons with similarly used synonymous ones. We retrieved the codon usage table for *C. botulinum* from online codon usage database (Nakamura et al., 2000). For the watermark sequence we used codons with similar usage to the ones originally found in the WT strain applying the following changes: codon GCA to GCT (frequency 0.47 and 0.44), GTA to GTT (frequency 0.49 and 0.42), GGT to GGA (frequency 0.34 and 0.5), ATT to ATA (frequency 0.40–0.54), and ACA to ACT (frequency 0.45–0.46).

The watermark lacks the PstI-recognized nucleotide sequence found within the original *spo0A* sequence. Altogether, the introduction of a watermark aims at leaving a genomic imprint in the restored gene copy and facilitates the distinction of complemented from the wild type strain by revealing different digestion patterns of the colony PCR generated DNA fragments. Following transfer of the requisite CRISPR-Cas9 complementation vector into the Δ*spo0A*::bm mutant, 12 antibiotic-resistant transconjugants were selected and all (100%) were shown by colony PCR (**Supplementary Figure S3**) and PstI digestion (**Supplementary Figure S4**) to be complemented watermark-harboring strains. Sanger sequencing of a random selection of clones confirmed that the expected region of the DNA had been inserted (**Supplementary Figure S5**). The watermark sequence applied in this study removes an already existing restriction enzyme site. However, this approach might not be readily applicable due to the limited restriction enzyme availability. In such case, the watermark-specific screening could be performed through PCR. One of the primers applied for the amplification should directly anneal to the watermarked genomic sequence with altered nucleotides at the 3' end, which would increase the specificity of the primer. That way, the watermark-specific primer, together with another regular primer annealing nearby, would yield a PCR product only in the case of successful watermark insertion. However, we would recommend validating the watermark-specific primer before the actual mutant screening to ensure the specificity of primer and expected amplification result.

The successful creation of *C. botulinum* Δ*spo0A*::bm and Δ*spo0A*::bm::*spo0A*-wm strains served to illustrate how the utility of the bookmark approach transcends simple complementation and can be employed to deliver derivatized genes to the genome. In this case, the modified gene contained five silent nucleotide changes, but could equally be applied to incorporate alterations in encoded amino acids for the purposes of structural or mechanistic studies. Successful complementation of *spo0A* harboring the pre-designed nucleotide changes provides a foundation for utilizing the bookmark approach to introduce other genomic alterations within a copy of a restored gene, i.e., modifications of putative protein binding boxes, single-amino-acid substitutions or deletions, or construction of reporter gene fusions.

Selection of Efficient Sporulation Media for *C. botulinum* Group II

An integral part of spore research is the availability of a laboratory medium that ensures sufficient and reproducible spore yields and which enables downstream processing. Roberts (1967) identified three different media likely support clostridial sporulation: (i) cooked meat medium; (ii) blood agar; and (iii) TPAY-GT composed of tryptone, peptone, ammonium sulfate, yeast extract, glucose, and sodium thioglycolate. No single medium yielded satisfactory results for all tested *C. botulinum* strains (Roberts, 1967). TPGY, a variation of TPAY-GT, eventually became the routine growth medium for *C. botulinum*, and also provided a functional sporulation medium for Group I strains (Eklund et al., 1969). However, while effectively supporting growth, TPGY is not an ideal sporulation medium for Group II strains (Mascher et al., 2017).

Different variations of cooked meat medium have been used to sporulate *C. botulinum* (Perkins, 1965; Roberts, 1967; Solomon and Kautter, 1979), despite difficulties in downstream processes caused by small meat particles floating in the medium (Perkins, 1965; Roberts, 1967). Peck (1992) used a biphasic sporulation medium consisting of a solid phase of CMM supplemented with agar and glucose, and a liquid phase of sterile water, to collect Group II spores for purification and heat inactivation assays (Peck et al., 1992). Other biphasic media with different solid phases and a liquid phase of sterile water have also been shown to support sporulation of Group II strains (Perkins, 1965; Bruch et al., 1968; Peck et al., 1992). The mechanism of a biphasic medium supporting efficient sporulation remains unclear, but it is possible that attachment to a solid surface induces sporulation (Vlamakis et al., 2008). Liquid phases of nutrient-rich broths likely support growth better than a water phase, and thus we expected nutrient-rich liquid phases primarily

to yield a larger number of sporulation-prone cells than water (Peck et al., 1992; Braconnier et al., 2001; Brunt et al., 2018).

Our goal was to find a *C. botulinum* Group II sporulation medium that would be straightforward to prepare, support efficient growth and sporulation, and enable easy and particle-free sampling. Therefore, we first screened the growth and sporulation of *C. botulinum* Group II strain Beluga in five different biphasic media: TPGY-H_2O, EYA-H_2O, RCM-H_2O, CMM-H_2O, and CMM-TPGY. Agar-TPGY and agar-H_2O were used as controls. All media except for agar-H_2O supported the growth and sporulation of Beluga, with concentrations of approximately 10^7–10^8 cells or spores/ml being reached in all of them (**Supplementary Figure S6**). However, the solid phase of TPGY-H_2O, EYA-H_2O, and RCM-H_2O was prone to crumbling, causing pipette tips to become clogged by small agar pieces during sampling and considerably interfered with downstream processes, preventing the collection of spores by centrifugation. CMM-H_2O, CMM-TPGY, and agar-TPGY all enabled effortless sampling, and CMM-TPGY resulted in better growth and subsequently higher spore concentration than the other two media (approximately 1 log, **Supplementary Figure S6**). This indicates that nutrient-rich solid and liquid phases support growth better than poorer media and, as a result, enable higher spore yields, even if there are no differences in the sporulation rate (maximum spore count divided by the maximum total cell count). Fulfilling our criteria better than the other media, CMM-TPGY was selected for further validation.

To evaluate the suitability of CMM-TPGY as a sporulation medium for different Group II *C. botulinum* strains, we used it to culture Beluga (toxin type E1), Eklund 17B (toxin type B4) and FT10F (toxin type F6) for 14 days at 30°C under anaerobic conditions. The standard TPGY broth (Solomon and Lilly, 2001) and DS used to sporulate *C. perfringens* (Duncan and Strong, 1968) were used as controls. We collected samples for viable cell and spore enumeration immediately after inoculation, day 1, day 7, and day 14. At the end of the growth period, we performed phase-contrast microscopy of the cultures. All three strains reached their maximum total viable cell counts 1 day after inoculation in all different media. The growth of Beluga and FT10F was best supported by TPGY broth and CMM-TPGY, with significantly lower ($p < 0.05$) total cell count detected in DS broth (**Figure 2A**). The growth of Eklund 17B was best supported by CMM-TPGY, with significantly lower total cell counts reached in TPGY and DS broths (**Figure 2A**). Although TPGY broth supported vegetative growth (**Figure 2A**, day 1), especially in case of strains Beluga and FT10F, sporulation in TPGY was generally poor. All three Group II strains reached significantly higher ($p < 0.05$) spore concentrations in the biphasic CMM-TPGY medium than in TPGY broth alone on days 1, 7, and 14 (**Figure 2B**). The sporulation rates were similar in CMM-TPGY and DS broths for all strains, but due to higher total cell counts, the maximum spore concentrations reached in CMM-TPGY were significantly higher ($p < 0.05$) than those in DS. The results indicate that abundant growth leading to a concentrated culture (Tyrrell et al., 1958) in combination with potential factors supporting sporulation, such as attachment (Vlamakis et al., 2008), biofilm formation (Branda et al., 2001), or sporulation-activating cell-to-cell communication due to cell crowding (Cooksley et al., 2010; Li et al., 2011) are prerequisites for a high spore yield.

Phase-contrast microscopy of Beluga, Eklund 17B and FT10F grown in CMM-TPGY, TPGY, and DS (**Figure 3**) supported the results of spore heating assays. Cultures grown in CMM-TPGY showed an abundance of phase-bright mature spores, as opposed to vast amounts of vegetative cell debris and only few spores observed in TPGY, and little cell debris and few spores observed in DS.

Characterization of Constructed CRISPR-Cas9 Mutants in CMM-TPGY Medium

To characterize the phenotype of the *spo0A* mutants, we performed a comprehensive sporulation test utilizing the biphasic CMM-TPGY medium for cultivating *C. botulinum* Group II Beluga strains Δ*spo0A*::bm and Δ*spo0A*::bm::*spo0A*-wm. Spo0A functions as a master regulator of sporulation in all spore-forming bacteria (Hoch, 1993; Al-Hinai et al., 2015), including *C. botulinum* Group II (Mascher et al., 2017). We therefore expected the *spo0A* deletion to prevent the cells from entering sporulation. Since spore formation can be readily traced in the laboratory, an asporogenous phenotype would serve as a reliable control for successful gene knock-out and for *in cis* or *in trans* complementation of the mutations. To compare the efficiency of a chromosomal gene complementation to the conventional plasmid complementation approach, we constructed a *C. botulinum* Beluga Δ*spo0A*::bm-pMTL82151::*spo0A* strain harboring shuttle vector with a WT copy of *spo0A* under the control of its native promoter (Mascher et al., 2017).

To provide suitable control strains capable of growing in a selective medium, along with the Δ*spo0A*::bm-pMTL82151::*spo0A* strain, we also characterized the WT-pMTL82151 and Δ*spo0A*::bm-pMTL82151 control strains, both carrying just the pMTL82151 shuttle vector. All the strains were cultivated for 2 weeks in CMM-TPGY medium in three biological replicates. The plasmid-carrying cultures were supplemented with 15 μg ml^{-1} thiamphenicol every 48 h to maintain sufficient concentration of antibiotic during the extended incubation.

As expected, the mutant Δ*spo0A*::bm was unable to enter sporulation, which was verified by the absence of heat-resistant spores (**Figure 4B**) and the visual absence of sporulating cells or free spores when subjected to microscopic examination (**Figure 5**). The maximum average cell count reached in the mutant cultures, 2.54×10^8 cells/ml, was slightly but significantly ($p = 0.039$) lower than that of the WT strain (1.78×10^9 cells/ml). As the applied total cell count assay enumerates the sum of vegetative cells and spores, a lower concentration of cells in the Δ*spo0A*::bm culture is likely due to lack of spores. The sporulating phenotype was successfully restored by returning a watermark-harboring *spo0A* copy in the original *spo0A* locus applying the CRISPR-Cas9 bookmark-targeting approach. The complemented Δ*spo0A*::bm::*spo0A*-wm mutant showed no

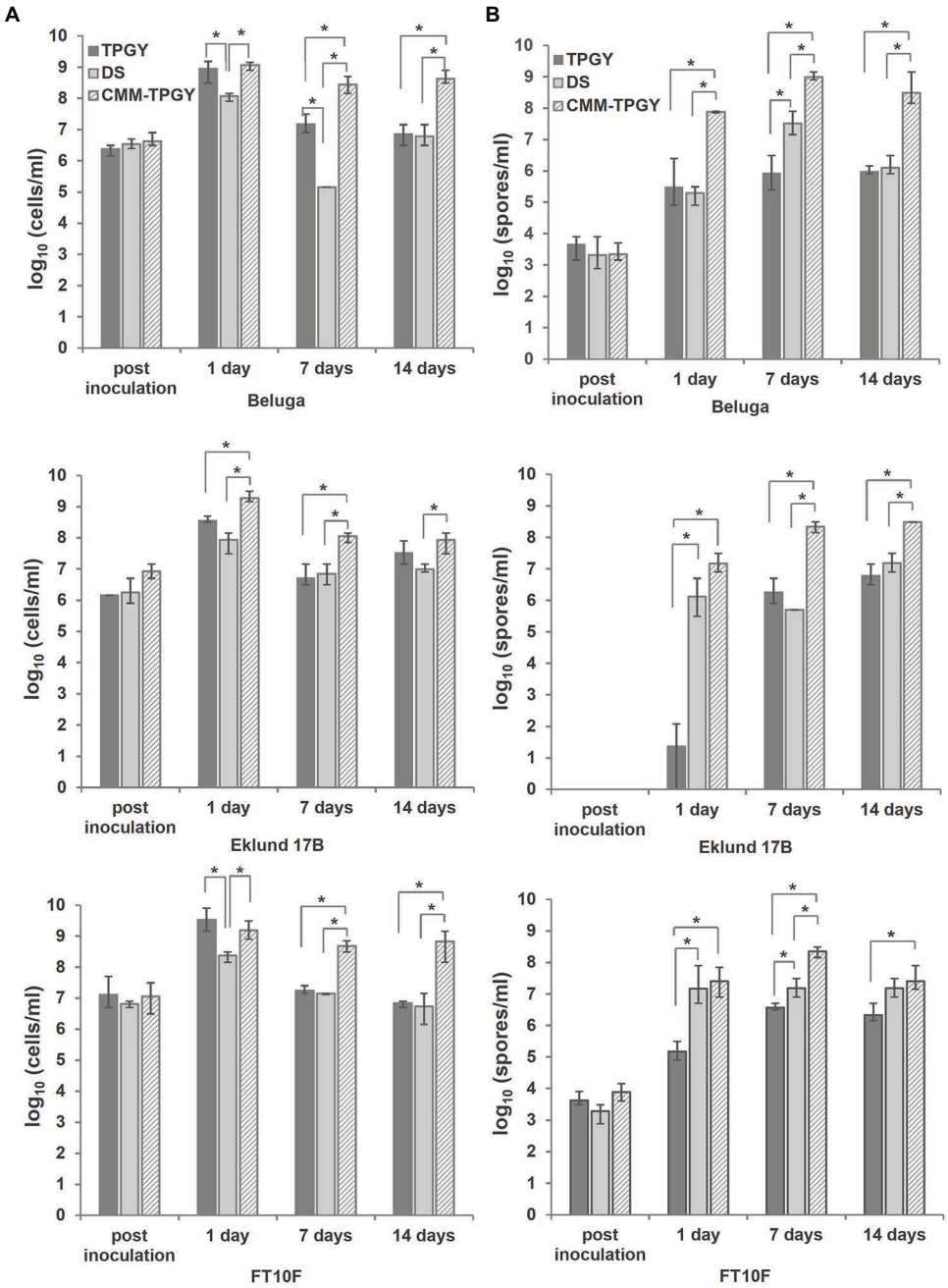


FIGURE 2 | Comparison of viable cell and spore counts produced by three different *C. botulinum* Group II strains (Beluga, Eklund 17B and FT10F) grown in three challenged media: standard tryptone-peptone-glucose-yeast extract (TPGY) broth, *Clostridium perfringens* sporulation medium Duncan-Strong (DS), and the newly described biphasic cooked meat medium-TPGY (CMM-TPGY) broth. Total viable cell and spore enumeration **(A)** and spore heating assays **(B)** were performed directly after inoculating the fresh media, and subsequently at day 1, day 7, and day 14 post inoculation. Biphasic CMM-TPGY medium promoted most efficiently bacterial growth **(A)** and sporulation **(B)** of all three tested *C. botulinum* Group II strains. The experiment was performed in three parallel biological replicates. The error bars represent the maximum and minimum values of three replicates. *Statistically significant ($p < 0.05$) difference in cell or spore concentration between different media.

significant differences to WT in the maximum cell concentrations, but the maximum spore counts were approximately 1 log lower in the complemented strains when compared to WT ($p < 0.001$; **Figure 4A**). The difference in sporulation rate between WT and complemented Δ*spo0A*::bm::*spo0A*-wm strain could arise from the limitations of the applied MPN enumeration method that is a statistical approach giving only an estimate number of cells in bacterial suspension, and it is known to

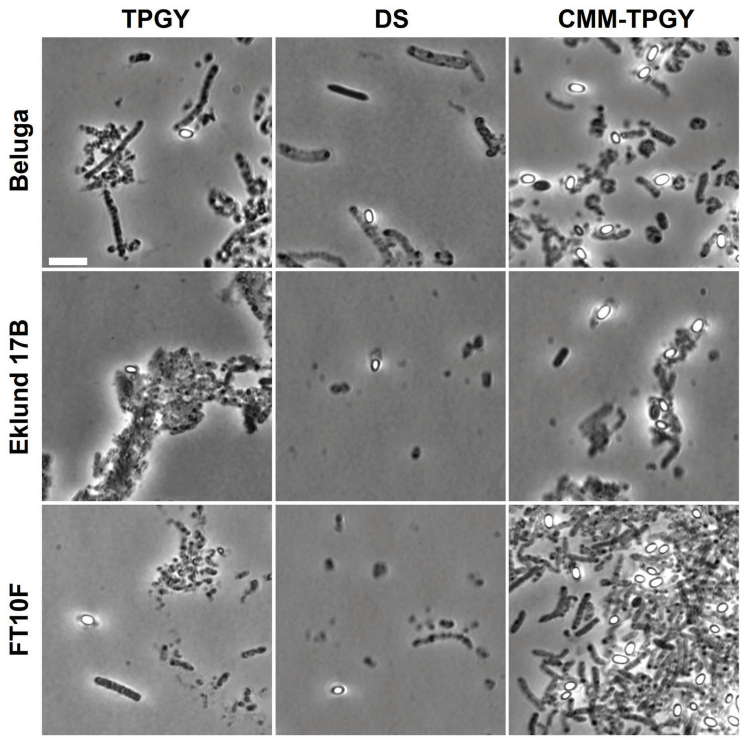


FIGURE 3 | Phase-contrast microscopy of *C. botulinum* Group II strains Beluga, Eklund 17B and FT10F after 14 days of growth in TPGY broth, DS broth, and in biphasic CMM-TPGY medium. From the three tested media, the biphasic CMM-TPGY provides the most favorable sporulation conditions for all the tested Group II strains. Scale bar 5 μm.

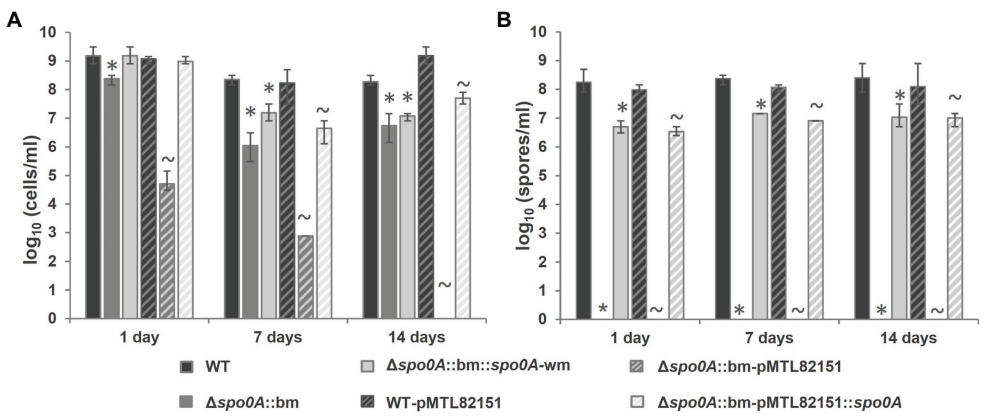


FIGURE 4 | Phenotype characterization in CMM-TPGY sporulation medium of the constructed Δ*spo0A*::bm *C. botulinum* Beluga strain compared to the WT parental and complemented strains, where a copy of functional *spo0A*-coding sequence was inserted back to the original chromosomal locus (Δ*spo0A*::bm::*spo0A*-wm) or expressed from a plasmid (Δ*spo0A*::bm-pMTL82151::*spo0A*). **(A)** Viable cell and **(B)** spore quantifications were performed at day 1, day 7, and day 14 post inoculation. The Δ*spo0A*::bm mutant strain produced significantly smaller total viable cell counts than the WT strain **(A)** due to the inability to form heat-resistant spores **(B)**. Both complementation strategies allowed comparable restoration of asporogenous phenotype for Δ*spo0A*::bm. All the data were obtained from three biological replicates. *Cell or spore concentration of a mutant or complemented strain was significantly ($p < 0.05$) lower than that of WT. ~Cell or spore concentration was significantly ($p < 0.05$) lower than that of the respective plasmid control.

be less accurate in case of dense bacterial populations (Chandrapati and Williams, 2014). Another explanation for the observed difference in spore number would be an effect of secondary SNPs in the genome, presumably caused by stress during the mutational process. Nevertheless, the spores of the complemented strains are produced in relatively high amount; they are heat-resistant and show regular morphology indicating that the entire sporulation pathway is restored and functional.

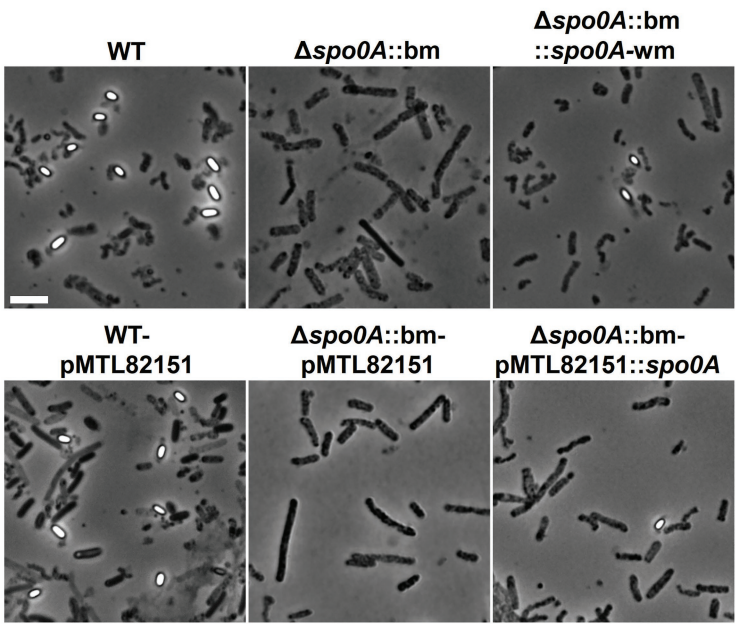


FIGURE 5 | Phase-contrast microscopy images at 14 days post inoculation demonstrating the morphological differences between WT, Δ*spo0A*::bm, and complemented mutant strains. Phase-bright endospores were detectable only in samples containing WT or complemented strains. The Δ*spo0A*::bm mutant did not display any spores or sporulating cells, suggesting that endospore formation ceased at the early stage. Scale bar 5 μm.

The plasmid-borne complementation of the *spo0A* deletion was equally efficient (reaching approximately 10^7 spores/ml) as the *spo0A*-wm insertion into the chromosomal *spo0A* locus, indicating that the introduction of a bookmark (i) does not interfere with the phenotype, (ii) is unlikely to cause polar effects, and (iii) can be safely used for routine deletion-tagging. No differences in growth or sporulation were detected between WT and its plasmid control strain WT-pMTL82151 (**Figures 4A,B**). However, the second control strain Δ*spo0A*::bm-pMTL82151 showed significantly lower total cell numbers than the respective mutant Δ*spo0A*::bm ($p < 0.001$) and no growth at all on day 14. A possible explanation behind this observation could be a gradual plasmid loss rendering the strain increasingly sensitive to thiamphenicol. In case of the sporulation-deficient strain incubated during extended periods, the control plasmid cannot be preserved inside the spores, therefore, the bacteria lose the ability to grow in antibiotic-supplemented broth.

CONCLUSION

The present study represents the first use of CRISPR-Cas9 in the generation of in-frame deletions in *C. botulinum* Group II. The system used relies on a truncated Cas9 nickase variant previously only exemplified in *Clostridium difficile* (Ingle et al., 2019). Using the *spo0A* gene as the target, we also provided the first practical demonstration of CRISPR-Cas9-mediated bookmark complementation technology (Seys et al., 2020), in which the deleted chromosomally located *spo0A* gene was restored to a functional copy *in situ*. This substitution was made possible by the prior incorporation into the mutant allele of a 24-nt bookmark sequence that should be widely applicable in other clostridial species. The functional copy of *spo0A* used for complementation was "watermarked" with five silent nucleotide changes that removed a WT PstI restriction site but still enabled production of active Spo0A. This was to allow the complemented strain to be easily distinguished from possible contamination with the WT, but also to demonstrate the potential of the bookmark technology in functional studies in which a chromosomal gene is replaced, in a two-step process, with a derivatized copy. In future studies such genes could include, for instance, variants with specific amino acid changes, homologs from other bacterial strains or species, and fusion proteins. Our analysis of the sporulation mutant and the generated complemented strain was made possible by the formulation of a biphasic CMM-TPGY medium, which supported the growth and efficient sporulation of three different *C. botulinum* Group II strains and enabled easy and particle-free sampling for microscopy and other downstream applications. Altogether, the findings described in this study constitute a solid base for future spore and sporulation studies of *C. botulinum* Group II strains.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualization: MBN, AM, GM, and ML. Formal analysis: MBN and AM. Investigation: MBN, AM, GM, and VH. Methodology: MBN, AM, GM, DG, and NPM. Writing the original draft—MBN, AM, and ML. Reviewing and editing the original draft: all authors. All the authors have read and agreed on the published version of the manuscript.

FUNDING

This research was funded by the European Research Council (ERC-CoG grant no. 683099), Marie Skłodowska-Curie Innovative Training Network CLOSPORE (642068), the Academy of Finland (299700), and by the Doctoral Programme in Food Chain and Health of the University of Helsinki and Doctoral Programme in Microbiology and Biotechnology of the University of Helsinki.

ACKNOWLEDGMENTS

We would like to thank Hanna Korpunen for her excellent technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.617269/full#supplementary-material

REFERENCES

Al-Hinai, M. A., Jones, S. W., and Papoutsakis, E. T. (2015). The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. *Microbiol. Mol. Biol. Rev.* 79, 19–37. doi: 10.1128/MMBR.00025-14

Blodgett, R. (2000). "Bacteriological analytical manual appendix 2: most probable number from serial dilutions" in *Bacteriological analytical manual*. U.S. Food and Drug Administration.

Braconnier, A., Broussolle, V., Perelle, S., Fach, P., Nguyen-The, C., and Carlin, F. (2001). Screening for *Clostridium botulinum* type a, B, and E in cooked chilled foods containing vegetables and raw material using polymerase chain reaction and molecular probes. *J. Food Prot.* 64, 201–207. doi: 10.4315/0362-028X-64.2.201

Branda, S. S., González-Pastor, J. E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001). Fruiting body formation by *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11621–11626. doi: 10.1073/pnas.191384198

Bruch, M. K., Bohrer, C. W., and Denny, C. B. (1968). Adaptation of biphasic culture technique to the sporulation of *Clostridium botulinum* type E. *J. Food Sci.* 33, 108–109. doi: 10.1111/j.1365-2621.1968.tb00895.x

Brunt, J., Carter, A. T., Pye, H. V., and Peck, M. W. (2018). The orphan germinant receptor protein GerXAO (but not GerX3b) is essential for L-alanine induced germination in *Clostridium botulinum* group II. *Sci. Rep.* 8:7060. doi: 10.1038/s41598-018-25411-x

Cañadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., and Minton, N. P. (2019). RiboCas: a universal CRISPR-based editing tool for *Clostridium*. *ACS Synth. Biol.* 8, 1379–1390. doi: 10.1021/acssynbio.9b00075

Chandrapati, S., and Williams, M. G. (2014). "Total viable counts: most probable number (MPN)" in *Encyclopedia of food microbiology: Second edition*. eds. C. A. Batt and M. L. Tortorello (Academic Press), 621–624.

Clauwers, C., and Lood, C., Van Noort, V., and Michiels, C. W. (2017). Canonical germinant receptor is dispensable for spore germination in *Clostridium botulinum* group II strain NCTC 11219. *Sci. Rep.* 7:15426. doi: 10.1038/s41598-017-15839-y

Clauwers, C., Vanoirbeek, K., Delbrassinne, L., and Michiels, C. W. (2016). Construction of nontoxigenic mutants of nonproteolytic *Clostridium botulinum* NCTC 11219 by insertional mutagenesis and gene replacement. *Appl. Environ. Microbiol.* 82, 3100–3108. doi: 10.1128/aem.03703-15

Cooksley, C. M., Davis, I. J., Winzer, K., Chan, W. C., Peck, M. W., and Minton, N. P. (2010). Regulation of neurotoxin production and sporulation by a putative *agrBD* signaling system in proteolytic *Clostridium botulinum*. *Appl. Environ. Microbiol.* 76, 4448–4460. doi: 10.1128/AEM.03038-09

Dahlsten, E., Kirk, D., Lindström, M., and Korkeala, H. (2013). Alternative sigma factor SigK has a role in stress tolerance of group I *Clostridium botulinum* strain ATCC 3502. *Appl. Environ. Microbiol.* 79, 3867–3869. doi: 10.1128/AEM.04036-12

Duncan, C. L., and Strong, D. H. (1968). Improved medium for sporulation of *Clostridium perfringens*. *Appl. Microbiol.* 16, 82–89.

Eklund, M. W., Poysky, F. T., and Boatman, E. S. (1969). Bacteriophages of *Clostridium botulinum* types a, B, E, and F and nontoxigenic strains resembling type E. *J. Virol.* 3, 270–274. doi: 10.1128/jvi.3.2.270-274.1969

Grossman, A. D., and Losick, R. (1988). Extracellular control of spore formation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 85, 4369–4373. doi: 10.1073/pnas.85.12.4369

Harris, A. (2015). "*Clostridium botulinum*" in *Encyclopedia of food and health*. eds. B. Caballero, P. M. Finglas and F. Toldrá (Academic Press), 141–145.

Hatheway, C. L. (1993). "*Clostridium botulinum* and other clostridia that produce botulinum neurotoxin" in *Clostridium botulinum. Ecology and control in foods*. eds. A. Hauschild and K. Dodds (New York: Marcel Dekker), 3–20.

Hawirko, R. Z., Naccarato, C. A., Lee, R. P. W., and Maeba, P. Y. (1979). Outgrowth and sporulation studies on *Clostridium botulinum* type E: influence of isoleucine. *Can. J. Microbiol.* 25, 522–527. doi: 10.1139/m79-076

Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., and Minton, N. P. (2007). The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* 70, 452–464. doi: 10.1016/j.mimet.2007.05.021

Hoch, J. A. (1993). Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. *Annu. Rev. Microbiol.* 47, 441–465. doi: 10.1146/annurev.mi.47.100193.002301

Ingle, P., Groothuis, D., Rowe, P., Huang, H., Cockayne, A., Kuehne, S. A., et al. (2019). Generation of a fully erythromycin-sensitive strain of Clostridioides difficile using a novel CRISPR-Cas9 genome editing system. *Sci. Rep.* 9:8123. doi: 10.1038/s41598-019-44458-y

Kirk, D. G., Zhang, Z., Korkeala, H., and Lindström, M. (2014). Alternative sigma factors SigF, SigE, and SigG are essential for sporulation in *Clostridium botulinum* ATCC 3502. *Appl. Environ. Microbiol.* 80, 5141–5150. doi: 10.1128/AEM.01015-14

Kuehne, S. A., Rood, J. I., and Lyras, D. (2019). Clostridial genetics: genetic manipulation of the pathogenic clostridia. *Microbiol. Spectr.* 7. doi: 10.1128/microbiolspec.gpp3-0040-2018

Li, J., Chen, J., Vidal, J. E., and McClane, B. A. (2011). The Agr-like quorum-sensing system regulates sporulation and production of enterotoxin and beta2 toxin by *Clostridium perfringens* type a non-food-borne human gastrointestinal disease strain F5603. *Infect. Immun.* 79, 2451–2459. doi: 10.1128/IAI.00169-11

Lindström, M., Kiviniemi, K., and Korkeala, H. (2006). Hazard and control of group II (non-proteolytic) *Clostridium botulinum* in modern food processing. *Int. J. Food Microbiol.* 108, 92–104. doi: 10.1016/j.ijfoodmicro.2005.11.003

Mascher, G., Derman, Y., Kirk, D. G., Palonen, E., Lindström, M., and Korkeala, H. (2014). The CLO3403/CLO3404 two-component system of *Clostridium botulinum* E1 Beluga is important for cold shock response and growth at low temperatures. *Appl. Environ. Microbiol.* 80, 399–407. doi: 10.1128/AEM.03204-13

Mascher, G., Mertaoja, A., Korkeala, H., and Lindström, M. (2017). Neurotoxin synthesis is positively regulated by the sporulation transcription factor Spo0A in *Clostridium botulinum* type E. *Environ. Microbiol.* 19, 4287–4300. doi: 10.1111/1462-2920.13892

McAllister, K. N., and Sorg, J. A. (2019). CRISPR genome editing systems in the genus *Clostridium*: a timely advancement. *J. Bacteriol.* 201, e00219–e00319. doi: 10.1128/JB.00219-19

Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., et al. (2016). A roadmap for gene system development in *Clostridium*. *Anaerobe* 41, 104–112. doi: 10.1016/j.anaerobe.2016.05.011

Nakamura, Y., Gojobori, T., and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28:292. doi: 10.1093/nar/28.1.292

Ng, Y. K., Ehsaan, M., Philip, S., Collery, M. M., Janoir, C., Collignon, A., et al. (2013). Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using *pyrE* alleles. *PLoS One* 8:e56051. doi: 10.1371/journal.pone.0056051

Peck, M. W., Fairbairn, D. A., and Lund, B. M. (1992). The effect of recovery medium on the estimated heat-inactivation of spores of non-proteolytic *Clostridium botulinum*. *Lett. Appl. Microbiol.* 15, 146–151. doi: 10.1111/j.1472-765X.1992.tb00749.x

Perkins, W. E. (1965). Production of clostridial spores. *J. Appl. Bacteriol.* 28, 1–16. doi: 10.1111/j.1365-2672.1965.tb02121.x

Perkins, W. E., and Tsuji, K. (1962). Sporulation of *Clostridium botulinum*. II. Effect of arginine and its degradation products on sporulation in a synthetic medium. *J. Bacteriol.* 84, 86–94. doi: 10.1128/JB.84.1.86-94.1962

Purdy, D., O'Keeffe, T. A. T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., et al. (2002). Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.* 46, 439–452. doi: 10.1046/j.1365-2958.2002.03134.x

Roberts, T. A. (1965). Sporulation of *Clostridium botulinum* type E in different culture media. *J. Appl. Bacteriol.* 28, 142–146. doi: 10.1111/j.1365-2672.1965.tb02136.x

Roberts, T. A. (1967). Sporulation of mesophilic clostridia. *J. Appl. Bacteriol.* 30, 430–443. doi: 10.1111/j.1365-2672.1967.tb00321.x

Seys, F. M., Rowe, P., Bolt, E. L., Humphreys, C. M., and Minton, N. P. (2020). A gold standard, CRISPR/Cas9-based complementation strategy reliant on 24 nucleotide bookmark sequences. *Gene* 11:458. doi: 10.3390/GENES11040458

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

Solomon, H. M., and Kautter, D. A. (1979). Sporulation and toxin production by *Clostridium botulinum* type G. *J. Food Prot.* 42, 965–967. doi: 10.4315/0362-028X-42.12.965

Solomon, H. M., and Lilly, T. J. (2001). "*Clostridium botulinum*" in *Bacteriological analytical manual*. U.S. Food and Drug Administration.

Strasdine, G. A., and Melville, J. (1968). Growth and spore production of *Clostridium botulinum* type E in chemically defined media. *J. Fish. Res. Board Can.* 25, 547–553. doi: 10.1139/f68-046

Tsuji, K., and Perkins, W. E. (1962). Sporulation of *Clostridium botulinum* l. *J. Bacteriol.* 84, 81–85.

Tyrrell, E. A., MacDonald, R. E., and Gerhardt, P. (1958). Biphasic system for growing bacteria in concentrated culture. *J. Bacteriol.* 75, 1–4. doi: 10.1128/JB.75.1.1-4.1958

Vlamakis, H., Aguilar, C., Losick, R., and Kolter, R. (2008). Control of cell fate by the formation of an architecturally complex bacterial community. *Genes Dev.* 22, 945–953. doi: 10.1101/gad.1645008

Whitmer, M. E., and Johnson, E. A. (1988). Development of improved defined media for *Clostridium botulinum* serotypes a, B, and E. *Appl. Environ. Microbiol.* 54, 753–759. doi: 10.1128/AEM.54.3.753-759.1988

Woods, D. R., and Jones, D. T. (1987). Physiological responses of *Bacteroides* and *Clostridium* strains to environmental stress factors. *Adv. Microb. Physiol.* 28, 1–64. doi: 10.1016/S0065-2911(08)60236-2

Zhang, Z., Dahlsten, E., Korkeala, H., and Lindström, M. (2014). Positive regulation of botulinum neurotoxin gene expression by CodY in *Clostridium botulinum* ATCC 3502. *Appl. Environ. Microbiol.* 80, 7651–7658. doi: 10.1128/AEM.02838-14

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 February 2021
doi: 10.3389/fmicb.2021.566908

The Distinctive Evolution of *orfX Clostridium parabotulinum* Strains and Their Botulinum Neurotoxin Type A and F Gene Clusters Is Influenced by Environmental Factors and Gene Interactions via Mobile Genetic Elements

Theresa J. Smith[1], Charles H. D. Williamson[1], Karen K. Hill[2], Shannon L. Johnson[2], Gary Xie[2], Fabrizio Anniballi[3], Bruna Auricchio[3], Rafael A. Fernández[4], Patricia A. Caballero[4], Paul Keim[1] and Jason W. Sahl[1]*

[1] Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, United States, [2] Los Alamos National Laboratory, Los Alamos, NM, United States, [3] Department of Food Safety, Nutrition and Veterinary Public Health, National Reference Centre for Botulism, Istituto Superiore di Sanità, Rome, Italy, [4] Área Microbiología, Departamento de Patología, Universidad Nacional de Cuyo, Mendoza, Argentina

OPEN ACCESS

Edited by:
Daniel Yero,
Autonomous University of Barcelona, Spain

Reviewed by:
Paul Stenmark,
Stockholm University, Sweden
César Rodríguez,
University of Costa Rica, Costa Rica

***Correspondence:**
Theresa J. Smith
Theresa.Smith@nau.edu;
terrys2much@comcast.net

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 29 May 2020
Accepted: 08 February 2021
Published: 26 February 2021

Citation:
Smith TJ, Williamson CHD, Hill KK, Johnson SL, Xie G, Anniballi F, Auricchio B, Fernández RA, Caballero PA, Keim P and Sahl JW (2021) The Distinctive Evolution of orfX Clostridium parabotulinum Strains and Their Botulinum Neurotoxin Type A and F Gene Clusters Is Influenced by Environmental Factors and Gene Interactions via Mobile Genetic Elements. Front. Microbiol. 12:566908. doi: 10.3389/fmicb.2021.566908

Of the seven currently known botulinum neurotoxin-producing species of *Clostridium*, *C. parabotulinum, or C. botulinum* Group I, is the species associated with the majority of human botulism cases worldwide. Phylogenetic analysis of these bacteria reveals a diverse species with multiple genomic clades. The neurotoxins they produce are also diverse, with over 20 subtypes currently represented. The existence of different *bont* genes within very similar genomes and of the same *bont* genes/gene clusters within different bacterial variants/species indicates that they have evolved independently. The neurotoxin genes are associated with one of two toxin gene cluster types containing either hemagglutinin (*ha*) genes or *orfX* genes. These genes may be located within the chromosome or extrachromosomal elements such as large plasmids. Although BoNT-producing *C parabotulinum* bacteria are distributed globally, they are more ubiquitous in certain specific geographic regions. Notably, northern hemisphere strains primarily contain *ha* gene clusters while southern hemisphere strains have a preponderance of *orfX* gene clusters. *OrfX C. parabotulinum* strains constitute a subset of this species that contain highly conserved *bont* gene clusters having a diverse range of *bont* genes. While much has been written about strains with *ha* gene clusters, less attention has been devoted to those with *orfX* gene clusters. The recent sequencing of 28 *orfX C. parabotulinum* strains and the availability of an additional 91 strains for analysis provides an opportunity to compare genomic relationships and identify unique toxin gene cluster characteristics and locations within this species subset in depth. The mechanisms behind the independent processes of bacteria evolution and generation of toxin diversity are explored through the examination of bacterial relationships relating to

source locations and evidence of horizontal transfer of genetic material among different bacterial variants, particularly concerning *bont* gene clusters. Analysis of the content and locations of the *bont* gene clusters offers insights into common mechanisms of genetic transfer, chromosomal integration, and development of diversity among these genes.

Keywords: *Clostridium parabotulinum*, neurotoxin, plasmids, *orfX*, *lycA*, *arsC*, *pulE*

INTRODUCTION

Botulinum neurotoxins (BoNTs) are a worldwide public health issue and are listed as Tier 1 Select Agents due to their potential to pose a severe threat to human and animal health (Federal Select Agent Program Select Agents and Toxins List)[1]. The recent availability of genomes from over 250 BoNT-producing clostridial strains has provided new opportunities to gain perspective about the toxins and their toxin gene clusters, and the relationships of the strains that produce these toxins (Giordani et al., 2015; Mazuet et al., 2016; Williamson et al., 2016).

The neurotoxins are extremely diverse. They have historically been categorized according to the neutralizing ability of serotype-specific antisera (Peck et al., 2017). Seven toxin types (A-G) have been determined using these methods. Four of these serotypes, BoNT/A, BoNT/B, BoNT/E, and BoNT/F, have been definitively linked with human botulism. While these toxins show commonalities in protein structure and activity, genetic sequencing of the toxin serotypes has revealed that they differ by 35–70% in amino acid sequence. A second level of diversity is seen within these toxins; over 40 toxin subtypes have been currently identified having amino acid sequence differences of 2–36% (Peck et al., 2017). The subtypes are labeled with a number following the toxin type, such as BoNT/A1 or BoNT/F5. This diversity among BoNT proteins contrasts with tetanus toxin, (TeNT), which is closely related to the BoNTs in structure and mechanism of action but differs in both its singularity and a lack of non-toxic complex proteins.

The BoNT-producing bacteria are also diverse in their composition. Historically any bacteria that produced BoNTs was considered to be the species "*Clostridium botulinum*" based on the single characteristic of their neurotoxicity, and various species were instead differentiated into four Groups (I-IV) based on their biochemical and metabolic properties. In 1988, Group IV organisms were genetically confirmed to be the species *Clostridium argentinense* (Suen et al., 1988a) and additional BoNT-producing strains were identified and genetically confirmed to be members of *Clostridium baratii* and C*lostridium butyricum* (Suen et al., 1988b). This created a confusing hybrid nomenclature system involving both Group and species designations. However, genomic studies have now confirmed that BoNT-producing strains are represented in seven distinct species (Smith et al., 2018). Within this manuscript the bacteria will be designated by Latin binomials: *C. parabotulinum* (proteolytic *C. botulinum* Group I*)* (Seddon, 1922; Meyer and Gunnison, 1929); *C. botulinum* (non-proteolytic Group II) (Smith, 1977); *C. novyi sensu lato* (*C. botulinum* Group III) (Skarin et al., 2011); BoNT/G-producing *C. argentinense*; BoNT/F-producing *C. baratii*; BoNT/E-producing *C. butyricum*; and BoNT/B-producing *C. sporogenes* (Weigand et al., 2015; Williamson et al., 2016). Examples of both BoNT-producing and non-neurotoxigenic bacteria are found within each species. **Table 1** lists information on the species, including the toxin types that they produce.

The discovery and study of BoNT-producing clostridia has a history spanning more than 100 years (van Ermengem, 1897; Landmann, 1904; Leuchs, 1910; Burke, 1919). *C. parabotulinum* bacteria are the most studied BoNT-producing species, as they produce toxins that are responsible for the vast majority of human botulism cases worldwide. They are most closely related to *C. sporogenes* (Hatheway, 1993; Collins and East, 1998; Stackebrandt et al., 1999), prompting some to consider them as variants of the same species. However, recent studies have indicated that they should be considered separate species (Weigand et al., 2015; Williamson et al., 2016).

The neurotoxins are naturally found in association with several non-toxic accessory proteins, known collectively as the toxin complex. The genes encoding these proteins are arranged in a cluster adjacent to the neurotoxin gene. There are two types of *bont* gene clusters, known as *ha* or *orfX*. The toxin gene clusters encode the neurotoxin protein and a variety of non-toxic proteins. Non-toxin/non-hemagglutinin (NTNH) proteins, expressed by the *ntnh* gene, are universally present in both types of toxin complexes. The *ha* gene clusters produce hemagglutinin proteins (HA70, HA17, and HA33) which form a complex that is directly linked to the NTNH protein. Within the *orfX* gene clusters are three open reading frames (*orfX1, orfX2, orfX3*) that have been shown to produce proteins that apparently form somewhat fragile toxin complexes (Mazuet et al., 2012; Lin et al., 2015; Kalb et al., 2017). The functions of these proteins are not presently known. Other genes that are found within the *bont* gene cluster include *botR*, a transcriptional regulator, and *p47*, whose protein was recently shown to share a structural domain

TABLE 1 | A listing of BoNT-producing species with Group nomenclature.

Species	Group designation	Toxins produced	Toxin gene cluster type(s)
C. parabotulinum	Group I	A, B, F	*orfX* +, *ha* +
C. sporogenes	Group I	B	*ha* +
C. botulinum	Group II	B, E, F	*orfX* +, *ha* +
C. novyi sensu lato	Group III	C, D	*ha* +
C. argentinense	Group IV	G	*ha* +
C. baratii	Group V	F	*orfX* +
C. butyricum	Group VI	E	*orfX* +

[1]https://www.selectagents.gov/selectagentsandtoxinslist.html

with the OrfX2 and OrfX3 proteins and that is related to the TULIP family of lipid-binding proteins (Gustafsson et al., 2017; Lam et al., 2018).

Each botulinum neurotoxin gene (*bont*) is part of one or the other of the above gene clusters. For example, *bont/B* genes are always found within *ha* clusters, while *bont/E* genes are always associated with *orfX* gene clusters. The *bont* genes within BoNT/A1 strains are unique in that they can be located within either of the two gene clusters. The *bont* genes in BoNT/A5 strains, most BoNT/A1 strains, and all BoNT/B strains are located within the smaller (11.7 kb) *ha* toxin gene cluster (Collins and East, 1998; Carter et al., 2011), while a few *bont/A1* genes, as well as *bont/A2-A4* and *bont/A6-A8* genes, and all *bont/F* genes are located within the larger (∼17 kb) *orfX* gene cluster (Luquez et al., 2005; Hill et al., 2009; Kull et al., 2015; Mazuet et al., 2016). In bivalent strains that express two toxins, such as BoNT/A2B5, BoNT/A1(B), and BoNT/B5F2 strains, both toxin gene clusters may be found within a single bacterial isolate. In those cases, the *bont/A* or *bont/F* genes are always within *orfX* gene clusters and the *bont/B* genes are within *ha* gene clusters (Hill et al., 2009; Kalb et al., 2017). The *orfX* gene clusters show a high degree of relatedness which contrasts with the variability of the associated toxin type or subtype. For example, gene clusters with identical or nearly identical *orfX*, *botR*, *p47*, and *ntnh* genes may contain *bont* genes with only ∼60% nucleotide identity.

C. parabotulinum strains are commonly isolated from environmental and clinical samples, particularly in association with human botulism cases. While *C. parabotulinum* strains are globally distributed, the distinct predominance of *C. parabotulinum* strains with *orfX* gene clusters in the southern hemisphere and, conversely, the predominance of *C. parabotulinum* having *ha* gene clusters in the northern hemisphere suggests a differential evolution of these bacteria and their toxin genes which may have been influenced by environmental factors.

Currently the genomes of over 150 *C. parabotulinum* strains have been publicly posted in the NCBI database, which provides an opportunity to examine genomic relationships and toxin gene cluster commonalities on a broad scale. An additional 28 *orfX C. parabotulinum* isolates that are mainly located in the southern hemisphere were sequenced and analyzed as part of this study. These genomes have provided an opportunity for in depth study of the relationships of this particular species subset that varies in its genomic characteristics, toxin genes and gene clusters, and geographic locations from their previously studied northern hemisphere counterparts.

A comprehensive approach was used that integrates molecular information and phylogeography in order to provide an understanding of these organisms from multiple perspectives, including an examination of the unique components, arrangements, and locations of their *orfX bont* gene clusters.

Genomic relationships among these *orfX C. parabotulinum* strains were analyzed and new insights associated with the components of their toxin gene clusters and associated co-located genes were revealed in this study. The plasmid and chromosomal locations of the *bont* gene clusters in these strains are dependent on both the species (*C. parabotulinum*) and the gene cluster type (*orfX*), and investigation of unique features within these *bont* gene clusters and surrounding genes suggests connections between the locations of these genes and the processes involved in their acquisition and chromosomal integration.

MATERIALS AND METHODS

Source attributions of BoNT-producing clostridia have become more important as we seek to understand links between these bacteria and their effects on humans, domestic animals, and the environment. Toxin subtypes differ in their sensitivities to treatments, and a knowledge of the prevalent toxin types in various geographical regions is key to prevention and treatment strategies. An extensive review of the literature, to include over 60 journal articles and book chapters, was undertaken to provide information relating to the geographic diversity of BoNT/A and BoNT/F producing strains, toxin subtypes, and locations. NCBI records associated with BoNT-producing clostridia and unpublished records from various collections were also examined for sources related to additional strains.

Whole genome sequencing of the 28 newly sequenced isolates was conducted on Illumina MiSeq or GAIIx platforms. Briefly, genomes were assembled with a pipeline that incorporates Trimmomatic (v0.30) (Bolger et al., 2014), BayesHammer (Nikolenko et al., 2013), SPAdes (v3.7.1) (Bankevich et al., 2012), and Pilon 1.17 (Walker et al., 2014). Contaminating contigs were identified with BLAST (Altschul et al., 1990) alignments against known contaminants in the NCBI nucleotide database. Any contig associated with contamination or containing anomalously low average coverage was removed. Closely-related, publicly-available genome assemblies and sequencing read data were downloaded from the National Center for Biotechnology Information (NCBI). If only sequencing read data were available for a genome, an assembly was generated with the approach described above or with SPAdes only. Genome assemblies were annotated with Prokka v1.11 (Seemann, 2014). Genomic data has been deposited in the appropriate NCBI databases. **Supplementary Table 1** provides information on all genomes in the study, including NCBI accession information, genome assembly statistics, and average nucleotide identity estimated with MASH (v2.2) (Ondov et al., 2016) compared to a reference genome (*C. parabotulinum* Kyoto-F – GCF_000022765.1).

Core-genome single nucleotide polymorphisms (SNPs) were called as described in Williamson et al. (2017). Sequencing reads were aligned to a reference genome, *C. parabotulinum* Kyoto-F (GCF_000022765.1), with BWA-MEM (v0.7.7) (Li, 2013) and SNPs were called with the Unified Genotyper method in GATK (v3.3) (McKenna et al., 2010; DePristo et al., 2011) within NASP (Sahl et al., 2016). If sequence reads were not available for a genome, reads were simulated from publicly-available genome assemblies with ART (MountRainier) (-ss MSv3 -l 250 -f 75 -m 300 -s 30) (Huang et al., 2012) in order to compare a uniform data type. SNPs were removed from the analysis if the depth of coverage was less than ten or if the allele proportion was less than 0.9. Duplicated regions of the reference genome were identified by a reference self-alignment with NUCmer (Delcher et al., 2002;

Kurtz et al., 2004) and SNPs falling in these regions were filtered from all downstream analysis. Maximum likelihood phylogenies were generated from the resulting SNP matrices (bestsnps) with IQ-TREE (v1.4.4) (Nguyen et al., 2015) using the K3Pu F ASC G4 model. The consistency index and retention index were calculated with Phangorn. Trees were viewed in FigTree[2].

Gene sequences and associated regions of *orfX* toxin gene clusters were extracted from Prokka output files for further analyses. BLASTn (Altschul et al., 1990) was used to identify and compare individual *bont* genes, toxin cluster genes and surrounding genes and intergenic sequences. The *lycA* genes and *arsC* genes (and pseudogenes) were aligned with MUSCLE (v3.8.31) (Edgar, 2004) and phylogenies were generated with IQ-TREE (v1.5.5) (Nguyen et al., 2015) using the following best-fit models (Kalyaanamoorthy et al., 2017): *lycA* gene sequences – TIM3 F G4, *arsC* gene sequences – TPM2u F R2.

Complete *orfX* BoNT/A and BoNT/F genomes were searched for evidence of bacteriophage DNA sequences using PHASTER (Zhou et al., 2011; Arndt et al., 2016). Complete, incomplete, and questionable phage sequences were located and compared in order to gain a better understanding of their movements and integration into chromosomes.

RESULTS

Distributions of BoNT-Producing *C. parabotulinum* Strains

The occurrence of human and animal botulism is directly linked to exposure to BoNTs or BoNT-producing bacteria that are resident in the environment or have been introduced via exported foods or other materials. A knowledge of the predominant BoNT serotypes and subtypes among clostridia resident within specific geographic locations is critical to providing effective treatment options.

Bacteria that contain *bont/A* genes and produce BoNT/A are the most commonly isolated BoNT-producing strains. While the geographic distribution of BoNT/A strains is known to be worldwide (Fernandez, 1994; Williamson et al., 2016), environmental studies, literature reviews, and examinations of records in public databases and private collections have determined that BoNT/A subtypes exhibit specific geographic ranges.

The major BoNT/A subtype in the northern hemisphere is BoNT/A1. BoNT/A1 strains with *ha* gene clusters are widespread, but BoNT/A1 within *orfX* gene clusters are quite rare. Exceptions to this would be BoNT/A1(B) strains that contain the *bont/A1* gene within an *orfX* gene cluster and a truncated *bont/B* gene within a complete *ha* gene cluster. These strains are commonly found within the United States (Dabritz et al., 2014; Raphael et al., 2014), and appear to be emerging pathogens in Japan (Kenri et al., 2014) and France (Mazuet et al., 2016).

In contrast, the predominant BoNT/A subtype in the southern hemisphere is BoNT/A2. BoNT/A2-producing *C. parabotulinum* are commonly found in the soils in Argentina (Luquez et al., 2005) and BoNT/A2 is the major causative agent in infant and foodborne botulism there (Sagua et al., 2009). Bacteria producing this toxin type are also found in Australia (McCallum et al., 2015) and they been associated with botulism cases in Africa (Mackay-Scollay, 1958; Smith et al., 1979; Frean et al., 2004; Luquez et al., 2012; Viray et al., 2014). Some *C. parabotulinum* strains are bivalent and contain *bont/B6*, *bont/F4*, or *bont/F5* genes in addition to *bont/A2* genes.

The first BoNT/A3-producing strain, known as the Loch Maree strain, was identified in 1922 in association with a foodborne botulism outbreak in Scotland (Leighton, 1922) and was considered to be unique. However, additional BoNT/A3 strains were isolated decades later in Argentina (Luquez et al., 2012) and recently there was an isolation of a BoNT/A3 strain in connection with a foodborne case in Slovakia (Mad'arova et al., 2017). *C. parabotulinum* strains containing *bont/A4*, *bont/A5*, *bont/A6*, *bont/A7*, and *bont/A8* genes are all quite rare and have the commonality of source locations in the northern hemisphere.

The first BoNT/F-producing strain was isolated in Denmark in 1960 (Moller and Scheibel, 1960) and since then eight distinct BoNT/F subtypes have been identified. BoNT/F strains are comparatively rare and are associated with a variety of clostridial strains, including *C. parabotulinum* (BoNT/F1-F5 and BoNT/F8), *C. botulinum* (BoNT/F6), and *C. baratii* bacteria (BoNT/F7). They have a global distribution, with occasional strains being isolated in North and South America, and across Eurasia from Denmark and Italy to China. *C. parabotulinum* strains producing BoNT/F3, BoNT/F4, and BoNT/F5 have only been isolated from Argentina, while BoNT/F1, BoNT/F2 and BoNT/F8 producers are confined in location to the northern hemisphere.

Genomic Diversity of OrfX BoNT/A and BoNT/F Strains

Phylogenies based on multiple genomic analysis methods support earlier findings that several clostridial species are capable of harboring *bont* genes and producing BoNTs (Popoff, 1995; Collins and East, 1998; Hill et al., 2009; Weigand et al., 2015; Williamson et al., 2016). One of these species, *C. parabotulinum*, includes a range of genomic variants. Strains containing *ha* and *orfX bont* gene clusters are dispersed throughout the *C. parabotulinum* phylogeny (Williamson et al., 2016). The genomes of 28 strains containing *bont/A* and/or *bont/F* genes within *orfX* gene clusters were sequenced as part of this study, and compared with 91 additional OrfX *C. parabotulinum* genomes using a core genome SNP phylogeny (consistency index with only parsimony informative SNPs – 0.43, retention index – 0.90) that was generated from an alignment of 119,955 SNP positions called from a core alignment of 1,714,645 positions. The core genome SNP analysis did not include SNPs within the toxin cluster genes or within plasmids, as they are not conserved across all genomes.

Examination of the core genome SNP phylogeny identified several defined clades within the *orfX C. parabotulinum* strains (**Figure 1**). One of these is a conserved clade composed of subclades containing closely related genomes from strains

[2]http://tree.bio.ed.ac.uk/software/figtree/

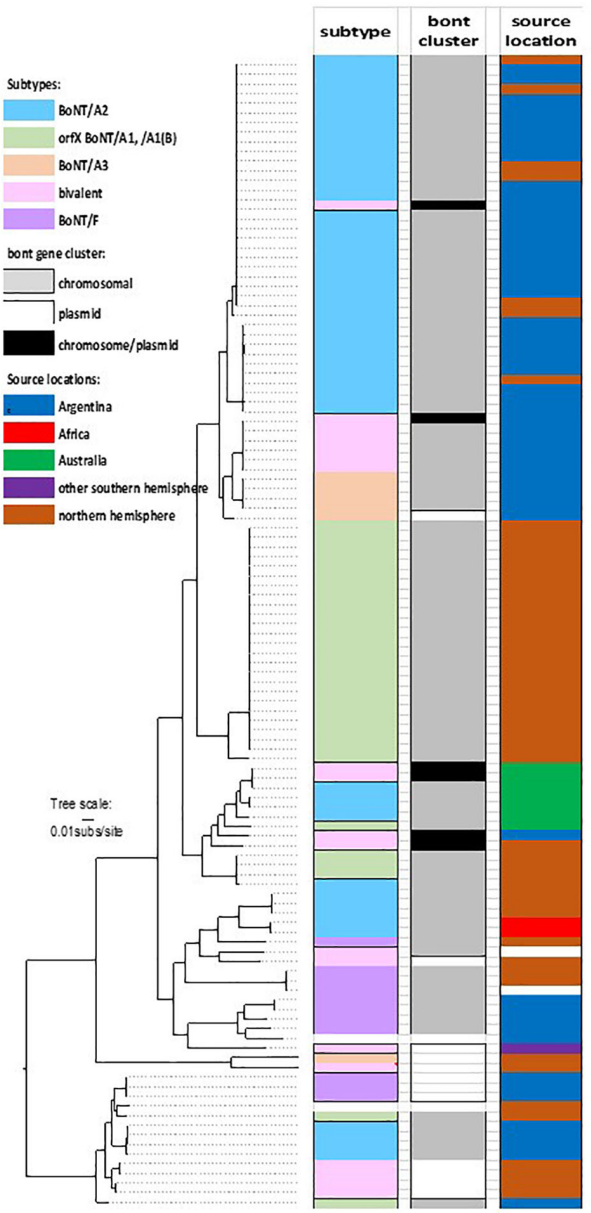


FIGURE 1 | Core genome SNP phylogeny of *C. parabotulinum* strains with *orfX* toxin clusters. The *bont* gene types, gene cluster locations, and geographic locations of the strains are color coded according to the legend. Additional information is listed in **Supplementary Table 1** with the genomes in the order they appear in this figure.

that were isolated from soils or foodborne botulism cases in northwest Argentina, and a few European strains. Within this major clade are three subclades: one subclade containing 37 genomes having *bont/A2* gene clusters, one subclade with six genomes having *bont/A2* and *bont/F4* gene clusters, and one subclade with five genomes containing *bont/A3* gene clusters. Notably, while BoNT/A2F4 strains form a highly conserved clade that is closely related to Argentinean BoNT/A2 and BoNT/A3 strains, the bacteria that contain only *bont/F4* genes are in a distinct, unrelated cluster of clades. Among these 48 strains, 41 were isolated from a specific region in Argentina. The close relationship of these lineages coupled with their common geographic location is an indication that these strains may have evolved from a single ancestral strain and subsequently acquired different toxin genes. The *bont* genes in these BoNT/A2 strains and several of the BoNT/A3 strains are located within the chromosome, which provides for a greater genetic stability than genes that are within extrachromosomal elements.

While one of the BoNT/A2-producing strains represented in this clade (Kyoto-F) was isolated from an infant botulism case in Japan, it is known that this case was associated with the ingestion of honey and that the honey was likely imported from Argentina (personal communication, Dr. Shunji Kozaki), providing a geographic link between these strains.

Two separate conserved clades that contain genomes of BoNT/A1(B) strains also exhibit localized geographic ranges. For example, in the largest clade of BoNT/A1(B) isolates 19 of 24 strains were located in the United States, four were from Japan and one was from Ecuador, while the smaller clade contained genomes from one Italian and three Japanese strains. With one exception, the strains within the two BoNT/A1(B) clades were isolated in the northern hemisphere.

A more variable clade contains Australian BoNT/A2 strains and one BoNT/A2B6 strain. While the *bont/A2* genes in these strains are located within the chromosome, *bont/B6* genes are universally located within large conjugative plasmids in both *C. parabotulinum* and *C. sporogenes* strains. The existence of *bont/B6* genes within plasmids in multiple clostridial species illustrates cross-species transfer of toxin genes by introduction via their plasmids.

Two African BoNT/A2 strains are part of a second variable genomic clade that includes Italian BoNT/A2 strains and BoNT/F8 It 357. Two subclades containing BoNT/F1 isolates, and BoNT/F4 strains plus the lone BoNT/F3 isolate indicate a relationship with the African strains. The Mexican isolate containing *bont/A6* and *bont/B1* genes is an outlier within this clade that is somewhat related to the BoNT/F3 and BoNT/F4 strains.

Rare BoNT/A1 strains having *orfX bont* gene clusters are found within one of two clades. One isolate from Australia is part of a clade that contains seven Australian BoNT/A2 strains, while two OrfX BoNT/A1 strains that are part of the "bivalent toxin" clade are from the United States and Argentina. Genomes from BoNT/A3 strains also sort into two clades according to location (Argentina and Scotland). While the Argentinian isolates show a definite relationship with geographically related Argentinean BoNT/A2 strains, the Scottish BoNT/A3 strain shows a relationship with the unusual Italian BoNT/A2B7 It 92 strain.

Genomes containing *bont/A2* genes form part of a final variable clade that also contains genomes of bivalent BoNT/A2B5, BoNT/B5A4, and BoNT/B5F2 strains, as well as several BoNT/F5 strains and the two OrfX BoNT/A1 strains previously mentioned. These strains have a global geographic range; many of its members are sourced from Argentina, but locations also include Sweden, Italy, and the United States. With

the exception of the BoNT/A2 and OrfX BoNT/A1 strains in this clade, the toxin genes within these genomes are located within large, highly conserved plasmids.

Within this phylogeny are examples of conserved clades containing geographically co-located strains having the same *bont* genes and gene clusters; conserved clades containing different *bont* genes; and variable clades containing unrelated strains having same *bont* genes. While the conserved clades emphasize an ability for a particular strain to become established in a certain region, the variable clades are indications that global movements of these strains do occur, and that transported strains are capable of exchanging their toxin genes with a variety of strain variants and species. While *bont* gene diversity is facilitated by the movement of extrachromosomal plasmids from strain to strain, stability of the genes may be ensured by subsequent integration into the chromosome. Novel toxins have been formed through homologous recombination events that place new toxin genes or gene fragments into existing *bont* gene clusters and, on a lesser scale, through single nucleotide mutations. A study of the composition and location of *orfX bont* genes and gene clusters aids in the understanding of the processes involved in gene cluster movements and integrations.

Toxin Gene Clusters and Their Locations

Toxin gene clusters in *C. parabotulinum* strains containing *ha bont/A1*, *bont/B*, and *bont/A5* genes are arranged in the following order: *ha70-ha17-ha33-botR-ntnh-bont*. Gene clusters from all other *C. parabotulinum* strains show the following arrangement: *orfX3-orfX2-orfX1-botR-p47-ntnh-bont*. Adjacent genes include the *lycA* gene and, in some cases, *arsC* genes.

OrfX bont gene clusters may be located within large, highly conserved plasmids or they may be present as pathogenicity islands (PAIs) within the chromosome. PAIs are gene clusters linked to virulence factors that endow the bacteria with pathogenic properties. They are derived from mobile genetic elements, such as conjugative plasmids or bacteriophage, and are subsequently integrated into the chromosome (Davis and Waldor, 2002). Within the chromosome, PAIs are typically present as discreet genetic units of 10–100 kb distinguished by a lower G + C content than the surrounding genomic DNA. They are often flanked by direct repeat sequences and mobility genes, such as IS elements, integrases, and transposases (Hacker et al., 1997). These general PAI characteristics are in agreement with those of *bont* gene clusters.

It is interesting to note that the non-toxic accessory genes within *orfX* toxin clusters show remarkable conservation, despite their association with multiple *bont/A* and *bont/F* genes. A discontiguous BLASTn comparison of the entire toxin gene cluster sequence (minus the toxin gene) from the BoNT/F1 Langeland strain results in 94–100% identity with gene clusters from over 60 strains located worldwide that contain *bont/A2*, *bont/A3*, *bont/HA*, *bont/F1*, *bont/F3*, *bont/F4*, and *bont/F5* genes. This high degree of toxin gene cluster identity ends at the terminal 50 nucleotides of the *ntnh* gene, where a homologous recombination (HR) event has placed *bont/A* neurotoxin genes within the *bont/F* gene cluster, or vice versa (**Figure 2A**). This is similar to a recombination event described within the *ntnh* gene that inserted the *bont/A1* gene into the *ha bont/B* toxin gene cluster, producing the *ha bont/A1* gene cluster (Hill et al., 2009).

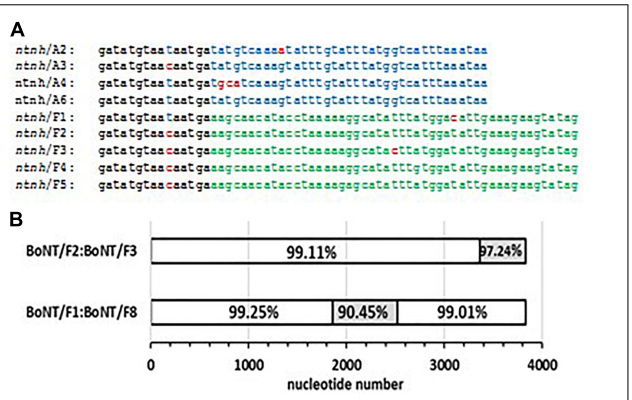


FIGURE 2 | Homologous recombination (HR) events that have exchanged or altered *bont* genes/gene clusters, resulting in the generation of diverse toxin subtypes. **(A)** C-terminal sequence of selected *orfX ntnh* genes showing homologous and disparate sequences that illustrate the recombination event placing *bont/A2/A3* genes within *bont/F* gene clusters, or vice versa. Sequences showing complete identity are in black font, *bont/A*-specific sequences are in blue font, *bont/F* sequences are in green font, and single mutational differences are shown in red font. **(B)** Comparisons of *bont/F2:bont/F3* and *bont/F1:bont/F8* gene sequences showing areas of identity, possibly gained through HR events, and regions where sequences differ (shaded in gray).

HR events within *bont* genes have also been responsible for the generation of novel toxin subtypes. It is known that the *bont/A2* gene is a mosaic of the *bont/A1* and *bont/A3* genes (Hill et al., 2007), and comparisons of *bont/F* genes as part of this study has revealed HR events among these subtypes as well. The *bont/F2* and *bont/F3* genes show greater than 99% identity until the final 470 nucleotides, where the identity decreases to 97.24%, indicating an HR event has occurred at its 3' terminus. Similarly, the *bont/F1* and *bont/F8* genes show >99% identity throughout the 5' half of the gene and also at final third of its nucleotide sequence; however, between these closely related DNA sections the percent identity decreases to slightly more than 90%, again signaling an HR event has occurred (**Figure 2B**).

The *orfX bont* gene clusters are found at four distinct genomic locations – three sites within the chromosome and one within plasmids (**Figure 3**). The locations are identified by their proximity to specific genes. Chromosomally located *orfX bont/A1*, *bont/A2, bont/F1,* and *bont/F8* gene clusters are co-located with the *ars* operon, while *bont/F3* and *bont/F4* gene clusters are inserted between fragments of a split *pulE* gene (Dover et al., 2013; Smith et al., 2020). Plasmid-borne *orfX bont* gene clusters and chromosomal *bont/A3* gene clusters are universally adjacent to *HepA/SNF2*, thermonuclease, and DNA helicase genes.

The Presence of the *lycA* Gene in *orfX C. parabotulinum* Gene Clusters

The *lycA* gene, while not historically considered a part of the toxin gene cluster, is universally situated alongside *orfX* gene

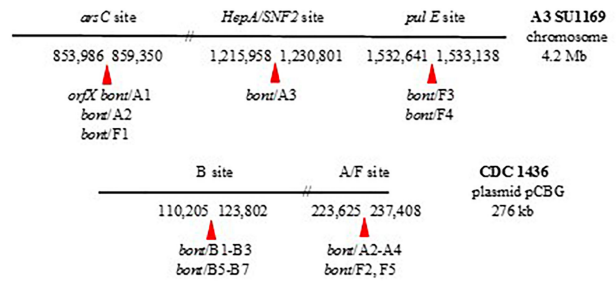


FIGURE 3 | Chromosomal and plasmid *C. parabotulinum orfX bont* gene cluster locations. The sites are numbered according to their base pair location within the BoNT/A3 SU1169 genome and plasmid pCBG from BoNT/A2B5 CDC 1436.

clusters. It is found in two locations: 1) following the *bont* gene with chromosomally located *orfX* toxin gene clusters that are co-located with the *ars* operon and 2) prior to the *orfX3* gene within plasmid-borne toxin gene clusters and chromosomally located *bont/A3*, *bont/F3* and *bont/F4* toxin gene clusters, where the *ars* operon is remotely located.

The *lycA* genes are highly conserved, with two basic variants that correspond to their locations relative to the *bont* and *ars* gene clusters, and are also generally linked to their chromosomal or plasmid location. A phylogeny comparing *lycA* gene sequences (**Figure 4**) illustrates these relationships. Nucleotide differences between the two *lycA* gene variants are approximately 6%, while differences of less than 2% are seen within the two variants.

The *lycA* gene product is a lysozyme that is homologous to lytic proteins found in *Lactobacillus* and *Streptococcus pneumoniae* bacteriophage (Henderson et al., 1977). These bacteriophage proteins lyse the bacterial cell walls, releasing phage during their lytic cycles. In similar fashion, the *lycA* lysozyme may be responsible for bacterial autolysis and subsequent release of toxin in proteolytic *C. parabotulinum* bacteria (Bonventre and Kempe, 1960; Dineen et al., 2002).

It should be noted that intact *lycA* genes are only found among *orfX* gene clusters from members of *C. parabotulinum*. The *lycA* gene's ubiquitous presence adjacent to this toxin cluster and the role its protein may play in toxin dissemination suggests that it could be considered a component of *orfX* toxin gene clusters.

Toxin Gene Clusters Within the Chromosome – Relationships With *ars* Genes

Chromosomally located *orfX bont* gene clusters, with the exception of those containing *bont/A3*, *bont/F3* or *bont/F4*, are co-located with the *ars* operon genes. The *ars* operon encodes proteins that confer resistance against inorganic arsenite that is found in anaerobic environments. The complete *ars* system (*arsR*, *arsD*, *arsA*, *arsB*, *arsC*, *arsM*) is present in most *C. parabotulinum* strains and provides the most efficient removal of arsenic, but operons that are devoid of some *ars* genes appear to provide a minimal level of protection (Lindstrom et al., 2009). The *ars* genes are found both within the chromosome and within mobile extrachromosomal elements in a wide variety of bacteria and are known to be involved in horizontal gene transfer events (Andres and Bertin, 2016). However, with

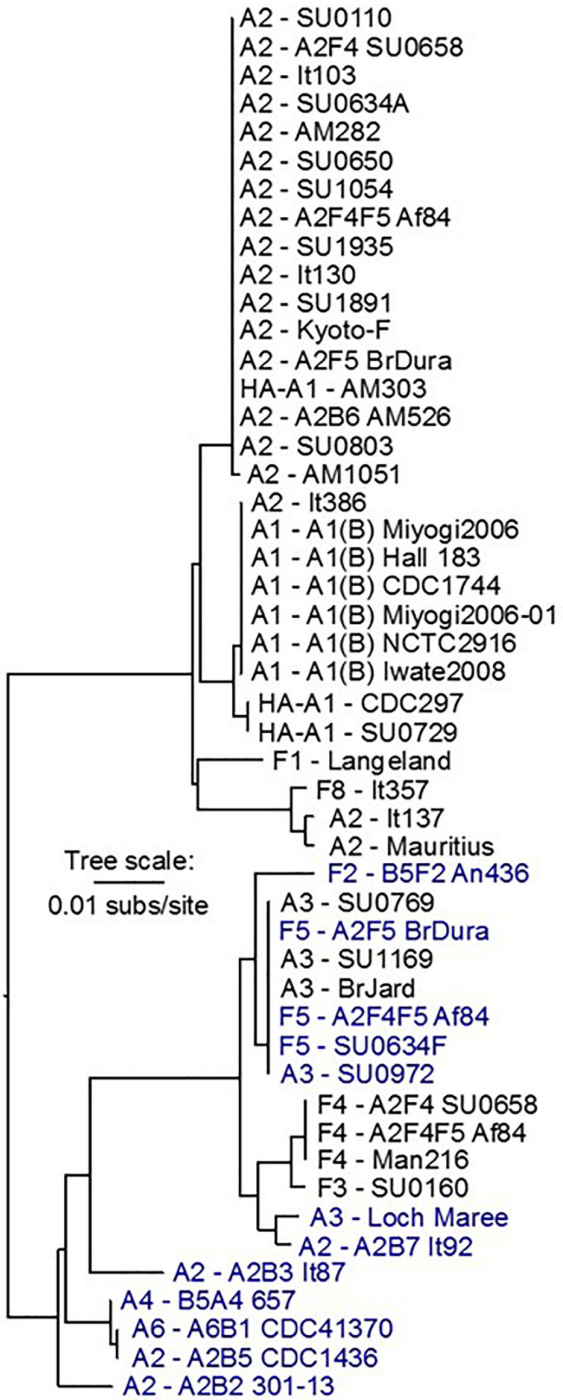


FIGURE 4 | Dendrogram showing the relationship between *bont/A* and *bont/F lycA* gene sequences located within the chromosome and plasmids. Chromosomally located *lycA* genes are in black font and plasmid-borne genes are in blue font. Two major *lycA* variants with nucleotide differences of ~6% can be discerned. The tree is midpoint rooted.

BoNT-producing clostridia, the *ars* operon appears to be strictly located within the chromosome.

Up to three copies of *arsC* genes form part of this operon (Dineen et al., 2004; Hill et al., 2009). The three *arsC* genes are identified here as *arsC1*, *arsC2*, or *arsC3* according to their locations relative to the *bont* and *ars* gene clusters. The *arsC1* gene is adjacent to the *orfX3* gene in the *bont* gene cluster, *arsC2* is found at the beginning of the *ars* gene cluster, and *arsC3,* when present, is located within the *ars* gene cluster. These genes vary in size, ranging from 337 to 392 bp with *arsC1* and 366–417 bp with *arsC2*. The *arsC* genes appear to have deteriorated in some cases, forming pseudogenes. A phylogeny comparing the *arsC1, arsC2,* and *arsC3* genes shows that the *arsC1* genes and the *arsC2* genes are each located within two distinct clades, while the *arsC3* genes form a single clade (**Figure 5**).

All three *arsC* genes are generally observed in strains containing chromosomally located *bont/A2* genes and with *orfX bont/A1* strain AM303. Where all three *arsC* genes are present, each *arsC* variant is nearly identical in sequence and the *arsC1* and *arsC2* genes are closely related, having greater than 96% identity. The *arsC3* gene is less closely related, with ∼80% identity when compared to *arsC1* or *arsC2*. In these genomes the *ars* operon follows the *bont* gene cluster (**Figure 6A**). However, with *orfX bont/A1* toxin gene clusters in BoNT/A1(B) strains and OrfX BoNT/A1 strains CDC 297 and SU0729; *bont/F1* gene clusters*;* the *bont/F8* gene cluster; and a few *bont/A2* toxin gene clusters the *arsC3* gene is missing and the location has been re-arranged so that the *ars* operon now precedes *arsC1* and the *bont* gene cluster (**Figures 6B–G**). Genomes that lack *arsC3* genes show greater variability within their *arsC1* and *arsC2* genes and their *arsC1* genes are within a distinct clade that differs by 6–7% from the clade in genomes having three *arsC* variants. In strains that lack both *arsC1* and *arsC3*, such as BoNT/A3, BoNT/F3, and BoNT/F4 strains, *bont* gene clusters are inserted into the chromosome at sites that are remote from the *ars* genes.

While the *bont* genes and gene arrangements immediately surrounding the *bont* gene clusters vary, the same conserved genes (#1–7 and #27–31 in **Figure 6**) flank these regions, placing them at a common location within the chromosome. **Figures 6E–G** show arrangements where *arsC1* and arsC2 are present but several other *ars* genes are absent, an indication of diminished arsenic resistance in these strains. **Supplementary Table 2** lists the genes/encoded proteins corresponding to the numbers shown in **Figure 6**.

The re-arrangement of the *bont* gene cluster related to the *ars* operon indicates an evolutionary shift, as does the loss of several *ars* genes seen with the arrangements of *bont* gene clusters in African and Italian strains. The arrangement in OrfX BoNT/A1 and BoNT/A1(B) strains suggests their *bont/A1* gene clusters may have evolved from a common ancestor, but several distinct genes that are lacking in the OrfX BoNT/A1 strains (**Figure 6B**) have been inserted upstream from the *bont/A1* gene cluster in the BoNT/A1(B) strains (**Figure 6C**), indicating an evolutionary divergence. Similarly, the arrangements in the African and Italian strains (**Figures 6D,E**, respectively) show a clear ancestral relationship, but additional genes have been inserted between the *ars* and *bont* cluster genes in the Italian strains, again signaling an evolutionary divergence. While these *bont* gene clusters show relationships within their non-toxin *bont* cluster genes and chromosomal integration arrangements, the

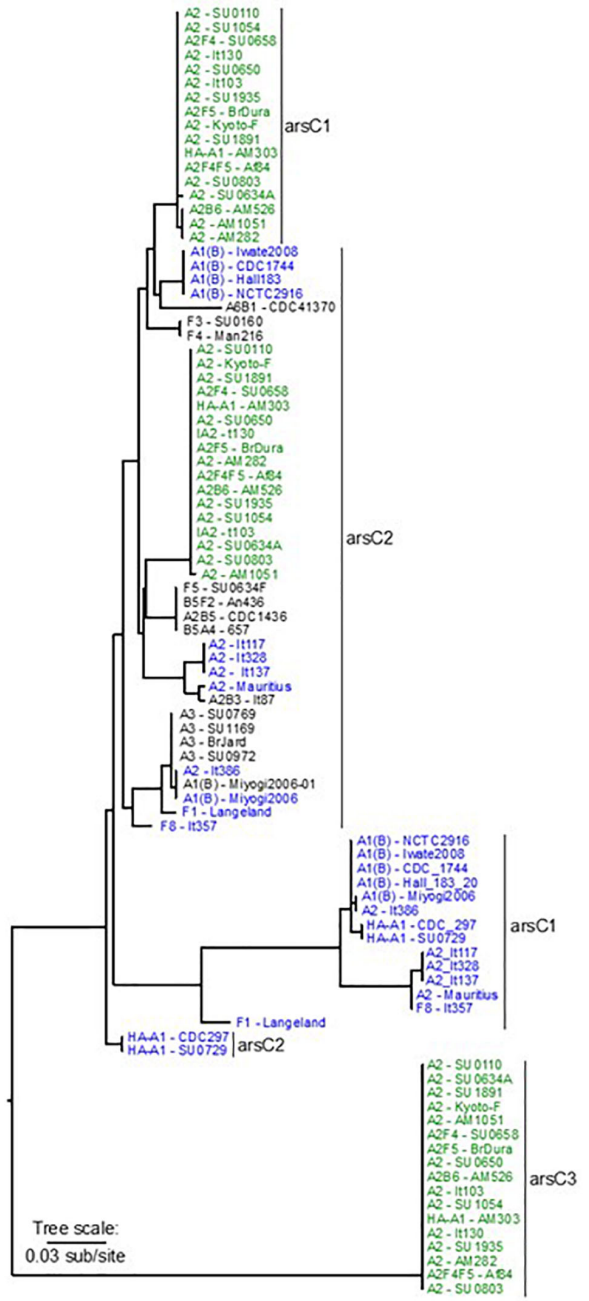


FIGURE 5 | Dendrogram comparing sequences from *arsC1*, *arsC2*, and *arsC3* genes associated with selected BoNT/A and BoNT/F toxin clusters. The *ars* genes are numbered 1, 2, or 3 according to their locations in relation to the *bont* gene cluster. Sequences from strains that contain all three *arsC* genes are shown in green font, sequences in blue font are from strains with *arsC1* and *arsC2* sequences, and sequences in black font are from strains with *arsC2* only. The *arsC3* gene differs by ∼20% in nucleotide residues from *arsC1* and *arsC2*. The tree is midpoint rooted.

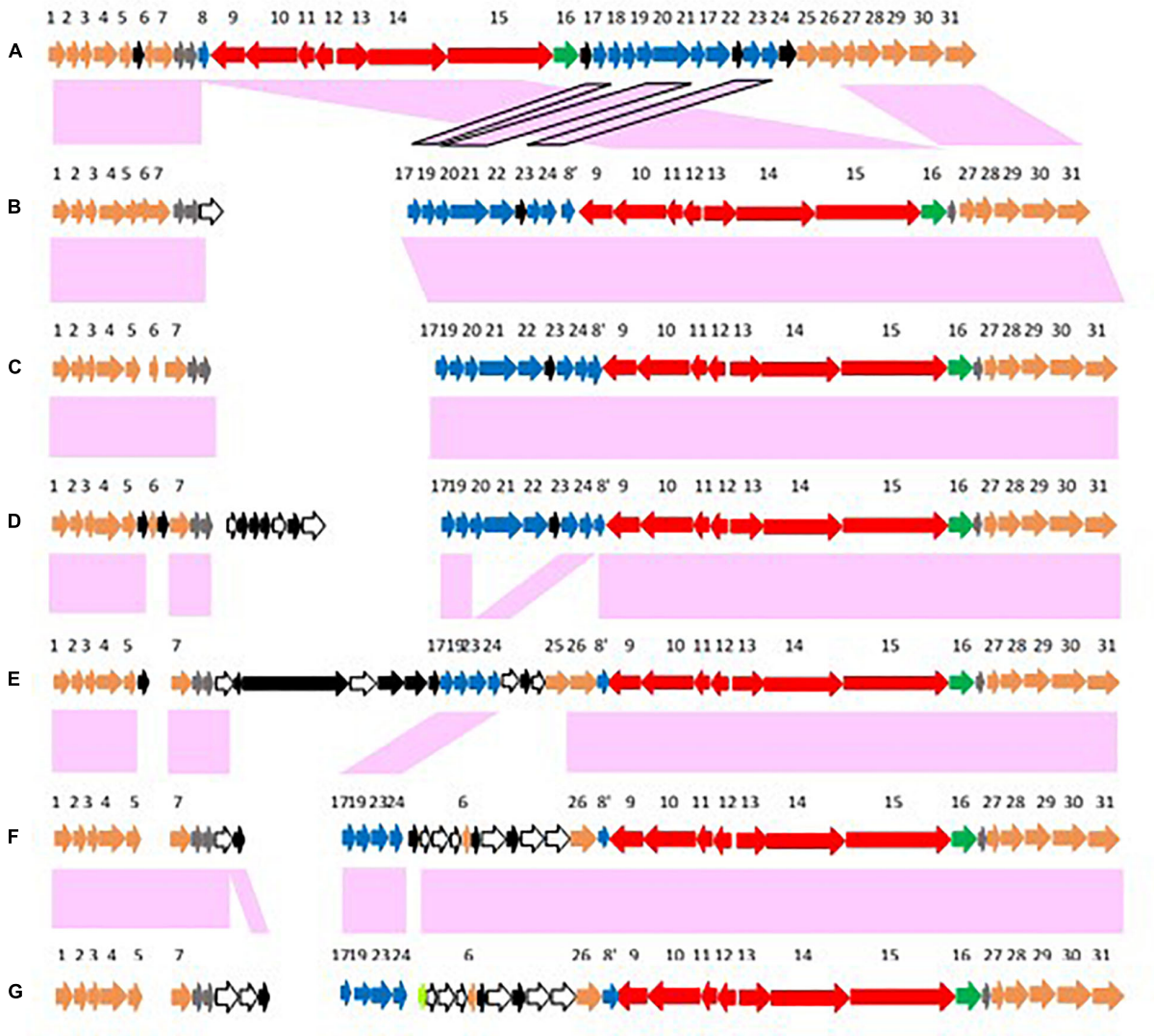


FIGURE 6 | Arrangements of conserved genes surrounding chromosomally located *bont* gene clusters that are adjacent to the *ars* operon. Conserved genes are numbered 1–31 with #9–16 representing the *bont* gene cluster (colored red with the *lycA* gene in green) and #17–24 representing the *ars* operon (colored blue). **Supplementary Table 2** lists the genes/encoded proteins corresponding to the numbers shown in **Figure 6**. Conserved genes that flank this location are colored orange; conserved hypothetical genes are gray; and non-conserved hypothetical genes are black. Panel **(A)** represents an arrangement where the *ars* operon is located downstream from the *bont* gene cluster while panels **(B–G)** represent arrangements where the *ars* operon precedes the *bont* gene cluster. Strains representing these arrangements include: **(A)** most Argentinean and Australian BoNT/A2 strains; **(B)** BoNT/F1 strains; **(C)** OrfX BoNT/A1 strains; **(D)** BoNT/A1(B) strains; **(E)** African BoNT/A2 strains; **(F)** Italian BoNT/A2 trains; and **(G)** Italian BoNT/F8 lt 357.

strains that harbor these genes are often unrelated, emphasizing the differential evolution of the bacteria and the *bont* genes.

Involvement of the *pulE* Gene in the Integration of *bont*/F Gene Clusters Into the Chromosome

As *bont/F3* and *bont/F4* gene clusters are remotely located (∼250 kb – 1.1 mb) from *ars* operons that lack both *arsC1* and *arsC3* genes, the chromosomal integration process differs from strains that contain *bont/A2* and *bont/F1* gene clusters. The insertion process for the *bont/F3* and *bont/F4* gene clusters involves excision of a *pulE* gene with insertion of *bont* gene cluster DNA within the *pulE* gene (**Figure 7**; Dover et al., 2013; Smith et al., 2020). The *pulE* gene is one of 15 genes that encode the *pul* secreton, which is type II secretion system for proteases and toxins in gram negative bacteria. This mechanism is similar to the insertion mechanisms for *bont/E* gene clusters that involve a split *rarA* gene (Hill et al., 2009) and *bont/F6* genes, where the split gene is *topB* (Carter et al., 2013); however, while a second intact *rarA* or *topB* gene is seen in these insertions, there is no duplicate intact *pulE* gene adjacent to the inserted *bont/F3* or *bont/F4* gene clusters. In clostridia the *pulE* gene may be non-essential, as most of the remaining

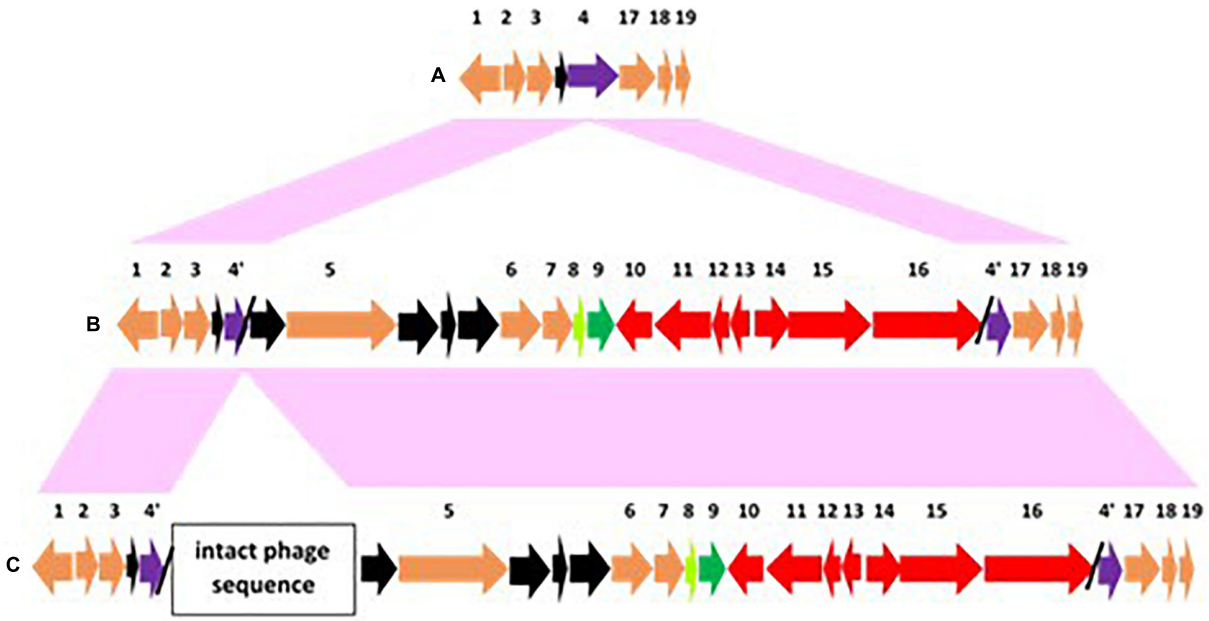


FIGURE 7 | Illustration of the integration location of chromosomally located *bont/F3* and *bont/F4* cluster genes, which is distinct from the *ars* operon location. At this site the *bont* gene clusters in panels **(B,C)** (#9–16, shown in red with the *lycA* gene in green) are inserted within a split *pulE* gene (#4 and 4', purple). **Supplementary Table 3** lists the genes/encoded proteins corresponding to the numbers shown in **Figure 7**. Additional conserved known genes are colored orange and conserved hypothetical genes are in black. Transposase genes are in light green. **(A)** Insertion region in BoNT F5 BrDura showing a complete *pulE* gene and an absence of the *bont* gene cluster. **(B)** The same region in BoNT F4 SU1425, showing an approximately 35 kb region between the truncated segments of the *pulE* gene. **(C)** The 92.5 kb DNA sequence in BoNT F3 SU0160 with the identical sequence found in BoNT F4 plus an inserted segment containing an intact bacteriophage.

genes encoding proteins that form this secretion system are absent. **Figure 7** shows the arrangement of the genes present within and flanking the split *pulE* gene for BoNT F3 SU0160 and multiple BoNT F4 strains, and the intact *pulE* gene in BoNT F5 strains, where the *bont* gene cluster resides within a plasmid. With both the *bont/F3* and *bont/F4* gene clusters the inserted genetic material between the two *pulE* gene fragments contains, in addition to the *bont* gene cluster, a 35 kb DNA fragment that includes a PHP domain protein gene, a KAP family p-loop domain protein gene, a helix-turn-helix protein gene (#5–7 in **Figures 7B,C**), and several hypothetical proteins. A complete listing of the genes/encoded proteins numbered in **Figure 7** is shown in **Supplementary Table 3**. However, with the *bont/F3* gene cluster, an additional 57 kb DNA fragment encoding an intact bacteriophage is present (Smith et al., 2020). The significance of this finding is as yet unknown. An interesting finding is a similar split *pulE* gene within the BoNT/A3 Loch Maree strain which is devoid of *bont* cluster genes but instead contains an intact prophage sequence that is distinct from the one associated with the *bont/F3* gene cluster (Dover et al., 2013; Smith et al., 2020), marking this as a general site for integration of prophage sequences as well as *bont* gene clusters.

Toxin Gene Clusters Within Plasmids

While most of the bacterial strains analyzed here contain *bont* genes within the chromosome, several strains contain one or more *bont* genes within large (∼250–280 kb) conserved conjugative plasmids. Characteristics of the plasmids for which complete sequence information is known are shown in **Table 2**. The strains that harbor plasmid-borne *bont* genes are relatively rare, contain a wider variety of *orfX bont/A* and *bont/F* genes, and are found in genomes that are within variable phylogenetic clades. Strains containing these plasmids have been isolated in North and South America, Europe, Asia, and Australia, indicating an ease with which these plasmids may spread among globally located strains. The ability of these plasmids to move between different clostridial strains and species has also been confirmed through laboratory experiments (Marshall et al., 2010).

These plasmids may contain one or two toxin gene clusters, located at two distinct sites. One site exclusively contains *ha bont/B* gene clusters and the second site contains *orfX bont/A* or *bont/F* gene clusters (**Figure 3**). Plasmids that contain only one *bont* gene cluster (*bont/A3*, *bont/F5*, or *bont/B* gene clusters) are known as well as those with two *bont* gene clusters (*bont/F2* plus *bont/B5* genes or *bont/A2, bont/A4*, or *bont/A6* paired with *bont/B* genes). A search for plasmid gene sequences within *C. parabotulinum* using BLASTn indicates that, with one exception, all of these related large *C. parabotulinum* plasmids contain *bont* genes and, because of their common toxin cluster locations within the plasmid, none contain both *bont/A* and *bont/F* gene clusters together.

Identical *bont* genes and gene clusters have been located either within plasmids or the chromosome of *C. parabotulinum* strains, which confirms the ability of these plasmid-borne genes to integrate into the chromosome and also affords us the

TABLE 2 | Characteristics of plasmids from *C. parabotulinum* containing *bont/A* and/or *bont/F* genes.

BoNT subtype(s)	Strain	Plasmid ID	Accession #	Size (bp)	Toxin cluster location
A2B3	It 87	p1_A2B3_87	AUZB01000012.1	275,568	Plasmid/plasmid
A2b5	CDC 1436	pCBG	CP006909.1	275,986	Plasmid/plasmid
A2B7	It 92	p_A2B7_92	AUZA01000014.1	260,807	Plasmid/plasmid
A2f4f5	Af84	pCLQ	AOSX01000021.1	246,124	A2, F4 chr; F5 plasmid
A2f5	BrDura	pRSJ20_1	CP014152.1	241,076	A2 chr; F5 plasmid
A3	SU0972		MWIV01000007.1	238,810	Plasmid
A3	Loch Maree	pCLK	NC_010418	266,785	Plasmid
B5a4	657	pCLJ	NC_012654	270,022	Plasmid/plasmid
B5a4	CFSAN034200	p1_CDC51232	CP031095	270,024	Plasmid/plasmid
B5f2	An436		LFON01000008	171,021	Plasmid/plasmid
F5	SU0634F	pRSJ3	CP013710	244,784	Plasmid
F5	SU0632		MWIY01000007.1	243,777	Plasmid

opportunity to study the extent of exchanged genetic material. This duplicate location of *bont* clusters in the chromosome or plasmids is not specific to *orfX C. parabotulinum* strains. Gene clusters containing *bont/B1* or *bont/B2* genes have been identified that are located within either the chromosome or plasmids (Franciosa et al., 2009), as well as *bont/E1*, *bont/E3*, and *bont/E10* genes (Zhang et al., 2013; Carter et al., 2016) and *bont*/F7 genes (Halpin et al., 2017; Mazuet et al., 2017).

BLASTn analysis of complete plasmids from twelve of these strains show 95–100% identity over large segments of the plasmid sequence, illustrating the close relationships among them (**Table 3**). However, they also show evidence of continuing insertion, deletion, and inversion events. For example, the plasmid from strain An436 (LFON01000008) that contains *bont/B5* and *bont/F2* gene clusters is closely related to the strain 657 plasmid (pCLJ) that contains *bont/B5* and *bont/A4*, but a large (99 kb) deletion can be seen and, in addition, the *bont/F2* gene cluster is reversed in orientation compared to the rest of its plasmid DNA (Hill et al., 2009).

In fact, the orientation of the *bont* gene cluster in relation to the surrounding genes varies from plasmid to plasmid. The *bont/A* gene cluster (#1–8 in **Figure 8**) in the BoNT/A2B3 It 87 strain, BoNT/A2B5 CDC 1436, BoNT/A6B1 CDC 41370, and BoNT/B5A4 strains is in the same orientation as the surrounding genes (**Figures 8A,B**), but with BoNT/A2B7 It 92, BoNT/A2F5 or BoNT/F5 strains, and BoNT/A3 Loch Maree the genes are in the opposite orientation from the surrounding genes (**Figures 8C–F**). The factors that govern the orientation of *bont* gene clusters are not understood; however, the presence of transposes adjacent to the *lycA* gene with most of these gene clusters suggests a possible role for them in gene integration and orientation.

In the plasmid-borne *orfX bont* cluster regions, the gene sequences that are downstream of the *bont* gene cluster are well-conserved, showing that all *orfX bont* clusters are located in the same position within the plasmid, regardless of the *bont* gene present. The *lycA* and *bont* clusters are followed by *HepA/SNF2*, which encodes a DNA helicase protein, plus thermonuclease (*Tnase*), and an additional DNA helicase gene (#9–11 in **Figure 8**). However, the upstream sequences are arranged in multiple patterns.

The upstream genes in the BoNT A2b3 It 87 strain differ from those in other plasmids and contain over a dozen unique hypothetical genes (**Figure 8A**). Several transposases flank the upstream and downstream genes. A conserved upstream gene sequence arrangement is seen with BoNT/A2b3 CDC 1436, BoNT/B5a4 657, and BoNT/A6b1 CDC 41370 strains that includes AraC family transcriptional regulators, a methionine adenosyltransferase gene, the chaperone ClpB gene, a cytidine deoxyribosyltransferase gene, and a thiamine biosynthesis protein gene (#14–20 in **Figures 8B,C**). These strains are geographically related, having been isolated in the southwestern U.S. and Mexico. A similar arrangement is seen in the North American BoNT/B5F2 strains.

A different arrangement is seen in the Italian A2b7 It 92 strain and strains containing bont/*F5* genes where the upstream genes have been replaced with five unique genes (genes encoding a single strand binding protein gene, a DNA polymerase III subunit gene, and A-type inclusion proteins) (#21–24 in **Figures 8D,E**). Variable numbers of transposase genes and genes encoding hypothetical proteins are also present within these regions.

The upstream region of the plasmid *bont/A3* gene cluster in the Loch Maree strain consists of nine unique hypothetical genes and genes encoding a putative ABC transporter lipoprotein and an adenylate and guanylate cyclase protein (#27 and #28 in **Figure 8F**). This region is similar to, but not identical with, that of the *bont/A3* gene cluster region in strain SU0972. A complete listing of the numbered genes/encoded proteins in **Figure 8** is shown in **Supplementary Table 4**.

The conserved downstream gene sequences identify a common *orfX bont* gene cluster location within the plasmid but the varied upstream DNA sequences may provide some clues as to why these *bont* gene clusters have not integrated into the chromosome. The *HepA/SNF2*, *Tnase*, and DNA helicase genes are also present in the absence of *orfX bont* gene clusters in plasmids that contain only *ha bont/B* clusters. Their presence may be an indication of the past existence of *orfX bont* gene clusters at that site or an opportunity for possible future acquisitions. Similar to the "chicken or the egg" quandary, it is not definitively known whether these *bont* gene clusters originated within plasmids followed by integration into the chromosome,

TABLE 3 | Comparison of large plasmids containing *bont* gene sequences.

	A2B5 CDC 1436	A2B6 AM282	A2B6 AM526	A2B7 It92	A2f5 BrDura	A2f4f5 Af84	A3 Loch Maree	A3 SU0972	B5a4 657	B5f2 An436	F5 SU0634F	F5 SU0632
A2B3	96%	97%	97%	97%	97%	97%	96%	97%	97%	97%	97%	98%
It87	76%	76%	76%	76%	73%	74%	71%	73%	74%	66%	70%	80%
A2B5	–	96%	96%	96%	95%	95%	96%	96%	98%	97%	96%	95%
CDC1436		70%	69%	73%	72%	73%	70%	70%	83%	83%	69%	78%
A2B6		–	100%	96%	95%	97%	96%	96%	95%	96%	97%	97%
AM282			100%	79%	72%	75%	72%	76%	72%	61%	74%	73%
A2B6			–	96%	97%	97%	95%	96%	96%	96%	97%	97%
AM526				73%	68%	69%	71%	69%	70%	54%	68%	67%
A2B7				–	98%	98%	98%	98%	97%	97%	98%	98%
It92					78%	78%	80%	79%	79%	68%	73%	80%
A2f5					–	99%	98%	99%	96%	92%	99%	99%
BrDura						99%	81%	92%	79%	64%	100%	99%
A2f4f5						–	98%	99%	96%	92%	99%	99%
Af84							79%	90%	79%	66%	97%	99%
A3 Loch							–	98%	96%	97%	98%	98%
Maree								79%	67%	63%	77%	80%
A3								–	97%	91%	98%	99%
SU0972									77%	63%	92%	91%
B5a4									–	99%	96%	96%
657										82%	75%	80%
B5f2										–	92%	92%
An436											64%	66%
F5											–	99%
SU0634												98%

Note that these plasmids share greater than 90% identity over at least 60% of their sequence. The percentage identity is in black font and the percent coverage between the plasmids is in grey font. Closely related plasmids (99–100% identity with >96% coverage), such as those containing bont/B6 or bont/F5 genes, are bordered. The comparisons were generated using discontiguous BLAST analysis (NCBI).

whether the opposite is true, or whether movements in both directions is possible.

An understanding of the gene components and arrangements surrounding identical *bont/A2* and *bont/A3* gene clusters within the chromosome and plasmids presents an opportunity to examine the extent of genetic material that is exchanged within these genetic regions and potentially identify genes or insertion elements that may aid in *bont* gene cluster integration. When chromosomal and plasmid *bont* gene cluster regions of *bont/A2* gene clusters are compared, only the *lycA* and *bont* gene cluster sequences are found to be present in both locations, so that the 17 kb of material transferred between the plasmid and chromosome is strictly composed of the *lycA/bont* gene cluster (**Figure 9**). In this figure, the *bont* gene cluster genes are colored red and the *lycA* gene is green. Known genes that are unique to each site are colored in purple in **Figure 9**, while conserved known genes are in orange, conserved hypothetical genes are in gray or black. Individual genes are identified in **Supplementary Tables 2,4**, as numbered in **Figure 6A** (BoNT/A2 chromosome) and **Figure 8B** (BoNT/A2 plasmid).

However, with BoNT/A3 strains the sequence that has been transferred is comprised of a 35 kb highly conserved region surrounding the inverted *bont/A3* gene cluster which includes two known genes and nine conserved hypothetical genes that are located upstream of the *bont/A3* gene cluster, and the downstream *HepA/SNF*, *Tnase*, and DNA helicase genes. Individual genes surrounding the *bont/A3* gene clusters are identified in **Supplementary Table 4**, as numbered in **Figure 8F** (BoNT/A3 chromosome and plasmid).

SNF family proteins contain helicase-like domains and often reside within large multi-protein complexes that may facilitate various DNA manipulations (Ryan and Owen-Hughes, 2011). Thus, it is possible that the proteins encoded by the *HepA/SNF2* and DNA helicase genes may function as facilitators for the transfer of the *bont/A3* gene cluster from plasmid to chromosome, or vice versa. While the plasmid-borne *bont/A2* and *bont/A3* gene clusters are located at the same site, these gene clusters are differentially placed at discrete sites within the chromosome implying interaction processes involving distinct target and facilitating genes.

The *bont/A2* gene clusters are commonly found within the chromosome, and only in rare cases are they located within plasmids. As there is an association between chromosomally located *bont/A2* gene clusters and *arsC* genes, an examination of the *ars* gene cluster regions within the genomes of several strains having plasmid-borne *bont/A2* gene clusters (BoNT/A2B3 It87, BoNT/A2B5 CDC 1436, and BoNT/A2B7 It92) was undertaken. The chromosomally located *ars* gene operons in these strains contained *arsC2* genes, however, none contained the *arsC1* or *arsC3* genes, which provides an indication by omission of the

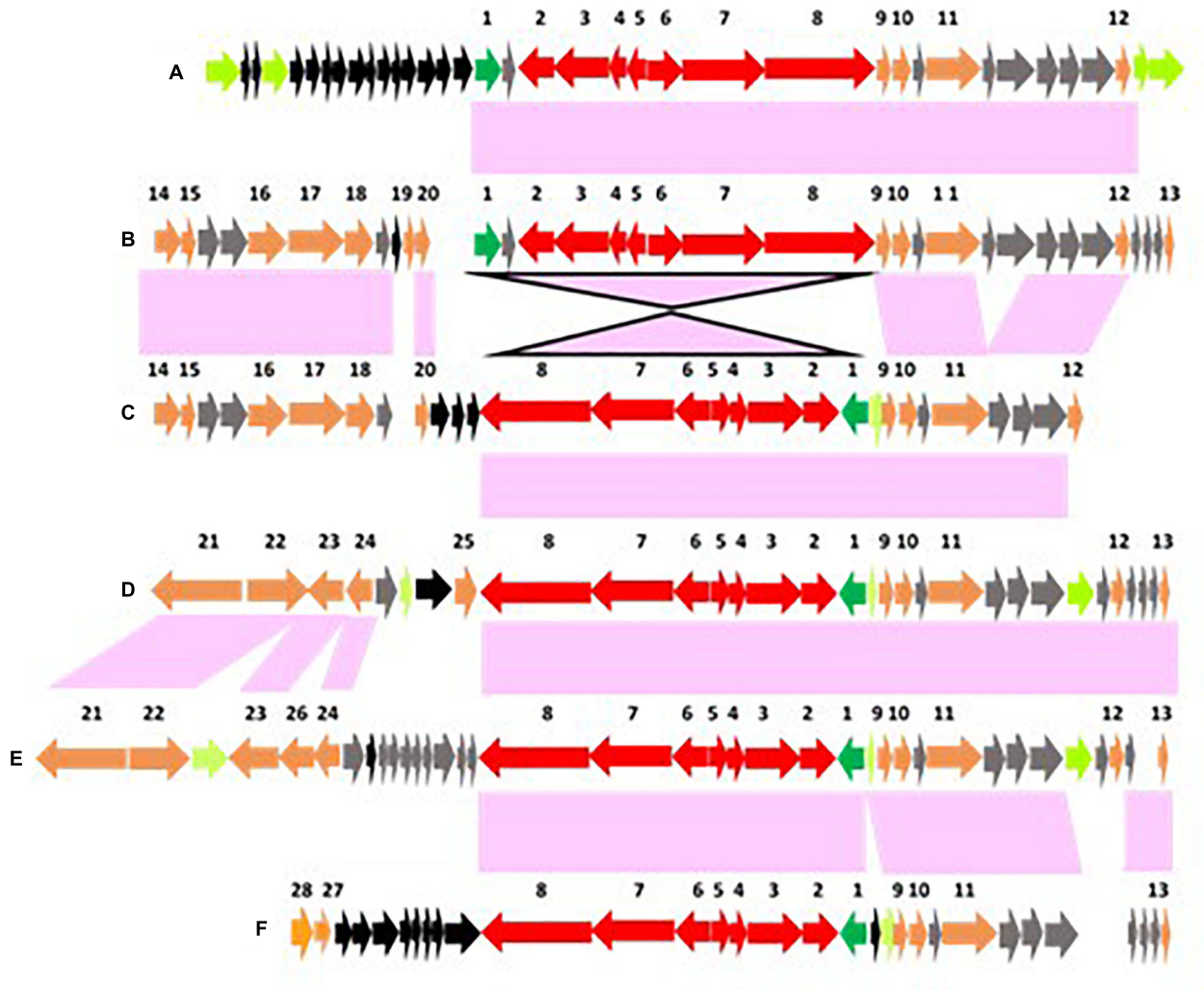


FIGURE 8 | Arrangements of conserved genes surrounding *orfX bont* gene clusters that are located within plasmids. Conserved genes representing the *bont* gene cluster are numbered #1–9, colored red with the *lycA* gene in green. Conserved genes that flank this location are colored orange, including the *HepA/SNF*, *Tnase*, and DNA helicase genes (#9–11); conserved hypothetical genes are gray; and non-conserved hypothetical genes are black. **Supplementary Table 4** lists the genes/encoded proteins corresponding to the numbers shown in **Figure 8**. Panels **(A,B)** illustrate arrangements where the *bont* gene cluster is in the same orientation as the surrounding genes, while panels **(C–F)** show arrangements where the *bont* genes are in opposite orientation. Examples of these arrangements are found with **(A)** the *bont/A2* gene cluster in strain BoNT/A2B3 lt 87; **(B)** the *bont/A2* gene cluster in BoNT/A2B5 CDC 1436; **(C)** the *bont/F2* gene cluster in BoNT/B5F2 An436; **(D)** the *bont/A2* gene cluster in BoNT/A2B7 lt 92; **(E)** the *bont/F5* gene cluster in BoNT/A2F5 BrDura, BoNT/F5 SU0634F, and BoNT/A2F4F5 AF84; and **(F)** the *bont/A3* gene cluster in BoNT/A3 Loch Maree.

possible involvement of the *arsC1* genes in the movement of *bont/A2* gene clusters between plasmid and chromosome.

Similarly, the sequence regions surrounding the *bont/A3* gene cluster within the chromosome of BoNT/A3 SU1169 were compared with the same region in BoNT/A3 Loch Maree, where the *bont/A3* gene cluster is found within a large plasmid. The Loch Maree chromosome is missing a 44.7 kb region found within SU1169 that includes the entire ∼35 kb region that is conserved among chromosomal and plasmid-borne *bont/A3* clusters plus additional genes, hinting that the as-yet-unknown facilitators for chromosomal insertion of the *bont*/A3 gene might be located within the 45 kb region that is absent in the Loch Maree strain. Further study of these DNA sequences is needed to enable a better understanding of how their *bont/A3* gene clusters are excised from plasmids and integrated into the chromosome.

Contributions From Bacteriophage

Both intact and incomplete bacteriophages are commonly found among the BoNT-producing clostridia. Intact active phages have been recovered from mitomycin C-induced lysed cultures from BoNT/A, BoNT/B, BoNT/E, and BoNT/F strains (Eklund et al., 1969). In addition, it is well known that the genes for type C and D toxins are associated with bacteriophage DNA, that "curing" of these bacteria of its phage results in a reversion from toxic to non-toxic status, and that subsequent infection with isolated phage preparations reactivate its toxicity (Eklund and Poysky, 1974).

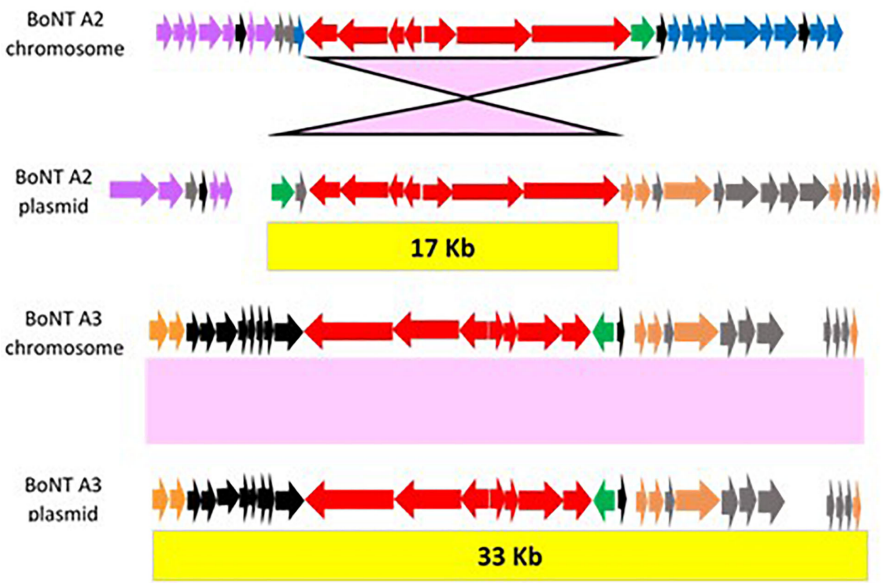


FIGURE 9 | Illustrations of conserved DNA sequences that are shared between chromosomal and plasmid located *bont* clusters. With BoNT A2 strains, conserved sequences comprise only the actual *bont* cluster and *lycA* genes, while the conserved sequence with BoNT A3 strains includes additional surrounding genes. The *bont* cluster genes are colored red, the *lycA* gene is green, and other conserved known genes are orange. Genes that are unique to chromosomal or plasmid-borne *bont/A2* gene clusters are shown in purple. Conserved hypothetical genes that are common to both plasmid-borne *bont/A2* and plasmid and chromosomally located *bont/A3* are in gray and conserved hypothetical genes that are exclusive to *bont/A3* are in black. Specific gene identifications are listed in **Supplementary Tables 2, 4** and are related to **Figure 6A** (*bont/A2* chromosome), 8B (*bont/A2* plasmid), and 8F (*bont/A3* chromosome and plasmid).

Bacteriophage are also known to be vehicles for the general transfer of PAIs.

Prophage DNA sequences have also been located within the genomes of numerous *orfX C. parabotulinum* organisms. A search for phage sequences in 15 complete *orfX* BoNT/A and BoNT/F genomes using PHASTER (Zhou et al., 2011; Arndt et al., 2016) uncovered more than 19 intact phages within the chromosomes, and numerous incomplete phage sequences were discovered within both the chromosome and the large highly conserved plasmids among these strains. The majority of the identified phages were found within Argentinean BoNT/A2, BoNT/A3, and BoNT/F strains and were related to phages associated with either *Clostridium difficile* or *Clostridium tetani* (**Table 4**). Some phages associated with *Clostridium sporogenes* and *Clostridium tyrobutyricum* were also identified. The additional finding of shared DNA sequences with prophage fragments from the phage Clostr_c_st (the phage that contains *bont/C* and *bont/D* gene clusters) in the *C. parabotulinum bont*-containing plasmids is not surprising, as these shared sequences were generally linked to non-toxic genes within *bont* gene clusters. However, some conserved transposases were noted as well.

A bacteriophage associated with *Bacillus megaterium*, Bacilli_Moonbeam, was found adjacent to the *bont/F3* gene cluster in the genome of BoNT/F3 SU0160. This is a unique association of an intact phage sequence with the *bont* cluster of an *orfX C. parabotulinum* strain, as in all other cases complete prophage DNA sequences are remotely located from the *bont* gene cluster. However, it is not known if the *bont/F3* gene cluster is actually part of the prophage, or if they are simply co-located.

Individual genes related to incomplete phage sequences have also been identified that are scattered among the conserved regions surrounding many *bont* clusters within both the chromosome and plasmids. These may represent remnants of ancient phages that previously integrated into the chromosome, possibly in association with *bont* gene clusters, or they may indicate that the present day conjugative plasmids responsible for the movement of *bont* gene clusters may have originated as bacteriophages where certain phage-specific genes have deteriorated while elements necessary for conjugation and *bont* genes have persisted. The significance of these bacteriophages may be underappreciated as contributors to overall genetic diversity in these bacteria through the placement of foreign genes into the chromosome via HGT.

DISCUSSION

Clostridium parabotulinum is a diverse species in their geographic range, in their genomic properties, and in their neurotoxins. BoNT-producing *C. parabotulinum* bacteria have a global distribution but are mainly found in temperate and subarctic regions in both the northern and southern hemispheres. It is noteworthy that, while the majority of *C. parabotulinum* strains isolated in the northern hemisphere harbor *bont* gene clusters containing *ha* genes, in the southern hemisphere the *bont* gene clusters predominantly contain *orfX* genes. As clostridia are

TABLE 4 | A sampling of bacteriophages found within *orfX* + BoNT/A and BoNT/F *C. parabotulinum* strains.

Phage	Phage type	Associated bacteria	*C. parabotulinum* strains containing intact phage sequences		*C. parabotulinum* strains with partial phage sequences*	
phiCT19406A = phiCTC2A	unclassified dsDNA virus	*C. tetani* ATCC 19406 *C. tetani* C2	A2b5 CDC 1436		B5f2 An436	
phiCT19406B = phiCT453B	Siphoviridae	*C. tetani* ATCC 19406 *C. tetani* ATCC 453	A2 Kyoto-F A2 CDC 53174 A2f5 SU0650 F1 Langeland	F4 SU1425	A2 SU0994 A2 SU0807 A2 SU0998 A2 SU1072	A2 SU1275 A2 SU0801 A2 SU0945 A2 SU1054
phiCT19406C	unclassified dsDNA virus	*C. tetani* ATCC 19406	A2 Kyoto-F A2 SU0634A A2b5 CDC 1436 A2f5 SU0650 A3 Loch Maree A3 SU1169 A3 CDC 54064 F1 Langeland	A2 SU1259 A2 SU0994 A2 SU1064 A2 SU1072 A2 SU1054 A2 SU1275 A2 SU1917 A2 SU1274	224-13 SU1074 SU0807 SU0801	
phiC2	Myoviridae	*C. difficile* CD27	F1 Langeland F1 Walls 8G	F1 230613		
phiCD27 = phiMMP02	Myoviridae	*C. difficile* CD27 = *C. difficile* CD119	A2 Kyoto-F A2 SU1275 A2 SU1274 A2 SU1072 A2f5 SU0650 A2f5 BrDura	A2 SU1259 A2 SU1917 A2 SU1891 A2 SU0807 A2 SU1064 A2 SU1054	A2 SU0998 A2 SU1112 A2 SU1887 A2 SU1934 A2 224-13	A2 SU1937 A2 SU0801 A2 SU0994 A2 SU1074
phiCD38_2	Siphoviridae	*C. difficile* CD38	A2 Mauritius			
phiCD119 = MMP04	Myoviridae	*C. difficile* CD119 = *C. difficile*	B5a4 657 AB CFSAN034200		CDC 69096	
phiCD506		*C. difficile* CD506	A2f5 BrDura		A3 CDC 54064	SU0635W
phiCD6356	Siphoviridae	*C. difficile* CD6356	A2f5 BrDura		A3 SU1169	A3 CDC 54064
phiSM101	Siphoviridae	*C. perfringens* SM101	F1 Langeland F1 Walls 8G	F1 230613		
phi8074_B1	Siphoviridae	*C. sporogenes* ATCC 8074	A2 0634A A2b5 CDC 1436 B5a4 657 CFSAN034200	F5 SU0634F CDC 69096		
phiCTP1	Siphoviridae	*C. tyrobutyricum*	B5a4 657 AB CFSAN034200			
Bacilli_Moonbeam	Myoviridae	*B. megaterium*	F3 SU0160			
Bacill_phIS3501	Siphoviridae	*B. thurgensis v israelensis*	F1 Langeland F1 Walls 8G	F1 230613		
OH2	Siphoviridae	*T. thermophilus* HB8	B5a4 657			

**Bacteria listed as containing partial prophage sequences show >65% coverage with >98% identity using BLAST analysis at the nucleotide level. Some partial sequences may represent overlapping phage sequences.*

spore-forming bacteria, this discrepancy in distribution may be due to differential movements of their bacterial spores. East-west movements among strains containing similar *bont* gene clusters, which could primarily be due to transport of bacterial spores on upper level wind currents, may be more common than north-south movements, which are more likely the results of carriage of spores by migrating birds or due to transfers of contaminated soil, plants, or foodstuffs.

There is evidence that certain geographic areas have provided conditions conducive to the widespread colonization of phylogenetically related bacterial strains, or clonal expansions. However, naturally occurring or manmade movements of bacterial spores from these areas have provided opportunities for these bacteria to become established in faraway regions and allowed for exchanges of novel genetic material between different bacterial strains via horizontal gene transfers (HGT). Comparisons of genomes and *bont* gene cluster sequences reveals that closely related strains have acquired different *bont* genes, often within related *bont* gene clusters, while unrelated strains may contain identical *bont* genes/gene clusters. The closely related strains isolated in northwest Argentina that contain *bont/A2*, *bont/A2F4*, or *bont/A3* genes might be considered clonal expansions, while the finding of phylogenetically diverse *C. parabotulinum* strains located in Argentina, Africa, and Australia that contain identical *bont/A2* gene clusters is an example of the apparent movements of these strains with subsequent exchanges of genetic material among them. Phylogenetic analysis of the genomes representing a subset

of *C. parabotulinum* strains that contain *orfX* gene clusters demonstrates these relationships, revealing the presence of several clades that are highly conserved, contrasting with others that are quite variable.

HGT events may be accomplished following active infection by bacteriophage or by movement of genes that are incorporated into mobile genetic elements such as temperate prophage DNA or conjugative plasmids. All three mechanisms may have been utilized in the movement of toxin genes among the BoNT-producing clostridia, but with *C. parabotulinum* strains the majority of such transfers appear to involve conjugative plasmids. However, bacteriophage gene remnants are frequently found adjacent to *bont* gene clusters within the chromosomes and extrachromosomal plasmids. These remnants contain DNA sequences that encode various viral components, including recombinases, integrases, and transposases. It is interesting to speculate that perhaps these large plasmids have originated from bacteriophage that subsequently lost DNA sequences encoding vital structural genes, so that ancient bacteriophages may have evolved to become modern conjugative plasmids.

Once the plasmid-borne *bont* gene clusters have established a presence in a bacterial strain, they may proceed to integrate into their host chromosome and/or exchange genetic material with existing *bont* genes or gene clusters. This is accomplished through homologous recombination (HR) events that involve alignment of paired gene sequences, excision of a particular gene or intergenic sequence, and insertion of a new or exchanged DNA sequence. HR events within *ntnh* genes have placed *orfX bont/A1* genes within *ha bont/B* gene clusters, and *bont/A2* genes within *bont/F1* gene clusters, and they are responsible for some of the toxin subtype diversity seen among *C. parabotulinum* strains. It is known that the *bont/A2* gene resulted from an HR event involving *bont/A1* and *bont/A3* genes, and there is evidence of HR interactions between *bont/F2* and *bont/F3* and *bont/F1* and *bont/F8* that may have shaped their individual genetic identities.

Chromosomal integrations of entire *bont* gene clusters also utilize HR processes, including the necessity for paired genes or intergenic sequences. For example, integrations of *bont/A2* and *bont/F1* gene clusters, that occur at a common site within the chromosome, are dependent on paired *arsC1* and *arsC2* genes. Chromosomal integration provides a more stable location within the bacteria than extrachromosomal locations - it has been shown both experimentally and through natural occurrence that extrachromosomal DNA is subject to deletion through plasmid loss or "curing" of bacteriophage, while loss of chromosomally-located *bont* genes has yet to be definitively demonstrated.

It is thought that movement of *bont* gene clusters into non-neurotoxigenic clostridia is a one-way process that occurs via introduction within extrachromosomal plasmids followed by chromosomal integration. Previously published information and results from analysis of *bont* gene cluster locations in this study provide some evidence for this hypothesis. The ease of movement of plasmids containing *bont* genes among *C. parabotulinum* has been demonstrated in the laboratory and plasmid losses, with subsequent reversion to non-neurotoxicity, have also been noted. With the exception of a few very recently published genomes, there is no evidence of the persistence of these large plasmids after loss of their *bont* genes. While identical *bont* gene clusters have been located within either plasmids or the chromosome, the majority of *C. parabotulinum* strains (∼90%) contain chromosomally located *bont* gene clusters, indicating an ease of integration and/or selective pressure to do so. In strains where *bont* genes remain within extrachromosomal plasmids, targeting or facilitating genes that are necessary for chromosomal integration appear to be lacking. For example, strains having chromosomally located *bont/A2* gene clusters contain paired *arsC* genes (*arsC1* and *arsC2*) but in strains where *bont/A2* genes are extrachromosomal, *arsC1* genes are missing.

However, a question arises as to whether chromosomally integrated genes may be excised from the chromosome and re-inserted into the plasmid. In some cases, there is evidence of deterioration of the facilitating genes or IS elements that may have participated in chromosomal integration, so that use of these facilitators to reverse the integration process is not possible. In others, the plasmids have been lost. There are very few examples where *bont* gene clusters are located within the chromosome but an extracellular plasmid remains. In BoNT/A2F5 strains and the Australian BoNT/A2B6 strain AM526, *bont/F5* and *bont/B6* gene clusters remain within the plasmids but the *bont/A2* gene clusters are within the chromosome. In the former case, the *bont/F5* gene cluster inhabits the plasmid site where the *bont/A2* gene would be located, presumably blocking re-integration, but in the BoNT/A2B6 strain, the *HepA/SNF* site is present in the plasmid. It is not possible to determine if the chromosomally located *bont/A2* gene cluster in the BoNT/A2B6 strain could be re-integrated into the plasmid or not, as currently only one such strain has been isolated. The rarity of isolates where there is even a possibility for plasmid re-integration suggests chromosomal integration of *bont* gene clusters is essentially a one-way process, but further investigations are needed to confirm or refute this hypothesis.

While the genomic diversity seen within this species may have evolved partly through individual genetic mutations that over time have resulted in minor changes in genes, the major evolutionary driver behind genomic diversity is likely wholesale movements of individual genes, gene clusters, and large segments of DNA via horizontal gene transfers and recombination events. Analysis of recently sequenced *orfX C. parabotulinum* genomes from southern hemisphere bacterial collections have added to our overall knowledge about the BoNT-producing clostridia, increased our understanding of the diversity seen among these strains, and provided insights into mechanisms behind the generation of this diversity. Bacterial strains have been transported across oceans and continents using various environmental means and interactions between diverse strains have produced novel genomic variants. Over time, gene transfers and exchanges have shaped *bont* gene diversity. While these observations are focused on a particular subset of a single bacterial species, insights may be applied to the study of diversity within other clostridial species and possibly other bacteria as well.

DATA AVAILABILITY STATEMENT

NCBI accession information for all genomic data presented in this study is available in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

TS performed analyses and wrote the manuscript. CW performed analyses and edited the manuscript. JS edited the manuscript and guided genomic analyses. KH, SJ, FA, BA, RF, PC, and PK contributed strains/DNA/genomic sequencing and source information for strains. All authors reviewed the manuscript.

FUNDING

The work was supported in part by the U.S. Department of Homeland Security Science and Technology Directorate via award HSHQDC-16-C-B0013.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.566908/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andres, J., and Bertin, P. N. (2016). The microbial genomics of arsenic. *FEMS Microbiol. Rev.* 40, 299–322. doi: 10.1093/femsre/fuv050

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

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

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bonventre, P. F., and Kempe, L. L. (1960). Physiology of toxin production by *Clostridium botulinum* types A and B. I. Growth, autolysis, and toxin production. *J. Bacteriol.* 79, 18–23.

Burke, G. S. (1919). Notes on *Bacillus botulinus*. *J. Bacteriol.* 4, 555–570.

Carter, A. T., Austin, J. W., Weedmark, K. A., and Peck, M. W. (2016). Evolution of chromosomal *Clostridium botulinum* type E neurotoxin gene clusters: evidence provided by their rare plasmid-borne counterparts. *Genome Biol. Evol.* 8, 540–555. doi: 10.1093/gbe/evw017

Carter, A. T., Pearson, B. M., Crossman, L. C., Drou, N., Heavens, D., Baker, D., et al. (2011). Complete genome sequence of the proteolytic *Clostridium botulinum* type A5(B3') strain H04402 065. *J. Bacteriol.* 193, 2351–2352.

Carter, A. T., Stringer, S. C., Webb, M. D., and Peck, M. W. (2013). The type F6 neurotoxin gene cluster locus of group II *Clostridium botulinum* has evolved by successive disruption of two different ancestral precursors. *Genome Biol. Evol.* 5, 1032–1037. doi: 10.1093/gbe/evt068

Collins, M. D., and East, A. K. (1998). Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *J. Appl. Microbiol.* 84, 5–17.

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.

Davis, B. M., and Waldor, M. K. (2002). "Mobile genetic elements and bacterial pathogenesis," in *Mobile DNA II*, ed. N. I. Craig (Washington, DC: ASM Press).

Delcher, A. L., Phillippy, A., Carlton, J., and Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30, 2478–2483. doi: 10.1093/nar/30.11.2478

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806

Dineen, S. S., Bradshaw, M., and Johnson, E. A. (2002). Neurotoxin gene clusters in *Clostridium botulinum* type A strains: sequence comparison and evolutionary implications. *Curr. Microbiol.* 46, 345–352.

Dineen, S. S., Bradshaw, M., Karasek, C. E., and Johnson, E. A. (2004). Nucleotide sequence and transcriptional analysis of the type A2 neurotoxin gene cluster in *Clostridium botulinum*. *FEMS Microbiol. Lett.* 235, 9–16. doi: 10.1016/j.femsle.2004.04.002

Dover, N., Barash, J. R., Hill, K. K., Davenport, K. W., Teshima, H., Xie, G., et al. (2013). *Clostridium botulinum* strain Af84 contains three neurotoxin gene clusters: bont/A2, bont/F4 and bont/F5. *PLoS One* 8:e61205. doi: 10.1371/journal.pone.0061205

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Eklund, M. W., and Poysky, F. T. (1974). Interconversion of type C and D strains of *Clostridium botulinum* by specific bacteriophages. *Appl. Microbiol.* 27, 251–258.

Eklund, M. W., Poysky, F. T., and Boatman, E. S. (1969). Bacteriophages of *Clostridium botulinum* types A, B, E, and F and nontoxigenic strains resembling type E. *J. Virol.* 3, 270–274.

Fernandez, R. A. (1994). *Distribution and Prevalence of Clostridium botulinum in Argentina. Implication in Human and Animal Pathology*. Ph.D. thesis Universidad Nacional de Cuyo, Mendoza.

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

Frean, J., Arntzen, L., van den Heever, J., and Perovic, O. (2004). Fatal type A botulism in South Africa, 2002. *Trans. R. Soc. Trop. Med. Hyg.* 98, 290–295. doi: 10.1016/S0035-9203(03)00069-5

Giordani, F., Fillo, S., Anselmo, A., Palozzi, A. M., Fortunato, A., Gentile, B., et al. (2015). Genomic characterization of Italian *Clostridium botulinum* group I strains. *Infect. Genet. Evol.* 36, 62–71. doi: 10.1016/j.meegid.2015.08.042

Gustafsson, R., Berntsson, R. P., Martinez-Carranza, M., El Tekle, G., Odegrip, R., Johnson, E. A., et al. (2017). Crystal structures of OrfX2 and P47 from a Botulinum neurotoxin OrfX-type gene cluster. *FEBS Lett.* 591, 3781–3792. doi: 10.1002/1873-3468.12889

Hacker, J., Blum-Oehler, G., Muhldorfer, I., and Tschape, H. (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* 23, 1089–1097. doi: 10.1046/j.1365-2958.1997.3101672.x

Halpin, J. L., Hill, K., Johnson, S. L., Bruce, D. C., Shirey, T. B., Dykes, J. K., et al. (2017). Finished whole-genome sequences of *Clostridium butyricum* toxin subtype E4 and *Clostridium baratii* toxin subtype F7 strains. *Genome Announc.* 5:e00375-17. doi: 10.1128/genomeA.00375-17

Hatheway, C. L. (1993). "Bacteriology and pathology of neurotoxigenic clostridia," in *Botulinum and Tetanus Neurotoxins: Neurotransmission and Biomedical Aspects*, ed. B. R. DasGupta (New York, NY: Plenum Press), 491–503.

Henderson, I., Davis, T., Elmore, M., and Minton, N. P. (1977). "The genetic basis of toxin production in *Clostridium botulinum* and *Clostridium tetani*," in *The Clostridia: Molecular Biology and Pathogenesis*, ed. J. Rood (New York, NY: Academic Press), 261–294.

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

Hill, K. K., Xie, G., Foley, B. T., Smith, T. J., Munk, A. C., Bruce, D., et al. (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. doi: 10.1186/1741-7007-7-66

Huang, W., Li, L., Myers, J. R., and Marth, G. T. (2012). ART: a next-generation sequencing read simulator. *Bioinformatics* 28, 593–594. doi: 10.1093/bioinformatics/btr708

Kalb, S. R., Baudys, J., Smith, T. J., Smith, L. A., and Barr, J. R. (2017). Characterization of Hemagglutinin negative Botulinum progenitor toxins. *Toxins* 9:193. doi: 10.3390/toxins9060193

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kenri, T., Sekizuka, T., Yamamoto, A., Iwaki, M., Komiya, T., Hatakeyama, T., et al. (2014). Genetic characterization and comparison of *Clostridium botulinum* isolates from botulism cases in Japan between 2006 and 2011. *Appl. Environ. Microbiol.* 80, 6954–6964. doi: 10.1128/AEM.02134-14

Kull, S., Schulz, K. M., Weisemann, J., Kirchner, S., Schreiber, T., Bollenbach, A., et al. (2015). Isolation and functional characterization of the novel *Clostridium botulinum* neurotoxin A8 subtype. *PLoS One* 10:e0116381. doi: 10.1371/journal.pone.0116381

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

Lam, K. H., Qi, R., Liu, S., Kroh, A., Yao, G., Perry, K., et al. (2018). The hypothetical protein P47 of *Clostridium botulinum* E1 strain Beluga has a structural topology similar to bactericidal/permeability-increasing protein. *Toxicon* 147, 19–26. doi: 10.1016/j.toxicon.2017.10.012

Landmann, G. (1904). Ueber die ursache der Darmstadter bohnenvergiftung. *Hyg. Rundschau* 14, 449–452.

Leighton, G. (1922). *Botulism and Food Preservation (the Loch Maree Tragedy)*. London: W. Collins Sons & Co.

Leuchs, J. (1910). Beitraege zur kenntnis des toxins und antitoxins des *Bacillus botulinus*. *Z. Hyg. Infektionskr.* 76, 55–84.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* [Preprint]. Available online at: http://arxiv.org/abs/1303.3997 (accessed July 20, 2019).

Lin, G., Tepp, W. H., Bradshaw, M., Fredrick, C. M., and Johnson, E. A. (2015). Immunoprecipitation of native botulinum neurotoxin complexes from *Clostridium botulinum* subtype A strains. *Appl. Environ. Microbiol.* 81, 481–491. doi: 10.1128/AEM.02817-14

Lindstrom, M., Hinderink, K., Somervuo, P., Kiviniemi, K., Nevas, M., Chen, Y., et al. (2009). Comparative genomic hybridization analysis of two predominant Nordic group I (proteolytic) *Clostridium botulinum* type B clusters. *Appl. Environ. Microbiol.* 75, 2643–2651. doi: 10.1128/AEM.02557-08

Luquez, C., Bianco, M. I., de Jong, L. I., Sagua, M. D., Arenas, G. N., Ciccarelli, A. S., et al. (2005). Distribution of botulinum toxin-producing clostridia in soils of Argentina. *Appl. Environ. Microbiol.* 71, 4137–4139. doi: 10.1128/AEM.71.7.4137-4139.2005

Luquez, C., Raphael, B. H., Joseph, L. A., Meno, S. R., Fernandez, R. A., and Maslanka, S. E. (2012). Genetic diversity among *Clostridium botulinum* strains harboring bont/A2 and bont/A3 genes. *Appl. Environ. Microbiol.* 78, 8712–8718. doi: 10.1128/AEM.02428-12

Mackay-Scollay, E. M. (1958). Two cases of botulism. *J. Pathol. Bacteriol.* 75, 482–485.

Mad'arova, L., Dorner, B. G., Schaade, L., Donath, V., Avdicova, M., Fatkulinova, M., et al. (2017). Reoccurrence of botulinum neurotoxin subtype A3 inducing food-borne botulism, Slovakia, 2015. *Euro Surveill.* 22:30591. doi: 10.2807/1560-7917.ES.2017.22.32.30591

Marshall, K. M., Bradshaw, M., and Johnson, E. A. (2010). Conjugative botulinum neurotoxin-encoding plasmids in *Clostridium botulinum*. *PLoS One* 5:e11087. doi: 10.1371/journal.pone.0011087

Mazuet, C., Ezan, E., Volland, H., Popoff, M. R., and Becher, F. (2012). Toxin detection in patients' sera by mass spectrometry during two outbreaks of type A botulism in France. *J. Clin. Microbiol.* 50, 4091–4094.

Mazuet, C., Legeay, C., Sautereau, J., Bouchier, C., Criscuolo, A., Bouvet, P., et al. (2017). Characterization of *Clostridium baratii* type F strains responsible for an outbreak of botulism linked to beef meat consumption in France. *PLoS Curr.* 9:ecurrents.outbreaks.6ed2fe754b58a5c42d0c33d586ffc606. doi: 10.1371/currents.outbreaks.6ed2fe754b58a5c42d0c33d586ffc606

Mazuet, C., Legeay, C., Sautereau, J., Ma, L., Bouchier, C., Bouvet, P., et al. (2016). Diversity of Group I and II *Clostridium botulinum* strains from France including recently identified subtypes. *Genome Biol. Evol.* 8, 1643–1660. doi: 10.1093/gbe/evw101

McCallum, N., Gray, T. J., Wang, Q., Ng, J., Hicks, L., Nguyen, T., et al. (2015). Genomic epidemiology of *Clostridium botulinum* isolates from temporally related cases of infant botulism in new South Wales, Australia. *J. Clin. Microbiol.* 53, 2846–2853. doi: 10.1128/JCM.00143-15

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Meyer, K. F., and Gunnison, J. B. (1929). South African cultures of *Clostridium botulinum* and *parabotulinum*. XXXVII with a description of *Cl. botulinum* type D, N. SP. *J. Infect. Dis.* 45, 106–118.

Moller, V., and Scheibel, I. (1960). Preliminary report on the isolation of an apparently new type of *Cl. botulinum*. *Acta Pathol. Microbiol. Scand.* 48:80.

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nikolenko, S. I., Korobeynikov, A. I., and Alekseyev, M. A. (2013). BayesHammer: bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* 14(Suppl. 1):S7. doi: 10.1186/1471-2164-14-S1-S7

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17:132. doi: 10.1186/s13059-016-0997-x

Peck, M. W., Smith, T. J., Anniballi, F., Austin, J. W., Bano, L., Bradshaw, M., et al. (2017). Historical perspectives and guidelines for Botulinum neurotoxin subtype nomenclature. *Toxins* 9:38. doi: 10.3390/toxins9010038

Popoff, M. R. (1995). "Ecology of neurotoxigenic strains of clostridia," in *Clostridial Neurotoxins, The Molecular Pathogenesis of Tetanus and Botulism*, ed. C. Montecucco (Berlin: Springer-Verlag).

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

Ryan, D. P., and Owen-Hughes, T. (2011). Snf2-family proteins: chromatin remodellers for any occasion. *Curr. Opin. Chem. Biol.* 15, 649–656. doi: 10.1016/j.cbpa.2011.07.022

Sagua, M. D., Luquez, C., Barzola, C. P., Bianco, M. I., and Fernandez, R. A. (2009). Phenotypic characterization of *Clostridium botulinum* strains isolated from infant botulism cases in Argentina. *Rev. Argent. Microbiol.* 41, 141–147.

Sahl, J. W., Lemmer, D., Travis, J., Schupp, J. M., Gillece, J. D., Aziz, M., et al. (2016). NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. *Microb. Genom.* 2:e000074. doi: 10.1099/mgen.0.000074

Seddon, H. R. (1922). Bulbar paralysis in cattle due to the action of a toxicogenic bacillus, with a discussion on the relationship of the condition to forage poisoning (botulism). *J. Comp. Pathol. Ther.* 35, 147–190.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Skarin, H., Hafstrom, T., Westerberg, J., and Segerman, B. (2011). *Clostridium botulinum* group III: a group with dual identity shaped by plasmids, phages and mobile elements. *BMC Genomics* 12:185. doi: 10.1186/1471-2164-12-185

Smith, D. H., Timms, G. L., and Refai, M. (1979). Outbreak of botulism in Kenyan nomads. *Ann. Trop. Med. Parasitol.* 73, 145–148. doi: 10.1080/00034983.1979.11687241

Smith, L. D. (1977). *Botulism: The organism, its Toxins, the Disease*. Springfield, IL: Charles C. Thomas.

Smith, T. J., Williamson, C. H., Hill, K., Sahl, J. W., and Keim, P. (2018). Botulinum neurotoxin -producing bacteria - isn't it time we called a species a species? *mBio* 9:e01469-18. doi: 10.1128/mBio.01469-18

Smith, T. J., Xie, G., Williamson, C. H. D., Hill, K. K., Sahl, J. W., Keim, P., et al. (2020). Genomic characterization of newly completed genomes of botulinum neurotoxin-producing species from Argentina, Australia and Africa. *Genome Biol. Evol.* 12, 229–242. doi: 10.1093/gbe/evaa043

Stackebrandt, E., Kramer, I., Swiderski, J., and Hippe, H. (1999). Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*. *FEMS Immunol. Med. Microbiol.* 24, 253–258.

Suen, J. C., Hatheway, C. L., Steigerwalt, A. G., and Brenner, D. J. (1988a). *Clostridium argentinense* sp. nov.: a genetically homogeneous group composed of all strains of *Clostridium botulinum* toxin type G and some nontoxigenic strains previously identified as *Clostridium subterminale or Clostridium hastiforme*. *Int. J. Syst. Bacteriol.* 38, 375–381.

Suen, J. C., Hatheway, C. L., Steigerwalt, A. G., and Brenner, D. J. (1988b). Genetic confirmation of identities of neurotoxigenic *Clostridium baratii* and *Clostridium butyricum* implicated as agents of infant botulism. *J. Clin. Microbiol.* 26, 2191–2192.

van Ermengem, E. (1897). A new anaerobic bacillus and its relation to botulism. E. van Ermengem. Originally published as "Ueber einen neuen anaëroben Bacillus und seine Beziehungen zum Botulismus" in Zeitschrift für Hygiene und Infektionskrankheiten 26: 1-56, 1897. *Rev. Infect. Dis.* 1, 701–719.

Viray, M. A., Wamala, J., Fagan, R., Luquez, C., Maslanka, S., Downing, R., et al. (2014). Outbreak of type A foodborne botulism at a boarding school, Uganda, 2008. *Epidemiol. Infect.* 142, 2297–2301. doi: 10.1017/S0950268814000387

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Weigand, M. R., Pena-Gonzalez, A., Shirey, T. B., Broeker, R. G., Ishaq, M. K., Konstantinidis, K. T., et al. (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

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

Williamson, C. H. D., Vazquez, A. J., Hill, K., Smith, T. J., Nottingham, R., Stone, N. E., et al. (2017). Differentiating botulinum-neurotoxin-producing clostridia with a simple, multiplex PCR assay. *Appl. Environ. Microbiol.* 83:e00806-17. doi: 10.1128/AEM.00806-17

Zhang, Z., Hintsa, H., Chen, Y., Korkeala, H., and Lindstrom, M. (2013). Plasmid-borne type E neurotoxin gene clusters in *Clostridium botulinum* strains. *Appl. Environ. Microbiol.* 79, 3856–3859. doi: 10.1128/AEM.00080-13

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 29 June 2021
doi: 10.3389/fmicb.2021.679377

Asymptomatic Carriage of *C. botulinum* Type D/C in Broiler Flocks as the Source of Contamination of a Massive Botulism Outbreak on a Dairy Cattle Farm

*Rozenn Souillard[1], Daniel Grosjean[2], Thibault Le Gratiet[3], Typhaine Poezevara[3], Sandra Rouxel[3], Loïc Balaine[1], Sabrina Macé[3], Laure Martin[3], Fabrizio Anniballi[4], Marianne Chemaly[3], Sophie Le Bouquin[1] and Caroline Le Maréchal[3]**

[1] ANSES, French Agency for Food, Environmental and Occupational Health Safety, Epidemiology, Health and Welfare Unit, Ploufragan, France, [2] DDCSPP de la Meuse, Departmental Authority in Charge of Veterinary Services for Meuse Department, Bar-le-Duc, France, [3] ANSES, French Agency for Food, Environmental and Occupational Health Safety, Hygiene and Quality of Poultry and Pig Products Unit, Ploufragan, France, [4] Department of Veterinary Public Health and Food Safety, National Reference Centre for Botulism, Istituto Superiore di Sanità, Rome, Italy

OPEN ACCESS

Edited by:
Javier Carballo, University of Vigo, Spain

Reviewed by:
Krzysztof Skowron, Nicolaus Copernicus University in Toruń, Poland
Frederique Pasquali, University of Bologna, Italy

***Correspondence:**
Caroline Le Maréchal
caroline.lemarechal@anses.fr

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *11 March 2021*
Accepted: *07 June 2021*
Published: *29 June 2021*

Citation:
Souillard R, Grosjean D, Le Gratiet T, Poezevara T, Rouxel S, Balaine L, Macé S, Martin L, Anniballi F, Chemaly M, Le Bouquin S and Le Maréchal C (2021) Asymptomatic Carriage of C. botulinum Type D/C in Broiler Flocks as the Source of Contamination of a Massive Botulism Outbreak on a Dairy Cattle Farm. Front. Microbiol. 12:679377. doi: 10.3389/fmicb.2021.679377

In winter 2018, a massive type D/C cattle botulism outbreak occurred on a mixed dairy and broiler farm in France. An investigation was conducted based on the hypothesis of asymptomatic carriage in poultry. We set out to identify the source of contamination of the dairy cattle and to monitor the contamination of broilers over time, including the hatchery delivering chicks to the farm. Environmental samples were collected on the farm during the cattle outbreak (n = 40), after the outbreak for three successive broiler flocks (n = 128), and once in the hatchery delivering the chicks (n = 58). These samples were analyzed using real-time PCR after an enrichment step to detect *Clostridium botulinum* type D/C. The results showed contamination in the manure from the broilers raised just before the onset of the cattle outbreak (5 + /5), as well as in some of the components of the cattle ration (3 + /17). This latter contamination is likely due to the use of the same tractor bucket to remove litter from the poultry house and to prepare the cattle ration on the same day. Contamination monitoring over several months revealed continuous asymptomatic carriage in the broilers (4 + /20 and 17 + /20 cloacal swabs in 2 successive flocks), a persistence of *C. botulinum* type D/C in the ventilation system of the poultry house (8 + /14), and contamination of the equipment coming from the hatchery used for delivering the chicks (3 + /18). Further investigations conducted in the hatchery demonstrated contamination in the hatchery by *C. botulinum* type D/C (6 + /58). Comparison of samples using a multilocus variable number tandem repeat analysis showed the same profile for samples collected on broilers, cattle and in the hatchery. This study highlighted the crucial role of the implementation of biosecurity measures in mixed farms to avoid cross-contamination between production units given the potential asymptomatic carriage of poultry. This study also revealed the contamination of the poultry hatchery. Further investigations are required to better understand the role of hatcheries in the epidemiology of animal botulism.

Keywords: botulism, cattle, poultry, epidemiology, hatchery, investigation, MLVA, PCR

INTRODUCTION

Botulism is a severe neurological disease caused by botulinum neurotoxins (BoNT) that prevent the release of acetylcholine at synaptic junctions and result in progressive symmetrical flaccid paralysis of muscles. There are nine different BoNTs (A, B, C, D, E, F, G, H, or H/A, or F/A, X) (Rasetti-Escargueil et al., 2020) and more than 40 subtypes have been described (Peck et al., 2017). Human botulism is a rare disease is mainly caused by BoNTs A, B, E, and to a lesser extent, F (Rasetti-Escargueil et al., 2019). Botulism is more common in animals than in humans and results in high mortality rate, raising significant animal welfare and economic concerns (Anniballi et al., 2013; Relun et al., 2017; Rasetti-Escargueil et al., 2019). Avian botulism is generally associated with BoNT C/D, whereas bovine botulism is more frequently associated with BoNT D/C, and to a lesser extent, BoNT C (Woudstra et al., 2012; Bano et al., 2017; Le Gratiet et al., 2020). Considering the serious consequences of botulism on bovine and avian species, a better understanding of this disease — particularly in terms of potential mechanisms of transmission — is crucial to improve prevention and management of animal botulism outbreaks in an efficient manner.

Poultry litter has been considered as a major source of contamination for cattle botulism outbreaks via contact or close proximity (Popoff, 1989; Payne et al., 2011; Relun et al., 2017). Cross-contamination from poultry to cattle has been widely reported in the literature (Senturk and Cihan, 2007; Payne et al., 2011; Ramírez-Romero et al., 2014; Relun et al., 2017; Souillard et al., 2017). Poultry is indeed considered as a reservoir and source of amplification of type D *Clostridium botulinum* and its toxin (Popoff and Argente, 1996). Poultry litter can be used as fertilizer, animal bedding or even as feed supplements (Payne et al., 2011). Surprisingly, no data is available on the prevalence of *C. botulinum* or BoNTs in poultry litter or more generally in healthy poultry. Recently, healthy carriage on a poultry farm was suspected as a source of two cattle botulism outbreaks due to a transfer of poultry manure to the cattle farms (Souillard et al., 2017). Consequently, asymptomatic carriage of *C. botulinum* in poultry can occur, which may represent a reservoir of *C. botulinum* (Rasetti-Escargueil et al., 2019). Animals are either resistant to some BoNT types or *C. botulinum* carriage occurs at low bacterial loads in the digestive tract (Rasetti-Escargueil et al., 2019). These low levels may explain the failure to detect *C. botulinum* carriage in previous studies, as they may be below the limit of detection of available methods (Popoff, 1989).

In 2018, a large BoNT type D/C cattle botulism outbreak occurred on a farm with both dairy and broiler production units in eastern France. Based on the hypothesis of asymptomatic *C. botulinum* carriage in poultry, the objectives of this study were (i) to identify the source of cattle botulism contamination using epidemiological investigations and strain tracking and (ii) to monitor the contamination of broilers over time including at the hatchery delivering chicks to the farm.

MATERIALS AND METHODS

Case History and Diagnosis of Botulism on the Farm

A botulism outbreak occurred in winter 2018 on a farm with dairy and broiler production in Meuse department, located in eastern France. **Figure 1** shows the different barns of the farm. Ninety-two dairy cows were being housed in Barn B and 25 heifers, calves and dry cows in Barn C. These two barns are separated by about 20 m. During winter, cattle are kept indoors and, beginning early spring, grazed on pasture. The Barn A poultry house is separated by about 10 m from Barn B. It is a "Louisiane"-style barn of 1000 m^2 with windows, natural and transversal ventilation, and a dirt floor. In this house, a flock of 20 000 broilers (Flock No. 1) was reared in January 2018. Moreover, the farm has two sheds: Shed A for the storage of the cattle ration ingredients and Shed B for straw storage. In addition, corn silage is also stored on a platform near Shed B. A second site for cattle production is located about 10 km away from this area.

On January 29, the first clinical signs of paralysis were observed in dairy cows in Barn B and mortality started (**Table 1**). On 30 January, samples were collected on two cows that died after showing clinical signs suggestive of botulism. The liver, gall bladder, rectal contents, ruminal contents, feces, and intestines were frozen before shipment to the laboratory. *C. botulinum* types C, D, C/D, and D/C in these samples were screened for as previously described (Le Maréchal et al., 2019). Ruminal contents from the two investigated cows as well as rectal contents from one cow were positive for *C. botulinum* type D/C using real-time PCR after enrichment in trypticase peptone-glucose-yeast extract (TPGY) broth. None of the samples were positive when using fortified-cooked meat medium (F-CMM) with a thermal treatment (70°C, 10 min). On 31 January, signs of paralysis and mortality also occurred in the heifers in Barn C. No clinical signs and no mortality were observed on the second site 10 km away. Vaccination using Ultravac® Botulinum (Zoetis, France) was performed on February 2 with a second injection on March 5 in both cattle barns.

Two phases of death were observed on the farm. Mortality was high within the first week (January 29 to February 5) with 68 dead cows in Barn B and 8 heifers in Barn C. A second mortality phase was observed from February 26 to March 5 with 15 cows and 1 heifer. Out of the 117 cows present on the farm at the beginning of the outbreak, 92 cows died, indicating a mortality rate of 79%. All animals with clinical signs eventually died (naturally or humanely euthanized), no recovery was reported.

Epidemiological Investigation and Collection of Environmental Samples on the Farm

An epidemiological investigation was conducted on the farm. A questionnaire was filled out with the farmer and the veterinarian to describe the history and chronology of the cases. It included questions on the overall management of cattle and poultry (farming practices and manure management, etc.) and

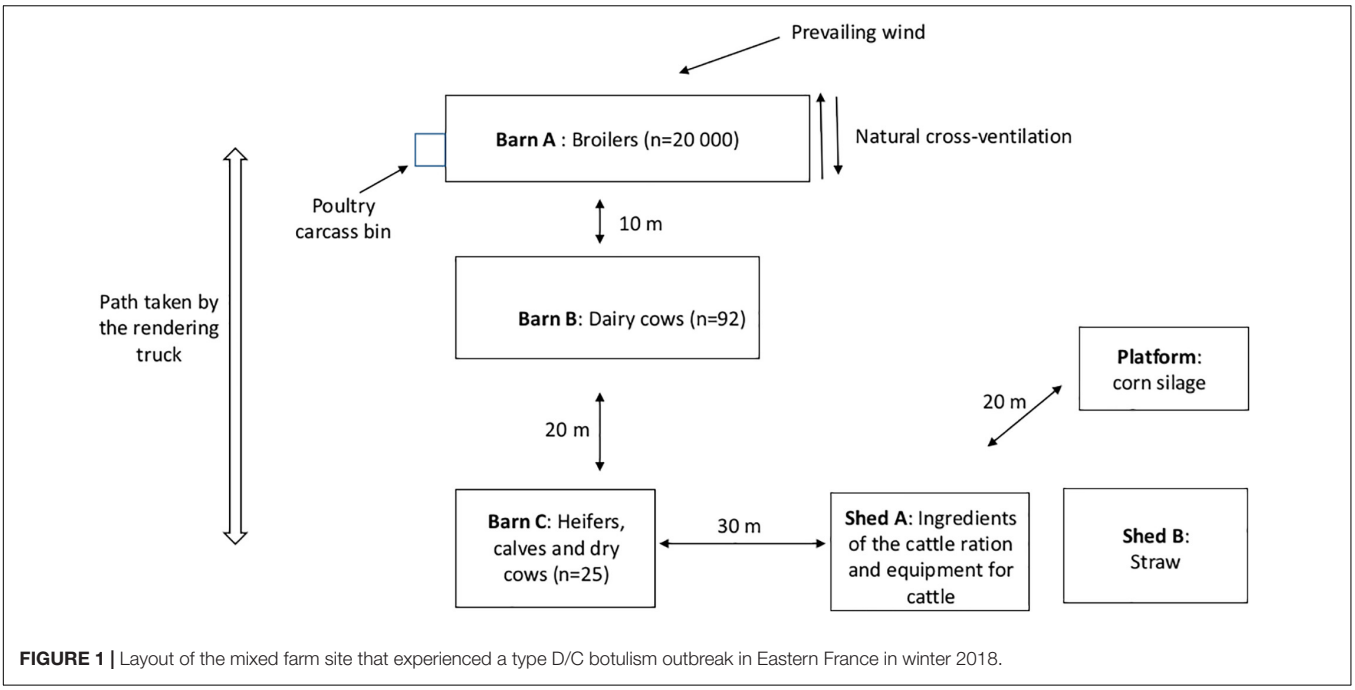


FIGURE 1 | Layout of the mixed farm site that experienced a type D/C botulism outbreak in Eastern France in winter 2018.

the movements of material, personnel, and vehicles between the poultry and dairy production areas.

During the outbreak, environmental samples (n = 40) were collected on the farm on three different days (January 30, February 5, and February 23) to investigate the source of contamination of the cattle (**Table 1**): 17 samples of the cattle ration stored in Shed A (n = 14) and on the platform (n = 3), 5 samples of the broiler manure from Flock No. 1 stored in a field since January 26, 3 samples in Barn B, 5 samples in Shed A and 10 samples in Barn A and its surroundings.

After the end of the outbreak, samples (n = 128) were also collected for several months in the poultry house to monitor the contamination of broilers over time (**Table 2**): 9 swabs after cleaning and disinfecting the house with 7 swabs inside the house and 2 swabs outside the house; 55 swabs on the day of chick delivery to the farm, including 11 swabs inside the house; 2 swabs outside the house, 20 cloacal swabs on chicks, 18 swabs on hatchery-origin material, and 4 swabs on the material returned to the hatchery; and finally 64 swabs at the end of the rearing period, with 4 swabs inside the poultry house and 60 cloacal swabs of broilers on the three following broiler flocks (Flocks No. 3, 4, and 5). It was not possible to sample the second flock of broilers reared just after Flock No. 1. Swabs used for sample collection in the farm were obtained from Sodibox (Nevez, France). Cloacal swabs were collected by veterinarians in accordance with the European and French regulation on farmed animal protection.

Epidemiological Investigation and Collection of Environmental Samples in the Hatchery

An epidemiological investigation was also conducted at the hatchery providing the chicks. A questionnaire was filled out with the hatchery owner to collect information on the hatchery operations (from egg hatching to the departure of chicks, cleaning and disinfection, waste management, and vehicle movements on the site, biosecurity measures, etc.). A breeder house was also located on a site near the hatchery. At the hatchery site, samples (n = 58) were collected (swabs taken on the walls, floor and equipment): 41 inside the hatchery (egg receiving room, incubation and hatcher rooms, chick sorting, and departure rooms), 9 outside the hatchery (truck platform, waste containers, central area and breeder house surroundings) and 8 in annex rooms (trolley storage rooms and refrigerated egg holding rooms).

Culture Conditions, DNA Extraction, and Real-Time PCR

Enrichment of 226 samples, DNA extraction and PCR were performed as previously described (Le Maréchal et al., 2019).

Multilocus Variable-Number of Tandem-Repeat Analysis (MLVA)

Thirteen DNA extracts positive for type D/C (with a Ct below 35) were selected for MLVA analysis using nine conventional PCRs (**Supplementary Table 1**), one for each variable number tandem repeat (VNTR) locus as described previously in Auricchio et al. (2019) and Scalfaro et al. (in preparation)[1]. The MLVA PCR mixture contained 10 μL of HotStarTaq Master mix (Qiagen, Courtaboeuf, France), 1 μL of forward and reverse primers (10 μM for a final concentration of 500 nM), 7 μL of water and 1 μL of template DNA. Amplification consisted of the following

[1]Scalfaro, C., Vicenza, T., Le Maréchal, C., Bano, L., Bilei, S., Chemaly, M., et al. (in preparation). Multilocus variable-number of tandem-repeat analysis as a tool for *Clostridium botulinum* group III sub-typing. *Microorganisms*

TABLE 1 | Chronology of the events on the farm affected by the cattle botulism outbreak type D/C and detection of *C. botulinum* type D/C during the outbreak.

	22/01/18	**26/01/18**	**27-28/01/18**	**29/01/18**	**30/01/18**	**31/01/18**	**02/02/18**	**05/02/18**	**23/02/18**	**05/03/18**
Chronology of the outbreak events	Departure of broilers (Flock No 1 Barn A) for slaughter	Removal of broiler litter using a tractor bucket and storage of manure in a field Transfer of the ingredients of the cattle ration (Shed A) with the same bucket in a mixing wagon to feed cows.	Heifers fed with leftovers from the cattle ration Cleaning and disinfection of the broiler house (Barn A)	Detection of paralysis and first cow mortalities (Barn B)	Diagnosis of botulism: *C. botulinum* type D/C in ruminal contents **(S3)** (2 + /2)* and rectal content (1 + /2)*	Paralysis and heifer mortalities (Barn C)	Vaccination of all animals (including calves, dry cows) with Utravac Botulinum (first injection)			Vaccination with Ultravac Botulinum (second injection)
Detection of *C. botulinum* type D/Con the farm* (n = 40)					Cattle ration Shed A (*n* = 3) Wrapped grass 1 + /3			**Cattle ration Shed A (*n* = 11)** Meslin (S1) **1 + /3** Rape **1 + /3** Brewery grains 0 + /3 Corn 0 + /2 **Cattle ration Platform (*n* = 3)** Corn silage 0 + /3 **Broiler manure stored since 26/01 (*n* = 5)** Manure (S2) in a field **5 + /5**	**Barn B-(*n* = 3)** Liquid manure **1 + /1** Swab of the floor of the stall **1 + /1** Swab of cattle feed table 0 + /1 **Shed A (*n* = 5)** Swab of the floor of the shed 0 + /1 Swab of the bucket **1 + /1** Swab of the tarpaulin covering rape **1 + /1** Swab of the rape storage area 0 + /1 Swab in the mixing wagon **1 + /1** **Barn A (n = 10)** *Inside the house* Swab of the ventilation system **2 + /4** Swab of the floor of the house 0 + /1 Swab of the shower? room 0 + /1 Feed from silo 0 + /1 *Outside the house* Swab of the surroundings 0 + /1 Swab of the carcass bin **1 + /1** Swab of the path 0 + /1	

**Number of positive samples/number of samples collected.*

Samples tested by MLVA are underlined ***and the sample code in Figure 2*** *is indicated in* ***parentheses****. Samples detected positive are in bold.*

TABLE 2 | Detection of *C. botulinum* type D/C in the poultry house after the cattle outbreak (*n* = 128) using swab samples.

		After extensive disinfection of the poultry house *n* = 9		**Upon chick delivery *n* = 55**			**Upon broiler departure *n* = 64**		
Flock No 1*	21/03/18	Inside the house *n* = 7	Ventilation system (S5) **2 + /4** Floor of the house 0 + /1 Changing room 0 + /1 Feed silo 0 + /1						
		Outside the house *n* = 2	Carcasses bin (S4) **1 + /1** Surroundings 0 + /1						
Flock No 2*						ND			ND
Flock No 3*				22/05/18	Inside the house *n* = 7	Ventilation system **2 + /4** Floor of the house 0 + /1 Changing room 0 + /1 Feed silo 0 + /1	13/06/18	Inside the house *n* = 1	Floor 0 + /1
					Outside the house *n* = 2	Surroundings 0 + /1 Carcasses bin 0 + /1		Broilers *n* = 20	Cloacal swabs **4 + /20** (S6 and S7)
Flock No 4*				9/07/18	Inside the house *n* = 2	Ventilation system 0 + /2	6/08/18	Inside the house *n* = 3	Ventilation system **2 + /2** Floor **1 + /1**
					Material from the hatchery *n* = 2	Chick box bottom **2 + /2**		Broilers *n* = 20	Cloacal swabs **17 + /20** (S11 to S13)
Flock No 5*				27/08/18	Inside the house *n* = 2	Ventilation system **2 + /2**	24/09/18	Broilers *n* = 20	Cloacal swabs 0 + /20
					Material from the hatchery *n* = 16	Chick box bottom 0 + /2 Article of the box bottom 0 + /10 10 chick boxes upon arrival 0 + /2 Trolley and its wheels upon arrival **1 + /2** (S8)			
					Chicks from the hatchery *n* = 20	Cloacal swab on chicks 0 + /20			
					Material returned to the hatchery *n* = 4	10 chick boxes upon departure 0 + /2 Trolleys and their wheels upon departure 0 + /2			

ND, no data.
**Number of positive samples/number of samples collected (positive samples are in bold).*
Samples tested by MLVA are underlined ***and the sample code in Figure 2*** *is indicated in* ***parentheses.***

cycle program: 95°C for 15 min, 35 cycles of 30 s at 94°C, of 1 min at 55°C, of 1 min at 72°C, and one final cycle at 72°C for 5 min.

Multilocus variable-number of tandem-repeat analysis typing was performed on a T100 ThermalCycler (Bio-Rad, Marnes-la-Coquette, France). PCR products were analyzed by capillary electrophoresis using a 2100 bioanalyzer system (Agilent, Les Ulis, France) to determine the number of repeats for each VNTR locus, deduced from each PCR product size. Fragment length and the corresponding number of repeat units was also checked by sequencing on a 3500 Series Genetic Analyzer (Applied Biosystems, Thermo Fischer Scientific, Illkirch-Graffenstaden, France).

The number of repeats obtained from each VNTR locus was imported into Bio-Numerics version 7.5 (Applied Math, Sint-Martens-Latem, Belgium) as character values. A dendrogram was calculated using the categorical coefficient and the UPGMA clustering algorithm to compare DNA extracts from samples tested in our study and available profiles from a previous study (Auricchio et al., 2019) or from other animal botulism outbreaks (**Supplementary Table 2**).

RESULTS

Chronology of the Events in the Mixed Farm

The chronology of the events is given in **Table 1**. Broiler Flock No. 1 was slaughtered on January 22, one week before the onset of the first clinical signs in cows. During their rearing period, an unusual event took place at 15 days of age: one side of the house collapsed due to a storm and caused the death of 120 broilers. The total mortality of this flock was 923 out of 20 000 broilers (4.6%). Otherwise, nothing unusual was reported regarding this flock nor for previous ones.

On January 26, poultry litter was removed from the house using a tractor bucket and manure was then stored in a field away from the farm (5 km from the farm). The same bucket was used later on the same day (after a quick rinse using a pressure cleaner with cold water) to prepare the cattle ration. Each ingredient of the cattle ration was taken one by one using the bucket and transferred to the mixing wagon to feed the cows later in the evening. Rape was taken first, then wrapped grass and meslin (all stored in Shed A as illustrated in **Figure 1**), then corn silage and finally brewery grains. Two days later, heifers were fed with the leftover dairy cattle ration. The day following the removal of poultry litter, Barn A was cleaned and disinfected as usual, with lime spread on the floor of the house and in the surroundings and a quaternary ammonium disinfectant sprayed (Spectragen®, Synthèse Élevage, France) in the house.

On January 29, three days after the distribution of the ration, paralysis and the first mortalities were noticed in the dairy cows in Barn B and the diagnosis of botulism type D/C was confirmed the next day. Two days later, on January 31, the same clinical signs occurred in heifers in Barn C.

The same farm personnel and equipment used for poultry and cattle were employed for both production units. Noteworthily, the poultry carcass bin was located next to the changing room of the poultry house and the rendering truck must drive by the cattle barns to reach the poultry house to collect poultry carcasses from the bin (**Figure 1**).

At the second cattle site 10 km away, no clinical signs were observed. This site was handled by the same farm personnel, using different equipment.

Detection of *C. botulinum* Type D/C on the Farm

During the outbreak, 40 samples were collected at different times to investigate the source of contamination of cattle on the farm (**Table 1**). *C. botulinum* type D/C was detected in the five samples collected on February 5 from Flock No. 1 manure stored in a field since January 26 as well as three ingredients of the cattle ration (number of positive samples/number of samples tested, 3 + /17), i.e., meslin, wrapped grass, and rape. Wrapped grass was sampled at the same time as the samples for botulism diagnosis on January 30, whereas meslin and rape were sampled on February 5, several days after the initiation of botulism outbreak on the farm. Investigations conducted on February 23 revealed two positive samples in Barn B (liquid manure and swab of the floor), three positive samples in Shed A (swabs from the bucket, the mixing wagon and a rape tarpaulin) and three positive samples in the poultry house (two swabs from the ventilation system and one swab from the carcass bin). Given this detection of *C. botulinum* type D/C in the poultry house, extensive decontamination was implemented on March 19, with a strong cleaning agent (Decagen® detergent, Synthèse Élevage, France), with disinfection (Virugen®, Synthèse Élevage, France) in the house and quicklime and caustic soda on the floor) and another disinfection using formaldehyde, with both procedures being carried out following the safety measures recommended when using such products.

After the outbreak, the 128 samples collected to monitor the contamination of broilers over time (**Table 2**) revealed the detection of *C. botulinum* type D/C in the ventilation system of the poultry house (8 + /14) (after the extensive disinfection of the house: 2 + /4; upon chick delivery of the following flocks, 4 + /8; and at the end of the rearing period, 2 + /2), contamination of some equipment coming from the hatchery (3 + /18) (2 swabs on chick box bottom and 1 swab of the egg transport trolleys and trolley wheels) and a continuous healthy carriage of broilers for several months detected in two consecutive flocks (4 + /20 and 17 + /20 cloacal swabs in two flocks and 1 swab of the floor litter). Moreover, the carcass bin was still detected positive after the extensive disinfection of the house.

Detection of *C. botulinum* Type D/C in the Poultry Hatchery

Detection of *C. botulinum* at the hatchery is presented in **Table 3**. *C. botulinum* was detected in 6 out of the 58 samples: in the machine that cuts the article chick box, in the container of sludge washing water, on the loading platform, in the storage rooms

TABLE 3 | Detection of *C. botulinum* type D/C at the hatchery delivering chicks on the farm (*n* = 58).

	Sampling location	Detection of *C. botulinum* type D/C
Inside the hatchery * *n* = 41	Egg receiving room	0 + /4
	Incubation room	0 + /12
	Transfer (to hatcher) room	0 + /2
	Hatcher room	0 + /11
	Changing room	0 + /3
	Chick sorting room	**1 + /5** Swab of the machine cutting the article box containing the chicks
	Chick departure room	0 + /4
vOutside the hatchery * *n* = 9	Trucks platforms	**1 + /3** Swab of the loading platform
	Waste containers	**1 + /3** Swab of the container of sludge washing water
	Central area	0 + /1
	Breeder house surroundings	0 + /2
Annex room* *n* = 8	Wash room (uncleaned egg trolleys)	**1 + /3** Swab of the room of the uncleaned egg trolleys (S9)
	Clean room (cleaned egg trolleys)	**2 + /3** Swab of the room of the cleaned egg trolleys
	Egg holding (cooler) room	and swab of the trolley wheels (S10)
		0 + /2

Number of positive samples/number of samples collected. Samples tested by MLVA are underlined* ***and the sample code in Figure 2* *is indicated in* ***parentheses****. Samples detected positive are in bold.*

of the uncleaned and cleaned egg trolleys, and on the wheels of the cleaned egg trolleys of hatching eggs. In addition, the epidemiological investigation revealed that the rendering truck entering the site to collect hatchery waste (silo of feathers and shells, breeder carcasses) passes hatchery vehicles or equipment (chick or egg delivery trucks, egg trolleys wheeled outside to be stored in annex rooms).

MLVA Analysis

Thirteen samples positive for *C. botulinum* type D/C were selected among the 291 samples analyzed in our study. Samples were selected so as to obtain an overview of the different investigated areas included in our study and to be as representative as possible of the encountered samples and situations, and finally to compare and evaluate the relatedness of strains detected in the different positive samples. Unfortunately, MLVA results obtained with samples initially detected positive for type D/C with a late Ct (above 35) using real-time PCR were not interpretable (such as for example the swab on the bucket).

Samples selected for MLVA analysis are indicated in **Tables 1**–**3** and detailed number of detected repeats for each VNTR locus in **Supplementary Table 3**.

A dendrogram was generated based on the VNTR repeat unit profiles (**Figure 2**). Two different profiles were detected among the selected samples. Three cloacal swabs collected in Flock No. 4 presented a profile different from the other samples (76.8% of similarities between the two groups). These results show that at least two different strains were detected during these investigations with the detection in Flock No. 4 broilers with a profile different from the one detected during the cattle outbreak. However, no differences could be detected between the other investigated samples (Flock No. 1 manure, ruminal contents from a dead cow, cloacal swabs collected in Flock No. 3, samples from the hatchery), demonstrating the presence of the same strain in all these samples.

DISCUSSION

An overview of the detection of *C. botulinum* type D/C on the farm (poultry and cattle barns) and in the hatchery as well as the reference of samples that were compared using MLVA is provided in **Figure 3**. This figure also presents the most likely scenario of the contamination pathways suggested by our investigations.

Identification of the Source of Cattle Contamination Through Epidemiological Investigations

Poultry manure has been reported to be a major source for cattle botulism outbreaks (Payne et al., 2011; Ramírez-Romero et al., 2014; Relun et al., 2017; Souillard et al., 2017). Poultry manure is also incriminated in this study, based on the chronology of the outbreak events, and our results showing positive samples (particularly the poultry manure samples) as well as MLVA profiles similar between the manure and the cows that suffered from botulism (Group 1, **Figure 2**). The positive manure samples from poultry Flock No. 1 (5 + /5) as well as the detection of *C. botulinum* in three samples collected within the poultry house (2 swabs of the ventilation system and 1 swab of the carcass bin) strongly suggest that *C. botulinum* type D/C was present in the poultry house when Flock No. 1 was present.

The tractor bucket used to remove broiler litter from the poultry house and then shortly after used to prepare the cattle ration was also positive for *C. botulinum* type D/C and may have transferred the contamination from poultry manure to the ration ingredients (3 + /17) and to the mixing wagon, which was positive as well. The three ingredients of the ration first taken with the bucket were all detected positive (rape, wrapped grass, and meslin). The MLVA results of the meslin moreover showed the same profile as Flock No. 1 manure and ruminal contents from a dead cow. The succession of these events seems therefore to have resulted in the contamination of dairy cattle fed with this ration, thereby initiating the outbreak. Ration leftovers were then

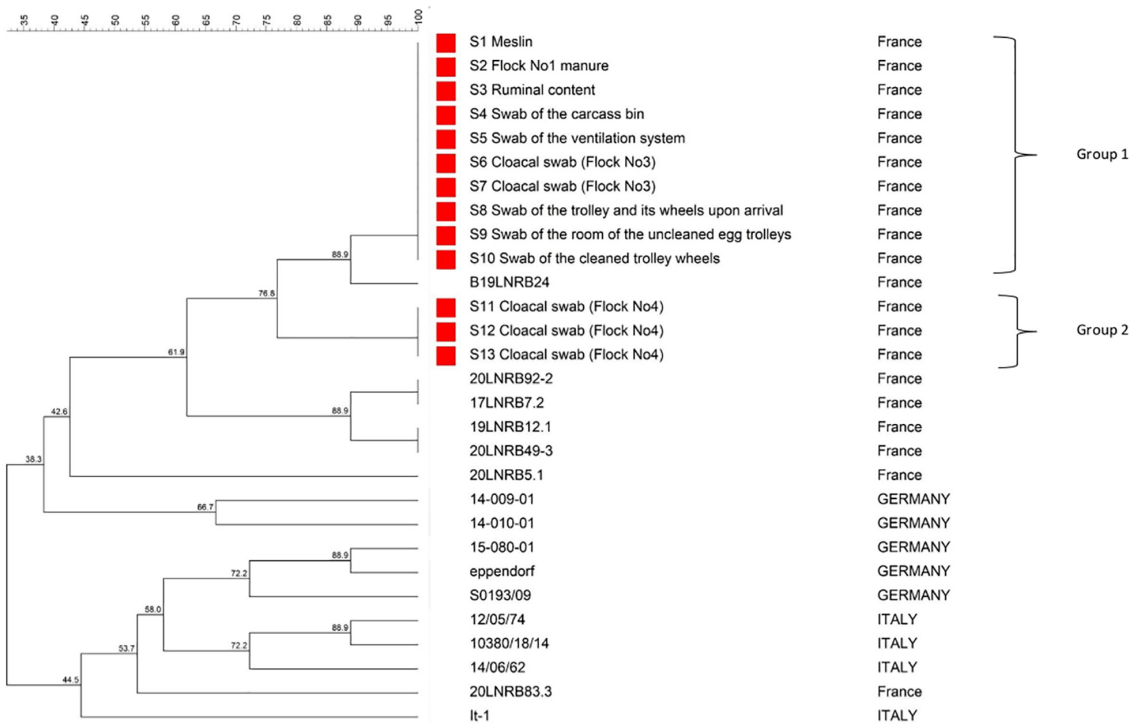


FIGURE 2 | Multilocus variable-number tandem repeat analysis (MLVA) dendrogram of 13 samples collected in this study and positive for *C. botulinum* type D/C (in red) and 16 other strains of *C. botulinum* type D/C (typed in other studies) constructed using the UPGMA method, implemented in BioNumerics. The name and code of samples from our study and names of strains from previous studies and their respective country are shown on the right. Samples S1–S13 were separated in 2 groups based on MLVA similarities.

distributed to heifers resulting in the onset of the disease in these animals as well. It is noteworthy that these samples were collected after the outbreak, except for the wrapped grass, therefore cross-contamination secondary to the outbreak cannot be completely ruled out. However, regarding the timing of the events and the analysis results, this seems unlikely. MLVA results also strongly support this scenario, the same profile being detected in Flock No. 1 manure, ruminal contents from a cow, ration ingredients as well as in samples collected in the poultry house after the Flock No. 1 rearing period.

An unusual event that occurred during the Flock No. 1 growth period may explain the manure contamination. The sudden death of 120 broilers due to the collapse of one side of the house during the rearing period may have increased the risk of carcasses being left in the litter despite daily carcass removal by the farmer. Carcasses as a decaying organic matter harboring high amounts of protein substrates and anaerobic conditions is known to support the growth of *C. botulinum* and BoNT production (Anniballi et al., 2013). Any stress factor in broilers can disturb the balance of the intestinal ecosystem (Tsiouris, 2016); therefore this event may have induced stress in broilers thereby providing favorable conditions for *C. botulinum* growth by modifying intestinal balance. The speed of the onset of clinical signs, 2 days after the distribution of contaminated ration, may be compatible with intoxication of the cattle by ingestion of preformed BoNTs. This hypothesis was supported by the detection of vegetative cells only, and not spores in the ruminal contents of two cows sampled at the beginning of the outbreak. Ruminal contents were only positive when analyzed using TPGY and not F-CMM, which includes a heat treatment.

In addition to this scenario, several biosecurity failures were also highlighted during the investigation. First, the storage conditions of the cattle ration ingredients were not appropriate, because they were stored uncovered, exposing them to potential contamination by wild birds or rodents. Providing safe, high-quality, and properly stored feed to animals is one of the key measures to minimize the risk of botulism (Anniballi et al., 2013). Rodents may be a reservoir of a variety of pathogens, in particular *C. botulinum* (Popoff, 1995; Popp et al., 2012; Skarin et al., 2013; Souillard et al., 2014). Moreover, the location of the rendering container (carcass bin), testing positive on two occasions, next to the poultry house can be a source of poultry contamination. Its location also obliges the rendering truck to drive alongside the cattle barns. Given the risk generated by carcasses in regard to botulism and the ability of spores to persist in the environment (Popoff, 1995), the management of the carcasses on a farm is a major critical point for the prevention of the disease. Third, equipment is also shared between the dairy and poultry production units, which can result in cross-contamination between the units. Finally, close proximity of poultry and dairy cows (less than 10 m between both barns with transversal ventilation in the poultry house) may also be a source of *C. botulinum* dissemination via dust and wind. Dust inside the ventilation system in the poultry house is frequently

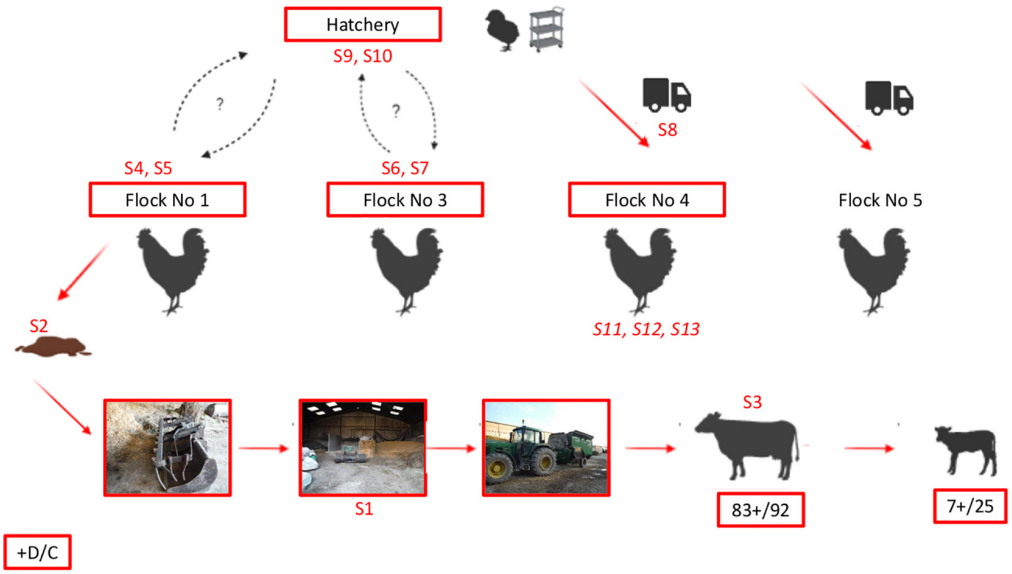


FIGURE 3 | Diagram summarizing the likely relationships between the broilers, cattle, and hatchery based on investigations and results obtained in this study. Samples collected during the study and positive for *C. botulinum* type D/C are shown in red. References of samples tested using MLVA (**Figure 2**) are mentioned in the diagram. Note that samples S11–S13 are shown in italics in contrast to the other samples (S1–S10), because they showed a different MLVA profile (**Figure 2**). " + D/C": sample positive for *bont* D/C using PCR are surrounded in Red. Detailed results of samples collected and analyzed in our study are given in **Table 1** (initial investigations during cattle botulism outbreak), **Table 2** (monitoring of successive broiler flocks), and **Table 3** (investigations in the hatchery that delivers chicks to the farm).

contaminated after a botulism outbreak (Souillard et al., 2014). Windborne transmission of spore- or BoNT- contaminated material has also been suggested (Hogg et al., 2008).

Since this outbreak, specific measures have been implemented to prevent recurrence of the disease: animals are vaccinated yearly, feed storage area has been reorganized, each component being now separated and protected; the bucket used to load feed in the mixing-wagon is now dedicated only to this activity, the tractor used for poultry manure is no longer used for cows feed; boots and clothes used for poultry and bovines by the farmer are now separated and dedicated to one activity. No new case has been reported since then.

Monitoring of Broilers Contamination Over Time

Few data are available regarding healthy carriage of *C. botulinum* in poultry and it remains a current issue to better understand the onset of animal botulism (Rasetti-Escargueil et al., 2019). Our monitoring of broilers on this farm using cloacal swabs revealed continuous healthy carriage of *C. botulinum* in successive broiler flocks over several months. After the first detection of *C. botulinum* in the poultry house in January, the following broiler flocks remained healthy carriers of *C. botulinum* at least until August as demonstrated by positive cloacal swabs detected at the end of the rearing period for two flocks.

At least two scenarios can be considered to explain this persistent contamination of broilers over time on the farm. First, our monitoring showed persistence of *C. botulinum* type D/C in the house, particularly in the ventilation system (8 + /14), despite the disinfection operations conducted between each flock. This detection is likely due to the high resistance of spores in the environment that can survive for several years (Notermans et al., 1981; Wobeser et al., 1987), particularly in the critical areas of the poultry house that are difficult to disinfect, such as the ventilation systems, as reported previously (Souillard et al., 2014, 2016). Consequently, this environmental persistence of *C. botulinum* in the poultry house is a source of recontamination and may have resulted in healthy carriage in successive broiler flocks. A second scenario that may explain *C. botulinum* carriage in broilers for several months is the re-introduction of *C. botulinum* via chick delivery, suggested by the detection of *C. botulinum* on equipment from the hatchery. Two swabs collected on chick box bottoms for Flock No. 4 and one swab collected on trolleys and their wheels for Flock No. 5 were already positive for *C. botulinum* type D/C upon their arrival on the farm, before any contact with the farm. The same MLVA profile was detected for sample S8 (see **Figure 2**) as previous samples collected in the farm (**Table 2** and **Figure 3**). Considering these findings, the question of a potential contamination of the hatchery thus arose and investigations were implemented in the hatchery delivering chicks to the farm to better explore this hypothesis.

Evaluation of Hatchery Contamination

Clostridium botulinum type D/C was detected in 6 of the 58 samples collected in the hatchery: on materials (machine cutting the article box containing the chicks and wheels of trolleys transporting the hatching eggs), in the annex room (room where trolleys of hatching eggs are stored) and on

outside surroundings (in the container of sludge washing water and on the platform where hatching eggs are unloaded). Contaminated areas were either located outside or in areas that are recognized as difficult to clean and disinfect. One sample positive for *C. botulinum* type D/C tested using MLVA showed the same profile (Group 1) as samples collected during the outbreak, demonstrating that the same strain was detected in both places.

During the hatchery visit and based on detection of *C. botulinum* type D/C in some samples, several risk factors were identified. First, similar vehicle routes were highlighted, in which rendering trucks and hatchery vehicles or equipment, for example, pass each other. Failures in cleaning and disinfection operations were also pointed out, as illustrated by the detection of *C. botulinum* type D/C on the system used to cut article boxes during chick delivery or in the room used to store cleaned trolleys or on wheels of cleaned trolleys.

In poultry breeding, the hatchery occupies a central position by being in daily contact with breeder farms to collect eggs and with broiler farms to deliver chicks. Hatcheries can serve as a reservoir and source of pathogenic microorganisms and via the movement of vehicles (delivery trucks), people or equipment (trays, trolleys, and chick boxes, etc.) can facilitate the dissemination of microorganisms (McMullin, 2009; Osman et al., 2018). Consequently, the contamination detected in our study in the hatchery can be also explained by this permanent, exchange of potentially healthy carriers of *C. botulinum* between the hatchery and broiler farms. It was not possible to identify the initial source of contamination, i.e., hatchery or poultry farms, here, because the hatchery was only taken into account after the botulism case investigation had begun.

Despite the hygiene and biosecurity procedures in the hatcheries, infectious agents can contaminate hatcheries by being transported on or within eggs, on hatchery personnel, on trolleys and trays, or as airborne contaminants (Wales and Davies, 2020). A wide range of microorganisms, such as *Salmonella, E. coli, Pseudomonas, Staphylococci, Mycoplasma,* or *Aspergillus*, can be detected in hatcheries and disseminated to chicks and subsequently to farms (Qureshi, 2002; McMullin, 2009). To the best of our knowledge, there is currently no data available regarding the contamination of hatcheries by *C. botulinum*. The risks of hatchery contamination arise from the hygiene of hatching eggs, the multiple exchanges between farms, and also the management of the hatchery, involving vehicles, people and equipment all along the process from the arrival of the hatching eggs to the delivery of chicks to customer farms (Qureshi, 2002; McMullin, 2009). In our study, *C. botulinum* type D/C was not detected in any of the samples collected in the incubation ($n = 12$) and hatching rooms ($n = 11$). Moreover, no asymptomatic carriage was detected in chicks delivered on the farm using cloacal swabs on Flock No. 5. The hatching process is considered as the most risky step for microbial dissemination (McMullin, 2009; Kim and Kim, 2010). This does not seem to be the case here based on detection results. No vertical transmission of *C. botulinum* from breeders to chicks was demonstrated here or in previous studies. The exact role of hatcheries in the epidemiology of animal botulism requires further investigation.

MLVA as a Useful Tool for Epidemiological Investigations

Genotyping methods such as pulsed-field gel electrophoresis (PFGE) and amplified fragment length polymorphism (AFLP) have been successfully used in the context of animal botulism outbreaks for tracking strains (Myllykoski et al., 2009; Skarin et al., 2015). However such approaches require isolation of pure strains, thus limiting the number of samples that can be analyzed. Despite improvements regarding *C. botulinum* group III isolation (Le Gratiet et al., 2020), it is still time-consuming and phages encoding BoNT are easily lost. MLVA has been available for *C. botulinum* group I and II for many years (Fillo et al., 2011; Umeda et al., 2013; Anniballi et al., 2016). It has been recently developed for *C. botulinum* group III subtyping (Auricchio et al., 2019). MLVA presents a major advantage because it does not require the isolation of the strain and can be directly used on DNA extracted from the initial samples (Kahlisch et al., 2010; Chen et al., 2011; Pereyre et al., 2012; Pailhoriès et al., 2015). Our study demonstrates the usefulness of the MLVA approach as a subtyping tool intended for tracing and tracking *C. botulinum* group III, in particular for investigations during animal botulism outbreaks.

In our study, profiles of samples encoded S1 to S10 were shown to be identical (Group 1). Three cloacal swabs collected on broilers from Flock No. 4 presented a different profile from the other samples collected during the study (76.8% similarity between Groups 1 and 2). The origin of this second profile was not explained here. Unfortunately, MLVA results from samples collected at the beginning of the Flock No. 4 rearing period were not interpretable (DNA amount was too low in these samples as shown by a late Ct obtained for these samples). Therefore, it was not possible to determine if this contamination could be linked to the hatchery or not. Samples from the hatchery (samples S8, S9, S10) were indeed all part of Group 1.

CONCLUSION

This study demonstrates that broilers can be healthy carriers of *C. botulinum* type D/C that can lead to cattle contamination and in the initiation of a botulism outbreak. This also shows that this contamination can last several months, spanning successive flocks. This study also shows that the environment of hatcheries can be contaminated by *C. botulinum* and may become a source of introduction of *C. botulinum* in poultry farms via chick delivery. As illustrated by a massive cattle botulism outbreak due to cross-contamination between poultry and cows, this study highlights the major importance of the implementation of appropriate biosecurity measures in mixed farms to avoid cross-contamination between the production units involving equipment (specific material or disinfection), vehicles (separate travel routes) and personnel (shoes and clothes in shower room). Further investigations are now required to evaluate *C. botulinum* contamination occurrence and level in hatcheries so as to better understand their potential role in the epidemiology of botulism.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Samples were collected by veterinarians for diagnostic purposes on dead animals (death due to the botulism outbreak). Cloacal swabs were collected on broilers in the farm by veterinarians for official analyses. This article presents non-experimental clinical veterinary studies. Animals on farms were treated in accordance with the European and French regulation on farmed animal protection. Samples were collected in accordance with Regulation EC/1099/2009 and Directive 2010/63/EU. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

CL, RS, SL, DG, and MC designed the study. CL, TP, SR, SM, and LM analyzed the samples for detection of *C. botulinum*. RS and CL analyzed the data and drafted the manuscript. DG, RS, and LB collected the epidemiological data and samples on field. TL and FA were in charge of the MLVA analysis. SL, FA, and MC provided their expertise feedback. All authors reviewed the draft and contributed significantly to the final manuscript.

FUNDING

This work was funded by the French Agency for Food, Environmental and Occupational Health & Safety (ANSES), the Côtes d'Armor Departmental Council, and the ANIBOTNET project (ANIHWA call). TL is the recipient of a Ph.D. grant from ANSES and the Bretagne Regional Council, and the recipient of a short stay grant funded by the MedVetNet Association for a training period regarding the MLVA method at the ISS laboratory in Italy under Fabrizio Anniballi's supervision.

ACKNOWLEDGMENTS

The authors would like to thank the farmers and veterinarians for their participation in the study; Concetta Scalfaro (ISS) for her help in MLVA analysis; and Valérie Rose (ANSES, Ploufragan Laboratory) and François-Xavier Briand (ANSES, Ploufragan Laboratory) for their help and expertise in the sequencing PCR products for MLVA analysis. **Figure 3** was created with BioRender.com.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.679377/full#supplementary-material

REFERENCES

Anniballi, F., Fillo, S., Giordani, F., Auricchio, B., Tehran, D. A., Di Stefano, E., et al. (2016). Multiple-locus variable number of tandem repeat analysis as a tool for molecular epidemiology of botulism: the Italian experience. *Infect. Genet. Evol.* 46, 28–32. doi: 10.1016/j.meegid.2016.10.014

Anniballi, F., Fiore, A., Löfström, C., Skarin, H., Auricchio, B., Woudstra, C., et al. (2013). Management of animal botulism outbreaks: from clinical suspicion to practical countermeasures to prevent or minimize outbreaks. *Biosec. Bioterrorism* 11, S191–S199. doi: 10.1089/bsp.2012.0089

Auricchio, B., Scalfaro, C., Vicenza, T., Bano, L., Bilei, S., Chemaly, M., et al. (2019). *Multilocus Variable-Number of Tandem-Repeat Analysis as a tool for Clostridium Botulinum Group III Subtyping ANIBOTNET Workshop Botulism, Paris, 28-29 march 2019.*

Bano, L., Drigo, I., Tonon, E., Pascoletti, S., Puiatti, C., Anniballi, F., et al. (2017). Identification and characterization of *Clostridium botulinum* group III field strains by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS). *Anaerobe* 48, 126–134. doi: 10.1016/j.anaerobe.2017.08.004

Chen, S., Li, J., Saleh-Lakha, S., Allen, V., and Odumeru, J. (2011). Multiple-locus variable number of tandem repeat analysis (MLVA) of *Listeria monocytogenes* directly in food samples. *Int. J. Food Microbiol.* 148, 8–14. doi: 10.1016/j.ijfoodmicro.2011.04.014

Fillo, S., Giordani, F., Anniballi, F., Gorgao, O., Ramisse, V., Vergnaud, G., et al. (2011). *Clostridium botulinum* group I strain genotyping by 15-locus multilocus variable-number tandem-repeat analysis. *J. Clin. Microbiol.* 49, 4252–4263. doi: 10.1128/JCM.05396-11

Hogg, R., Livesey, C., and Payne, J. (2008). Diagnosis and implications of botulism. *In Practice* 30, 392–397. doi: 10.1136/inpract.30.7.392

Kahlisch, L., Henne, K., Draheim, J., Brettar, I., and Höfle, M. G. (2010). High-resolution in situ genotyping of *Legionella pneumophila* populations in drinking water by multiple-locus variable-number tandem-repeat analysis using environmental DNA. *Appl. Environ. Microbiol.* 76, 6186–6195. doi: 10.1128/aem.00416-10

Kim, J. H., and Kim, K. S. (2010). Hatchery hygiene evaluation by microbiological examination of hatchery samples. *Poultry Sci.* 89, 1389–1398. doi: 10.3382/ps.2010-00661

Le Gratiet, T., Poezevara, T., Rouxel, S., Houard, E., Mazuet, C., Chemaly, M., et al. (2020). Development of an innovative and quick method for the isolation of *Clostridium botulinum* strains involved in avian botulism outbreaks. *Toxins (Basel)* 12:42. doi: 10.3390/toxins12010042

Le Maréchal, C. L., Hulin, O., Macé, S., Chuzeville, C., Rouxel, S., Poëzevara, T., et al. (2019). A case report of a botulism outbreak in beef cattle due to the contamination of wheat by a roaming cat carcass: from the suspicion to the management of the outbreak. *Animals* 9:1025. doi: 10.3390/ani9121025

McMullin, P. F. (2009). Hygiene and microbiological control in hatcheries. *Avian Biol. Res.* 2, 93–97. doi: 10.3184/175815509X431885

Myllykoski, J., Lindström, M., Keto-Timonen, R., Söderholm, H., Jakala, J., Kallio, H., et al. (2009). Type C bovine botulism outbreak due to carcass contaminated non-acidified silage. *Epidemiol. Infect.* 137, 284–293. doi: 10.1017/S0950268808000939

Notermans, S., Dufrenne, J., and Oosterom, J. (1981). Persistence of Clostridium botulinum type B on a cattle farm after an outbreak of botulism. *Appl. Environ. Microbiol.* 41, 179–183. doi: 10.1128/aem.41.1.179-183.1981

Osman, K. M., Kappell, A. D., Elhadidy, M., Elmougy, F., El-Ghany, W. A. A., Orabi, A., et al. (2018). Poultry hatcheries as potential reservoirs for antimicrobial-resistant *Escherichia coli*: a risk to public health and food safety. *Sci. Rep.* 8:5859. doi: 10.1038/s41598-018-23962-7

Pailhoriès, H., Buzelé, R., Picardeau, M., Robert, S., Mercier, E., Mereghetti, L., et al. (2015). Molecular characterization of *Leptospira sp* by multilocus variable number tandem repeat analysis (MLVA) from clinical samples: a case report. *Int. J. Infect. Dis.* 37, 119–121. doi: 10.1016/j.ijid.2015.06.026

Payne, J. H., Hogg, R. A., Otter, A., Roest, H. I., and Livesey, C. T. (2011). Emergence of suspected type D botulism in ruminants in England and Wales (2001 to 2009), associated with exposure to broiler litter. *Vet. Rec.* 168:640. doi: 10.1136/vr.d1846

Peck, M. W., Smith, T. J., Anniballi, F., Austin, J. W., Bano, L., Bradshaw, M., et al. (2017). Historical perspectives and guidelines for botulinum neurotoxin subtype nomenclature. *Toxins* 9:38. doi: 10.3390/toxins9010038

Pereyre, S., Charron, A., Hidalgo-Grass, C., Touati, A., Moses, A. E., Nir-Paz, R., et al. (2012). The spread of *Mycoplasma pneumoniae* is polyclonal in both an endemic setting in France and in an epidemic setting in Israel. *PLoS One* 7:e38585. doi: 10.1371/journal.pone.0038585

Popoff, M. (1989). Revue sur l'épidémiologie du botulisme bovin en France et analyse de sa relation avec les élevages de volailles. *Rev. Sci. Tech. Off. Int. Epiz.* 8, 129–145. doi: 10.20506/rst.8.1.404

Popoff, M. R. (1995). "Ecology of neurotoxigenic strains of clostridia," in *Clostridial Neurotoxins. Current Topics in Microbiology and Immunology*, ed. C. Montecucco (Berlin: Springer-Verlag), 1–29. doi: 10.1007/978-3-642-85173-5_1

Popoff, M. R., and Argente, G. (1996). Le botulisme animal est-il une menace pour l'homme ? *Bull. l'Académie Vétérinaire France* 69, 373–382. doi: 10.4267/2042/63881

Popp, C., Hauck, R., Gad, W., and Hafez, H. M. (2012). Type C botulism in a commercial Turkey farm: a case report. *Avian Dis.* 56, 760–763. doi: 10.1637/10274-061212-Case.1

Qureshi, A. (2002). Hatchery sanitization and chick mortality. *World Poultry* 18, 24–25.

Ramírez-Romero, R., Ramírez-Hernández, C., García-Márquez, L. J., Macedo-Barragán, R. J., Martínez-Burnes, J., and López-Mayagoitia, A. (2014). Bovine diseases causing neurological signs and death in Mexican feedlots. *Trop. Anim Health Prod.* 46, 823–829. doi: 10.1007/s11250-014-0572-y

Rasetti-Escargueil, C., Lemichez, E., and Popoff, M. R. (2019). Public health risk associated with botulism as foodborne zoonoses. *Toxins (Basel)* 12:17. doi: 10.3390/toxins12010017

Rasetti-Escargueil, C., Lemichez, E., and Popoff, M. R. (2020). Human botulism in France, 1875–2016. *Toxins (Basel)* 12:338. doi: 10.3390/toxins12050338

Relun, A., Dorso, L., Douart, A., Chartier, C., Guatteo, R., Mazuet, C., et al. (2017). A large outbreak of bovine botulism possibly linked to a massive contamination of grass silage by type D/C *Clostridium botulinum* spores on a farm with dairy and poultry operations. *Epidemiol. Infect.* 145, 3477–3485. doi: 10.1017/s0950268817002382

Senturk, S., and Cihan, H. (2007). Outbreak of botulism in a dairy herd in Turkey. *Ir. Vet. J.* 60, 481–484. doi: 10.1186/2046-0481-60-8-481

Skarin, H., Lindgren, Y., and Jansson, D. S. (2015). Investigations into an outbreak of botulism caused by *Clostridium botulinum* Type C/D in laying hens. *Avian Dis.* 59, 335–340. doi: 10.1637/10861-051214-Case

Skarin, H., Tevell Åberg, A., Woudstra, C., Hansen, T., Löfström, C., Koene, M., et al. (2013). The workshop on animal botulism in Europe. *Biosec. Bioterrorism* 11, S183–S190. doi: 10.1089/bsp.2012.0076

Souillard, R., Le Marechal, C., Ballan, V., Mahe, F., Chemaly, M., and Le Bouquin, S. (2017). A bovine botulism outbreak associated with a suspected cross-contamination from a poultry farm. *Vet. Microbiol.* 208, 212–216. doi: 10.1016/j.vetmic.2017.07.022

Souillard, R., Le Marechal, C., Ballan, V., Rouxel, S., Leon, D., Balaine, L., et al. (2016). Investigation of a type C/D botulism outbreak in free-range laying hens in France. *Avian Pathol.* 46, 195–201. doi: 10.1080/03079457.2016.1240355

Souillard, R., Woudstra, C., Le Marechal, C., Dia, M., Bayon-Auboyer, M. H., Chemaly, M., et al. (2014). Investigation of Clostridium botulinum in commercial poultry farms in France between 2011 and 2013. *Avian Pathol.* 43, 458–464. doi: 10.1080/03079457.2014.957644

Tsiouris, V. (2016). Poultry management: a useful tool for the control of necrotic enteritis in poultry. *Avian Pathol.* 45, 323–325. doi: 10.1080/03079457.2016.1154502

Umeda, K., Wada, T., Kohda, T., and Kozaki, S. (2013). Multi-locus variable number tandem repeat analysis for *Clostridium botulinum* type B isolates in Japan: comparison with other isolates and genotyping methods. *Infect. Genet. Evol.* 16, 298–304. doi: 10.1016/j.meegid.2013.02.022

Wales, A., and Davies, R. (2020). Review of hatchery transmission of bacteria with focus on *Salmonella*, chick pathogens and antimicrobial resistance. *World's Poultry Sci. J.* 76, 517–536. doi: 10.1080/00439339.2020.1789533

Wobeser, G., Marsden, S., and Macfarlane, R. J. (1987). Occurrence of toxigenic Clostridium botulinum type C in the soil of wetlands in Saskatchewan. *J. Wildlife Dis.* 23, 67–76. doi: 10.7589/0090-3558-23.1.67

Woudstra, C., Skarin, H., Anniballi, F., Fenicia, L., Bano, L., Drigo, I., et al. (2012). Neurotoxin gene profiling of *Clostridium botulinum* types C and D native to different countries within Europe. *Appl. Environ. Microbiol.* 78, 3120–3127.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 July 2021
doi: 10.3389/fmicb.2021.688240

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

OPEN ACCESS

Edited by:
Fabrizio Anniballi, Istituto Superiore di Sanità, Italy

Reviewed by:
Michael William Peck, Quadram Institute, United Kingdom
Concetta Scalfaro, National Institute of Health (ISS), Italy

****Correspondence:***
Lori Gladney
lgladney@cdc.gov

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *30 March 2021*
Accepted: *25 May 2021*
Published: *13 July 2021*

Citation:
Gladney L, Halpin JL and Lúquez C (2021) Genomic Characterization of Strains From a Cluster of Infant Botulism Type A in a Small Town in Colorado, United States. Front. Microbiol. 12:688240. doi: 10.3389/fmicb.2021.688240

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

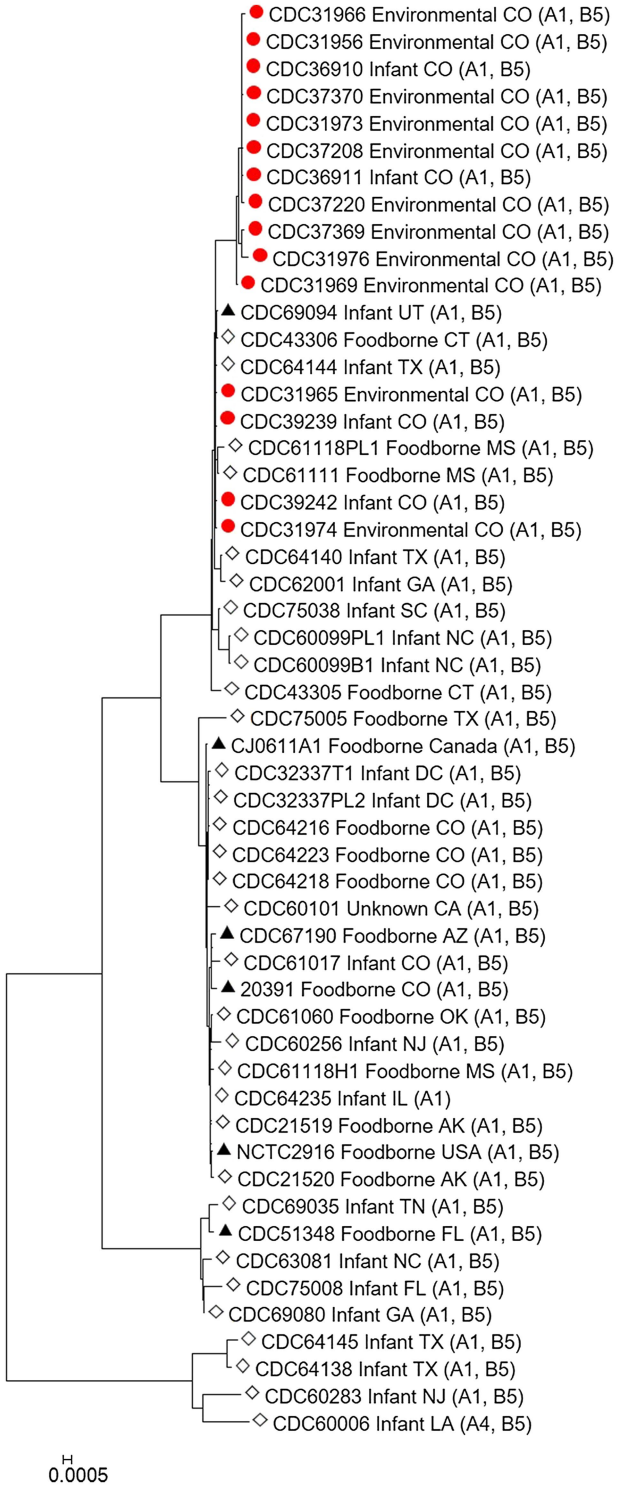
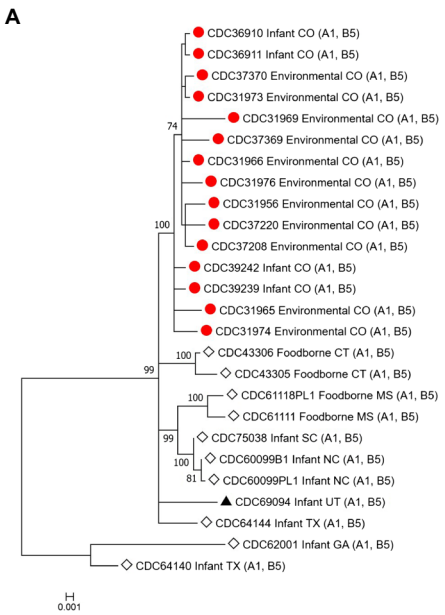


FIGURE 1 | Neighbor-joining tree of *Clostridium botulinum* type A(B) isolates. Tree was drawn using Mashtree and includes study strains harboring both *bont*/A1 and *bont*/B5 genes as described in **Table 1** (red circles), reference sequences from NCBI (black triangles), and other unpublished sequences from the CDC reference collection (white diamonds), as described in **Supplementary Table 1.** Scale bar represents the Mash distance.



B

	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	3	5	7	9	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	21	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	22	90	35	18	21	16
CDC36911	8	9	3	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	9	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	17	23	20	0	1	17	17	93	38	21	26	2
CDC60099PL1	21	22	17	26	17	21	20	18	17	18	20	22	20	18	18	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	24	26	25	17	19	0	8	98	44	27	31	16
CDC61118PL1	25	25	21	30	21	24	24	22	21	20	24	26	24	22	22	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	19	18	16	16	19	21	19	17	17	23	20	21	22	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

LG, JH, and CL wrote this manuscript and conceived the project study design. LG and JH selected the datasets. LG analyzed the sequence data and generated figures and tables. JH prepared DNA and sequenced the strains in this study. All authors contributed to the article and approved the submitted version.

FUNDING

Centers for Disease Control and Prevention and Center for Preparedness and Response provided support for this project. 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.

ACKNOWLEDGMENTS

We thank Charles L. Hatheway and Loretta M. McCroskey for their previous work on this outbreak and contributions to the investigation. We would also like to thank our reviewers for thoughtful critique of this manuscript.

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

MINI REVIEW
published: 16 July 2021
doi: 10.3389/fmicb.2021.713101

Foodborne Botulism Outbreaks in the United States, 2001–2017

Carolina Lúquez, Leslie Edwards, Chelsey Griffin and Jeremy Sobel*

Centers for Disease Control and Prevention, Atlanta, GA, United States

Foodborne botulism is an intoxication caused by ingestion of food containing botulinum neurotoxin. Cases of foodborne botulism are usually sporadic (single, unrelated) but outbreaks of two or more cases occur. In this mini-review we will examine the following for the period 2001–2017, in the United States: botulism surveillance data, outbreaks of botulism affecting 10 or more people, and the public health preparedness and response approach.

Keywords: *Clostridium botulinum*, foodborne botulism, botulism outbreak, epidemiology, public health

OPEN ACCESS

Edited by:
John W. Austin,
Health Canada, Canada

Reviewed by:
Dario De Medici,
National Institute of Health (ISS), Italy
Michael William Peck,
Quadram Institute, United Kingdom

****Correspondence:***
Carolina Lúquez
CLuquez@cdc.gov

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *21 May 2021*
Accepted: *23 June 2021*
Published: *16 July 2021*

Citation:
Lúquez C, Edwards L, Griffin C and Sobel J (2021) Foodborne Botulism Outbreaks in the United States, 2001–2017. Front. Microbiol. 12:713101. doi: 10.3389/fmicb.2021.713101

INTRODUCTION

Botulism is a rare, potentially fatal illness caused by botulinum neurotoxins (BoNT), produced by *Clostridium botulinum*, and rare strains of *C. butyricum* and *C. baratii* (Sobel, 2005). Seven antigenically distinct BoNTs have been identified, denominated serotypes A, B, C, D, E, F, and G (Sobel, 2005). Serotypes A, B, E, and F cause botulism cases in humans. Botulism is characterized by cranial nerve palsies which may be followed by flaccid, symmetrically descending paralysis. Four forms of naturally occurring botulism have been described: foodborne botulism, infant botulism, wound botulism, and adult intestinal colonization. Foodborne botulism is caused by the ingestion of foods contaminated with BoNT. Wound botulism is caused by spores of BoNT-producing species of *Clostridium* that germinate in a contaminated wound; the vegetative cells then multiply and produce BoNT *in situ*. Infant botulism is caused by BoNT-producing species of *Clostridium* that colonize the intestinal tract of infants (children under 1 year of age). Adult intestinal colonization is similar to infant botulism, but it affects persons older than 1 year (Sobel, 2005).

EPIDEMIOLOGY OF FOODBORNE BOTULISM IN THE UNITED STATES, 2001–2017

According to CDC's National Botulism Surveillance, from 2001 to 2017, 326 laboratory confirmed foodborne botulism cases were reported in the United States (**Table 1**; CDC, 2017). The median number of confirmed cases per year was 19 (range: 2 to 39 cases). In 2013, a case definition for additional probable foodborne botulism was implemented by the Council of State and Territorial Epidemiologists (CSTE) that includes patients whose clinical presentation is compatible with botulism and who had an epidemiological risk factor, such as consumption of home-canned food in the 48 h preceding illness onset (CDC, 2011). From 2013 to 2017, 29 probable foodborne botulism cases were identified in the United States and the median number of cases per year was 6 (range: 2 to 8 cases).

The median age of patients with laboratory confirmed cases was 49 years (range: 6 months to 92 years). Most cases occurred in males (183 cases, 56%). Death was reported in 17 cases (5% case fatality ratio) and the median age of patients who died was 76 years (range: 53 to 91 years).

TABLE 1 | Laboratory confirmed foodborne botulism cases in the United States, 2001–2017.

Year	Type A	Type B	Type E	Type F	NI[(1)]	Total
2001	20	2	10	1	0	33
2002	5	2	14	0	0	21
2003	5	0	2	1	0	8
2004	12	0	2	0	0	14
2005	7	1	10	0	0	18
2006	12	1	1	0	5	19
2007	15	4	7	0	0	26
2008	10	0	6	0	2	18
2009	10	1	0	0	0	11
2010	3	3	2	0	1	9
2011	14	0	5	1	0	20
2012	19	3	3	0	0	25
2013	1	0	1	0	0	2
2014	4	4	7	0	0	15
2015	34	0	5	0	0	39
2016	25	1	3	0	0	29
2017	15	0	4	0	0	19
TOTAL	211	22	82	3	8	326

[(1)] *Serotype not identified.*

A food or beverage was implicated in the epidemiological investigation for 277 (85%) laboratory confirmed botulism cases, and a food or beverage sample was laboratory confirmed as the source of BoNT in 156 (47%) cases. The method of food preparation implicated among the cases included foods prepared at home other than home canned foods (154 cases, 47%), home canned foods (94 cases, 29%), commercially canned foods (31 cases, 10%) and commercially prepared foods (20 cases, 6%) (**Figure 1**). The method of food preparation was not available for 27 (8%) cases. Pruno, an illicit prison-brewed alcoholic beverage (45 cases, 14%), fish (30 cases, 9%), and seal (24 cases, 7%) were the food or beverage items associated with the largest number of laboratory confirmed botulism cases.

Botulinum neurotoxins type A (211 cases, 65%) was the toxin type most frequently identified, followed by toxin type E (82 cases, 25%) and toxin type B (22 cases, 7%). States with the largest number of reported cases were Alaska (90 cases, 28%), California (50 cases, 15%) and Ohio (40 cases, 12%). Toxin type E was identified in 93% of Alaska cases in which the toxin type was reported. Nationwide, most cases were reported between April and August (**Supplementary Figure 1**).

OUTBREAKS OF FOODBORNE BOTULISM IN THE UNITED STATES, 2001–2017

Because of botulism's potential to cause outbreaks through the foodborne route and the potential for severe illness for any case that occurs, a single case of botulism constitutes a public health emergency, as it may herald a larger event. From 2001 to 2017, 53 foodborne outbreaks of two or more cases were reported, and 223 (68%) of all reported foodborne botulism cases were part of an outbreak. Foodborne botulism cases are usually sporadic (single cases) or involve few cases, but large outbreaks do occur. For the purpose of this report, "large outbreak" was defined as a laboratory confirmed outbreak of foodborne botulism that affected 10 or more people. This number was based on the average number of four cases per outbreak (United States, 2001–2017). We focused on these large outbreaks to illustrate how the occurrence of botulism by the foodborne route can cause many cases of severe disease, which can strain local or regional medical resources. From 2001 to 2017, five large outbreaks of foodborne botulism were reported in the United States (**Table 2**; Kalluri et al., 2003; Ginsberg et al., 2007; Juliao et al., 2013; McCarty et al., 2015; McCrickard et al., 2017; Rosen et al., 2018, 2020).

Commercially Prepared Chili in Texas, 2001

This outbreak occurred among attendees of a church dinner. Fourteen of 24 (58%) attendees who ate commercially prepared frozen chili developed signs and symptoms of botulism. Stool specimens from 8 of the 14 patients (56%) tested positive for BoNT type A. A stool sample from an attendee who ate chili and remained asymptomatic tested positive for BoNT type A, an exceedingly rare finding. An additional botulism case occurred in Dallas, in a person who did not attend the church event but had consumed chili of the same brand as the chili served at the church. BoNT type A was identified in, and *C. botulinum* type A isolated from, chili leftovers from the church supper and in the opened container of frozen chili eaten by the patient who did not attend the church event. An unopened plastic tub of the same brand of frozen chili, found in the church's kitchen, tested positive for BoNT type A; the level of BoNT was estimated as 10,000 mouse lethal dose 50%/g (mLD_{50}/g), i.e., 2×10^7 mLD_{50}/container. Both frozen chili packages had been purchased at the same salvage food store (discount grocery store) on sequential days. The chili served at the church event had been prepared by heating the chili without bringing it to a boil, then adding hot dogs and a different, commercially canned chili. The patient who did not attend the church event also heated the frozen chili without boiling it. Inspection of the salvage store revealed gross mishandling of foods, including opened plastic containers of food, and foods that required refrigeration or freezing but were displayed for sale at room temperature. Inspections of the chili manufacturing plant did not reveal deficiencies of production practices or product mishandling.

Commercially Prepared Hot Dog Chili, Multi-State, 2007

This outbreak was associated with 10 cases in California (1 case), Indiana (2 cases), New Mexico (1 case), Ohio (3 cases), and Texas (3 cases). All patients developed signs and symptoms of botulism; five were treated with botulinum antitoxin. The three patients in Texas had shared a meal, which included commercially canned hot dog chili sauce. The two Indiana patients had also shared a meal, which included commercially manufactured chili of the same brand as the patients in Texas. Although the three

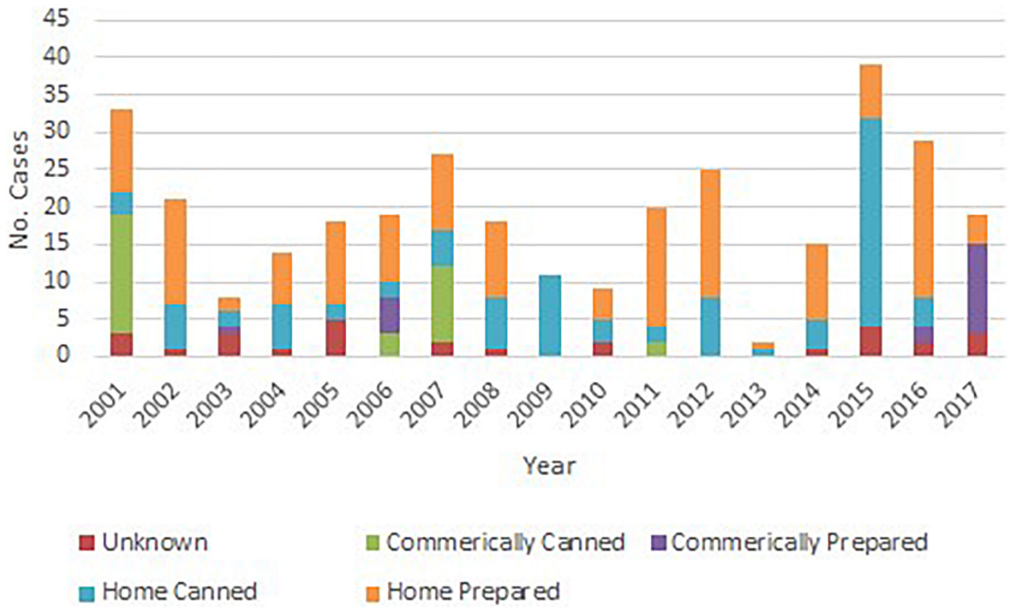


FIGURE 1 | Number of foodborne botulism cases by year and type of food preparation, United States, 2001–2017.

Ohio patients were not related to each another, strong evidence indicated that they had consumed hot dog chili sauce of the same brand as the patients in Texas and Indiana. The case in California and the case in New Mexico manifested clinical signs consistent with botulism but could not be definitively linked to the implicated product through interviews or product collection from the homes. Serum and stool specimens from two of the eight patients tested positive for BoNT Type A and *C. botulinum* type A, respectively. Leftover chili mixture obtained from the home of the Indiana patients tested positive for BoNT type A and the toxin level was estimated as 4,000 mLD_{50}/g. BoNT type A was also identified in leftover chili sauce collected from the home of one of the Ohio patients. No food leftovers from the cases in Texas were available for testing and an unopened can of the same product found in the patients' home was negative for BoNT. An investigation of the cannery identified violations of canned food regulations that may have permitted spores of *C. botulinum* to survive the canning process: heating vessels used to sterilize cans were improperly maintained and operated; cooling water valves were malfunctioning; faulty alarm indicator lights and improperly calibrated temperature monitoring devices were observed. Twenty of 23 swollen cans of hot dog chili sauce collected during the investigation tested positive for BoNT type A. The manufacturer issued a nationwide recall of all products manufactured on the same production line; more than 100 million cans were recalled. No additional botulism cases were identified.

Homemade Potato Salad in Ohio, 2015

This outbreak occurred among people who attended a church potluck meal. Twenty-nine of 77 persons (38%) who consumed potluck food had signs or symptoms consistent with botulism. Nineteen cases were laboratory confirmed, and 10 were probable cases (patients showed signs or symptoms of botulism but cases were not laboratory confirmed). One patient died of respiratory failure shortly after arriving at the hospital. Twenty-five patients were treated with botulinum antitoxin. Serum and stool specimens tested positive for BoNT type A and/or *C. botulinum* type A. Symptoms began a median of 2 days after the meal (range: 1 to 6 days). More than 50 foods were reportedly served at the event; interviews conducted among 75 participants suggested potato salad as the source. All foods served at the event had been discarded; 12 food samples were collected from the church dumpster. Six samples were positive for BoNT type A; of these, five contained potato salad and one contained macaroni and cheese that might have been cross-contaminated after being discarded. Levels of toxin in the food samples were not determined. The potato salad had been prepared with home-canned potatoes that were canned using a boiling water canner.

TABLE 2 | Outbreaks of laboratory confirmed botulism affecting 10 or more people, United States, 2001–2017.

Year	State	Food source	Serotype	No. of cases	References
2001	Texas	Chili	A	16	Kalluri et al., 2003
2007	Multi-state	Hot dog chili sauce	A	10	Ginsberg et al., 2007; Juliao et al., 2013
2015	Ohio	Potato salad	A	29	McCarty et al., 2015
2016	Mississippi	Pruno	A	31	McCrickard et al., 2017
2017	California	Nacho cheese	A	10	Rosen et al., 2018, 2020

Boiling water's temperature is not sufficient to destroy spores of *C. botulinum* that might have been present in the raw potatoes. In addition, the canned potatoes were not heated before preparing the salad.

Pruno (Prison Brew) in Mississippi, 2016

This outbreak occurred among inmates at a federal prison. The suspected source of BoNT was pruno, an illicit alcoholic beverage made by inmates by fermenting fruit or vegetables, sugar, water, and other ingredients. Thirty-one of 33 inmates who had consumed pruno presented with signs and symptoms of botulism. Twenty patients were treated with botulinum antitoxin. No deaths were reported. Serum and stool samples from 19 patients tested positive for BoNT type A or *C. botulinum* type A. Symptoms began a median of 3 days after first consuming pruno (range: several hours to 11 days). Reportedly, the pruno was made by combining honey, potatoes, apples, and tomato paste from a bulging can, then fermenting the mixture in a sealed plastic bag at room temperature for three to 5 days. About 20 gallons of pruno were confiscated during an investigation, but no leftovers of the batch implicated in this outbreak were found. The first two reported foodborne botulism outbreaks from pruno occurred among inmates at two prisons in California, in 2004 and 2005 (Vugia et al., 2009). Since then, four additional botulism outbreaks from pruno have been reported in the United States; the outbreak in Mississippi affected the largest number of inmates. This was the largest foodborne botulism outbreak in the United States since 1978.

Commercially Prepared Nacho Cheese Sauce in California, 2017

Ten people were associated with an outbreak of foodborne botulism linked to nacho cheese sauce served from a dispenser at a gas station market. Symptoms began a median of 3 days after consuming cheese sauce (range: 2 to 9 days). All patients were hospitalized and were treated with botulinum antitoxin. Seven patients required mechanical ventilation and one patient died. Eight of the patients confirmed eating nacho cheese sauce at the same gas station food market; one patient died before an interview could be conducted, but had been at the gas station in the week before illness onset; the other patient worked at the gas station but did not confirm eating cheese sauce. Serum and/or stool specimens from all patients were positive for BoNT type A and/or *C. botulinum* isolates type A. A remaining pouch of nacho cheese sauce tested positive for BoNT type A and *C. botulinum* type A. Although the mechanism of contamination is unclear, inspection of the gas station indicated that the cheese in the dispenser was about 30 days past expiration date. The approximately 20 ounces of sauce found by investigators in the dispenser exhibited apparent oil separation, a temperature of 111°F, and pH of 5.95, which would allow growth of *C. botulinum* and toxin production. According to the manufacturer's instructions, the cheese sauce should have been consumed or discarded within 5 days of placing in the dispenser, and the sauce temperature should have been maintained at 140°F. An unopened bag of cheese from different lot obtained at the gas station tested negative for BoNT. No other botulism cases associated with this commercial cheese sauce were identified.

PUBLIC HEALTH PREPAREDNESS AND RESPONSE APPROACH IN THE UNITED STATES

Every case of suspected botulism is a public health emergency, because of the potential for severe disease, and the possibility that it could be the herald of a larger outbreak from contaminated food. Therefore, public health preparedness for botulism rests most fundamentally on early clinical diagnosis by the astute clinician and immediate notification of public health authorities (Sobel, 2005).

In the United States, diagnosis of botulism, provision of botulinum antitoxin, and epidemiologic investigation are closely linked. Physicians suspecting botulism in a patient should immediately contact their state health departments' emergency phone number (CDC, 2021). The physician will be put in touch with an expert from CDC's Botulism Clinical Consultation Service, or, in California or Alaska, at the state health department. When indicated, CDC will urgently dispatch botulinum antitoxin from the federal stockpiles at no charge, and the state health department will facilitate collection of specimens and their testing at a specialized public health laboratory. State health departments will quickly initiate an epidemiologic investigation to identify and control the source of BoNT and determine if there are other cases (Sobel, 2005).

Since botulism can result in respiratory collapse and ventilator dependence lasting weeks to months, an outbreak involving many cases could strain or overwhelm intensive care resources locally or regionally. For example, an outbreak of 29 cases exceeded the resources of a local, modern American hospital and strained those of its surrounding metropolitan area (McCarty et al., 2015); a contaminated, widely consumed food could produce a far larger outbreak. In a mass event, worried asymptomatic persons and those with symptoms that might be due to botulism may seek medical care; the need to rapidly identify patients in danger of respiratory failure must be anticipated. In a large event, the time required to ship botulinum antitoxin, which can limit the extent of botulism-induced paralysis and prevent respiratory collapse, might result in temporary delays in administering antitoxin. Emergency preparedness planning by hospitals and local and state public health officials should incorporate consideration of such concerns into emergency planning. Clinical, laboratory diagnostic, and emergency preparedness considerations for botulism are comprehensively addressed in CDC's botulism clinical guidelines (Sobel and Rao, 2017; Rao et al., 2021).

DISCUSSION

Botulism is a rare disease seen by few medical providers, which produces severe and potentially fatal illness. Botulism should always be considered a public health emergency because it can cause severe disease and potentially large outbreaks by the

foodborne route; contamination of widely consumed food could cause many cases, which in turn could strain local or regional medical resources. Public health investigation and interventions to remove suspect foods can prevent additional cases and deaths. Results of investigations can help guide food manufacturers, environmental health specialists, and the general public about safe preparation and handling of foods.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This project was supported in part by an appointment to the Research Participation Program at the Centers for Disease Control and Prevention administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, Center for Preparedness and Response provided support for this project. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.713101/full#supplementary-material

Supplementary Figure 1 | Distribution of foodborne botulism cases per month of the year, United States, 2001–2017.

REFERENCES

CDC (2011). *Botulism (Clostridium botulinum) 2011 Case Definition*. Available online at: https://wwwn.cdc.gov/nndss/conditions/botulism/case-definition/2011/2011 (accessed May, 2021).

CDC (2017). *National Botulism Surveillance, Annual Summaries*. Available online at: https://www.cdc.gov/botulism/surveillance.html (accessed May, 2021).

CDC (2021). *Botulism, Information for Health Professionals*. Available online at: https://www.cdc.gov/botulism/health-professional.html2021 (accessed May, 2021).

Ginsberg, M. M., Granzow, L., Teclaw, R. F., Gaul, L. K., Bagdure, S., Cole, A., et al. (2007). Botulism associated with commercially canned chili sauce — Texas and Indiana, July 2007. *MMWR* 56, 767–769.

Juliao, P. C., Maslanka, S., Dykes, J., Gaul, L., Bagdure, S., Granzow-Kibiger, L., et al. (2013). National outbreak of type a foodborne botulism associated with a widely distributed commercially canned hot dog chili sauce. *Clin. Infect. Dis.* 56, 376–382. doi: 10.1093/cid/cis901

Kalluri, P., Crowe, C., Reller, M., Gaul, L., Hayslett, J., Barth, S., et al. (2003). An outbreak of foodborne botulism associated with food sold at a salvage store in Texas. *Clin. Infect. Dis.* 37, 1490–1495. doi: 10.1086/379326

McCarty, C. L., Angelo, K., Beer, K. D., Cibulskas-White, K., Quinn, K., de Fijter, S., et al. (2015). Large outbreak of botulism associated with a church potluck meal–Ohio, 2015. *MMWR Morb. Mortal. Wkly. Rep.* 64, 802–803. doi: 10.15585/mmwr.mm6429a6

McCrickard, L., Marlow, M., Self, J. L., Watkins, L. F., Chatham-Stephens, K., Anderson, J., et al. (2017). Notes from the field: botulism outbreak from drinking prison-made illicit alcohol in a federal correctional facility - Mississippi, June 2016. *MMWR Morb. Mortal. Wkly. Rep.* 65, 1491–1492. doi: 10.15585/mmwr.mm6552a8

Rao, A. K., Sobel, J., Chatham-Stephens, K., and Luquez, C. (2021). Clinical guidelines for diagnosis and treatment of botulism: sporadic cases and outbreaks. *MMWR* 70, 1–30. doi: 10.15585/mmwr.rr7002a1

Rosen, H. E., Kimura, A. C., Crandall, J., Poe, A., Nash, J., Boetzer, J., et al. (2020). Foodborne botulism outbreak associated with commercial nacho cheese sauce from a gas station market. *Clin. Infect. Dis.* 70, 1695–1700. doi: 10.1093/cid/ciz479

Rosen, H. E., Kimura, A. C., Mukhopadhyay, R., Nash, J., Boetzer, J., Poe, A., et al. (2018). An outbreak of botulism associated with nacho cheese sauce from a gas station in California. *Open Forum Infect. Dis.* 5:S53. doi: 10.1093/cid/ciz479

Sobel, J. (2005). Botulism. *Clin. Infect. Dis.* 41, 1167–1173.

Sobel, J., and Rao, A. K. (2017). Making the best of the evidence: toward national clinical guidelines for botulism. *Clin. Infect. Dis.* 66(suppl. 1), S1–S3.

Vugia, D. J., Mase, S. R., Cole, B., Stiles, J., Rosenberg, J., Velasquez, L., et al. (2009). Botulism from drinking pruno. *Emerg. Infect. Dis.* 15, 69–71. doi: 10.3201/eid1501.081024

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 July 2021
doi: 10.3389/fmicb.2021.678473

Clostridium botulinum Type B Isolated From a Wound Botulism Case Due to Injection Drug Use Resembles Other Local Strains Originating From Hawaii

Jessica L. Halpin, Victoria Foltz, Janet K. Dykes, Kevin Chatham-Stephens and Carolina Lúquez*

Centers for Disease Control and Prevention, Atlanta, GA, United States

OPEN ACCESS

Edited by:
Fabrizio Anniballi, Sanità Pubblica Veterinaria e Sicurezza Alimentare, Istituto Superiore di Sanità, Italy

Reviewed by:
Sabine Pellett, University of Wisconsin-Madison, United States
Concetta Scalfaro, National Institute of Health (ISS), Italy

***Correspondence:**
Jessica L. Halpin
JLHalpin@cdc.gov

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *09 March 2021*
Accepted: *09 June 2021*
Published: *22 July 2021*

Citation:
Halpin JL, Foltz V, Dykes JK, Chatham-Stephens K and Lúquez C (2021) Clostridium botulinum Type B Isolated From a Wound Botulism Case Due to Injection Drug Use Resembles Other Local Strains Originating From Hawaii. Front. Microbiol. 12:678473. doi: 10.3389/fmicb.2021.678473

Clostridium botulinum produces botulinum neurotoxin (BoNT), which can lead to death if untreated. In the United States, over 90% of wound botulism cases are associated with injection drug use of black tar heroin. We sought to determine the phylogenetic relatedness of *C. botulinum* isolated from an injection drug use wound botulism case and isolates from endogenous infant botulism cases in Hawaii. Nineteen *C. botulinum* type B isolates from Hawaii and one type B isolate from California were analyzed by whole-genome sequencing. The botulinum toxin gene (*bont*) subtype was determined using CLC Genomics Workbench, and the seven-gene multi-locus sequence type (MLST) was identified by querying PubMLST. Mashtree and pairwise average nucleotide identity were used to find nearest neighbors, and Lyve-SET approximated a phylogeny. Eighteen of the isolates harbored the *bont*/B5 gene: of those, 17 were classified as sequence type ST36 and one was classified as ST104. A single isolate from Hawaii harbored *bont*/B1 and was determined to belong to ST110, and the isolate from California harbored *bont*/B1 and belonged to ST30. A tree constructed with Lyve-SET showed a high degree of homology among all the Hawaiian *C. botulinum* isolates that harbor the *bont*/B5 gene. Our results indicate that the *bont*/B-expressing isolates recovered from Hawaii are closely related to each other, suggesting local contamination of the drug paraphernalia or the wound itself with spores rather than contamination of the drug at manufacture or during transport. These findings may assist in identifying interventions to decrease wound botulism among persons who inject drugs.

Keywords: wound botulism, bont/B5, BoNT, *Clostridium botulinum*, heroin use, skin popping, injection drug use

INTRODUCTION

Botulism is a life-threatening disease caused by botulinum neurotoxins (BoNT) which are produced by *Clostridium botulinum* and rare strains of *Clostridium butyricum* and *Clostridium baratii*. *C. botulinum* is an anaerobic spore-forming bacterium commonly found in soil. Currently, there are seven well-characterized serotypes for BoNT (A–G), classically determined by polyclonal

antibody neutralization assays (Hill et al., 2007). Additional serotypes have been proposed (BoNT/X,/En, and/H) but have not yet reached consensus within the scientific community (Barash and Arnon, 2014; Mansfield et al., 2015; Brunt et al., 2018; Popoff, 2018). BoNT are encoded by the *bont* gene, which is part of a cluster with a regulator (*botR*), and non-toxic accessory genes: non-toxic non-hemagglutinin (*ntnh*) and either open reading frame X (*orfX*) or hemagglutinin (*HA*) genes (Rossetto et al., 2014). BoNT serotypes can have multiple gene subtypes, which are determined by differences in the amino acid sequence. Over 40 subtypes of BoNT serotypes A, B, E, and F have been identified to date (Peck et al., 2017). Serotypes A and B cause most human botulism cases within the United States.

Botulinum neurotoxins are metalloproteases that target motor neurons, preventing the release of acetylcholine at neuromuscular junctions, which can lead to a flaccid paralysis (Rossetto et al., 2014). There are four naturally occurring forms of human botulism: infant botulism, foodborne intoxication, wound colonization, and adult intestinal colonization. Infant botulism is the most common form of the disease within the United States. It occurs when a baby under 1 year of age ingests *C. botulinum* spores, most likely by swallowing dust particles that carry locally acquired spores; the spores then germinate within the intestinal tract and produce BoNT *in situ* [Chin et al., 1979;Centers for Disease Control and Prevention (CDC), 1998]. Foodborne botulism occurs when food becomes contaminated with pre-formed BoNT and is ingested [Centers for Disease Control and Prevention (CDC), 1998]. Wound botulism occurs when *C. botulinum* spores enter a wound or necrotic tissue; the spores germinate, multiply, and release BoNT [Centers for Disease Control and Prevention (CDC), 1998]. A very rare form of the disease, adult intestinal colonization, occurs when an adult becomes colonized with *C. botulinum* spores, which germinate and produce BoNT *in situ* [Centers for Disease Control and Prevention (CDC), 1998].

CDC, in partnership with state and local public health laboratories, endeavors to conduct a surveillance for every confirmed case of botulism in the United States. Botulism has been a nationally required notifiable disease since 1947, and a case is confirmed when toxin is detected in clinical samples (e.g., serum, stool) or a suspected food sample or when *C. botulinum* organisms are isolated from a stool sample[1]. CDC's National Botulism Laboratory serves as a reference laboratory, a central testing laboratory for states that do not do botulism testing, and an overflow laboratory for states who may typically conduct their own botulism testing but require assistance due to capacity or other needs.

The Centers for Disease Control and Preventions' surveillance data from 1981 to 2016 totals 4,807 laboratory-confirmed cases of botulism within the United States[2]—70% infant botulism, 11% wound botulism, and 19% foodborne or "other" (adult intestinal colonization or unknown route of transmission). Once a rare form of the disease, wound botulism has become more common in the United States starting in the 1990s with the increased use of black tar heroin (BTH) (Passaro et al., 1998; Werner et al., 2000; Davis and King, 2008; Peak et al., 2019). From 1981 to 2016, there was an average of 15 wound botulism cases per year. During this 35-year period, 74.6% of wound botulism cases were due to injection drug use, and 25.4% of wound cases were associated with other types of trauma (e.g., lacerations, abscesses, necrotic tissue, gunshot wounds, or unknown)[2].

Users of heroin, particularly those who inject subcutaneously (i.e., skin popping), have a higher incidence of wound botulism. Skin popping can create an anaerobic environment under the skin, which facilitates the germination and release of BoNT. Because it is linked to skin popping, black tar heroin is the suspected source of *C. botulinum* spores. Currently, it is unknown exactly how BTH becomes contaminated with *C. botulinum* spores, but there are many opportunities: the drug is often cut with other substances, the drug travels long distances from source to user, the use of dirty needles or equipment, and the drug may be manipulated by the user prior to injection (Davis and King, 2008). Culturing *C. botulinum* from BTH is challenging, as it has a sticky composition that makes it difficult to solubilize in buffers and standard culture media. One of the major challenges for testing BTH is that, because it is a schedule I controlled substance by the United States Drug Enforcement Agency (DEA), a laboratory intending to perform microbiological testing of heroin requires a DEA license and adherence to a strict chain of custody and inventory documentation (United States Department of Justice Drug Enforcement Administration Office of Diversion Control, 2006).

The CDC's National Botulism Laboratory received isolates from a wound botulism case in Hawaii for confirmation testing. The patient had a history of BTH use, specifically *via* skin-popping daily for 2 weeks prior to hospital admittance. Due to the patient's history, it is believed that the wound botulism occurred from injection drug use. Based on publicly available CDC surveillance data[2], this was the first reported laboratory-confirmed case of wound botulism from Hawaii. The isolates were confirmed as *C. botulinum* serotype B. Our analysis of CDC surveillance data for the United States between 1981 and 2016 showed that about 77% (n = 420) of reported wound cases in the United States were due to serotype A, 12% (n = 65) were attributed to serotype B, and 11% did not have a serotype reported (n = 61) (9). Due to the novel report of wound botulism from Hawaii, the isolated geography of the case, and the rarity of serotype B wound botulism, we sought to determine whether the *C. botulinum* type B isolated from this wound botulism case resembled other *C. botulinum* type B isolates from infant botulism cases (approximately 0–4 confirmed cases per year, 1981–2016) in Hawaii (Nevas et al., 2005). *C. botulinum* isolates from infant cases were chosen, as it has been long recognized to be caused by spores from the local environment which become ingested by babies under 1 year old [Centers for Disease Control and Prevention (CDC), 1998]. As a result, these isolates could be considered representative of the local geography in which they occur.

[1]https://wwwn.cdc.gov/nndss/conditions/botulism/case-definition/2011/

[2]http://www.cdc.gov/nationalsurveillance/botulism-surveillance.html

A single *C. botulinum* type B isolate from an infant botulism case in California, United States, was included, as it is a common thoroughfare for travel to Hawaii, and it was the geographically closest serotype B environmental isolate found within our collection.

MATERIALS AND METHODS

Microbiology

The strains used in this study are described in **Table 1**. We recovered each strain from long-term storage by inoculating 0.5–1.0 ml of sporulation media (20 g/L peptone, Difco, Franklin Lakes, NJ, United States; beef brain, Pel-Freez Biologicals; Rogers, AR, United States) into chopped meat glucose starch broth (CMGS) (Remel; Lenexa, KS, United States), and each grew in a Coy (Grass Lake, MI, United States) anaerobic chamber at 35 ± 2°C for 24–48 h. We examined CMGS cultures for growth and quadrant streaked ∼0.5 ml of CMGS culture for isolation onto McClung Toabe egg yolk agar with yeast extract agar plates (McClung Toabe agar base, 75 g/L; yeast extract, 5 g/L; egg yolk enrichment, 100 ml/L—all from Difco, Franklin Lakes, NJ, United States). The plates grew anaerobically at 35 ± 2°C for 24–48 h and were examined for purity. We picked single colonies that exhibited lipase activity and inoculated them into trypticase peptone glucose yeast extract broth (TPGY) (Remel, Lenexa, KS, United States). These cultures grew anaerobically at 35 ± 2°C for 16–24 h prior to genomic DNA extraction.

Genomic DNA Extraction

We used a modified Epicenter (Madison, WI, United States) MasterPure Complete DNA and RNA Purification kit to extract genomic DNA from 8 to 9 ml of turbid TPGY culture. Briefly, cells were pelleted at 4 ± 1°C for 10 min at 4,000 rpm, and the supernatant was discarded. We resuspended cell pellets in lysozyme stock solution (25 mM Tris–HCl, pH 8.0, Invitrogen, Waltham, MA, United States; 2.5 mM 0.5 M EDTA, Invitrogen, Waltham, MA, United States; 10 ml Triton X-100, Sigma, St. Louis, MO, United States; and 20 mg/ml lysozyme from chicken egg white, Sigma, St. Louis, MO, United States) and incubated them for a minimum of 15 min in a 37 ± 1°C water bath. We added 300 μl of undiluted 2X T&C buffer (Epicenter, Madison, WI, United States) and 3 μl of RNase A (Qiagen, Redwood City, CA, United States) to the cell suspension, gently mixed it, and then incubated it in a 57 ± 1°C water bath for 10 min. To this mixture, we added 3 μl of Proteinase K (Invitrogen, Waltham, MA, United States) and incubated it in the water bath for another 10 min at 57 ± 1°C. After incubation with proteinase, we added 350 μl of MPC protein precipitation buffer (Epicenter, Madison, WI, United States) to the solution and centrifuged it at 4,000 rpm at 4 ± 1°C for 10 min. We collected the supernatants and added each to 1 ml of 99% isopropanol to precipitate the DNA. We collected the precipitates, washed them with 1 ml of 70% ethanol, and rehydrated them at least overnight (up to 1 week) in 200 μl of 10 mM Tris–HCl (Invitrogen, Waltham, MA, United States). We filtered the rehydrated gDNA through 0.1-μM centrifugal filters (MilliporeSigma, Burlington, MA, United States) to remove any spores or unlysed cells. We used the Nanodrop 2000 (NanoDrop

TABLE 1 | Summary of *Clostridium botulinum* producing toxin serotype B strains used in this study: year isolated, botulism type, originating state, specimen source, toxin subtype, and seven-gene multi-locus sequence type (MLST).

Isolate ID #	Year	Botulism type	State	Specimen type	Toxin subtype	Seven-gene MLST
CDC21601	1976	Infant	CA	Stool	B1	30
CDC31747	1986	Infant	HI-Oahu	Stool	B5	36
CDC34293	1989	Infant	HI-Maui	Stool	B5	36
CDC36757	1981	Infant	HI-Oahu	Stool	B5	36
CDC37391	1982	Infant	HI-Oahu	Stool	B5	36
CDC38839	1983	Infant	HI-Oahu	Stool	B5	36
CDC39168	1984	Infant	HI-Maui	Stool	B5	36
CDC40176	1995	Infant	HI-Oahu	Stool	B5	36
CDC41623	1996	Infant	HI-Oahu	Stool	B5	36
CDC45459	1990	Infant	HI-Maui	Stool-enema	B5	36
CDC47455	1992	Infant	HI-Maui	Stool	B5	36
CDC48611	1993	Infant	HI-Kaua'i	Stool	B1	110
CDC49917	1994	Infant	HI-Oahu	Stool	B5	36
CDC53044	2008	Infant	HI-Maui	Stool	B5	36
CDC54117	2009	Infant	HI-Oahu	Stool	B5	36
CDC54250	2009	Infant	HI-Oahu	Stool	B5	36
CDC59947	2004	Infant	HI-Maui	Stool	B5	36
CDC60225	2015	Infant	HI-Hawaii	Stool	B5	36
CDC61035	2016	Wound	HI-Oahu	Wound	B5	36
CDC65069	2010	Infant	HI-Maui	Stool	B5	104

Technologies, Wilmington, DE, United States) and the Qubit 4 (Invitrogen, Waltham, MA, United States) fluorometer high-sensitivity assay to assess genomic DNA quality and quantity according to the manufacturers' instructions.

Whole-Genome Shotgun Sequencing

We constructed barcoded shotgun libraries using NextFlex DNA barcodes (BIOO Scientific, Austin, TX, United States) and the 400-bp Kapa Biosciences (Wilmington, MA, United States) kit for Ion Torrent chemistry (KAPA Biosystems, 2016). We performed size selection using the e-gel system to select for 500-bp fragments (Life Technologies, 2017). We diluted and pooled completed libraries to obtain an equimolar concentration of 200 pM and then templated and enriched using the Ion Chef instrument (Life Technologies), followed by sequencing on Ion Torrent S5 (Life Technologies, 2019).

Bioinformatics and Quality Control

We assessed read quality using FastQC v.0.11.5 (Andrews, 2014) and assembled reads using SPAdes v.3.10.1 with the following parameters set: sc, iontorrent, single end, and careful (Bankevich et al., 2012). We assessed the resulting assemblies with Quast v.4.3 (Gurevich et al., 2013). We used the map reads to reference tool in CLC Genomic Workbench v.10.1.1 to determine toxin gene subtypes as well as identify accessory genes HA, BotR, and ntnH. We identified the legacy seven-gene multi-locus sequence types (MLST) by querying draft genomes against the PubMLST (Jacobson et al., 2008) database with the Center for Genomic Epidemiology website[3]. We used Mashtree v.0.37 (Katz et al., 2019) to determine pairwise mash distances (Ondov et al., 2016) and place the isolates in a neighbor-joining tree of reference sequences and other serotype B sequences from the CDC short read sequence collection. We used pairwise average nucleotide identity (ANI) (Goris et al., 2007) to identify the closest reference sequences as well as to compare pairwise sequence homologies. We identified the nearest neighbor by ANI, and this reference sequence was used in Lyve-SET v.1.1.4f (Katz et al., 2017) to determine high-quality single-nucleotide polymorphism (hqSNP) sites across the study sequences and then draw a tree with bootstrap support to approximate a phylogeny (Lyve-SET settings enabled: single end, allowed Flanking 5, min_alt_frac 0.75, min_coverage 10, mask-phages, mask-cliffs, read_cleaner = CGP, mapper = smalt). Using these settings, Lyve set defines high-quality SNPs as those in regions of contiguous coverage of at least 10, present in at least 75% of the reads, and not in known phage sequences.

Data Availability

The Short Read Archive accession numbers are as follows: SHBZ00000000, SHCA00000000, SHCB00000000, SHCC00000000, SHCD00000000, SHCE00000000, SHCF00000000, SHCG00000000, SHCH00000000, SHCI00000000, SHCJ00000000, SHCK00000000, SHCL00000000, SHCM00000000, SHCN00000000, SHCO00000000, SHCP00000000, SHCQ00000000, SHCR00000000, and SHCS00000000.

[3]http://www.genomicepidemiology.org/

RESULTS

The resulting number of reads, GC content, and estimated genome size are found in **Table 2**. The number of resulting reads was 227,559 to 2,461,814; GC content ranged from 27.7 to 28.0%; reads assembled into 35–257 contigs; N50s ranged from 26,682 to 277,264; average coverage ranged from 20.4× to 227.7×; and approximate assembled genome size was 3.9 to 4.2 mega-bases. Subtypes of Hawaii *C. botulinum* infant or wound isolates were determined to be *bont*/B5 (n = 18) and *bont*/B1 (n = 1), with an average coverage across the gene of 23× (3–51×). Querying the PubMLST database, which generates an allelic profile or seven-gene MLST sequence type (ST), revealed that the strains harboring *bont*/B5 gene were all members of ST-36 except one, which was ST-104. The *C. botulinum* isolate from Hawaii harboring *bont*/B1 gene was a member of ST-110, and the *C. botulinum* isolate from California harboring a *bont*/B1 gene was a member of ST-30 (**Table 1**).

Mashtree placed the *C. botulinum* bont/B5 isolates from Hawaii into a single cluster, distant from bivalent strains and other strains that harbor the *bont*/B5 gene (**Figure 1**). The sole California isolate as well as the isolate from Hawaii harboring *bont*/B1 were members of the *bont*/B1 cluster. ANI corroborated the mashtree results by identifying CDC67071 (RefSeq accession numbers NZ_CP013242.1 and NZ_CP013241.1) as the closest closed reference strain to the Hawaii cluster and verifying that the *C. botulinum* type B isolates from Hawaii are more closely related to each other than to other isolates tested by pairwise ANI.

We used Lyve-SET with CDC67071 as the reference to approximate a phylogeny of the *C. botulinum* isolates from Hawaii that harbor *bont*/B5 (**Figure 2**). Lyve-SET used 10,424 hqSNP sites to approximate a phylogeny which resulted in 4–10,103 pairwise hqSNP differences between the sequences. Within this tree, two clades formed: clade 1 containing CDC65069 (ST-104, harboring bont/B5 gene) from an infant botulism case and clade 2 containing CDC31747, CDC34293, CDC36757, CDC37391, CDC38839, CDC38168, CDC40176, CDC41623, CDC45459, CDC47455, CDC49917, CDC53044, CDC54117, CDC54250, CDC59947, CDC60225, and CDC61035 (wound isolate). Clade 2 contains the remainder of the *bont*/B5 isolates from Hawaii, all belonging to ST36 and all from infant cases except for the single wound botulism isolate (as noted). Clade 1 is separated from clade 2 by 10,103 hqSNP differences, and 4–29 hqSNP differences are found within clade 2.

For these 18 *C. botulinum* type B isolated from botulism cases from Hawaii, the *bont*/B5 gene is located on a plasmid. We found nothing notable about the hemagglutinins or the *bot*R gene, but the *ntnh* gene has a 39-bp insert. This insertion was not in the *ntnh* gene of isolates harboring bont/B1 that we investigated herein nor in reference strains Okra, CDC67071, CDC69094, or CDC1436. This addition does not disrupt the frame of the gene.

TABLE 2 | Summary of the sequencing statistics for *Clostridium botulinum* strains isolated from infant and wound botulism cases in Hawaii and California.

Isolate ID #	GC (%)	# contigs	Largest contig (bp)	Total length (bp)	N50 (bp)	# reads	Average read length (bp)	Average coverage	# reads mapping to *bont*	Average coverage across *bont*
CDC21601	28.0	55	686,068	3,978,905	277,264	1,916,257	284	143.2	965	69.31
CDC31747	27.7	64	594,898	4,15,0368	224,881	2,461,814	293	180.3	649	44.26
CDC34293	27.7	85	594,976	4,227,167	118,018	1,593,628	293	116.7	333	23.02
CDC36757	27.7	35	653,290	4,156,210	255,842	2,567,944	337	227.7	642	50.47
CDC37391	27.7	79	278,989	4,139,909	128,587	522,334	333	45.8	120	9.36
CDC38839	27.7	41	501,509	4,158,619	301,466	849,846	307	68.7	209	15.24
CDC39168	27.8	193	177,384	4,141,306	46,127	895,235	282	63.1	158	10.25
CDC40176	27.7	116	416,804	4,158,067	75,335	626,019	207	32.4	61	3.24
CDC41623	27.7	59	540,802	4,217,736	194,214	1,283,745	227	72.9	368	20.62
CDC45459	27.7	87	322,199	4,198,117	111,949	2,068,237	254	131.3	498	29.75
CDC47455	27.7	80	273,668	4,199,837	114,010	1,361,983	262	89.2	310	19.07
CDC48611	28.0	122	253,643	3,843,066	53,585	624,594	274	42.8	286	18.95
CDC49917	27.7	252	104,953	4,163,377	26,682	757,736	200	37.9	888	44.96
CDC53044	27.8	86	587,411	4,331,526	225,901	1,788,944	308	137.7	452	33.91
CDC54117	27.7	97	213,589	4,203,614	90,518	648,044	232	37.6	172	9.66
CDC54250	27.7	257	148,760	4,162,820	32,556	227,559	340	20.4	43	3.51
CDC59947	27.7	131	215,257	4,187,706	65,567	615,768	273	42.0	181	12.55
CDC60225	27.7	68	439,777	4,333,055	142,003	1,031,444	334	90.7	254	12.63
CDC61035	27.7	78	430,766	4,153,245	133,216	845,541	288	60.9	209	13.69
CDC65069	27.7	41	636,062	4,186,206	255,806	896,025	334	78.8	227	15.53

DISCUSSION

The results herein suggest that the heroin that is presumed to be the causative agent of wound botulism was contaminated with *C. botulinum* spores locally rather than in manufacturing or transit. This was the first reported wound case from Hawaii and was confirmed in the laboratory as toxin type B. The CDC surveillance data (1981–2016) reports a total of 61 confirmed cases of botulism from Hawaii (53 infant botulism, seven foodborne botulism, and one wound botulism). Serotype B was confirmed for 56 out of the 61 total cases; the remaining five were infant botulism cases due to serotype A^{N1}.

Black tar heroin was first introduced into the United States from Mexico in the 1970s and was predominantly used in California. According to the Hawaii Drug Threat Assessment, the main distribution hub for BTH to Hawaii is through Los Angeles, CA (National Drug Intelligence Center, 2002). Heroin is typically cut where it is distributed; however, it is still unclear if contamination with *C. botulinum* spores occurs from local or distant geographic sources (Passaro et al., 1998; Werner et al., 2000; Peak et al., 2019).

Because of the high number of type B botulism cases in Hawaii, we had the opportunity to select and compare 18 *C. botulinum* type B isolates from the CDC strain collection associated with infant botulism. These isolates were selected for this study to aid in determining if this *C. botulinum* type B isolate from a wound botulism case is phylogenetically related to other locally acquired Hawaiian *C. botulinum* isolates. Infants may acquire *C. botulinum* spores by swallowing dust particles, and thus *C. botulinum* isolated from infant botulism cases can be viewed as environmental representatives of endogenous strains (Nevas et al., 2005; Fleck-Derderian et al., 2017). We also included a California *C. botulinum* type B isolate because the import path for BTH to Hawaii is through California. This isolate is also the only "environmentally acquired" *C. botulinum* type B isolate from the west coast in the CDC historic strain collection.

Eighteen out of the 19 *C. botulinum* type B isolates from botulism cases from Hawaii harbored the *bont*/B5 gene, including the wound isolate (CDC61035). Previous studies investigating *C. botulinum* genetic diversity (Hill et al., 2007; Luquez et al., 2009; Williamson et al., 2016) indicate that a large portion of *bont*/B5-harboring strains has been identified in bivalent strains (types Ab, Ba, and Bf) or type A isolates that harbor a silent B gene [designated as A(B) strains]. These earlier studies indicate that A(B) strains typically cluster together due to their highly conserved sequences, and bivalent strains display more variability within their sequences; however, they still cluster closely. It is known that the *bont*/B5 gene can be found alone (Franciosa et al., 2009; Kenri et al., 2014), and our observation is that the occurrence of *bont*/B5 with no other toxin gene present is more common than once thought (unpublished observations). Williamson et al. (2016) reported that a *C. botulinum* type B strain that harbors only a *bont*/B5 gene still clusters closely to bivalent strains in core genome phylogenies. There have not been extensive studies published studying the evolution of *C. botulinum* strains over time as they persist in the environment.

For these 18 *C. botulinum* type B isolated from botulism cases from Hawaii, the *bont*/B5 gene is located on a plasmid. Each of the contigs containing the *bont* gene was compared against the NCBI nucleotide database using blastn, and the group most

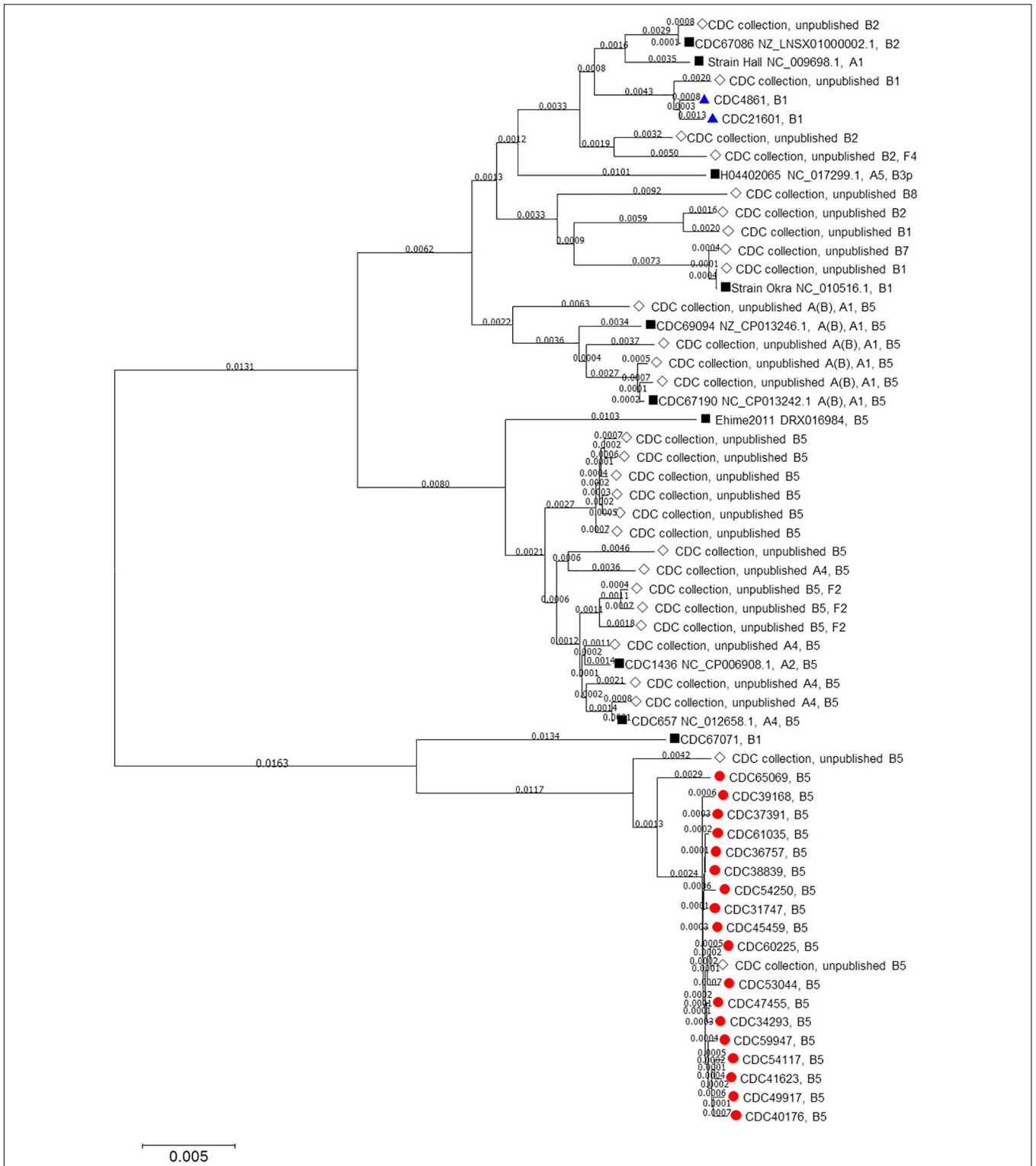


FIGURE 1 | Neighbor-joining tree drawn using MASH distances. The tree includes *Clostridium botulinum* isolates harboring *bont*/B5 (red circles) and *bont*/B1 (blue triangles) genes as described in **Table 1** as well as NCBI reference sequences (black squares) and other unpublished isolates from the CDC collection (white diamonds).

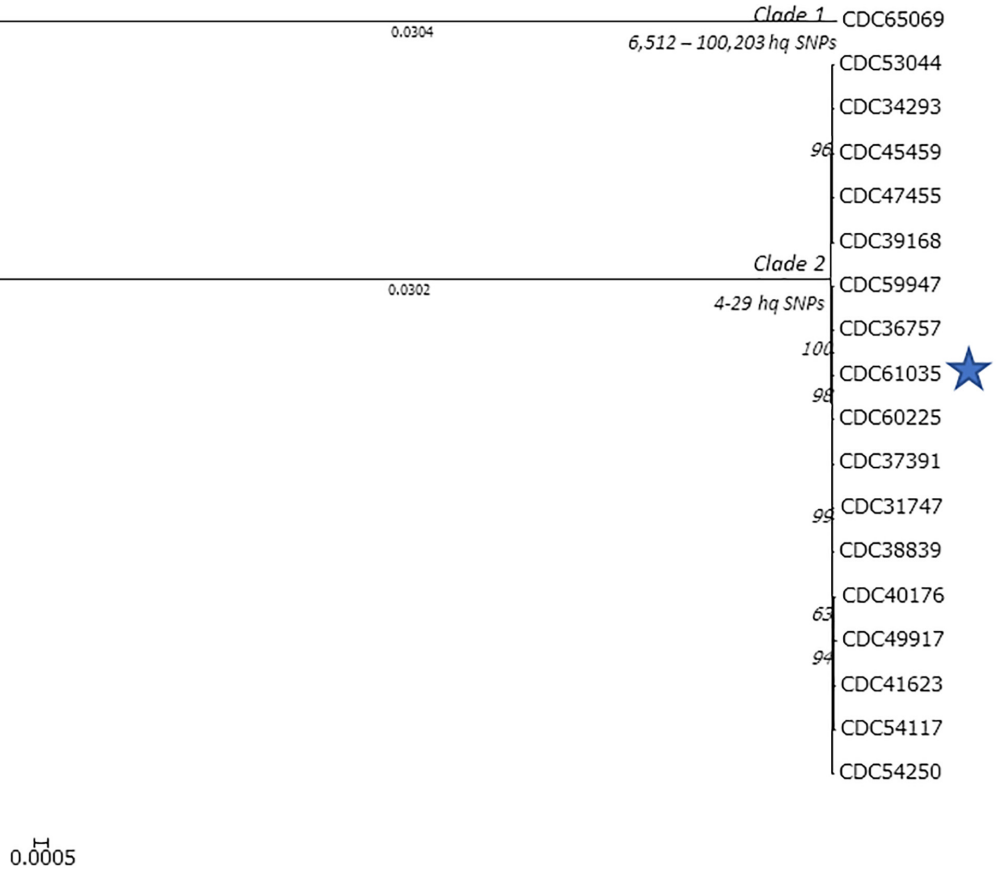


FIGURE 2 | Lyve-SET phylogenetic tree of the *Clostridium botulinum* isolates harboring *bont*/B5 from Hawaii. The blue star represents the isolate from the wound botulism case. Branch lengths < 0.0001 are hidden, and bootstrap values are shown in italics. Lyve-SET was run with the following settings enabled: single end, allowed flanking = 5, min alt frac = 0.75, minimum coverage = 10, mask-phages, mask-cliffs, read cleaner = CGP, mapper = smalt. The reference used to identify high-quality single-nucleotide polymorphism sites was CDC67071 (NZ_CP013242.1).

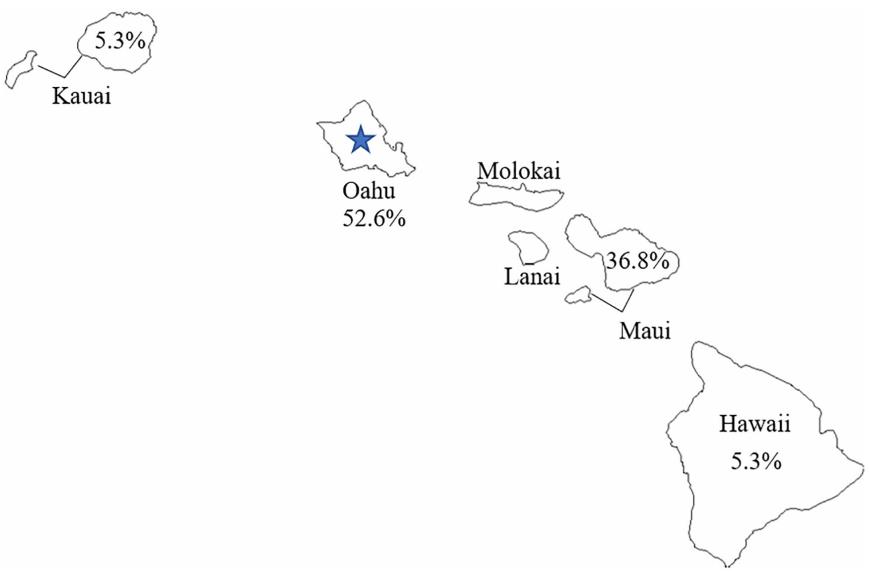


FIGURE 3 | Geographic distribution of *Clostridium botulinum* type B used in this study. The blue star represents the reported origin of the wound botulism case.

closely resembled pCLJ from bivalent strain CDC657, which harbors the *bont*/A4 and *bont*/B5 genes. There was nothing notable about most of the accessory genes (hemagglutanins and BotR), but there is a 39-bp insert in the *ntnh* gene that will not disrupt the reading frame. This insertion is also present in *C. botulinum* strain CDC657 *bont*/B-associated *ntnh* gene (accession #EU341304) but not in the *ntnh* gene of isolates harboring *bont*/B1 that we investigated herein nor in reference strains Okra, CDC67071, CDC69094, or CDC1436. This addition does not disrupt the frame of the gene, allowing a fully functional protein to be produced.

A neighbor-joining tree (**Figure 1**) of all the isolates used in this study was constructed using their Mash distances along with publicly available *C. botulinum* reference sequences and sequences of other *C. botulinum* type B isolates from the CDC collection. As indicated in **Figure 1**, **17** isolates harboring *bont*/B5 gene, including the wound isolate (CDC61035), formed their own cluster away from other *bont*/B5 harboring strains. The MLST analysis revealed that these 17 isolates belonged to ST-36. The other isolate harboring *bont/B5* gene, CDC65069, clustered just outside the other 17 isolates and belonged to ST-104. These two sequence types, ST-36 and ST-104, differ at the *recA* and *hsp* loci. These results show that seven-gene MLST and MASH tree clustering provided congruent subtyping results.

One *C. botulinum* type B isolated from an infant botulism case from Hawaii (CDC48611) and the *C. botulinum* type B isolated from an infant botulism case from California (CDC21601) were determined to be subtype *bont*/B1. Both isolates clustered together and away from isolates harboring *bont*/B5 gene. The seven-gene MLST analysis showed CDC48611 as belonging to ST110 and CDC21601 as belonging to ST30. These two sequence types differ by only one locus, at *aceK*. Based on these limited data, it is unclear whether isolates harboring *bont*/B1 are common within Hawaii or whether the Hawaii *bont*/B1 case was due to travel from mainland United States.

LYVE-Set was used to approximate the phylogeny among the 18 *C. botulinum* isolates from Hawaii harboring *bont*/B5 gene (**Figure 2**). LYVE-Set resulted in two separate clades: one consisting of 17 isolates with 4–29 high -quality SNP differences among them; all these isolates belong to MLST ST-36. The small number of SNPs observed among members of clade 1 indicates a high homology. Clade 2 was separated from clade 1 by 10,103 high-quality SNPs, with only one member, CDC65069, which belongs to ST-104. The LYVE-set results indicate a high homology between the wound isolate (CDC61035) and the other isolates from infant botulism cases, which suggests that the *C. botulinum* type B isolated from a wound botulism case is closely related to other *C. botulinum* isolates present in the environment in Hawaii.

Attempts to identify how BTH becomes contaminated with *C. botulinum* spores have been unsuccessful. Previous studies conducted by the California Department of Public Health suggest that contamination of BTH with *C. botulinum* spores may occur during the cutting and diluting process (Passaro et al., 1998), and a study by Peak et al. suggests that BTH is typically cut where it is distributed (Peak et al., 2019). The Hawaii Drug Threat Assessment indicates that Oahu is the point of entry for Hawaii and is also a major distribution hub. It is also suggested that heroin is cut in Oahu before supplying to sellers and/or users (National Drug Intelligence Center, 2002).

The geographic distribution of *C. botulinum* isolates used in this study is relatively broad. **Figure 3** shows the percentage of isolates used in the study per Hawaiian island. Oahu had the highest representation of isolates from infant botulism cases used in this study, and it was also the origin of the wound isolate (CDC61035). The close phylogenetic relationship between the *C. botulinum* type B isolated from the wound botulism case and the isolates from local infant botulism cases and the geographic distribution of these isolates in Hawaii strongly supports the hypothesis that this wound botulism case from Hawaii was due to locally acquired *C. botulinum* spores rather than from contamination during BTH production from a distant source.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by CDC Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JH and CL conceived the study. JH and VF performed the sequencing and data analysis. JD performed the original microbiology and identification of strains. KC-S provided the epidemiological review. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

Centers for Disease Control and Prevention, Center for Preparedness Response kindly provided support for this project. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.678473/full#supplementary-material

REFERENCES

Andrews, S. (2010). *FastQC A Quality Control Tool for High Throughput Sequence Data [Online]*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 2020).

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

Barash, J. R., and Arnon, S. S. (2014). A novel strain of Clostridium botulinum that produces type B and type H botulinum toxins. *J. Infect. Dis.* 209, 183–191. doi: 10.1093/infdis/jit449

Brunt, J., Carter, A. T., Stringer, S. C., and Peck, M. W. (2018). Identification of a novel botulinum neurotoxin gene cluster in Enterococcus. *FEBS Lett.* 592, 310–317. doi: 10.1002/1873-3468.12969

Centers for Disease Control and Prevention (CDC) (1998). *Botulism In The United States (1899–1996). Handbook For Epidemiologists, Clinicians, And Laboratory Workers*. Atlanta: U.S. Department of Health and Human Services, CDC.

Chin, J., Arnon, S. S., and Midura, T. F. (1979). Food and environmental aspects of infant botulism in California. *Rev. Infect. Dis.* 1, 693–697. doi: 10.1093/clinids/1.4.693

Davis, L. E., and King, M. K. (2008). Wound botulism from heroin skin popping. *Curr. Neurol. Neurosci. Rep.* 8, 462–468. doi: 10.1007/s11910-008-0074-2

Fleck-Derderian, S., Shankar, M., Rao, A. K., Chatham-Stephens, K., Adjei, S., Sobel, J., et al. (2017). The Epidemiology of Foodborne Botulism Outbreaks: a Systematic Review. *Clin. Infect. Dis.* 66, S73–S81.

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

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

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

Jacobson, M. J., Lin, G., Raphael, B., Andreadis, J., and Johnson, E. A. (2008). Analysis of neurotoxin cluster genes in Clostridium botulinum strains producing botulinum neurotoxin serotype A subtypes. *Appl. Environ. Microbiol.* 74, 2778–2786. doi: 10.1128/aem.02828-07

KAPA Biosystems. (2016). *KAPA Library Preparation Kit; Ion Torrent Platforms*. Wilmington: KAPA Biosystems.

Katz, L. S., Griswold, T., Morrison, S., Caravas, J., Zhang, S., Bakker, H., 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

Kenri, T., Sekizuka, T., Yamamoto, A., Iwaki, M., Komiya, T., Hatakeyama, T., et al. (2014). Genetic characterization and comparison of Clostridium botulinum isolates from botulism cases in Japan between 2006 and 2011. *Appl. Environ. Microbiol.* 80, 6954–6964. doi: 10.1128/aem.02134-14

Life Technologies (2017). *Ion 520 & Ion 530 Kit – Chef USER GUIDE Revision D.0 [Online]. Thermo Fisher Scientific*. Available online at: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0010846_Ion_520_530_Kit_Chef_UG.pdf (Accessed August 9, 2018).

Life Technologies (2019). *Ion 510™ & Ion 520™ & Ion 530™ Kit – Chef USER GUIDE. Instructions for automated template preparation, chip loading, and sequencing*. Carlsbad: Life Technologies Corporation.

Luquez, C., Raphael, B. H., and Maslanka, S. E. (2009). Neurotoxin gene clusters in Clostridium botulinum type Ab strains. *Appl. Environ. Microbiol.* 75, 6094–6101. doi: 10.1128/aem.01009-09

Mansfield, M. J., Adams, J. B., and Doxey, A. C. (2015). Botulinum neurotoxin homologs in non-Clostridium species. *FEBS Lett.* 589, 342–348. doi: 10.1016/j.febslet.2014.12.018

National Drug Intelligence Center. (2002). *Hawaii Drug Threat Assessment*. Pennsylvania: National Drug Intelligence Center.

Nevas, M., Lindstrom, M., Virtanen, A., Hielm, S., Kuusi, M., Arnon, S. S., et al. (2005). Infant botulism acquired from household dust presenting as sudden infant death syndrome. *J. Clin. Microbiol.* 43, 511–513. doi: 10.1128/jcm.43.1.511-513.2005

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17:132.

Passaro, D. J., Werner, S. B., Mcgee, J., Mac Kenzie, W. R., and Vugia, D. J. (1998). Wound botulism associated with black tar heroin among injecting drug users. *JAMA* 279, 859–863. doi: 10.1001/jama.279.11.859

Peak, C. M., Rosen, H., Kamali, A., Poe, A., Shahkarami, M., Kimura, A. C., et al. (2019). Wound Botulism Outbreak Among Persons Who Use Black Tar Heroin - San Diego County, California, 2017-2018. *MMWR Morb. Mortal. Wkly. Rep.* 67, 1415–1418. doi: 10.15585/mmwr.mm675152a3

Peck, M. W., Smith, T. J., Anniballi, F., Austin, J. W., Bano, L., Bradshaw, M., et al. (2017). Historical Perspectives and Guidelines for Botulinum Neurotoxin Subtype Nomenclature. *Toxins* 9:38. doi: 10.3390/toxins9010038

Popoff, M. R. (2018). Botulinum Neurotoxins: still a Privilege of Clostridia?. *Cell Host Microbe* 23, 145–146. doi: 10.1016/j.chom.2018.01.014

Rossetto, O., Pirazzini, M., and Montecucco, C. (2014). Botulinum neurotoxins: genetic, structural and mechanistic insights. *Nat. Rev. Microbiol.* 12, 535–549. doi: 10.1038/nrmicro3295

United States Department of Justice Drug Enforcement Administration Office of Diversion Control. (2006). *Practitioner's Manual: An Informational Outline of the Controlled Substances Act*. Washington: United States Department of Justice.

Werner, S. B., Passaro, D., Mcgee, J., Schechter, R., and Vugia, D. J. (2000). Wound botulism in California, 1951-1998: recent epidemic in heroin injectors. *Clin. Infect. Dis.* 31, 1018–1024. doi: 10.1086/318134

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fmicb.2021.711413

Exploration of the Diversity of Clustered Regularly Interspaced Short Palindromic Repeats-Cas Systems in *Clostridium novyi sensu lato*

Thibault Le Gratiet[1,2†], Caroline Le Marechal[1†], Marie Devaere[1], Marianne Chemaly[1] and Cédric Woudstra[3]*

[1]Hygiene and Quality of Poultry and Pig Products Unit, ANSES, French Agency for Food, Environmental and Occupational Health Safety, Ploufragan, France, [2]UFR of Life Sciences and Environment, University of Rennes, Rennes, France, [3]Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg, Denmark

OPEN ACCESS

Edited by:
Theresa Smith,
Northern Arizona University,
United States

Reviewed by:
Raul Raya,
Centro de Referencia para Lactobacilos (CERELA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina
Maria Carolina Quecine,
University of São Paulo, Brazil

***Correspondence:**
Caroline Le Marechal
caroline.lemarechal@anses.fr

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 18 May 2021
Accepted: 16 August 2021
Published: 13 September 2021

Citation:
Le Gratiet T, Le Marechal C, Devaere M, Chemaly M and Woudstra C (2021) Exploration of the Diversity of Clustered Regularly Interspaced Short Palindromic Repeats-Cas Systems in Clostridium novyi sensu lato. Front. Microbiol. 12:711413. doi: 10.3389/fmicb.2021.711413

Classified as the genospecies *Clostridium novyi sensu* lato and distributed into four lineages (I–IV), *Clostridium botulinum* (group III), *Clostridium novyi*, and *Clostridium haemolyticum* are clostridial pathogens that cause animal diseases. *Clostridium novyi sensu lato* contains a large mobilome consisting of plasmids and circular bacteriophages. Here, we explored clustered regularly interspaced short palindromic repeats (CRISPR) arrays and their associated proteins (Cas) to shed light on the link between evolution of CRISPR-Cas systems and the plasmid and phage composition in a study of 58 *Clostridium novyi sensu lato* genomes. In 55 of these genomes, types I-B (complete or partial), I-D, II-C, III-B, III-D, or V-U CRISPR-Cas systems were detected in chromosomes as well as in mobile genetic elements (MGEs). Type I-B predominated (67.2%) and was the only CRISPR type detected in the Ia, III, and IV genomic lineages. Putative type V-U CRISPR Cas14a genes were detected in two different cases: next to partial type-IB CRISPR loci on the phage encoding the botulinum neurotoxin (BoNT) in lineage Ia and in 12 lineage II genomes, as part of a putative integrative element related to a phage-inducible chromosomal island (PICI). In the putative PICI, Cas14a was associated with CRISPR arrays and restriction modification (RM) systems as part of an accessory locus. This is the first time a PICI containing such locus has been detected in *C. botulinum*. Mobilome composition and dynamics were also investigated based on the contents of the CRISPR arrays and the study of spacers. A large proportion of identified protospacers (20.2%) originated from *Clostridium novyi sensu lato* (p1_Cst, p4_BKT015925, p6_Cst, CWou-2020a, p1_BKT015925, and p2_BKT015925), confirming active exchanges within this genospecies and the key importance of specific MGEs in *Clostridium novyi sensu lato*.

Keywords: CRISPR-Cas, type V-U CRISPR, Cas14a, phage-inducible chromosomal island, protospacers, mobilome, *Clostridium botulinum* group III

INTRODUCTION

Clostridium novyi sensu lato is a genospecies, containing the closely related species *Clostridium botulinum* (group III), *Clostridium novyi*, and *Clostridium haemolyticum* (Skarin et al., 2011; Skarin and Segerman, 2014). This proposed genospecies (Skarin et al., 2011) is based on genomic comparisons: their chromosome is highly conserved and the differences between these three species can be attributed to the content of their mobile genetic elements (MGEs; Skarin and Segerman, 2011). These three clostridia are ubiquitous, anaerobic, Gram-positive, and spore-forming pathogens that can affect both humans and animals. Botulinum neurotoxins (BoNTs) produced by *C. botulinum* group III are responsible for animal botulism. Alpha, beta, and gamma toxins produced by *Clostridium novyi* cause gas gangrene in humans and animals and black disease in animals, mostly in sheep. The beta toxin produced by *C. haemolyticum* is responsible for bacillary hemoglobinuria affecting ruminants (Skarin and Segerman, 2014). These virulence genes are all carried on MGEs (plasmids and phages) as part of their mobilome. The *Clostridium novyi sensu lato* mobilome is highly diverse: More than 60 plasmids or circular bacteriophages (called phages hereafter) categorized into 13 groups were detected upon sequencing only 24 genomes (Skarin and Segerman, 2014). Considering the few genome sequences available in the public databases, exploring further *Clostridium novyi sensu lato* MGE diversity is difficult. Yet, clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas), together making up the CRISPR-Cas system, which records MGEs past encounters, could help to shed light on a putative larger mobilome within *Clostridium novyi sensu lato*.

Found in bacteria and archaea, CRISPR–Cas is a prokaryotic adaptive immunity system that interferes with invading phages and plasmids (Koonin and Makarova, 2019). CRISPR systems are composed of CRISPR arrays, divided into repeats and spacers (short variable DNA sequences), and *cas*-associated genes. Spacers, which are of foreign origin (protospacers), are integrated into the CRISPR locus, thereby acting as a memory of previous infections. Analysis of these spacers can give clues as to the mobile elements encountered in the past by a given organism (Mcginn and Marraffini, 2019), thus providing information on the identity of horizontally transferable elements and their interaction with their bacterial hosts in their environment. The immune function of CRISPR is carried out by Cas proteins in three successive steps (adaptation, expression, and interference; Briner et al., 2015; Makarova et al., 2020; Pan et al., 2020). Briefly, during the adaptation stage, a part of the foreign DNA (i.e., protospacers) is incorporated into CRISPR arrays. Then, these CRISPR arrays are transcribed to generate short processed CRISPR RNA (crRNA). The last stage consists of crRNA guiding Cas proteins to their DNA target and finally cleavage of the target. With the ever-increasing number of available bacterial genomes, the classification of CRISPR-Cas systems is constantly evolving (Makarova et al., 2020). Currently, two classes, six types, and more than 20 subtypes of CRISPR-Cas systems have been described based on *cas* gene content (Mcginn and Marraffini, 2019; Makarova et al., 2020).

Up to now, only one study has explored *in silico C. botulinum* CRISPR-Cas systems, with an analysis of 20 genomes (mainly draft sequences) demonstrating 80% prevalence of CRISPR loci in *C. botulinum* (Negahdaripour et al., 2017). Most of the studied strains contained several CRISPR loci, with one-fifth located on plasmids, although CRISPR loci are commonly found on circular chromosomes in bacteria (Negahdaripour et al., 2017). Once again, this particularity attests to the importance of the *C. botulinum* mobilome. Only eight strains of *C. botulinum* group III (including seven BoNT type C/D and one type D, from genomic lineages Ia, Ib, III, and IV) were included in this study and led to the detection of CRISPR loci and CRISPR arrays, demonstrating the existence of CRISPR systems in *C. botulinum* group III strains (Negahdaripour et al., 2017). However, with only a few genome sequences analyzed, the full diversity of the CRISPR content of *C. botulinum* group III has not been fully explored. Further investigations are therefore required to better characterize the CRISPR-Cas systems and their spacers memory in *C. botulinum* group III genomes and more generally in the *Clostridium novyi sensu lato* genospecies. The objective of our study was thus to explore the CRISPR-Cas systems in the *Clostridium novyi sensu lato* genospecies to evaluate their presence, determine their characteristics, and explore the protospacer origins to gain insight into the MGEs interacting in this taxon.

MATERIALS AND METHODS

Genome Selection

Fifty-eight genomes (**Table 1**) available in the GenBank database (National Center for Biotechnology Information, NCBI; https://www.ncbi.nlm.nih.gov/) as of January 2021 were retrieved to explore the CRISPR-Cas systems in the *Clostridium novyi sensu lato* genospecies. Seven of them were complete genomes (BKT015925, C-Stockholm, BKT2873, 1873, 3859/11, 150557, and NT) and the remaining 51 were draft genomes: 45 belong to *C. botulinum* (BoNT/C, C/D, /D, and D/C), 10 belong to *Clostridium novyi*, and three belong to *C. haemolyticum*, showing high variability in sampling year, location, country, sample origin, and genomic lineage. The *Clostridium novyi sensu lato* genospecies, defined according to a previously published classification (Skarin and Segerman, 2011), fall into four different lineages, based on genomic comparisons using Gegenees (Agren et al., 2012), a software tool using a fragmented alignment approach to compare bacterial genomes (Skarin and Segerman, 2014).

CRISPR-Cas Systems Identification

The CRISPRCasFinder server (https://crisprcas.i2bc.paris-saclay.fr/, University of Paris-Saclay, France) was used with default parameters (Couvin et al., 2018) to identify CRISPR loci, as well as to determine the presence and content of *cas* genes. The genetic environment of solitary CRISPR arrays with no predicted CRISPR locus was further explored using the HHpred server (https://toolkit.tuebingen.mpg.de/tools/hhpred/, Max Planck

Institute, Germany; Zimmermann et al., 2018) with default parameters to search for undetected *cas* genes. Most of the available genomes were draft sequences, which may have resulted in incomplete or partial CRISPR-Cas loci identification.

In silico Analysis of the Spacer Content in CRISPR Arrays

Clustered regularly interspaced short palindromic repeats arrays were identified using the CRISPR Recognition tool (CRT), software designed to detect CRIPSR arrays (Bland et al., 2007). Settings were chosen so as to have at least three repeats in the CRISPR array with a repeat length between 19 and 38 base pairs (bp) and a spacer length between 19 and 48 bp. The spacer content of each CRISPR array was then extracted manually. Identical spacers were detected in several genomes and were removed to keep only one copy of each spacer that will be called "unique spacer" in the rest of the study. Complementary spacers in the reverse strand or harboring one or two mismatches were considered as unique spacers. Unique spacers were then screened with a Basic Local Alignment Search Tool (BLAST) search to identify protospacers using the CRISPRTarget tool (http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html, University of Otago, New Zealand; Biswas et al., 2013). Only matches in the GenBank-Phage and Refseq-plasmid databases with at least 80% homology were considered. In addition, only the available complete (not draft) genomes were used in these databases. Parameters for the initial BLAST screen were gap open = −10 and extend = −2, minimum BLAST score = 21 with nucleotide match = 1 and mismatch = −2, E value = 1, and Word = 7. Species targeted, target name and type, location in the genome, and gene function (for the main MGEs detected) were reported. Protospacer-adjacent motif (PAM) sequences associated with identified protospacers were reported when identified. In these databases, some MGEs had different names although they were identical (e.g., pCN2 is identical to pCLG2). Unnamed plasmids were designated with a "p_" followed by their strain name and number.

Anti-CRISPR Genes Search

Anti-CRISPR proteins were searched on *Clostridium novyi sensu lato* plasmidome using AcRanker,[1] a machine learning system for direct identification of anti-CRISPR proteins (Eitzinger et al., 2020).

RESULTS

Overview of CRISPR-Cas Types Detected in *Clostridium novyi sensu lato*

Six different CRISPR-Cas system types belonging to class 1 and 2 CRISPR loci were identified in 55 out of the 58 *Clostridium novyi sensu lato* strains included in our study, located either on the chromosome or on MGEs (BoNT phages, plasmids p2BKT015925, pCLG2/pCN2; **Table 1**; **Figure 1**). Twenty to 209 spacers were detected per strain. *In silico*-identified PAM sequences associated with type I-B, I-D, and II-C CRISPRs matched with canonical PAMs (Kieper et al., 2018; **Figure 1**). Type I-B CRISPR predominated and was detected in all lineages (**Figure 1**). Type I-D was found in 10 *C. botulinum* BoNT/D and D/C strains and in two *Clostridium novyi* strains (lineages I-b and II). Type III-B was detected in four *C. botulinum* BoNT D/C and one *Clostridium novyi* strains (lineage I-b; **Table 1**). The type III-B CRISPR-Cas locus appeared to be devoid of the Cas proteins involved in spacer acquisition (Cas1 and Cas2) and located in close vicinity to the type I-B CRISPR locus (among the complete available genomes). *Clostridium novyi* strain 150557 (lineage II) seems to possess a type III-D CRISPR locus. Among class 2 CRISPR loci, canonical type II-C was present in seven *C. botulinum* BoNT/D and D/C strains (lineage Ib) and type V-U in 12 *C. botulinum* BoNT/C, *Clostridium novyi*, and *C. haemolyticum* strains (lineage II).

Clustered regularly interspaced short palindromic repeats-Cas genes were found either in the chromosome or in MGEs (in which case, they were located on BoNT phages and plasmids, as illustrated in **Figure 2**). When located on MGEs, the CRISPR loci were of type I-B. Some strains harbored up to three different type I-B CRISPRs (located in the chromosome, BoNT phages, and plasmids). They varied among MGEs. For example, the type I-B CRISPR locus from p1_BKT015925 was different from that of p2_BKT015925, whereas none were detected for p1_Eklund phage and p1_3859/11 phage carried instead a bacteriophage exclusion system (BREX; Goldfarb et al., 2015). Of the available p1 BoNT phage and plasmid sequences, two type I-B CRISPR-Cas loci with major differences were identified (**Figure 2**): one shared by p1_Stockholm, p1_16868 BoNT phages, and the p2_BKT015925 and pCLG2/pCN2 plasmids and a second one shared by the p1_BKT015925 and p1_BKT2873 and p1_1873 BoNT phages. Therefore, two different CRISPR-Cas loci were acquired independently in these MGEs. CRISPR loci were also shared between plasmids and phages: we detected a related CRISPR locus in the p1_16868 phage and p2_BKT015925 plasmid (**Figure 2**). Furthermore, the CRISPR loci were unrelated to chromosomal type I-B CRISPR loci (data not shown) and exhibited no link with the bacterial genomic lineages.

These CRISPR loci located on MGEs were mostly incomplete, except in the p1_16868 BoNT phage, which contain an entire type I-B CRISPR (effector and adaptation module; Koonin et al., 2017). Incomplete CRISPR loci had lost Cas4, Cas1, and Cas2, which constitute the adaptation module (in MGEs p1_BKT015925, p2_BKT015925, p1_BKT2873, and p1_1873), as well as the Cas3 endonuclease that cleaves the DNA target in the effector module (in the p1_Stockholm and pCLG2/pCN2 MGEs). Noteworthily, transposases interrupting the CRISPR loci were detected in p1_BKT015925, p1_BKT2873, p1_1873, and p1_Stockholm. In p1_BKT015925, up to five transposases were detected in close vicinity to CRISPR genes. Additionally, they showed similarities with TnpB and Cas14a [one of 376 amino acids (a.a.) with a 100% HHPRED probability of being Cas14a, E-value 3.6e-43, score 327.87; another of 470 a.a. with a Cas14a probability of 100%, E-value 2.9e-40, and

[1]http://acranker.pythonanywhere.com/

TABLE 1 | Strain metadata and clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated proteins types.

Strain	Species	Estimated genome Size (Mbp)	Location	Year	Origin	BoNT type	Lineage[a]	GenBank access number	CRISPR type	CRISPR location	Number of spacers in the genome
Eklund	*Clostridium botulinum*	2.96	United States	/	/	C	III	ABDQ00000000.1	I-B	chr	20
K25	*Clostridium* sp.	2.60	South-Korea	2012	Pig	NT	II	JENU00000000.1	I-B, V-U	/	91
Stockholm	*Clostridium botulinum*	2.70	Sweden	1946	Mink	C	II	CP063816-CP063821	I-B, V-U	chr, p1, pCLG2/ pCN2	85
IFR 18/084	*Clostridium botulinum*	2.61	/	/	/	C	II	SXEF00000000	I-B, V-U	chr	118
Colworth BL165	*Clostridium botulinum*	2.49	United Kingdom	1970	/	C	II	SWNS00000000	I-B, V-U	chr	135
571C	*Clostridium botulinum*	2.62	United States	/	/	C	II	SWUK00000000	I-B, V-U	chr	121
IFR 18/049	*Clostridium botulinum*	2.41	/	/	/	(C)	II	SWNT00000000	I-B, V-U	chr	91
7221C	*Clostridium botulinum*	2.68	/	/	/	C	II	SXDK00000000	I-B, V-U	chr	133
Davies AO	*Clostridium botulinum*	2.56	/	/	/	C	II	SXEV00000000	I-B, V-U	chr	119
IFR 18/078	*Clostridium botulinum*	2.46	/	/	/	(C)	II	SXES00000000	I-B, V-U	chr	51
12LNR10	*Clostridium botulinum*	3.04	France	2012	Turkey	C/D	Ia	LGVQ00000000	I-B	p1, p2	83
12LNR13	*Clostridium botulinum*	3.07	France	2012	Chicken	C/D	Ia	LGVT00000000	I-B	chr, p1	81
12LNRI	*Clostridium botulinum*	3.00	France	2012	Duck	C/D	Ia	LGVR00000000	I-B	p1, p2	73
29401	*Clostridium botulinum*	3.04	France	2008	Chicken	C/D	Ia	LGVP00000000	I-B	p1, p2	76
38028	*Clostridium botulinum*	3.12	France	2008	Chicken	C/D	Ia	LGVO00000000	I-B	p1, p2	75
43243	*Clostridium botulinum*	3.00	France	2009	Guinea Fowl	C/D	Ia	LGVU00000000	I-B	p1, p2	72
48212	*Clostridium botulinum*	3.00	France	2008	Duck	C/D	Ia	LGVS00000000	I-B	p1, p2	90
49511	*Clostridium botulinum*	3.09	France	2008	Chicken	C/D	Ia	LHYP00000000	I-B	chr, p1, p2	65
50867	*Clostridium botulinum*	3.08	France	2008	Chicken	C/D	Ia	LHYQ01000025	I-B	p1, p2	84
55741	*Clostridium botulinum*	3.04	France	2008	Turkey	C/D	Ia	LHYR00000000	I-B	p1, p2	83
58272	*Clostridium botulinum*	3.07	France	2008	Chicken	C/D	Ia	LHYS00000000	I-B	p1, p2	87
58752	*Clostridium botulinum*	2.82	France	2008	Chicken	(C/D)	Ia	LHYT00000000	I-B	p2	45
69285	*Clostridium botulinum*	2.94	France	2008	Chicken	(C/D)	Ia	LHYU00000000	I-B	p2	45
71840	*Clostridium botulinum*	3.00	France	2008	Chicken	C/D	Ia	LHYV00000000	I-B	p1, p2	84
BKT2873	*Clostridium botulinum*	3.19	Sweden	2007	Chicken	C/D	Ib	CP063965-CP063968	I-B	chr, p1	80
BKT12695	*Clostridium botulinum*	2.75	Sweden	2010	Chicken	(C/D)	III	JENP00000000	I-B	chr	68
BKT015925	*Clostridium botulinum*	3.20	Sweden	2008	Chicken	C/D	Ia	CP002410.1	I-B	p1, p2	83
BKT028387	*Clostridium botulinum*	2.83	Sweden	2007	Chicken	(C/D)	Ia	AESB00000000	I-B	p2	45
BKT75002	*Clostridium botulinum*	3.14	Denmark	2010	Chicken	C/D	Ib	JENS00000000	I-B	chr, p1	80

(Continued)

TABLE 1 | Continued

Strain	Species	Estimated genome Size (Mbp)	Location	Year	Origin	BoNT type	Lineage[a]	GenBank access number	CRISPR type	CRISPR location	Number of spacers in the genome
It1	*Clostridium botulinum*	2.50	Italy	/	Bovine	(D/C?)	IV	JENO00000000	/	/	0
Sp77	*Clostridium botulinum*	3.06	Spain	2011	Duck	C/D	Ia	JENQ00000000	I-B	p1, p2	77
V891	*Clostridium botulinum*	3.17	Sweden	2007	Gull	C/D	Ia	AESC00000000	I-B	chr, p1	71
1873	*Clostridium botulinum*	2.57	Chad	1958	Ham	D	II	CP063822-CP063828	I-B	chr, p1, pCLG2/pCN2	22
16868	*Clostridium botulinum*	3.08	Netherlands	2001	Bovine	D	Ia	JENR00000000	I-B	p1, p2	91
CCUG7971	*Clostridium botulinum*	2.81	South Africa	1926	/	(D)	Ib	JDRZ00000000	I-D, II-C	/	26
3859/11	*Clostridium botulinum*	2.89	Italy	2011	Bovine	D/C	Ib	CP063959-CP063964	I-B, I-D, III-B	chr, p2	69
1274	*Clostridium botulinum*	2.94	Brazil		Vaccine strain	D/C	Ib	MVIY00000000	I-D, II-C	/	70
1275	*Clostridium botulinum*	2.94	Brazil	/	Vaccine strain	D/C	Ib	MVIZ00000000	I-D, II-C	/	70
1276	*Clostridium botulinum*	2.91	Brazil	/	Vaccine strain	D/C	Ib	MVJA00000000	I-D, II-C	/	70
1277	*Clostridium botulinum*	2.57	Brazil	/	Vaccine strain	D/C	Ib	MVJB00000000	II-C	/	32
47295	*Clostridium botulinum*	3.18	France	2009	Bovine	D/C	Ib	LHYX00000000	I-B, I-D, III-B	/	43
51714	*Clostridium botulinum*	3.18	France	2009	Bovine	D/C	Ib	LHYY00000000	I-B, I-D, III-B	/	48
CP05	*Clostridium botulinum*	2.88	Brazil	/	Vaccine strain	D/C	Ib	MVJC00000000	I-D, II-C	/	46
DC5	*Clostridium botulinum*	3.32	Italy	/	Bovine	D/C	Ib	JDRY00000000	I-B, I-D, III-B	/	85
LNC5	*Clostridium botulinum*	2.89	New Caledonia	2013	Bovine	D/C	Ib	LHYW00000000	I-D, II-C	/	39
4540	*Clostridium novyi*	2.50	Great Britain	2000	Human	NT	IV	JENL00000000	/	/	0
4552	*Clostridium novyi*	2.80	Great Britain	2000	/	NT	III	JENJ00000000	I-B	/	126
4,570	*Clostridium novyi*	2.32	Great Britain	2001	/	NT	IV	JDRX00000000	I-B	/	200
BKT29909	*Clostridium novyi*	2.46	Sweden	2007	Chicken	NT	IV	JENM00000000	/	/	4
GD211209	*Clostridium novyi*	2.46	Netherlands	2009	Bovine	NT	IV	JENN00000000	/	chr	25
NCTC538	*Clostridium novyi*	2.52	Great Britain	1920	Human	NT	IV	JENK00000000	I-B	/	94
ATCC27606	*Clostridium novyi*	2.61	Germany	/	/	NT	II	JENW00000000	I-B, I-D, III-B	/	112
NCTC9691	*Clostridium novyi*	2.64	Great Britain	1955	Sheep	NT	II	JENV00000000	I-B, I-D, V-U	/	77
150557	*Clostridium novyi*	2.30	South Korea	2015	Pig	NT	II	NZ_CP029458	I-B, III-D, V-U	chr, pCLG2/pCN2	78
NT	*Clostridium novyi*	2.55	/	/	/	NT	IV	NC_008593.1	I-B	chr	96
NCTC9693	*Clostridium haemolyticum*	2.61	United States	1955	/	NT	II	JENX00000000	I-B	/	39
KFSHRC_CH1	*Clostridium haemolyticum*	2.87	Saudi Arabia	2014	Human	NT	II	NZ_LSZB01000000	I-B, V-U	/	209

(Continued)

TABLE 1 | Continued

Strain	Species	Estimated genome Size (Mbp)	Location	Year	Origin	BoNT type	Lineage[a]	GenBank access number	CRISPR type	CRISPR location	Number of spacers in the genome
NCTC8350	*Clostridium haemolyticum*	2.47	United States	1946	/	NT	II	NZ_ JDSA00000000	I-B	/	38

A slash indicates that no information could be retrieved. For the Clostridium botulinum strains, the botulinum neurotoxin (BoNT) type is indicated in parentheses when the BoNT phage has been lost and is not present in the genome sequences. The Clostridium novyi and Clostridium haemolyticum strains are non-toxigenic (NT). Complete genomes are indicated in gray, all others are draft sequences.

[a]*According to Skarin et al. (2011).*

(), incomplete or lost BoNT phage p1; /, no information retrieved; NT, non-BoNT-producing strain.

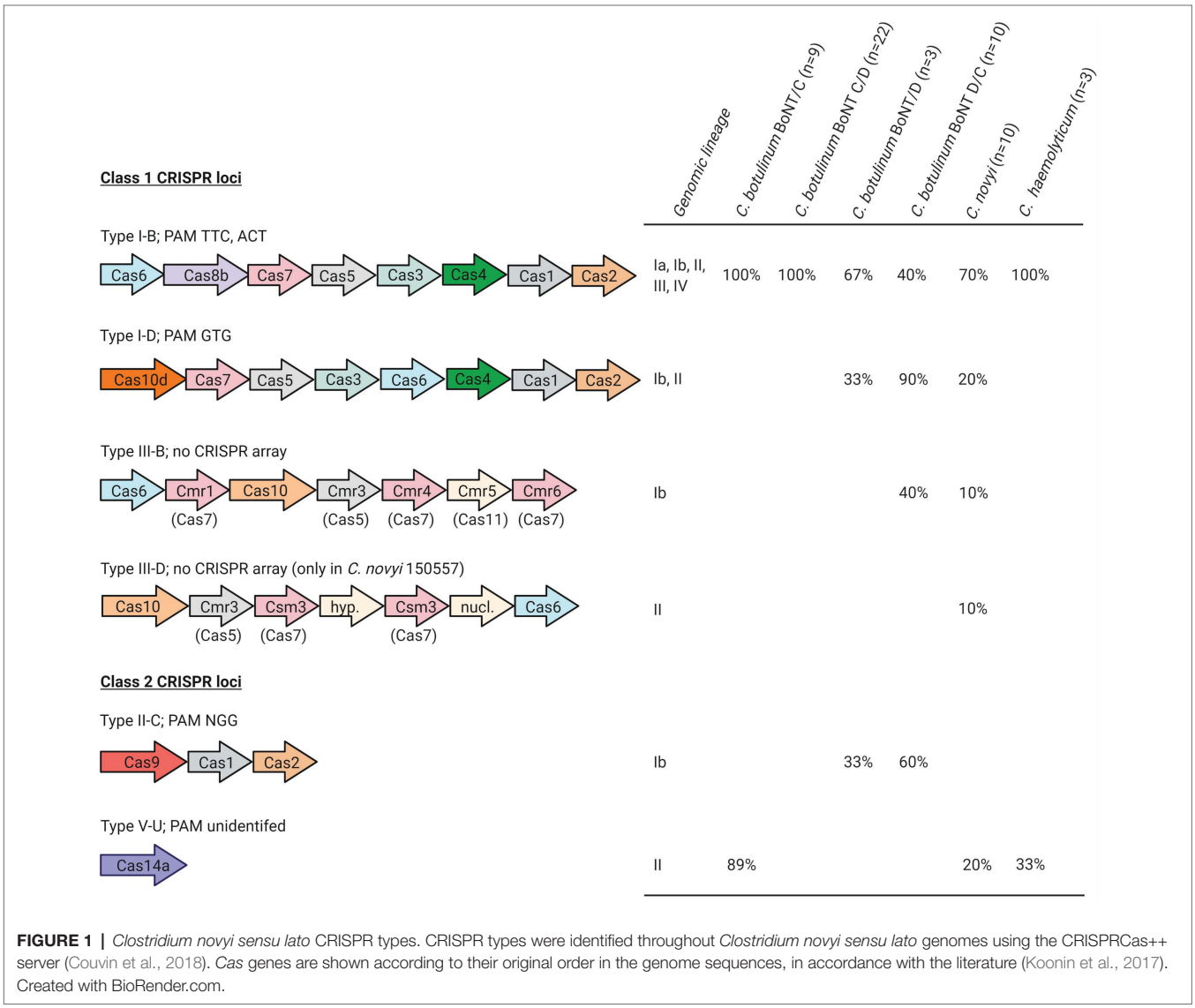


FIGURE 1 | *Clostridium novyi sensu lato* CRISPR types. CRISPR types were identified throughout *Clostridium novyi sensu lato* genomes using the CRISPRCas++ server (Couvin et al., 2018). *Cas* genes are shown according to their original order in the genome sequences, in accordance with the literature (Koonin et al., 2017). Created with BioRender.com.

score 342.39]. Detection of TnpB transposases next to Cas proteins (in BKT015925, BKT2873, 1873, and C-Stockholm located on the *bont* phage) raises the question of their potential role in the CRISPR-Cas system and possible connection with the evolution of their CRISPR loci.

A Putative Phage-Inducible Chromosomal Island Found in *Clostridium novyi sensu lato*

Two other putative TnpB transposases (TnpB_IS605) were detected on the chromosome of *Clostridium novyi sensu lato* genomic lineage II, next to solitary CRISPR arrays (**Table 1**;

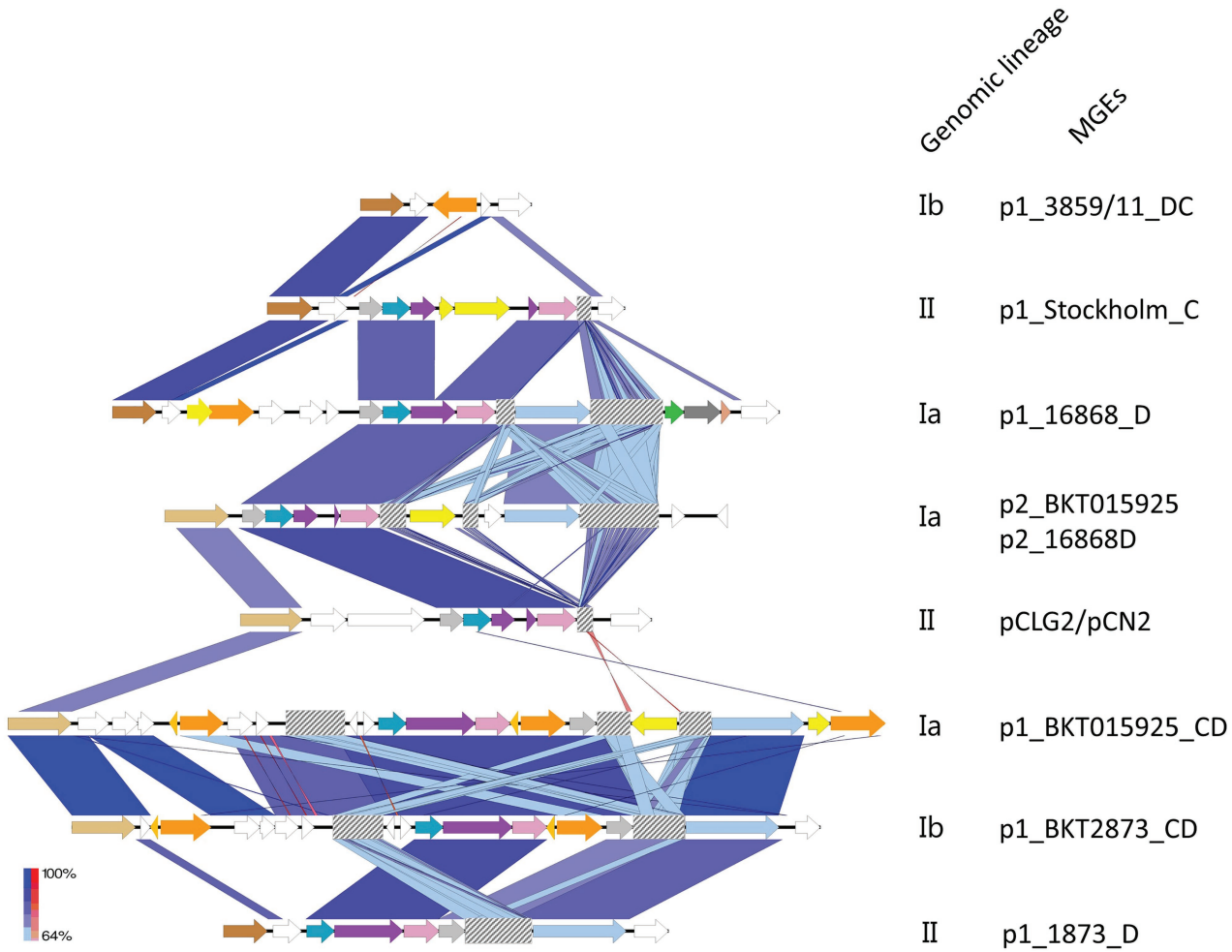


FIGURE 2 | Comparison of the architecture of the CRISPR loci located in MGEs. Annotated CDS are shown in white. CRISPR loci were always found in close proximity to a replication initiation protein (first CDS on the left, dark and light brown). TnpB Cas14a related transposases are shown in orange; hypothetical transposase regulators, light orange; the IS256-like element ISCbo4 family transposase integrated into the CRISPR arrays of p1_BKT015925 and p2_BKT015925, yellow. CRISPR-Cas type Ib, arrows color coded as in **Figure 1** (light gray = Cas5, Blue indigo = Cas6, Purple = Cas8b, Pink = Cas7, Light blue = Cas3, Green = Cas4, Dark grey = Cas1, and Orange = Cas2). CRISPR arrays are shown as gray-checkered boxes. BLAST hits are colored on a gradient ranging from minimum (light blue) to maximum BLAST scores (dark blue). Reverse matches are shown in a red gradient. Created with Easyfig (Sullivan et al., 2011).

Figure 3). HHPRED predicted that the TnpB transposases were homologous to class 2 type V-U CRISPR Cas14a with a probability of 100%, E-value 3.1e-38–1.6e-46, and score 327.01–367.87 (present in 12 strains of *C. botulinum* BoNT/C, *Clostridium novyi*, *C. haemolyticum*, and *Clostridium* sp.; **Table 1**). The two putative chromosomal *cas14a* genes encode homologous proteins (35.6% homology at the C terminus between the two genes) conserved among 11 strains, but associated with distinct CRISPR arrays with different repeats (gttgagaatcaacaaaggatatgtttaagc and gttttagtttaactatgtgaaatgtaaat) and diverse spacers. Strain *Clostridium novyi* 150557 had only one putative *cas14a* gene and one CRISPR array. Noteworthily, three different types of restriction modification (RM) systems were identified among the 12 strains, located between the two identified *cas14a* genes (**Supplementary Figure S1**). Furthermore, *cas14a* genes were found to be part of an integrative element splitting the RNA component of RNase P (*rnpB*; **Figure 3**; **Supplementary Figure S1**). The detailed analysis of this integrative element revealed a genetic content similar to a phage-inducible chromosomal island (PICI), with the presence of a *Staphylococcus aureus* pathogenicity island (SaPI) repressor homolog, a terminase small subunit, a phage interference redirecting packaging protein (*rppC*) and a minor capsid protein, which are characteristic of PICIs (Penadés and Christie, 2015; **Figure 3**). This putative PICI was detected in 13 strains (**Supplementary Figure S1**). In the *Clostridium novyi* NCTC 9693 strain, it was not integrated in *rnpB*, but located 30 kb away, upstream from a type I-D CRISPR-Cas system. Furthermore, in *C. botulinum* BoNT/C strain IFR 18/078, this PICI was split by another larger integrative element, which appeared to be plasmid pCN1 from the C-Stockholm strain (**Supplementary Figure S1**). This putative

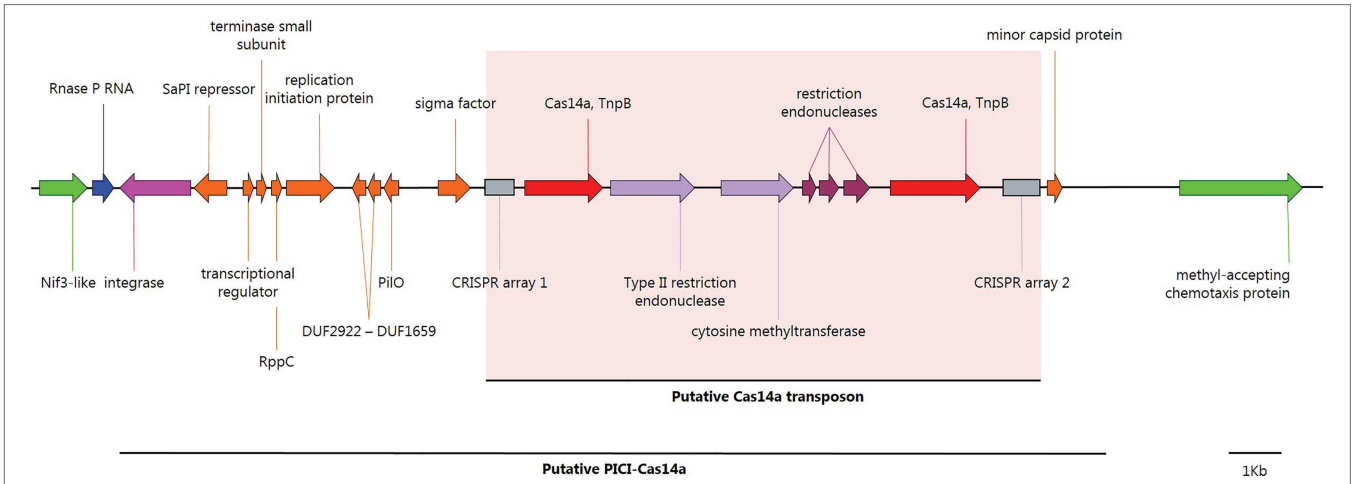


FIGURE 3 | Annotated map of genes in a putative genetic phage-inducible chromosomal island (PICI; C-Stockholm strain) from *Clostridium novyi sensu lato* lineage II. Nif3-like and the methyl-accepting chemotaxis proteins surrounding the PICI site of insertion in the *C. botulinum* C-Stockholm (NZ_CP063816.1) strain are shown in bright green; integration targeted RNase P ncRNA, recognized by the PICI integrase, blue; PICI tyrosine-recombinase integrase, pink; PICI-related CDS, orange; type V-U CRISPR Cas14a, red; CRISPR arrays, gray boxes; and restriction-modification related CDS, light purple, and dark red. This figure was created with SnapGene.

PICI is the first reported in *C. botulinum*; further investigations are therefore required to confirm its nature and role.

Identification of Protospacers to Trace Past Encounters of *Clostridium novyi sensu lato* With MGEs

The spacers of *Clostridium novyi sensu lato* CRISPR arrays were also investigated. Spacer arrays indeed represent the memory of the past encounters with MGEs and, by tracing back the spacer origin (the protospacer); they can be used to better understand the mobilome composition/interaction among *Clostridium novyi sensu lato* strains. Spacer sequences were extracted to build a representative database from the 58 strains used in this study (**Table 1**). We recovered 4,320 spacers, which mostly belonged to *C. botulinum* BoNT C/D (35.8%), *C. botulinum* BoNT/C (20.2%), and *Clostridium novyi* (18.8%), proportional to the number of strains used in our study (**Figure 4A**; **Supplementary Table S1**). Shared identical spacers were investigated first (**Supplementary Table S2**). Interestingly, the distribution of strains based on shared spacers was very similar to the distribution previously obtained using Gegenees (Skarin and Segerman, 2014) or based on SNP phylogeny profiling (Woudstra et al., 2016). For example, 17 *C. botulinum* BoNT C/D strains from lineage Ia shared more than 90% of their spacers (**Supplementary Table S2**). Analysis of the spacers in strains BKT2873 and BKT75002 (99% whole genome similarity using Gegenees; Skarin and Segerman, 2014) revealed 100% identical spacer content (**Supplementary Table S2**).

After exploring the shared spacers, we focused on the analysis of the 2,045 unique spacers (47.3%; **Supplementary Table S3**). Among the unique spacers, the distribution was no longer proportional to the number of strains in each analyzed species, with *Clostridium novyi* spacers now representing up to 34.6% of the unique spacer content and *C. botulinum* BoNT C/D spacers accounting for 20.6% (**Supplementary Table S3**). This pattern is in concordance with our above observation on spacers widely shared in the *C. botulinum* BoNT C/D lineage Ia, resulting in a lower number of unique spacers, with many of them being present multiple times in genetically related strains.

We then screened the unique spacers *via* a BLAST search using the CRISPRTarget tool (Biswas et al., 2013) to determine their origin (the protospacers) and found 474 spacers matching with at least 80% homology in the GenBank-Phage and Refseq-plasmid databases (**Figure 4B**; **Supplementary Table S4**). The remaining 1,571 unidentified spacers (**Supplementary Table S5**) belong to unknown or uncharacterized MGEs, the so-called "spacer dark matter" (Mcginn and Marraffini, 2019).

A BLAST search on the 474 spacers returned 1,503 protospacer hits in the Refseq-Plasmid and GenBank-Phage databases (**Figure 4C; Supplementary Table S6**). Most of the identified protospacers originally belonged to spacers from *Clostridium novyi* (51.7%, with 20.9% located in plasmids and 30.8% in phages) and *C. botulinum* BoNT C/D (18.8%, 12.6% located in plasmids, and 6.1% in phages; **Supplementary Table S7**). The high number of protospacers identified relative to the number of spacer sequences reflects the acquisition of protospacers from homologous genes present in different or related MGEs.

The 1,503 BLAST hits retrieved from the identified protospacers corresponded to 776 unique MGEs (**Supplementary Table S8**), composed of 379 plasmids (48.8%) and 397 phages (51.2%). Phages were mostly temperate *Siphoviridae* and *Herelleviridae* (among the most common occurring phage species; **Supplementary Table S8**). The phages ranged in size from 497,513 bp for the largest (*Myoviridae* phage G) to 17,637 bp for the smallest (*Podoviridae* vB_SauP_phiAGO1.9). MGEs originating from the *Staphylococcus*, *Streptococcus*, *Clostridium*, *Bacillus*, *Escherichia*, and *Borrelia* bacterial genera represented 66.8% of the unique MGEs (**Figure 4D; Supplementary Table S9**).

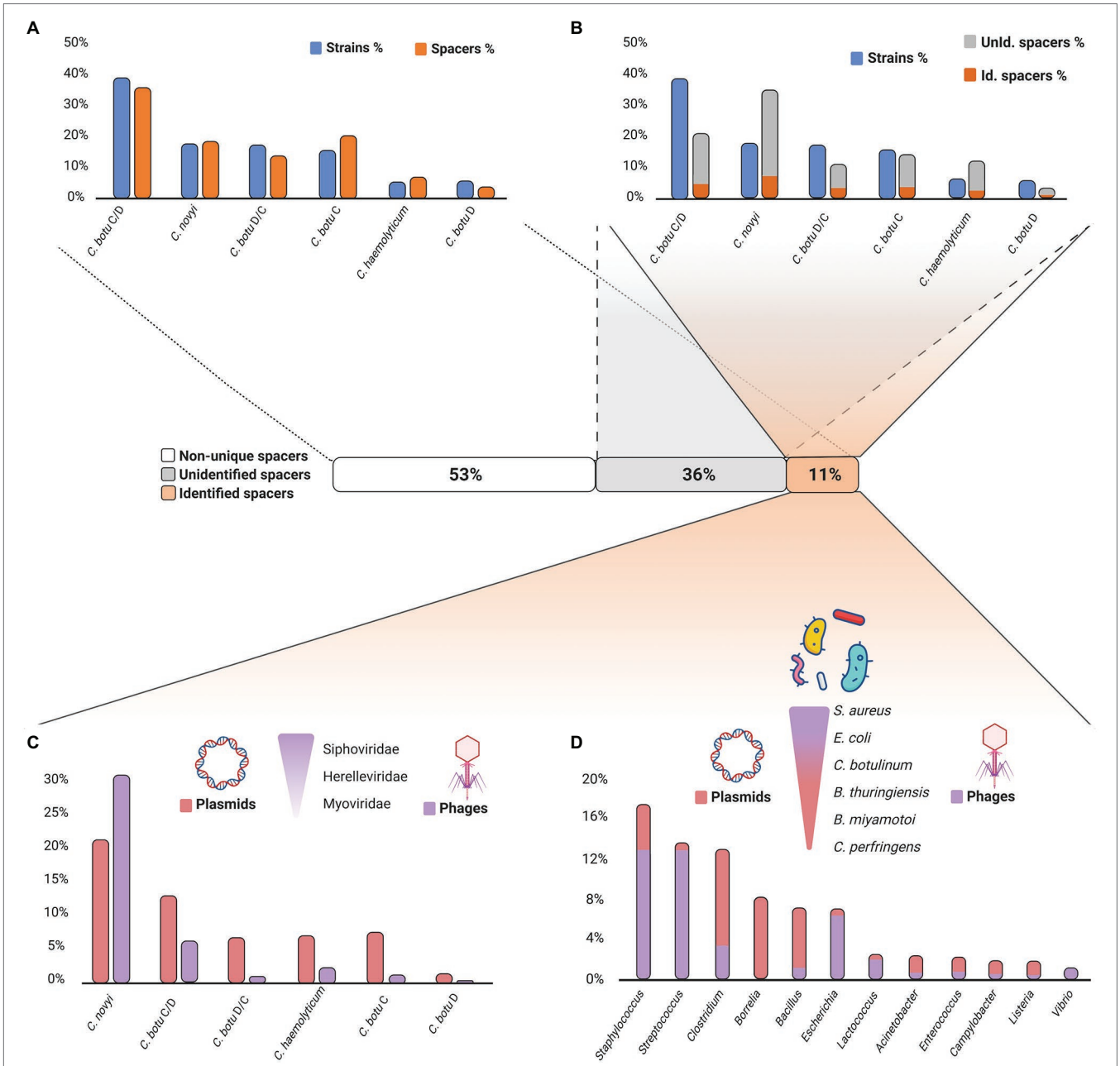


FIGURE 4 | Exploration of *Clostridium novyi sensu lato* spacer content. Created with BioRender.com. **(A)** Distribution of spacers among the analyzed *Clostridium novyi* s.l. strains. Comparison of the distribution of spacers among the analyzed *Clostridium novyi sensu lato* strains. Overall, each species contributed equally to the total number of spacers. Raw data are available in **Supplementary Table S1**. **(B)** Identified and unidentified unique spacer distribution among the analyzed *Clostridium novyi sensu lato* strains. Comparison of the distribution of identified and unidentified spacers and their species origin based on unique spacer sequences. *Clostridium novyi* spacers are the main contributor to the unique spacer content. The distribution of the identified spacers is not proportional to the number of strains in each species. Raw data are available in **Supplementary Tables S3** and **S4**. **(C)** Distribution of the MGE origins among *Clostridium novyi sensu lato* spacers. Plasmids and phages are equal sources of protospacers. Among the phage protospacer origins, *Siphoviridae*, *Myoviridae*, and *Herelleviridae* predominate. *Clostridium novyi* and *Clostridium botulinum* BoNT C/D are the main analyzed species carrying the protospacers. Raw data are available in **Supplementary Tables S6** and **S7**. **(D)** Targeted MGE host origins among the *Clostridium novyi* s.l. spacers. Exploration of the hosts of the MGEs identified as targets of the CRISPR spacers found in *Clostridium novyi sensu lato*. The MGEs originate mainly from *Clostridium*, *Bacillus*, *Streptococcus*, and *Borrelia* genera. The main species in these genera are *C. botulinum*, *Bacillus thuringiensis*, *Bacillus cereus*, *Staphylococcus aureus Borrelia burgdorferi*, and *Borrelia afzelii*. Raw data are available in **Supplementary Tables S8** and **S9**.

Interestingly, the bacterial genera varied in the type of MGE contribution (plasmid or phage). *Clostridium*, *Bacillus*, and *Borrelia* mainly contributed plasmids (32.3%), whereas *Staphylococcus*, *Streptococcus*, and *Escherichia* mainly contributed phages (32.5%; **Supplementary Table S9**). At the species level, *S. aureus*, *Escherichia coli*, *C. botulinum*, *Bacillus thuringiensis*,

TABLE 2 | The score value represents the percentage of homology between the spacer and its target (Biswas et al., 2013).

Table 2A: Spacer BLAST scores of the main targeted bacterial species.

Genus	Protospacer species origin	% BLAST results	Average score	St. dev.	Min.	Max.
Clostridium	*Clostridium botulinum*	28%	28.76	4.97	21	38
Staphylococcus	*Staphylococcus aureus*	11%	21.65	0.80	21	26
Streptococcus	*Streptococcus thermophilus*	7%	24.38	1.16	22	26
Escherichia	*Escherichia coli*	4%	22.25	0.54	22	24
Borrelia	*Borrelia miyamotoi*	2%	22.19	0.68	21	23
Bacillus	*Bacillus thuringiensis*	2%	23.84	1.95	21	28
Clostridium	*Clostridium perfringens*	2%	23.38	2.88	21	32
Lactococcus	*Lactococcus lactis*	2%	23.50	2.15	21	30
Clostridium	*Clostridium difficile*	1%	23.00	0.82	22	24
Staphylococcus	*Staphylococcus epidermidis*	1%	22.25	0.83	21	23
Acinetobacter	*Acinetobacter baumannii*	1%	22.67	0.94	22	24
Streptococcus	*Streptococcus pyogenes*	1%	23.27	2.21	21	28
Bacillus	*Bacillus cereus*	1%	22.86	1.28	21	25
Streptococcus	*Streptococcus agalactiae*	1%	23.87	0.88	22	26
Streptococcus	*Streptococcus mitis*	1%	24.36	2.38	22	28

Table 2B: Spacer BLAST scores of the main targeted *C. botulinum* mobile genetic elements (MGEs).

Mobile genetic element	Protospacer species origin	% BLAST results	Average score	St. dev.	Min.	Max.
p1_Cst	*Clostridium botulinum*	6%	31.36	4.98	21	38
p4_BKT015925	*Clostridium botulinum*	5%	29.42	4.43	21	38
p6_Cst	*Clostridium botulinum*	4%	30.57	4.69	22	37
CWou-2020a	*Clostridium botulinum*	3%	29.03	5.33	22	38
p1_BKT015925	*Clostridium botulinum*	2%	30.46	3.71	24	36
p2_BKT015925	*Clostridium botulinum*	1%	27.58	3.66	22	34

The higher the score, the better the homology, indicating a putative recent encounter. A: Clostridium botulinum is the main target of the spacers with up to 28% of the BLAST results, together with the highest average score. This illustrates the active dissemination of C. botulinum mobile genetic elements (MGEs) throughout Clostridium novyi sensu lato. B: The most encountered C. botulinum MGEs are p1_Cst>p4_BKT015925>p6_Cst, accounting for 15% of the BLAST results. Other C. botulinum MGEs (CWou-2020a, p1_BKT015925, and p2_BKT015925) account for 6% of the BLAST results, showing less frequent encounters. The average BLAST scores are roughly equivalent among the C. botulinum MGEs, suggesting concomitant dissemination.

TABLE 3 | Spacers targeting the restricted mobilome.

MGE name	MGE type	Localization of spacers	BoNT-producing strains	% of all unique spacers	% of BoNT+ – strains containing spacers	Number of spacers/ strain
p1_Cst, p1_ BKT015925	BoNT phage	97.8% chromosomal, 2.2% MGEs	+	4.2%	84.1%	1–13
			–	3.8%	50.0%	2–16
p4_BKT015925, p6_ Cst	Prophage/plasmid hybrid	93.1% chromosomal, 6.9% MGEs	+	5.2%	65.9%	1–11
			–	3.7%	78.6%	2–14
CWou-2020a	Prophage	100% chromosomal	+	1.2%	47.7%	1–5
			–	1.3%	50.0%	1–8
p2_BKT015925	Plasmid	100% chromosomal	+	0.5%	25.0%	1–4
			–	0.3%	28.6%	1

Origin of spacers specifically targeting the restricted mobilome p1_Cst, p1_BKT015925, p4_BKT015925, p6_Cst, CWou-2020a, and p2_BKT015925 were analyzed. BoNT+ strains refer to C. botulinum BoNT C, C/D, D, and D/C; BoNT– strains refer to Clostridium novyi, C. haemolyticum, and Clostridium sp. Distribution of spacers with regard to the entire BLAST results, to the proportion of BoNT +/– strains and the number of spacers targeting this restricted mobilome are given. BoNT, botulinum neurotoxin; MGE, mobile genetic element; BoNT-, C novyi, C. haemolyticum, and Clostridium sp.; BoNT+, others.

Streptococcus thermophilus, *Clostridium perfringens*, *Lactococcus lactis*, *Staphylococcus epidermidis*, *Acinetobacter baumannii*, *Streptococcus pyogenes*, *Borrelia miyamotoi*, *Clostridium difficile*, *Bacillus cereus*, *Streptococcus agalactiae*, and *Streptococcus mitis* were the main sources of MGEs (**Supplementary Table S9B**).

A Limited Number of *C. Botulinum* MGEs Predominate *Clostridium novyi sensu lato* Spacer Sequences

Protospacers identified from MGEs belonging to *C. botulinum* represented 27.6% of all BLAST results, with the highest average BLAST score of 28.8 (**Table 2A**), therefore suggesting that they are the most recent and frequently encountered ones. Of these MGEs, p1_Cst, p4_BKT015925, p6_Cst, C_Wou-2020a, p1_BKT015925, and p2_BKT015925 represented up to 20.2% of the BLAST results (hereafter called the "restricted mobilome"), showing a roughly similar average BLAST score, suggesting parallel and concomitant dissemination (**Table 2B**). Although these MGEs belong to different groups of mobile elements (p1_Cst and p1_BKT015925 are pseudolysogenic BoNT phages; Skarin et al., 2011, p4_BKT015925 and p6_Cst have the characteristics of prophage-plasmid hybrids; Pfeifer et al., 2021, CWou-2020a is a circular prophage and p2_BKT015925 is a plasmid; Skarin et al., 2011), they appeared to be genuinely recognized by the CRISPR systems. In a few cases, CRISPR protection was incomplete (e.g., against p4_BKT015925), because the strain had spacers as well as the immunized MGE as part of their genome (Woudstra et al., 2017). This may suggest that these MGEs persist even in the presence of spacers matching with 100% homology, using an unidentified strategy (screening for anti-CRISPR using AcRanker was negative, data not shown; Eitzinger et al., 2020). On the contrary, full CRISPR protection was also observed, e.g., strains that had acquired protospacers against p2_BKT015925 were devoid of the corresponding plasmid (Woudstra et al., 2017).

In *Clostridium novyi sensu lato*, BoNT-producing strains showed slightly more encounters with the restricted mobilome with 11.1% of BLAST results, whereas non-BoNT-producing strains accounted for 9.1% (**Table 3**). The localization of spacers targeting the restricted mobilome was mainly chromosomal, but also found in MGEs (**Table 3**; **Supplementary Table S6**), indicating that MGEs containing CRISPR arrays may be involved in the protection against these elements. Phage-plasmid hybrids p4_BKT15025 and p6_Cst were the main source of acquired protospacers (8.9%), with 65.9% of BoNT-producing, and 78.6% of non-BoNT-producing strains being immunized. A high number of protospacers was detected in some strains, with up to 16 spacers for *C. haemolyticum* KFSHRC_CH1 targeting the BoNT phages (**Table 3**; **Supplementary Table S6**). The protospacers in these MGEs originated mainly from hypothetical genes and intergenic regions (39.4%) with high BLAST scores (**Supplementary Table S10**). Other protospacers were located in phage-related genes (e.g., capsid, tail, baseplate, terminase, and holin), but also plasmid-partition and replication genes (e.g., ParM, ParB, and replication initiation protein; **Supplementary Table S10**).

Spacer Sequences in *C. botulinum* Group III MGE CRISPR Arrays

The presence of a type I-B CRISPR-Cas system associated with CRISPR arrays was detected in BoNT phages p1_Cst, p1_1873, p1_BKT015925, p1_BKT2873, p1_16868, and plasmids p2_BKT015925 and p?CLG2/pCN2 (**Figure 2**; **Supplementary Table S11**). From the 171 spacers extracted (11.4% of the unique spacers), surprisingly none were shared among any of the different MGEs. Protospacer identification was possible for 28.6%, yielded 133 BLAST hits at an equivalent ratio between phages and plasmids, and belonged mostly to phages from *Escherichia* and *Enterococcus* (41.4%) and to plasmids from *Clostridium* and *Bacillus* (40.6%). *Clostridium botulinum* MGEs (26.3%), with BLAST scores ranging from 22 to 36, predominated, especially p4_BKT015925, p6_Cst, p1_Cst, p1_BKT015925, CWou-2020a, p5_BKT015925, and p2_BKT015925 (18.0%; **Supplementary Table S11**), in accordance with our previous results (**Table 2B**). Interestingly, protospacers from p4_BKT015925 were found to have been frequently acquired by CRISPR arrays located on chromosomes or on MGEs, suggesting that *C. botulinum* MGEs have also protected themselves against this potentially invasive element. In addition, BoNT phage acquisition of protospacers from BoNT phages was also detected (p1_16868 vs. p1_BKT015925; p1_BKT015925 vs. p1_Cst; and p1_BKT2873 vs. p1_BKT015925), suggesting that BoNT phages to use the CRISPR system to target other BoNT phages.

PICI-Related Spacer Content

A type V-U CRISPR was detected in a putative PICI element present in 12 *Clostridium novyi sensu lato* strains from lineage II, containing two distinctive CRISPR arrays (namely CRISPR 1 and CRISPR 2). In the CRISPR arrays, CRISPR 1 contained only 7–9 spacers, and CRISPR 2 had 9–28 spacers. Interestingly, the main spacer diversity was found in *Clostridium novyi*, *C. haemolyticum*, and *Clostridium* sp., which did not share any spacers (**Supplementary Table S12**). Among the eight *C. botulinum* BoNT/C strains, the PICI sequences were highly conserved (as illustrated in **Supplementary Figure S1** for the PICI-Cas14a elements) and had almost exactly the same CRISPR arrays (differences in 1–2 spacers), suggesting that they are highly phylogenetically related. This close relationship was confirmed by comparing whole genome sequences (data not shown), therefore suggesting that all *C. botulinum* BoNT/C strains having a PICI are clonal. When considering the spacers extracted from the two CRISPR arrays, 51.8% were unique. Of the unique spacers, 32.2% gave BLAST results and could be attributed to a protospacer (**Supplementary Table S12**). Among the identified protospacers, 18 with the best BLAST score (25–36) belonged to the above-identified restricted mobilome (**Table 2B**). Interestingly, one spacer located in the *C. botulinum* BoNT/C Stockholm strain in the PICI CRISPR array one targets its own BoNT phage (BLAST score 30) as well as plasmid pCLG2/pCN2 (BLAST score 28; **Supplementary Table S12**). Other identified spacers were mainly acquired by the CRISPR arrays from PICIs in *Clostridium novyi* (7), *C. haemolyticum* (5), and *Clostridium* sp. strains (5), which did

not possess these MGEs as part of their genomes (**Supplementary Table S12**).

DISCUSSION

Clostridium novyi sensu lato Species Harbor a Range of CRISPR-Cas Systems

The first objective of our study was to investigate the diversity and composition of *Clostridium novyi sensu lato* CRISPR-Cas systems, which had not been fully explored yet (Woudstra et al., 2016; Negahdaripour et al., 2017). We detected a high prevalence of CRISPR-Cas systems (complete or partial) throughout the 58 strains included in our study, with only three strains lacking CRISPR-Cas systems (*C. botulinum* It1, *Clostridium novyi* BKT29909, and *Clostridium novyi* 4540, all lineage IV). Accordingly, a CRISPR prevalence of 80% was detected in *C. botulinum* strains, in a study including only eight *Clostridium novyi sensu lato* genomes (Negahdaripour et al., 2017). This prevalence is much higher than the rate of 46% reported for bacteria (Grissa et al., 2007) or other clostridia species such as *C. perfringens*, which showed a CRISPR prevalence of 53.15% (Long et al., 2019). The location of some CRISPR-Cas systems on MGEs may play a role in this high prevalence. Some strains, in particular from lineage Ia, only have CRISPR-Cas systems in MGEs and none in their chromosome.

In our study, the number of CRISPR loci varied between one and three per strain, which seems to be common in pathogenic bacteria (Cady et al., 2011; Shariat et al., 2013; Yin et al., 2013; Karah et al., 2015), and belonged to class 1 and 2. Class 1 CRISPR-Cas systems (types I-B, I-D, III-B, and III-D) were identified in the *Clostridium novyi sensu lato* strains investigated in our study, similar to previous studies that have shown that the type I CRISPR-Cas system is the most common type found in clostridia (Makarova et al., 2020) and that type III-B, followed by type I-B, is the most commonly identified types in *C. botulinum* groups I and II (Negahdaripour et al., 2017). It is noteworthy that only one type of CRISPR-Cas system (I-B) was detected in lineages Ia, III, and IV, but several types were detected in lineages Ib and II. Previous studies have shown that the number of MGEs is higher in *C. botulinum* BoNT C/D (lineage Ia) than in BoNT D/C (lineage Ib) containing several different CRISPR types (Woudstra et al., 2016, 2017). In *B. cereus*, strains with active CRISPR-Cas systems have fewer MGEs than strains with partial CRISPR-Cas systems or without any CRISPR-Cas systems (Zheng et al., 2020). Likewise, CRISPR-Cas systems may play a barrier role against horizontal gene transfer by preventing the acquisition of MGEs, resulting in a lower occurrence of plasmids and phages in strains from lineage Ib, and the opposite in lineage Ia strains.

Type I-B CRISPR loci (either complete or incomplete) were the most common CRISPR type detected in *Clostridium novyi sensu lato*. Type I-B seems to be typical of clostridia. It has indeed been detected in *C. perfringens* (Long et al., 2019) as well as in *C. chauvoei* (Thomas et al., 2017), *C. tetani* (Cohen et al., 2017), and in all *C. difficile* genomes (Hargreaves et al., 2014; Andersen et al., 2016). Here, several type I-B CRISPR arrays were also located on MGEs such as the BoNT phages and the plasmid p2_BKT015925. The localization in prophages has also been reported in *C. difficile* (Hargreaves et al., 2014; Boudry et al., 2015). Prophage localization of CRISPR arrays may play a role in preventing infection by related competing phages (Sorek et al., 2008). It may also play a role in the mobilome dynamics in *Clostridium novyi sensu lato*, in particular in lineage Ia for which the localization of a partial type I-B CRISPR in the pseudolysogenic BoNT phage was preponderant, whereas no CRISPR locus was detected chromosomally. Noteworthily, type I-B CRISPR loci located on BoNT phages in BoNT C/D strains was incomplete, lacking the adaptation modules (Cas1, Cas2, and Cas4). In a previous study, putative type I-related operons, derivatives of subtype I-F or subtype I-B encoding interference *cas* genes, but not associated with *cas1*, *cas2*, or *cas3* genes, have been detected (Makarova et al., 2015). They are located either on plasmids or associated with transposon-related genes (Makarova et al., 2015). Here, they were located in plasmids and in the BoNT phages sometimes associated with transposases. The functionality of these CRISPRs was not experimentally investigated in our study, but the presence of various spacers in CRISPR arrays suggests they are active. The absence of the adaptation module (Cas1, Cas2, and Cas4) would only be detrimental to the acquisition of new spacers, and it would not prevent the effector complex from performing its function, except in the absence of Cas3. Without Cas3, the effector complex is not able to cleave its target – but only bind –, which may result in transcription interference (CRISPRi; Hawkins et al., 2015). The presence of complete CRISPR-IB systems in the chromosome for some strains may also compensate for the missing Cas proteins, allowing these incomplete systems to be fully functional (e.g., in the *C. botulinum* group III Stockholm strain). Tranposases showing similarities with TnpB and Cas14a and interrupting the CRISPR loci were detected in some strains. This is intriguing because Cas14a has been shown to be capable of both transposase- and CRISPR-related activities (Shmakov et al., 2017). Further investigations are required to determine the exact role of these transposases.

There may be other biological significance for these incomplete CRISPR-Cas systems located on MGEs, as roles other than immunological protection have been revealed for CRISPR-Cas systems, such as signal transduction and gene regulation (Bikard et al., 2012). Considering that some CRISPR locus located on MGEs can lose their endonuclease, the cascade complex would only be able to bind its target, therefore promoting steric transcriptional inhibition (Zhang et al., 2021). Further experimental investigations are required to better understand the role and mechanisms of these incomplete type I-B systems located in MGEs in the biology of *Clostridium novyi sensu lato*.

In addition to type I-B, other CRISPR-Cas system subtypes were found in *Clostridium novyi sensu lato* strains, except *C. botulinum* BoNT C/D (lineage Ia) and strains from lineages III and IV. Type I-D was found in *C. botulinum* BoNT/D (33%) and D/C (90%) and in *Clostridium novyi* (20%; lineages Ib and II, **Figure 1**). Considering the low number of available

BoNT D/C genomes (currently 10 genomes in public databases), it is difficult to draw conclusions and establish generalities. However, based on the analysis of these genomes, a complete type I-D system appears to be preponderant mostly in *C. botulinum* BoNT D/C strains (lineage Ib). This CRISPR-Cas system is a hybrid of type I and type III systems (Lin et al., 2020; Mcbride et al., 2020), most often reported in cyanobacteria and archaea (Lin et al., 2020). So far, little is known about this subtype and its biological properties (Lin et al., 2020).

Type II-C was the second-most common CRISPR type found in *C. botulinum* BoNT D/C strains (60%) in lineage Ib. This class II CRISPR system is detected in diverse bacterial species in various environments, especially in pathogenic and commensal bacteria (Mir et al., 2018). A possible role in virulence for type II-C CRISPR-Cas systems has been suggested (Mir et al., 2018), but there is no experimental evidence for such function in *Clostridium novyi sensu lato*. Because *C. botulinum* genome engineering has been successful using CRISPR type II-B Cas9 (Mertaoja et al., 2021), type II-C can also be explored as a native alternative for genome modification using a mini-plasmid strategy (Pyne et al., 2016). This strategy may be advantageous, as modification of clostridia is cumbersome.

Finally, CRISPR type III was also found in *Clostridium novyi sensu lato* and is represented in numerous archaea but is less frequent in bacteria (Koonin et al., 2017), except in *C. botulinum* groups I and II where type III-B has been reported to be preponderant (Negahdaripour et al., 2017). Type III systems degrade both DNA and RNA (Pyenson and Marraffini, 2017; Tamulaitis et al., 2017; Terns, 2018). Here in our study, type III-B and III-D systems were detected in four strains (lineages Ib and II). They lacked *cas1* and *cas2* genes, which seems to be common in type III-B, C, and D systems (Makarova et al., 2015), especially in genomes also containing type I CRISPR-Cas loci. Due to their plasticity in crRNA selection, type III-B interference machinery can use type I spacers to counter infection from bacteriophages that have escaped the type I defenses through PAM mutations (Silas et al., 2017). Here, the four genomes in which these CRISPR systems were detected also had type I-B and type I-D systems, which suggest such a mechanism for these strains.

Lastly, two putative type V-U CRISPR *cas14a* genes encoding homologous proteins associated with two distinct CRISPR arrays were detected in 81.25% of *Clostridium novyi sensu lato* strains belonging to genomic lineage II, and not in other lineages. Type V-U CRISPR targets single-stranded DNA (ssDNA) sequences without requiring restrictive signal sequences (Khan et al., 2019). The ability of Cas14a to specifically target ssDNA suggests a role in defense against ssDNA viruses or MGEs that propagate through ssDNA intermediates, which may be an advantage in some ecosystems containing ssDNA viruses (Harrington et al., 2018), such as soils (Kim et al., 2008), water (Angly et al., 2006; Rosario et al., 2009), or human and farm-animal feces (Sikorski et al., 2013). Here no ssDNA viruses were identified among the spacers located in the two CRISPR arrays from the putative type V-U CRISPR Cas14a systems with a BLAST match (**Supplementary Table S12**). This absence does not exclude the possibility of finding such ssDNA targets among the spacers with no BLAST matches, but there may be other roles for these putative Cas14a systems. For example, the localization of CRISPR-arrays close to the Cas14a genes may suggest that Cas14a plays a role in the spacer acquisition step.

The presence of this putative type V-U CRISPR only in strains from genomic lineage II suggests a specific mechanism of horizontal transfer for this lineage. Moreover, it was included in a putative PICI. PICIs are a recently discovered family of pathogenicity islands that contribute to horizontal gene transfer, host adaptation, and virulence in Gram-positive cocci, but are also widespread among Gram-negative bacteria (Fillol-Salom et al., 2018). These highly evolved molecular parasites can hijack phage-packaging systems, thus allowing its transfer to other bacterial cells (Crestani et al., 2020). In doing so, PICIs protect the host resident population against specific phages. Although the insertion region was present in all *Clostridium novyi sensu lato* genomic lineages (data not shown), only lineage II possessed the putative PICI-Cas14a element, which suggests a narrow host-range dissemination. Furthermore, the presence of diverse spacer sequences in the PICI CRISPR arrays indicates that the system actively acquires protospacers from invading MGEs. The high protospacer identification BLAST score suggests recent acquisition, especially in the identified restricted mobilome. Interestingly, all *C. botulinum* BoNT/C strains having a PICI were clonal with highly conserved PICI sequences; despite being isolated from different countries and years (1946 in Sweden for strain Stockholm and 1970 in United Kingdom for strain Colworth BL165 for example). This might suggest a high stability of these genomes, including the PICI sequences but more genomic data of *C. botulinum* BoNT/C lineage II strains are required to better explore and understand this result.

Noteworthily, an RM system was associated with the putative *cas14a* genes that may add another layer of defense for the PICI host. Transfer of genes encoding non-essential or accessory functions by PICIs has been previously reported (Penadés and Christie, 2015), but to our knowledge, this is the first time that CRISPR and RM systems have been encountered in a PICI. Further investigations are required to study this putative PICI element and its role in *Clostridium novyi sensu lato* lineage II ecology in more detail.

Protospacer Origins Reveal a Handful of MGEs That Are Widespread Among *Clostridium novyi sensu lato* Species

The second objective of our study was to investigate the spacers found in the CRISPR arrays of *Clostridium novyi sensu lato*. Their analysis revealed a high correlation with strain distribution previously obtained using other tools, in particular the distribution of genetic lineages (Skarin and Segerman, 2014; Woudstra et al., 2016). Strains previously identified as being closely related by genomic comparison indeed harbor the same or almost the same spacers (Skarin and Segerman, 2014; Woudstra et al., 2016). Genotyping spacers thus holds promise for identifying common ancestors among strains, which would allow strain tracking with high resolution and accuracy or

highlight potential relatedness between isolates. This approach has been used for several pathogens such as *Campylobacter* (Kovanen et al., 2014), *Salmonella* (Xie et al., 2017), or *C. difficile* (Andersen et al., 2016).

A large proportion of the identified unique spacers originated from *Clostridium novyi* and *C. botulinum* BoNT C/D strains. However, the current number of available genomes of *Clostridium novyi sensu lato* is quite low (58) and the lineages, BoNT types and species are unequally represented. Detection of most of the unique spacers in *Clostridium novyi* and *C. botulinum* BoNT C/D genomes may thus be linked to the high number of strains or high diversity observed in both species. Among the few previously analyzed genomes (Skarin and Segerman, 2014), *Clostridium novyi* and *C. botulinum* BoNT C/D were distributed throughout three different lineages, whereas *C. botulinum* BoNT/D strains were classified into two lineages and *C. botulinum* type C, type D/C and *C. haemolyticum* in only one single lineage. This may explain why the detection of unique spacers was higher in *Clostridium novyi* and *C. botulinum* BoNT C/D than in the other genomes. However, more genomes are needed to confirm this observation. The high diversity of spacers detected in *Clostridium novyi* genomes despite the low number of available genomes (10) may also be linked to the high diversity of strains included in our study. They were isolated between 1920 and 2015, from five different countries and five different animal species.

Among the spacers whose origin was identified, plasmids and phages were found in equivalent proportions (respectively 51.2 and 48.8% of all MGEs), as already previously reported in *C. botulinum* group I, II, and III (Negahdaripour et al., 2017). This distribution is different from what is commonly reported in prokaryotes, where 85% of spacers usually map to phages (Mcginn and Marraffini, 2019), including in other clostridia species such as *C. perfringens* (Long et al., 2019). In addition, some MGEs of *Clostridium novyi sensu lato* contain proteins characteristic of phages as well as plasmids (e.g., p1_BKT015925, p1_Cst, p4_BKT015925, and p6_Cst), resulting in misleading annotations in public databases, which may have artificially inflated the proportion of spacers from plasmids.

The exploration of CRISPR spacer sequences allowed the identification of their protospacers and thus past encounters of *Clostridium novyi sensu lato* with MGEs. The hosts from which the targeted MGEs originate provided partial information on the environmental microbiome composition in which *Clostridium novyi sensu lato* thrived and on the bacterial species with which MGEs are commonly exchanged. The main bacterial species from which the MGEs targeting *Clostridium novyi sensu lato* originated included *S. aureus*, *E. coli*, *C. botulinum*, *B. thuringiensis*, *B. miyamotoi*, *C. perfringens*, *S. epidermidis*, and *S. thermophilus* (38.1% of the total BLAST results). BLAST scores were lower for *Borrelia* than for *Clostridium* and *Bacillus* reflecting putative distant matches. *Clostridium botulinum* was the primary source of the main MGEs detected in this study, with the best BLAST scores, reflecting active dissemination of their MGEs throughout *Clostridium novyi sensu lato*.

The analysis of protospacers also provides information on the ability of MGEs to disseminate successfully within a bacterial species. Of the 776 unique MGEs identified from the identified protospacers, only six of them represented up to 20.2% of the total BLAST results with the best BLAST scores, namely, phage-plasmid hybrids (p4_BKT015925, p6_Cst), BoNT pseudolysogenic phages (p1_Cst, p1_BKT015925), prophages (C_Wou-2020a), and a plasmid (p2_BKT015925) originating from *C. botulinum*. Considering the large mobilome of *Clostridium novyi sensu lato* (Skarin and Segerman, 2014), it may be surprising that only a handful of MGEs are responsible for such large protospacer acquisition. This may imply that these elements are highly transferable throughout *Clostridium novyi sensu lato*. Although spacers targeting p1_Cst, p1_BKT015925 were detected in 84.1% of *C. botulinum* group III strains and 50% of non-toxic *Clostridium novyi sensu lato* strains (**Table 3**), no spacer targeting the BoNT phage from strains 1873, BKT2873, or 3859 was detected, suggesting no encounters between available sequenced strains and these MGEs. Half of the non-toxic strains did not have any spacers targeting p1_Cst, p1_BKT015925, and neither did six out of the seven lineage IV strains, suggesting no encounter between these strains and p1_Cst and p1_BKT015925, perhaps due to their presence in different ecological niches. This also suggests that these non-toxic strains do not have CRISPR-Cas systems trained against the BoNT phage from the genomes currently available in the database and may therefore potentially be infected by these MGEs. Movements of plasmids and toxin genes across lineage boundaries have been reported in *Clostridium novyi sensu lato* strains (Skarin and Segerman, 2014).

Limited Acquisition of Mobilome Protospacers Results in Variable Protection Outcomes

Regarding spacers from CRISPR arrays located either in the chromosome or in MGEs, two main scenarios were encountered: (1) the genome containing the spacer(s) is devoid of the MGE carrying the protospacer; (2) the genome containing the spacer(s) also contains the protospacer MGE. A good example of the first situation is the plasmid p2_BKT015925. Protospacers from p2_BKT015925 were indeed only present in strains devoid of the plasmid. In that case, it seems that the CRISPR system efficiently prevents the acquisition of p2_BKT015925 and plays its expected immunological role, demonstrating the active part CRISPR-Cas systems play to regulate the presence or absence of a specific MGE in *Clostridium novyi sensu lato* strains. This was also the case for prophage CWou-2020a, which was only detected in strains that had not acquired CWou-2020a protospacers. Interestingly, several *C. botulinum* BoNT C/D strains had a CWou-2020a homologous prophage (NZ_AESB01000027) present in their genome, but only strains that had not acquired CWou-2020a protospacers, with incomplete type I-B CRISPR system located on MGEs. The absence of spacers targeting CWou-2020a in strains harboring NZ_AESB01000027 may be related to the mutual exclusion of both prophages, either by integrating in the same location or by encoding multiple infection exclusion proteins (Bondy-Denomy et al., 2016), therefore avoiding the need for a CRISPR system defending specifically against CWou-2020a.

The second situation was observed for p1_BKT015925, p4_BKT015925 as well as p6_BKT015925 and may

be considered a paradox, because both targets and the corresponding spacers were detected at the same time. Induction of autoimmunity and elimination of the targeted MGE should be expected from such situations. Previous studies have reported the detection of self-targeting spacers (targeting host chromosomal DNA) in many bacterial species (Wimmer and Beisel, 2020). The exact impact of self-targeting is unknown and may be minor. Moreover, some studies have reported examples in which a self-targeting spacer can be tolerated, for example through the activation of effective DNA repair or through mutations or even deletions of targets from the genome (Wimmer and Beisel, 2020). Alternatively, these self-targeting spacers may have biological roles other than immunity such as evolution or RNA degradation (Wimmer and Beisel, 2020). For example, three spacers from *C. botulinum* BoNT C/D strains localized in the chromosome target p1_BKT015925 and had this BoNT phage within their genome. Interestingly, these spacers specifically target variable regions of p1_BKT015925, suggesting that this may be part of a mechanism involved in the prevention of superinfection by several related BoNT phages at the same time. The existence of such a mechanism to prevent superinfection by multiple BoNT phages is also supported by the detection of protospacers from BoNT phages and located on other BoNT phages. For example, spacer 16868_D_90 (present in strains *C. botulinum* BoNT D 16868, BoNT D/C 1274, 1275, and 1276) located on the BoNT D and D/C phages target BoNT C/D phage p1_BKT015925 on the CRISPR Cas7 gene (with four mismatches). The detection of spacers targeting p4_BKT015925 found on BoNT phage p1_BKT015925 and plasmid p2_BKT015925 suggests this kind of competition between MGEs (**Supplementary Tables S10** and **S11**).

CONCLUSION

In the present study, the exploration of the CRISPR-Cas systems in the *Clostridium novyi sensu lato* genospecies found six different CRISPR-Cas system types (mainly I-B) belonging to class 1 and 2 CRISPR loci, protecting against up to 776 MGEs, but only a handful of them being widespread. Although the CRISPR mechanism and CRISPR-MGE interaction have been largely investigated, there is much less information on CRISPR systems used by MGEs targeting other MGEs, whose functions remain largely uncharacterized (Faure et al., 2019). This study highlighted that CRISPR-Cas systems are numerous in *Clostridium novyi sensu lato* strains, and may be present either in the bacterial chromosome or in MGEs. The components carrying CRISPR-Cas systems seemed to have been recruited as anti-MGE systems and for inter-MGE conflicts, to protect mainly against a restricted number of MGEs. The role and function of these CRISPR-Cas systems in the bacterial life cycle and more generally in pathogenesis, in particular when located on MGEs, now need to be explored. Furthermore, studying this restricted mobilome may prove very useful for biocontrol strategies against the pathogenic *Clostridium novyi sensu lato* species.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

CW designed and coordinated the study. CM is the recipient of a grant that funded this study. TG, CW, and MD performed the bioinformatics analysis. TG, CM, MD, MC, and CW participated in data analysis and interpretation of the results. CM, MC, and CW supervised the study. CW, TG, and CM wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the "Institut Carnot AgriFood Transition" (TRACK'SPORE project), the French National Research Agency, and the Bretagne Regional Council of Région Bretagne. TG is the recipient of a PhD grant from ANSES and "Région Bretagne."

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.711413/full#supplementary-material

Supplementary Table S1 | *Clostridium novyi sensu lato* total spacer content.

Supplementary Table S2 | *Clostridium novyi sensu lato* duplicated spacer content.

Supplementary Table S3 | *Clostridium novyi sensu lato* unique spacer content.

Supplementary Table S4 | *Clostridium novyi sensu lato* unique identified spacer origin.

Supplementary Table S5 | *Clostridium novyi sensu lato* unique unidentified spacer origin.

Supplementary Table S6 | BLAST results on *Clostridium novyi sensu lato* unique spacer origins.

Supplementary Table S7 | Proportion of MGEs among the BLAST results on the recognized spacers.

Supplementary Table S8 | Composition of MGE host genus origins.

Supplementary Table S10 | Main targeted mobile genetic element (MGE) genes.

Supplementary Table S11 | Identification of mobile genetic element (MGE) spacer (protospacer) origin.

Supplementary Table S12 | Identification of phage-inducible chromosomal island (PICI) spacer (protospacer) origin.

REFERENCES

Agren, J., Sundström, A., Håfström, T., and Segerman, B. (2012). Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One* 7:e39107. doi: 10.1371/journal.pone.0039107

Andersen, J. M., Shoup, M., Robinson, C., Britton, R., Olsen, K. E., and Barrangou, R. (2016). CRISPR diversity and microevolution in *Clostridium difficile*. *Genome Biol. Evol.* 8, 2841–2855. doi: 10.1093/gbe/evw203

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., et al. (2006). The marine viromes of four oceanic regions. *PLoS Biol.* 4:e368. doi: 10.1371/journal.pbio.0040368

Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12, 177–186. doi: 10.1016/j.chom.2012.06.003

Biswas, A., Gagnon, J. N., Brouns, S. J., Fineran, P. C., and Brown, C. M. (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* 10, 817–827. doi: 10.4161/rna.24046

Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., et al. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. doi: 10.1186/1471-2105-8-209

Bondy-Denomy, J., Qian, J., Westra, E. R., Buckling, A., Guttman, D. S., Davidson, A. R., et al. (2016). Prophages mediate defense against phage infection through diverse mechanisms. *ISME J.* 10, 2854–2866. doi: 10.1038/ismej.2016.79

Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., et al. (2015). Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. *mBio* 6, e01112–e01115. doi: 10.1128/mBio.01112-15

Briner, A. E., Lugli, G. A., Milani, C., Duranti, S., Turroni, F., Gueimonde, M., et al. (2015). Occurrence and diversity of CRISPR-Cas systems in the genus *Bifidobacterium*. *PLoS One* 10:e0133661. doi: 10.1371/journal.pone.0133661

Cady, K. C., White, A. S., Hammond, J. H., Abendroth, M. D., Karthikeyan, R. S. G., Lalitha, P., et al. (2011). Prevalence, conservation and functional analysis of *Yersinia* and *Escherichia* CRISPR regions in clinical *Pseudomonas aeruginosa* isolates. *Microbiology* 157, 430–437. doi: 10.1099/mic.0.045732-0

Cohen, J. E., Wang, R., Shen, R.-F., Wu, W. W., and Keller, J. E. (2017). Comparative pathogenomics of *clostridium tetani*. *PLoS One* 12:e0182909. doi: 10.1371/journal.pone.0182909

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

Crestani, C., Forde, T. L., and Zadoks, R. N. (2020). Development and application of a prophage integrase typing scheme for group B *streptococcus*. *Front. Microbiol.* 11:1993. doi: 10.3389/fmicb.2020.01993

Eitzinger, S., Asif, A., Watters, K. E., Iavarone, A. T., Knott, G. J., Doudna, J. A., et al. (2020). Machine learning predicts new anti-CRISPR proteins. *Nucleic Acids Res.* 48, 4698–4708. doi: 10.1093/nar/gkaa219

Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., et al. (2019). CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* 17, 513–525. doi: 10.1038/s41579-019-0204-7

Fillol-Salom, A., Martínez-Rubio, R., Abdulrahman, R. F., Chen, J., Davies, R., and Penadés, J. R. (2018). Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J.* 12, 2114–2128. doi: 10.1038/s41396-018-0156-3

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:638. doi: 10.1186/1471-2105-8-172

Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., et al. (2015). BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* 34, 169–183. doi: 10.15252/embj.201489455

Hargreaves, K. R., Flores, C. O., Lawley, T. D., and Clokie, M. R. J. (2014). Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium* difficile strains and prophages target multiple phage types within this pathogen. *mBio* 5, e01045–e01013. doi: 10.1128/mbio.01045-13

Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., et al. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* 362:839. doi: 10.1126/science.aav4294

Hawkins, J. S., Wong, S., Peters, J. M., Almeida, R., and Qi, L. S. (2015). Targeted transcriptional repression in bacteria using CRISPR interference (CRISPRi). *Methods Mol. Biol.* 1311, 349–362. doi: 10.1007/978-1-4939-2687-9_23

Karah, N., Samuelsen, Ø., Zarrilli, R., Sahl, J. W., Wai, S. N., and Uhlin, B. E. (2015). CRISPR-cas subtype I-fb in *Acinetobacter baumannii*: evolution and utilization for strain subtyping. *PLoS One* 10:e0118205. doi: 10.1371/journal.pone.0118205

Khan, M. Z., Haider, S., Mansoor, S., and Amin, I. (2019). Targeting plant ssDNA viruses with engineered miniature CRISPR-Cas14a. *Trends Biotechnol.* 37, 800–804. doi: 10.1016/j.tibtech.2019.03.015

Kieper, S. N., Almendros, C., Behler, J., Mckenzie, R. E., Nobrega, F. L., Haagsma, A. C., et al. (2018). Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep.* 22, 3377–3384. doi: 10.1016/j.celrep.2018.02.103

Kim, K.-H., Chang, H.-W., Nam, Y.-D., Roh, S. W., Kim, M.-S., Sung, Y., et al. (2008). Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl. Environ. Microbiol.* 74:5975. doi: 10.1128/AEM.01275-08

Koonin, E. V., and Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 374:20180087. doi: 10.1098/rstb.2018.0087

Koonin, E. V., Makarova, K. S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37, 67–78. doi: 10.1016/j.mib.2017.05.008

Kovanen, S. M., Kivistö, R. I., Rossi, M., and Hänninen, M. L. (2014). A combination of MLST and CRISPR typing reveals dominant *campylobacter jejuni* types in organically farmed laying hens. *J. Appl. Microbiol.* 117, 249–257. doi: 10.1111/jam.12503

Lin, J., Fuglsang, A., Kjeldsen, A. L., Sun, K., Bhoobalan-Chitty, Y., and Peng, X. (2020). DNA targeting by subtype I-D CRISPR–Cas shows type I and type III features. *Nucleic Acids Res.* 48, 10470–10478. doi: 10.1093/nar/gkaa749

Long, J., Xu, Y., Ou, L., Yang, H., Xi, Y., Chen, S., et al. (2019). Diversity of CRISPR/Cas system in *Clostridium perfringens*. *Mol. Gen. Genomics.* 294, 1263–1275. doi: 10.1007/s00438-019-01579-3

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi: 10.1038/s41579-019-0299-x

Mcbride, T. M., Schwartz, E. A., Kumar, A., Taylor, D. W., Fineran, P. C., and Fagerlund, R. D. (2020). Diverse CRISPR-Cas complexes require independent translation of small and large subunits from a single gene. bioRxiv [Preprint]. doi: 10.1101/2020.04.18.045682

Mcginn, J., and Marraffini, L. A. (2019). Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat. Rev. Microbiol.* 17, 7–12. doi: 10.1038/s41579-018-0071-7

Mertaoja, A., Nowakowska, M. B., Mascher, G., Heljanko, V., Groothuis, D., Minton, N. P., et al. (2021). CRISPR-Cas9-based toolkit for *clostridium botulinum* group II spore and sporulation research. *Front. Microbiol.* 12:617269. doi: 10.3389/fmicb.2021.617269

Mir, A., Edraki, A., Lee, J., and Sontheimer, E. J. (2018). Type II-C CRISPR-Cas9 biology, mechanism, and application. *ACS Chem. Biol.* 13, 357–365. doi: 10.1021/acschembio.7b00855

Negahdaripour, M., Nezafat, N., Hajighahramani, N., Rahmatabadi, S. S., and Ghasemi, Y. (2017). Investigating CRISPR-Cas systems in *clostridium botulinum* via bioinformatics tools. *Infect. Genet. Evol.* 54, 355–373. doi: 10.1016/j.meegid.2017.06.027

Pan, M., Nethery, M. A., Hidalgo-Cantabrana, C., and Barrangou, R. (2020). Comprehensive mining and characterization of CRISPR-Cas systems in *Bifidobacterium*. *Microorganisms* 8:720. doi: 10.3390/microorganisms8050720

Penadés, J. R., and Christie, G. E. (2015). The phage-inducible chromosomal islands: a family of highly evolved molecular parasites. *Annu. Rev. Virol.* 2, 181–201. doi: 10.1146/annurev-virology-031413-085446

Pfeifer, E., Moura De Sousa, J. A., Touchon, M., and Rocha, E. P. C. (2021). Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res.* 49, 2655–2673. doi: 10.1093/nar/gkab064

Pyenson, N. C., and Marraffini, L. A. (2017). Type III CRISPR-Cas systems: when DNA cleavage just isn't enough. *Curr. Opin. Microbiol.* 37, 150–154. doi: 10.1016/j.mib.2017.08.003

Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *clostridium*. *Sci. Rep.* 6:25666. doi: 10.1038/srep25666

Rosario, K., Nilsson, C., Lim, Y. W., Ruan, Y., and Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environ. Microbiol.* 11, 2806–2820. doi: 10.1111/j.1462-2920.2009.01964.x

Shariat, N., Dimarzio, M. J., Yin, S., Dettinger, L., Sandt, C. H., Lute, J. R., et al. (2013). The combination of CRISPR-MVLST and PFGE provides increased discriminatory power for differentiating human clinical isolates of *salmonella enterica subsp. enterica* serovar Enteritidis. *Food Microbiol.* 34, 164–173. doi: 10.1016/j.fm.2012.11.012

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., et al. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* 15, 169–182. doi: 10.1038/nrmicro.2016.184

Sikorski, A., Argüello-Astorga, G. R., Dayaram, A., Dobson, R. C., and Varsani, A. (2013). Discovery of a novel circular single-stranded DNA virus from porcine faeces. *Arch. Virol.* 158, 283–289. doi: 10.1007/s00705-012-1470-0

Silas, S., Lucas-Elio, P., Jackson, S. A., Aroca-Crevillén, A., Hansen, L. L., Fineran, P. C., et al. (2017). Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *eLife* 6:27601. doi: 10.7554/eLife.27601

Skarin, H., Hafström, T., Westerberg, J., and Segerman, B. (2011). *Clostridium botulinum* group III: A group with dual identity shaped by plasmids, phages and mobile elements. *BMC Genomics* 12:185. doi: 10.1186/1471-2164-12-185

Skarin, H., and Segerman, B. (2011). Horizontal gene transfer of toxin genes in *clostridium botulinum*: involvement of mobile elements and plasmids. *Mob. Genet. Elem.* 1, 213–215. doi: 10.4161/mge.1.3.17617

Skarin, H., and Segerman, B. (2014). Plasmidome interchange between *clostridium botulinum*, *clostridium novyi* and *clostridium haemolyticum* converts strains of independent lineages into distinctly different pathogens. *PLoS One* 9:e107777. doi: 10.1371/journal.pone.0107777

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186. doi: 10.1038/nrmicro1793

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Tamulaitis, G., Venclovas, Č., and Siksnys, V. (2017). Type III CRISPR-Cas immunity: major differences brushed aside. *Trends Microbiol.* 25, 49–61. doi: 10.1016/j.tim.2016.09.012

Terns, M. P. (2018). CRISPR-based technologies: impact of RNA-targeting systems. *Mol. Cell* 72, 404–412. doi: 10.1016/j.molcel.2018.09.018

Thomas, P., Semmler, T., Eichhorn, I., Lübke-Becker, A., Werckenthin, C., Abdel-Glil, M. Y., et al. (2017). First report of two complete clostridium chauvoei genome sequences and detailed in silico genome analysis. *Infect. Genet. Evol.* 54, 287–298. doi: 10.1016/j.meegid.2017.07.018

Wimmer, F., and Beisel, C. L. (2020). CRISPR-Cas systems and the paradox of self-targeting spacers. *Front. Microbiol.* 10:3078. doi: 10.3389/fmicb.2019.03078

Woudstra, C., Le Marechal, C., Souillard, R., Anniballi, F., Auricchio, B., Bano, L., et al. (2017). Investigation of *clostridium botulinum* group III's mobilome content. *Anaerobe* 49, 71–77. doi: 10.1016/j.anaerobe.2017.12.009

Woudstra, C., Le Marechal, C., Souillard, R., Bayon-Auboyer, M. H., Mermoud, I., Desoutter, D., et al. (2016). New insights into the genetic diversity of *clostridium botulinum* group III through extensive genome exploration. *Front. Microbiol.* 7:757. doi: 10.3389/fmicb.2016.00757

Xie, X., Hu, Y., Xu, Y., Yin, K., Li, Y., Chen, Y., et al. (2017). Genetic analysis of *salmonella enterica serovar Gallinarum biovar Pullorum* based on characterization and evolution of CRISPR sequence. *Vet. Microbiol.* 203, 81–87. doi: 10.1016/j.vetmic.2017.02.010

Yin, S., Jensen, M. A., Bai, J., Debroy, C., Barrangou, R., and Dudley, E. G. (2013). The evolutionary divergence of Shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. *Appl. Environ. Microbiol.* 79, 5710–5720. doi: 10.1128/AEM.00950-13

Zhang, R., Xu, W., Shao, S., and Wang, Q. (2021). Gene silencing through CRISPR interference in bacteria: current advances and future prospects. *Front. Microbiol.* 12:635227. doi: 10.3389/fmicb.2021.635227

Zheng, Z., Zhang, Y., Liu, Z., Dong, Z., Xie, C., Bravo, A., et al. (2020). The CRISPR-Cas systems were selectively inactivated during evolution of *Bacillus cereus* group for adaptation to diverse environments. *ISME J.* 14, 1479–1493. doi: 10.1038/s41396-020-0623-5

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Public Health

MINI REVIEW
published: 16 December 2021
doi: 10.3389/fpubh.2021.744179

Wound Botulism Among Persons Who Inject Black Tar Heroin in New Mexico, 2016

Nicole Middaugh[1,2], *Leslie Edwards*[1]*, *Kevin Chatham-Stephens*[1] and *D. Fermin Arguello*[2]

[1] Centers for Disease Control and Prevention, Atlanta, GA, United States, [2] New Mexico Department of Health, Santa Fe, NM, United States

Outbreaks of wound botulism are rare, but clinicians and health departments should maintain suspicion for signs, symptoms, and risk factors of wound botulism among persons who inject drugs in order to initiate treatment quickly. This report describes an outbreak of three wound botulism cases among persons in two adjacent counties who injected drugs. Provisional information about these cases was previously published in the CDC National Botulism Surveillance Summary. All three cases in this outbreak were laboratory-confirmed, including one case with detection of botulinum toxin type A in a wound culture sample taken 43 days after last possible heroin exposure. Findings highlight the delay in diagnosis which led to prolonged hospitalization and the persistence of botulinum toxin in one patient.

Keywords: wound, botulism, heroin, black tar heroin, outbreak, intubation

OPEN ACCESS

Edited by:
Fabrizio Anniballi,
Istituto Superiore di Sanità, Italy

Reviewed by:
Azzurra Schicchi,
Poison Control Center of Pavia, Italy
Michael Adler,
USAMRICD, United States

***Correspondence:**
Leslie Edwards
fjr2@cdc.gov

Specialty section:
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

Received: *22 July 2021*
Accepted: *17 November 2021*
Published: *16 December 2021*

Citation:
Middaugh N, Edwards L, Chatham-Stephens K and Arguello DF (2021) Wound Botulism Among Persons Who Inject Black Tar Heroin in New Mexico, 2016. Front. Public Health 9:744179. doi: 10.3389/fpubh.2021.744179

INTRODUCTION

Botulism is a potentially fatal illness caused by a neurotoxin produced most frequently by the bacterium *Clostridium botulinum*. There are six different types of botulism, distinguished by the way in which neurotoxin exposure occurs—foodborne, infant, wound, adult intestinal colonization, iatrogenic, and inhalational. In 2017, 182 laboratory-confirmed botulism cases were reported in the United States (1). Botulism typically manifests with symmetrical descending paralysis and, if untreated, the illness can lead to respiratory failure and death. *C. botulinum* is ubiquitous in soil and produces seven botulinum toxins (A–G). Human illnesses are caused by types A, B, E and, rarely, F (2).

Wound botulism is caused by contamination of a wound with *C. botulinum* spores. Under anaerobic conditions present in a wound, the spores germinate and produce botulinum toxin. Wound botulism in the United States occurs mostly among persons who inject black tar heroin (BTH) subcutaneously or intradermally (i.e., skin popping) (3). Typically, wound botulism accounts for <10% of all U.S. botulism cases annually, and most wound botulism cases are reported from California (1). **Table 1** includes epidemiologic information about outbreaks of wound botulism linked to injection drug use. An increase in wound botulism cases among persons who inject drugs (PWID) that used BTH was reported in California from 1988 to 1998 (3). From 2002 to 2018, 376 laboratory confirmed wound botulism cases in the U.S. were among PWID and 10% of cases reported contacts with other cases (1). In 2008, a wound botulism outbreak among six PWID that used heroin was reported in Dublin, Ireland (4). There were no epidemiological links among cases in this outbreak and only one case was laboratory confirmed due to delayed suspicion of botulism among hospital clinicians and challenges with specimen collection and testing. In 2009,

TABLE 1 | Wound botulism outbreaks among persons who inject drugs (PWID).

Year(s)	Location	Number of cases*	Number of lab confirmed cases	Botulism toxin identified	Botulism testing types used	Drug(s) used	Reference
1988–1998	California, USA	105	105	A (most cases)	MBA	BTH	Werner
2005	Germany	16	4	B	Culture, PCR, PFGE, MBA	Heroin	Schroeter
2008	Dublin, Ireland	6	1	B	MBA	Heroin	Barry
2013	Norway	6	4	Not listed	MBA	Heroin	Macdonald
2017	Scotland	25	9	B	MBA, culture, PCR	Heroin	Martin
2017–2018	California, USA	9	8	A	MBA, MS	BTH and Heroin	Peak

BTH, black tar heroin; MBA, mouse bioassay; MS, mass spectrometry; PCR, polymerase chain reaction; USA, United States of America.

a wound botulism outbreak among six PWID that used heroin was reported in North Rhine, Germany (5). Mouse bioassay testing was used to confirm four cases. In late 2013, a cluster of six wound botulism cases among PWID that used heroin was reported in Norway (6). Two cases knew one another however none reported sharing drugs with other cases. From 2014 to 2015, an outbreak of 25 cases of wound botulism among PWID that used heroin was reported in Glasgow, Scotland (7). One married couple was identified among cases; however, no other social links were reported among cases. These outbreaks highlight the geographic variability of wound botulism outbreaks among PWID although it remains challenging to understand the connection among cases and the precise aspects of the injection drug use process such as contamination of drugs and or drug paraphernalia that led to botulism transmission.

This report describes an outbreak of wound botulism among three epidemiologically linked persons that injected drugs in New Mexico in 2016. From 1997 to 2015, four confirmed cases of wound botulism were reported in New Mexico, with three cases linked to injection drug use and one following a motor vehicle accident. This investigation highlights the need for enhanced surveillance for wound botulism among close contacts of persons who inject drugs and the importance of health alerts to the medical community to raise clinicians' awareness of wound botulism.

METHODS

During June–July 2016, the New Mexico Department of Health (NMDOH) investigated an outbreak of three wound botulism cases among persons who injected drugs in two neighboring counties in New Mexico. Patients and family members were interviewed by NMDOH in order to collect information about foods eaten and injection drug use in the 3 days prior to illness. Hospital clinicians consulted with NMDOH and with the Centers for Disease Control and Prevention (CDC) Botulism Clinical Consultation Service in order to determine whether treatment with heptavalent botulism antitoxin (BAT) was appropriate for each patient with suspected botulism (8). The CDC operates a botulism clinical consultation service 24 h a day/7 days a week that hospital based clinicians and their respective state health departments can use to obtain BAT. CDC Quarantine Station Officers coordinated shipment of BAT with hospital pharmacists. In adults, BAT is administered using intravenous infusion beginning at an infusion rate of 0.5 mL/min and may be increased to an 2 mL/min infusion rate if the medication is well-tolerated. Infusion in completed in 3.5–6.7 h depending on the rate of infusion used. Clinical specimens were tested for botulinum toxin using mouse bioassay (9) and EndoPEP-MS mass spectrometry (10). Limited details about this outbreak were previously published in the CDC 2016 National Botulism Surveillance Summary (1). CDC reviewed this report for human subjects' protection and deemed it not to constitute research.

RESULTS

On July 1, 2016, Patient 1 presented to the emergency department with shortness of breath, dysphagia and double vision which began on June 29. The patient experienced respiratory arrest, requiring intubation and mechanical ventilation, and was immediately transferred to a tertiary hospital where wound botulism was suspected based on clinical presentation, history of skin-popping black tar heroin, and presence of a right hip abscess. The patient reported multiple occasions of skin-popping in the days before onset of symptoms. After evaluating the patient, the consulting neurologist contacted the NMDOH to request BAT. NMDOH contacted CDC on July 2, 2016, for consultation and to arrange BAT shipment and specimen testing. The patient received BAT and penicillin on July 2, 2016. Botulinum toxin type A was detected by mouse bioassay in serum collected on July 1. The patient was transferred to a long-term care facility until they were extubated, and subsequently required inpatient physical rehabilitation.

NMDOH issued a Health Alert Network (HAN) notification to clinicians throughout New Mexico on July 12, 2016, communicating the potential for *C. botulinum*-contaminated black tar heroin and advising clinicians to consider botulism in patients with neurological signs and symptoms, particularly those who skin-pop black tar heroin. NMDOH collaborated with the New Mexico Harm Reduction Program to distribute information to participants of New Mexico's needle exchange program, urging

persons with signs and symptoms of botulism to seek immediate medical attention.

On July 14, a regional emergency department physician notified NMDOH of a new patient with suspected botulism ("Patient 2") who presented with blurry vision, dysphagia, dysarthria, neck weakness and extremity weakness which began on July 11. The physician had read the HAN notification and obtained the patient's history, including skin-popping black tar heroin. The patient reported being an intimate partner of Patient 1 and had recently shared heroin with the patient. Because of their neurologic symptoms, connection to Patient 1 and black tar heroin use, Patient 2 was transferred to a tertiary hospital where specimens were collected for botulism testing. NMDOH immediately contacted CDC for clinical consultation and to obtain BAT. The patient received BAT 9 h after presenting to the regional emergency department. Botulinum toxin type A was detected by EndoPEP-MS testing in serum collected on July 14. Patient 2 was discharged home 4 days after receiving BAT, never requiring intubation.

On June 23, 2016, Patient 3 presented to a local hospital complaining of a spider bite, dysphagia, diplopia, lip swelling and tongue swelling with onset the same day. They had an indurated area on the right buttock with reddish-purple discoloration but no visible abscesses and was discharged home with antibiotics. On June 24, Patient 3 was found unresponsive at home and was taken to another hospital by emergency medical services and was intubated shortly after arrival due to respiratory failure initially attributed to illicit substance toxicity. On July 4, 11 days after onset, Patient 3 was transferred to a tertiary hospital after their inability to wean from mechanical ventilation and significant extremity weakness. After an extensive workup that included an inconclusive nerve conduction study and needle electromyography, but did not include a CT scan for abscesses, the patient was diagnosed with Miller-Fischer variant Guillain-Barré Syndrome. Botulism was on the differential diagnosis; however, botulinum toxin testing was not conducted. The patient was treated with intravenous immunoglobulin (IVIG) and transferred to a long-term care facility 5 days later.

NMDOH discussed the possibility of the patient having botulism with the long-term care facility (LTCF) physician. The LTCF physician was also caring for Patient 1 and suspected that Patient 3 also had botulism, not Guillain-Barré Syndrome. On July 14, NMDOH interviewed Patient 3 as a potential third case in the outbreak, because of their continued symptoms and failure to improve after IVIG. Patient 3 reported skin-popping black tar heroin on June 24 and that they knew the index patient, Patient, 1 but that they had not shared drugs or drug paraphernalia. NMDOH worked with the physicians to obtain serum samples previously drawn on July 4 (12 days after onset) and newly drawn on July 17 (25 days after onset) for botulinum toxin testing. On July 29, botulinum toxin type A was detected by EndoPEP-MS in serum collected on July 4. Because more than 3 weeks had passed since that serum had been collected, and it was unclear if the patient still had circulating botulinum toxin, the LTCF physician opted to wait to request BAT, pending test results of the July 17 drawn serum. On August 5, botulinum toxin type A was detected in the July 17 serum tested by EndoPEP-MS. The patient's neurological signs and symptoms had still not improved; therefore, a CT scan of the hips was conducted to identify a persistent source of toxin. The CT revealed two occult abscesses deep in the patient's right and left buttocks. On August 6, the right-side abscess was drained, serum and wound cultures were obtained, and BAT was administered. Wound cultures tested positive for botulinum toxin type A-−43 days after symptom onset and the last possible exposure to black tar heroin. Botulinum toxin type A was not detected by mouse bioassay in serum drawn on August 6, which was obtained before BAT was administered. Patient 3 required mechanical ventilation for 110 days. They were eventually transferred to an inpatient physical rehabilitation facility. **Table 2** lists the onset date, time from presentation to BAT administration, duration of mechanical ventilation, and outcome for the three patients.

DISCUSSION

NMDOH investigated an outbreak of three laboratory-confirmed, epidemiologically linked, type A wound botulism cases among persons who injected black tar heroin. The persons presented to emergency departments with neurologic signs and symptoms associated with botulism (e.g., diplopia, dysphagia). Patient 2 was promptly diagnosed with botulism and treated with BAT within 9 h; this patient did not develop respiratory compromise and hospital discharge occurred within a few

TABLE 2 | Time to botulism antitoxin (BAT) administration, duration of mechanical ventilation, and outcome among wound botulism patients ($N = 3$)[a].

	Onset date	Time from illness onset date to mechanical ventilation	Time from illness onset date to botulism antitoxin	Laboratory confirmed	Botulism testing types used	Botulism toxin identified	Duration of mechanical ventilation	Patient outcome
Patient 1[b]	6/29/2016	2 days	5 days	Yes	MBA	A	30 days	Survived
Patient 2	7/11/2016	n/a	9 h	Yes	MS	A	Not required	Survived
Patient 3[b]	6/23/2016	1 day	43 days	Yes	MS, wound culture	A	110 days	Survived

[a] *Time from exposure to onset of symptoms is unavailable due to multiple instances of injection drug use prior to onset of signs and symptoms.*

[b] *Required inpatient physical rehabilitation.*

MBA, mouse bioassay; MS, mass spectrometry.

days. Wound botulism was not considered during patient 1 and patient 3's initial emergency department evaluations and consequently wound care and BAT administration were delayed. Both suffered respiratory arrest and required intubation and mechanical ventilation for weeks or months. All three patients engaged in behaviors associated with wound botulism, including skin-popping and using black tar heroin (3, 11, 12). Two patients shared drugs and needles, and each had an occult or visible abscess.

Wound botulism diagnosis among patients who inject drugs continues to be challenging. Since this outbreak was reported in 2016, a wound botulism outbreak among nine persons who injected black tar heroin and/or heroin was reported in San Diego, California (13). Mechanical ventilation was required for 6 (67%) patients and the median duration of hospitalization was 15 days (range: 9–67 days). There was a delay in botulism diagnosis for four (44%) of patients who were initially diagnosed with drug intoxication. Both the New Mexico outbreak and the San Diego outbreak demonstrate the need for increased awareness among hospital clinicians about the possibility of wound botulism among persons who inject drugs.

BAT is a well-tolerated medication with a low incidence of adverse events. Among 249 patients treated with BAT, the most frequently reported adverse events included fever (4%), rash (2%), chills, agitation, edema or nausea (1% each) (14). Mild serum sickness was reported in one patient 11 days after receiving BAT. Patients treated with BAT should be monitoring for signs of hypersensitivity during BAT infusion and afterwards for adverse events (8). In this outbreak, there was an association between the duration of ventilator use and the amount of time between onset of symptoms and BAT receipt. This observation is consistent with prior reports that earlier treatment with botulism antitoxin is linked to better clinical outcomes including decreased duration of hospitalization (7, 15) and decreased mortality (16).

From 2002 to 2018 in the United States, 95% of lab confirmed wound botulism cases were reported among PWID (17). Traumatic injury was reported among 3% of cases with the remainder of cases having no identified risk factor for wound botulism. Traumatic injuries reported among cases include motor vehicle accident, gunshot wound, or other injury resulting in an open fracture. Wound botulism is a severe and often life-threatening illness that requires rapid identification and proper treatment, including identifying and debriding abscesses, providing antimicrobial therapy when indicated, and administering BAT as soon as possible. Among 94 patients with wound botulism from injecting drugs, Werner et al. (3) reported that 14 (15%) patients had wounds that did not appear grossly infected. Detailed examination or injection site imaging might be needed to identify deep abscesses, as in Patient 3. The implementation of new botulism testing including EndoPEP-mass spectrometry has been helpful in decreasing the amount of time needed to complete botulism clinical testing, although mass spectrometry is available in very few public health laboratories. In many ways wound botulism may be more challenging to diagnose than foodborne botulism due to the sensitive nature of exposures among PWID. Wound botulism remains a challenge to rapidly diagnose and treat as patients may be hesitant to seek medical care and to report use of illegal drugs including heroin. Skilled interviewers are needed to help gather information about exposures among PWID. An additional challenge in diagnosing wound botulism among PWID is that certain signs and symptoms of botulism may be misdiagnosed as opiate overdose. Clinicians suspecting botulism in a patient should immediately call the emergency telephone number at their state health department to request a clinical consultation for botulism. The CDC Botulism Clinical Consultation Program is available 24 h a day, 7 days a week with staff available to help consult with hospital-based clinicians and health department staff in order to facilitate treatment with BAT and botulism testing.

CONCLUSIONS

Outbreaks of wound botulism are rare, but clinicians and health departments should maintain suspicion for signs, symptoms, and risk factors of wound botulism among persons who inject drugs—especially among persons who skin-pop black tar heroin—to initiate treatment quickly. Minimizing the time from presentation to abscess debridement and to BAT administration, has been shown to reduce the need for mechanical ventilation and decrease the duration of intensive care (18, 19). In this outbreak, we detected botulinum toxin type A in a wound culture 43 days after last possible heroin exposure. Early detection and debridement of abscesses, use of imaging, and treatment with antibiotics and BAT are essential to prevent ongoing toxin production. Communication between clinicians, health departments, and CDC is key to rapid identification of cases, especially in outbreaks.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was provided by the New Mexico Department of Health and the Centers for Disease Control and Prevention.

ACKNOWLEDGMENTS

This investigation would not have been possible without the help of CDC's Botulism Consultation Service, specifically Julie Self, Lindsay McCrickard, Kelly Gambino Shirley, Louise Francois Watkins, and laboratorians from New Mexico's Scientific Laboratory Division (Nicole Espinoza and Lisa Onischuk) and from CDC (Jan Dykes, Suzy Kalb, and Carolina Luquez). New Mexico Public Health Officials, Julianna Fererra, Francella Jojola, Michael Landen, and Chad Smelser were instrumental to this investigation.

REFERENCES

1. Centers for Disease Control and Prevention. *Botulism Annual Summary, 2017*. (2019). Available online at: https://www.cdc.gov/botulism/pdf/Botulism-2016-SUMMARY-508.pdf (accessed May 10, 2021).
2. Sobel J. Botulism. *Clin Infect Dis.* (2005) 41:1167–73. doi: 10.1086/444507
3. Werner SB, Passaro D, Mcgee J, Schechter R, Vugia DJ. Wound botulism in California, 1951-1998: recent epidemic in heroin injectors. *Clin Infect Dis.* (2000) 31:1018–24. doi: 10.1086/318134
4. Barry J, Ward M, Cotter S, MacDiarmada J. Botulism in injecting drug users, Dublin, Ireland, November-December 2008. *Eurosurveillance.* (2009) 14:1–3. doi: 10.2807/ese.14.01.19082-en
5. Schroeter M, Alpers K, Van Treeck U, Frank C, Rosenkoetter N, Schaumann R. Outbreak of wound botulism in injecting drug users. *Epidemiol Infect.* (2009) 137:1602–8. doi: 10.1017/S0950268809002544
6. Macdonald E, Arnesen TM, Brantsaeter AB, Gerlyng P, Grepp M, Hansen BA, et al. Outbreak of wound botulism in people who inject drugs, Norway, October to November 2013. *Euro Surveill.* (2013) 18:20630. doi: 10.2807/1560-7917.ES2013.18.45.20630
7. Martin SJ, Penrice G, Amar C, Grant K, Gorrie GH. Wound botulism, its neurological manifestations, treatment and outcomes: a case series from the Glasgow outbreak, 2015. *Scott Med J.* (2017) 62:136–41. doi: 10.1177/0036933017707165
8. Food and Drug Administration. *BAT (Botulism Antitoxin Heptavalent (A, B, C, D, E, F, G) - (Equine)*. (2018). Available online at: https://www.fda.gov/vaccines-blood-biologics/approved-blood-products/bat-botulism-antitoxin-heptavalent-b-c-d-e-f-g-equine (accessed April 27, 2021).
9. Food and Drug Administration. *Bacteriological Analytical Manual, Chapter 17*. (2001). Available online at: https://www.fda.gov/food/laboratory-methods-food/bam-chapter-17-clostridium-botulinum (accessed July 15, 2021).
10. Barr JR, Moura H, Boyer AE, Woolfitt AR, Kalb SR, Pavlopoulos A, et al. Botulinum neurotoxin detection and differentiation by mass spectrometry. *Emerg Infect Dis.* (2005) 11:1578–83. doi: 10.3201/eid1110.041279
11. Passaro DJ, Werner S, Mcgee J, Mac Kenzie WR, Vugia DJ. Wound botulism associated with black tar heroin among injecting drug users. *JAMA.* (1998) 279:859–63. doi: 10.1001/jama.279.11.859
12. Gordon RJ, Lowy FD. Bacterial infections in drug users. *N Engl J Med.* (2005) 353:1945–54. doi: 10.1056/NEJMra042823
13. Peak CM, Rosen H, Kamali A, Poe A, Shahkarami M, Kimura AC, et al. Wound botulism outbreak among persons who use black tar heroin - San Diego County, California, 2017-2018. *MMWR Morb Mortal Wkly Rep.* (2019) 67:1415–8. doi: 10.15585/mmwr.mm675152a3
14. Yu PA, Lin NH, Mahon BE, Sobel J, Yu Y, Mody RK, et al. Safety and improved clinical outcomes in patients treated with new equine-derived heptavalent botulinum antitoxin. *Clin Infect Dis.* (2017) 66:S57–S64. doi: 10.1093/cid/cix816
15. Chang GY, Ganguly G. Early antitoxin treatment in wound botulism results in better outcome. *Eur Neurol.* (2003) 49:151–3. doi: 10.1159/000069073
16. O'Horo JC, Harper EP, El Rafei A, Ali R, Desimone DC, Sakusic A, et al. Efficacy of antitoxin therapy in treating patients with foodborne botulism: a systematic review and meta-analysis of cases, 1923-2016. *Clin Infect Dis.* (2017) 66:S43–S56. doi: 10.1093/cid/cix815
17. Centers for Disease Control and Prevention. *National Botulism Surveillance*. (2018). Available online at: https://www.cdc.gov/botulism/surveillance.html (accessed September 5, 2021).
18. Sandrock CE, Murin S. Clinical predictors of respiratory failure and long-term outcome in black tar heroin-associated wound botulism. *Chest.* (2001) 120:562–6. doi: 10.1378/chest.120.2.562
19. Offerman SR, Schaefer M, Thundiyil JG, Cook MD, Holmes JF. Wound botulism in injection drug users: time to antitoxin correlates with intensive care unit length of stay. *Western J Emerg Med.* (2009) 10:251–6.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 February 2022
doi: 10.3389/fmicb.2021.787726

Endogenous CRISPR-Cas Systems in Group I *Clostridium botulinum* and *Clostridium sporogenes* Do Not Directly Target the Botulinum Neurotoxin Gene Cluster

*Travis G. Wentz[1,2,3], Benjamin J. M. Tremblay[4], Marite Bradshaw[3], Andrew C. Doxey[4], Shashi K. Sharma[2], John-Demian Sauer[5] and Sabine Pellett[3]**

[1] Microbiology Doctoral Training Program, University of Wisconsin–Madison, Madison, WI, United States, [2] Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD, United States, [3] Department of Bacteriology, University of Wisconsin–Madison, Madison, WI, United States, [4] Department of Biology, University of Waterloo, Waterloo, ON, Canada, [5] Department of Medical Microbiology and Immunology, University of Wisconsin–Madison, Madison, WI, United States

OPEN ACCESS

Edited by:
Fabrizio Anniballi, Department of Food Safety, Nutrition and Veterinary Public Health, National Institute of Health, Istituto Superiore di Sanità (ISS), Italy

Reviewed by:
Changyi Zhang, University of Illinois Urbana-Champaign, United States
Enas Newire, University of Liverpool, United Kingdom

***Correspondence:**
Sabine Pellett
Sabine.pellett@wisc.edu

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *01 October 2021*
Accepted: *27 December 2021*
Published: *09 February 2022*

Citation:
Wentz TG, Tremblay BJM, Bradshaw M, Doxey AC, Sharma SK, Sauer J-D and Pellett S (2022) Endogenous CRISPR-Cas Systems in Group I Clostridium botulinum and Clostridium sporogenes Do Not Directly Target the Botulinum Neurotoxin Gene Cluster. Front. Microbiol. 12:787726. doi: 10.3389/fmicb.2021.787726

Most strains of proteolytic group I *Clostridium botulinum* (G1 *C. botulinum*) and some strains of *Clostridium sporogenes* possess genes encoding botulinum neurotoxin (BoNT), a potent neuroparalytic agent. Within G1 *C. botulinum,* conserved *bont* gene clusters of three major toxin serotypes (*bont*/A/B/F) can be found on conjugative plasmids and/or within chromosomal pathogenicity islands. CRISPR-Cas systems enable site-specific targeting of previously encountered mobile genetic elements (MGE) such as plasmids and bacteriophage through the creation of a spacer library complementary to protospacers within the MGEs. To examine whether endogenous CRISPR-Cas systems restrict the transfer of *bont* gene clusters across strains we conducted a bioinformatic analysis profiling endogenous CRISPR-Cas systems from 241 G1 *C. botulinum* and *C. sporogenes* strains. Approximately 6,200 CRISPR spacers were identified across the strains and Type I-B, III-A/B/D *cas* genes and CRISPR array features were identified in 83% of the strains. Mapping the predicted spacers against the masked strain and RefSeq plasmid dataset identified 56,000 spacer–protospacer matches. While spacers mapped heavily to targets within *bont*(+) plasmids, no protospacers were identified within the *bont* gene clusters. These results indicate the toxin is not a direct target of CRISPR-Cas but the plasmids predominantly responsible for its mobilization are. Finally, while the presence of a CRISPR-Cas system did not reliably indicate the presence or absence of a *bont* gene cluster, comparative genomics across strains indicates they often occupy the same hypervariable loci common to both species, potentially suggesting similar mechanisms are involved in the acquisition and curation of both genomic features.

Keywords: CRISPR-Cas, CRISPR, *botulinum*, botulinum neurotoxin, conjugative plasmids, horizontal gene transfer, *Clostridium botulinum, Clostridium sporogenes*

INTRODUCTION

Botulinum neurotoxins are potent proteinaceous toxins that are horizontally distributed throughout multiple species of *Clostridium*, a genus of anaerobic, Gram-positive bacteria (Collins and East, 1998). Species differ in the serotypes they produce and serotypes vary in their ability to cause the disease botulism in humans. Eight antigenically distinct botulinum neurotoxin (BoNT) serotypes (A–G, X) have been identified in certain strains across multiple *Clostridium* species, including four species groups of *C. botulinum* (G1–4), *C. sporogenes*, *C. butyricum*, and *C. baratii* (Collins and East, 1998; Mansfield et al., 2015, 2019; Zhang et al., 2017, 2018; Brunt et al., 2018, 2020b; Contreras et al., 2019). Human botulism, systemic flaccid paralysis caused by BoNT-mediated blockade of neurotransmitter release at the neuromuscular junction, is caused by BoNT serotypes A, B, E, and F (Johnson and Montecucco, 2008). The primary BoNT serotypes produced by proteolytic group I *Clostridium botulinum* (G1 *C. botulinum*) strains are BoNT A/B/F and can occur within a conjugative plasmid or as part of a chromosomally integrated genomic island (Brunt et al., 2020a). In contrast, *C. sporogenes* only produces BoNT/B despite being the nearest neighbor species to G1 *C. botulinum* with ∼93% shared nucleotide identity between the two species. The *bont/B* gene is generally plasmid-borne in most toxigenic *C. sporogenes* strains (Weigand et al., 2015; Brunt et al., 2020a). Phylogenetic analysis based on whole genome assemblies indicates non-toxic G1 *C. botulinum* strains are rare while non-toxic *C. sporogenes* strains are relatively common (Brunt et al., 2020a). Despite G1 *C. botulinum* being responsible for a significant portion of foodborne botulism cases and being the predominant source of infant botulism due to colonization of the infant's intestine by the toxin-producing bacteria (Arnon et al., 1979; Nevas et al., 2005), numerous questions remain regarding the means of, and restrictive barriers to, the horizontal transfer of the *bont* gene cluster in these two species. The *bont* gene carrying plasmid pCLJ from G1 *C. botulinum* has been experimentally transferred via conjugation to *C. sporogenes*, *C. butyricum*, and G3 *C. botulinum* (Marshall et al., 2010; Nawrocki et al., 2018), demonstrating inter- and intra-species plasmid transfer can occur. While plasmids are the primary *bont* associated mobile genetic element (MGE) in most *bont*(+) species, phage carry the *bont/C* and *D* genes in G3 *C. botulinum* (Eklund and Poysky, 1974). The *bont/C* and *D* genes in G3 *C. botulinum* reside within a prophage region that has been experimentally cured, ablating toxicity (Eklund and Poysky, 1974). Similar dynamics have not been observed in other species groups. While prophage have rarely been identified near specific *bont* insertion sites in several G1 *C. botulinum* strains (Smith et al., 2021b), there is no additional evidence that currently suggests phage as a driver of the propagation of the *bont* virulence factor in G1 *C. botulinum* or *C. sporogenes*.

In G1 *C. botulinum*, the mechanism by which chromosomal *bont* gene clusters are established remains unknown. However, conserved integration sites have been identified across strains and several types of MGE have been observed nearby. Several recurring integration sites for *bont* gene clusters have been identified within G1 *C. botulinum*. Within the chromosome, these include sites within *arsC*, *pulE*, disrupting the respective genes, and sites near *brnq* (Hill et al., 2009, 2015; Dover et al., 2013; Smith et al., 2021b). Several *bont* gene clusters are integrated at distinct sites. The chimeric *bont/FA(H)* gene cluster occurs at a non-standard insertion site and possesses several unique characteristics (Dover et al., 2014; Gonzalez-Escalona et al., 2014) and the unique *bont/X* gene cluster is integrated between a putative chitinase (RSJ2_770) and copper chaperone (RSJ2_773). Together, the *arsC*, *brnq*, and chitinase sites occur within 90 kbp of each other and together account for the vast majority of chromosomally integrated *bont* gene clusters. Insertion sequences (IS) are frequently present in the vicinity of *bont* gene clusters and are also a candidate for *bont* propagation. Alone, IS are simple MGEs, under 2.5 kbp and in possession of the bare minimum gene contingent necessary to facilitate their insertion and sometimes excision from a genomic site (Chandler and Mahillon, 2002). IS flanked genomic regions may become co-mobilized as a composite transposon (Siguier et al., 2014). In addition, some IS elements are adept at causing genomic rearrangements and deletions (Vandecraen et al., 2017). A recent study has demonstrated that ISs potentially play a major role in transferring virulence associated genes from conjugative plasmids to the chromosome (Che et al., 2021). Both intact and degraded ISs are known to occur within the vicinity of and, in some cases, flanking *bont* gene clusters (Smith et al., 2007, 2015; Hill et al., 2009; Dover et al., 2014). However, IS activity has not been experimentally validated in *C. botulinum* and other *bont* gene cluster carrying *Clostridia*, and it remains unknown whether IS play a role in horizontal *bont* gene cluster mobilization. Regardless of genomic localization, the ∼3.9-kbp *bont* gene is adjacent to a catalytically inactive ∼3.6-kbp paralog non-toxic non-hemagglutinin (*ntnh*) and either the *hemagglutinin* genes (*ha-33*, *ha-17*, *ha-70*) or *p47/orfX* genes (*p47*, *orfX1–3*) (Tsuzuki et al., 1990; Willems et al., 1993; Henderson, 1996), forming the *bont* gene cluster. The gene cluster ∼11–14 kbp effectively constitutes the minimal transferrable unit that an integrative mechanism would need to be able to accommodate to introduce the toxin inside the chromosome in G1 *C. botulinum*. Existing studies have predominantly focused on the chromosomal integration sites associated with *bont* genes rather than on horizontal gene transfer on a genome-wide level. Comparative genomics can be leveraged to gain additional resolution regarding whether an examined horizontal gene transfer event is species, lineage, or strain specific.

In their role as host adaptive immune modules, clustered regularly interspaced short palindromic repeats and CRISPR-associated protein (CRISPR-Cas) systems can be used to gain direct insight into horizontal gene transfer events. CRISPR-Cas systems, composed of CRISPR spacer arrays and *cas* gene clusters, are present in a wide range of bacterial species and enable the hosts to engage in sequence-specific targeting and cleavage of DNA and/or RNA (Van Der Oost et al., 2014). CRISPR-Cas systems are utilized by the bacterial host in adaptive immune and regulatory roles (Bhaya et al., 2011). In the former role,

transcribed CRISPR arrays are processed by Cas6 into CRISPR RNAs (crRNA) consisting of a direct repeat and a spacer, which generally complement a fragment of foreign DNA encountered at some point in the past and initiates degradation of recognized invading DNA (Carte et al., 2008; Makarova et al., 2015).

Several bioinformatic studies have investigated CRISPR-Cas systems in G1 *C. botulinum* to varying degrees and, to the best of our knowledge, none have investigated *C. sporogenes*. As part of a survey on CRISPR-Cas systems in pathogenic bacteria, Hatoum-Aslan and Marraffini reported 14/14 closed *C. botulinum* genomes possessed type III-B CRISPR-Cas systems (Hatoum-Aslan and Marraffini, 2014). A 2017 report indicated the presence of type I-B and III-B systems in G1-3 *C. botulinum* (Negahdaripour et al., 2017). Finally, a comparative genomics study investigating recombination events at *cas* gene clusters additionally identified the presence of a type III-D CRISPR-Cas system in a subset of strains (Puigbò et al., 2017). Type I systems are generally composed of the structural proteins Cas5, Cas8, Cas7, and nuclease Cas3, and require the presence of a short protospacer adjacent motif in addition to the protospacer for cleavage of DNA to occur (Sinkunas et al., 2011; Makarova et al., 2015). Type III systems are composed of structural proteins Csm (III-A and III-D), Cmr (III-B and III-C), nuclease Cas10, and target DNA or RNA without a PAM requirement (Makarova et al., 2015; Plagens et al., 2015; Samai et al., 2015). In instances where a bacterial host possesses both type I and III systems, processed spacers may be shared between the systems and have been observed to provide functional redundancy against viral escape mutants (Silas et al., 2017). Finally, Cas1, 2, and 4 gene products have roles associated with the generation and insertion of new spacers derived from recently encountered mobile actors (Zhang et al., 2012; Nuñez et al., 2014; Lee et al., 2019). As a result, even closely related strains may have vastly different spacer arrays depending on what plasmids or phage were encountered by that strain in the environment. While CRISPR-Cas systems are generally employed against large MGEs such as plasmids and prophage, some strains of bacteria such as *Porphyromonas gingivalis* have been reported to employ CRISPR-Cas systems against IS elements, which are highly active within this species and play a major role in inter-strain diversification (Watanabe et al., 2013).

CRISPR arrays represent a library of past encounters with horizontally mobile entities including phage, plasmids, and other MGEs. Previous studies have identified type I and type III CRISPR-Cas systems in a selection of G1 *C. botulinum* strains, and none have examined their presence in *C. sporogenes* (Hatoum-Aslan and Marraffini, 2014; Carter et al., 2016; Woudstra et al., 2016; Negahdaripour et al., 2017). The number of G1 *C. botulinum* strains with fully sequenced genomes has more than doubled since these analyses, now enabling comprehensive analyses of CRISPR-Cas systems and the library of past encounters left by the systems within the genomes. We sought to examine whether the *bont* gene cluster, itself horizontally distributed, was targeted by CRISPR-Cas systems directly or indirectly through associated MGEs including plasmids, bacteriophage, ISs, or group II introns. Our in-depth analyses of 241 G1 *C. botulinum* and *C. sporogenes* strains indicate that the two species possess and utilize the same types of CRISPR-Cas systems, which do not directly target the *bont* gene clusters but target *bont*(+) conjugative plasmids.

MATERIALS AND METHODS

Strain Selection and Phylogenetic Analysis

All available *Clostridium* refseq strains as of 3/1/2021 were downloaded from NCBI/GenBank and typed via an established MLST scheme for G1 *C. botulinum* and *C. sporogenes* (G1C)[1] (Jacobson et al., 2008; Jolley and Maiden, 2010; Jolley et al., 2018). The pangenome and a core genome SNP (cgSNP) phylogenetic tree of 250 refseq annotated *Clostridium* genomes were determined and constructed via PanX on default settings using refseq annotations (O'Leary et al., 2016; Ding et al., 2018). Following examination of the cgSNP phylogenetic tree, eight strains of *C. sporogenes* and one unnamed strain (GCF_001276215.1, GCF_011015155.1, GCF_011016125.1, GCF_011017215.1, GCF_011017365.1, GCF_011019515.1, GCF_011020825.1, GCF_011021555.1, GCF_011021645.1) were withheld from analysis due to extreme observed distance from the major *C. sporogenes* and G1 *C. botulinum* groups. The resulting dataset consisted of 241 *Clostridium* strains [146 *C. botulinum* (including one assembly of *C. combessii*), 95 *C. sporogenes*]. BoNT serotype and subtype were determined through alignment via ClustalΩ (default, automatic) and phylogenetic analysis via raxml (-PROTGAMMAAUTO) (Sievers et al., 2011; Stamatakis, 2014; **Supplementary File 1**). All strains were passed through RFPlasmid (*Clostridium* model) to obtain predictions of chromosomal or extrachromosomal origin at the contig level (Van Bloois et al., 2020; **Supplementary File 2**).

CRISPR Spacer Array and *cas* Gene Prediction

CRISPR arrays and *cas* genes for all assemblies were predicted via CRISPR-Cas finder under default settings with Cas subtyping enabled (**Figure 1A**; Couvin et al., 2018). Identified *cas* genes were directly examined within the assembly for *cas* gene clusters based on established CRISPR-Cas system families (Makarova et al., 2015, 2018). Type I-B was coded as complete (I-B) if possessing cas5/6/7/8/3 and cas1/2/4 and a conserved subset possessing only cas5/6/7/8 were typed as I-B*. An additional subset of type I-B strains (I-B**) that were cas5/6/7/8/3(+) and cas1/2/4(−) were identified via examination of the pan-genomic data. Type III *cas* systems were coded as such if in possession of non-pseudogenized majority of the genes associated with Type III-A (cas6/10, csm 2/3/4/5/6), III-B (cas6/10, cmr1/3/4/5/6), and III-D (cas10, csm2/3/5, csx10). All detected CRISPR arrays, and 1,000 bp upstream and downstream flanking them, were utilized to mask direct to array self-matches (**Figure 1B**). Duplicate

[1] Seemann T, mlst Github https://github.com/tseemann/mlst

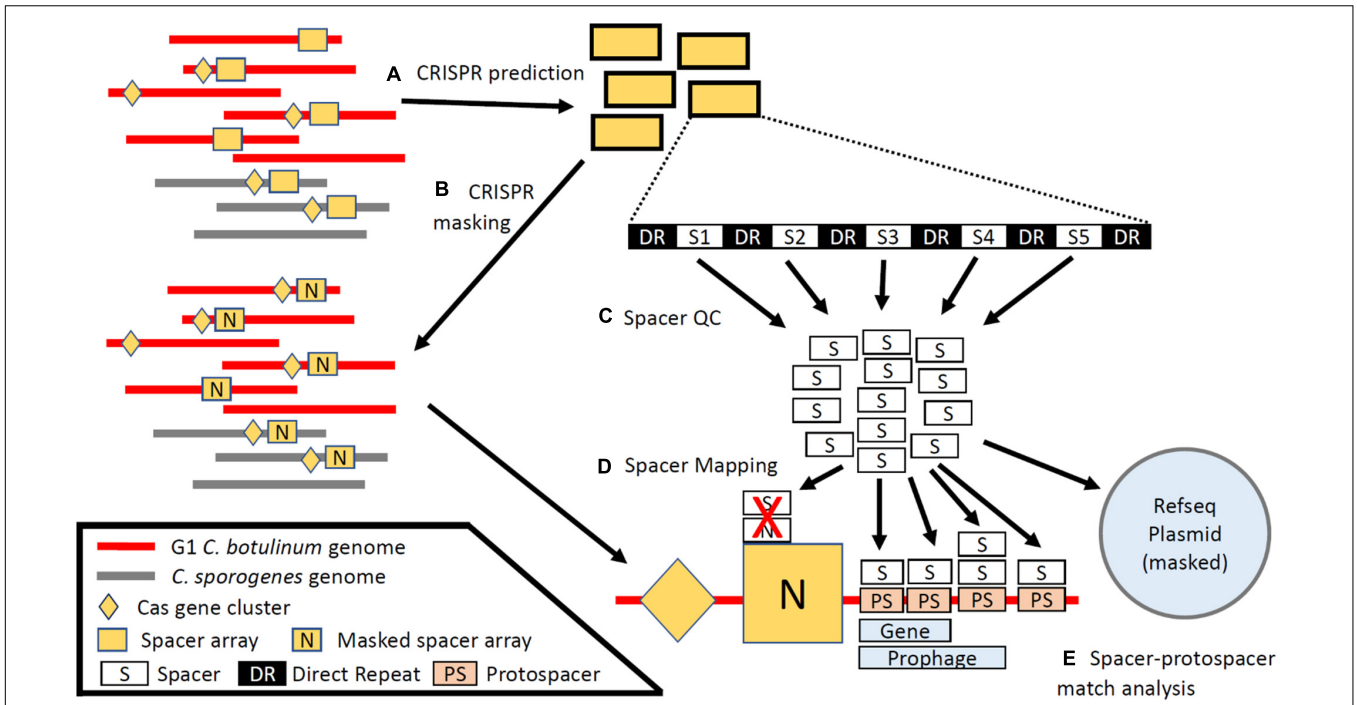


FIGURE 1 | Schematic of Cas and CRISPR array identification and protospacer prediction. **(A)** CRISPR arrays and Cas genes were predicted from 241 contig-level RefSeq assemblies of group I *C. botulinum* and *C. sporogenes*. **(B)** Predicted CRISPR arrays were used to mask the assemblies and the RefSeq plasmid datasets. **(C)** Spacers were parsed by length, array length, and distance to end of the contig. **(D)** Spacers were mapped back to the masked genomes and the RefSeq plasmid dataset to identify cognate protospacer sequences. **(E)** Spacer–protospacer matches were analyzed to determine protospacer localization and predicted function.

spacers predicted from the same assembly were dropped and spacers were assigned unique and non-redundant identifiers based on exact, directional sequence. To select for high-quality spacers and reduce spurious hits, spacers were parsed for those between 20 and 70 nt, part of an array containing five or more spacers, and not predicted within 1,000 bp of the end of a contig (**Figure 1C**). Parsed spacers were mapped as short reads by bowtie2 as a local alignment with allowance of up to two mismatches (-a –local -D 20 -R 3 -N 1 -L 20 -i S,1,0.5 –no-unal –no-sq –no-hd –mp 4,4 –ma 2) against protospacer datasets including Refseq Plasmid (March 1, 2021) and the 241 assemblies making up the investigative dataset (**Figure 1D**; Langmead and Salzberg, 2012). In addition, a subset of spacers erroneously predicted from a family of leucine-rich repeat proteins were excluded from analysis. BEDTools was utilized to obtain relevant annotations overlapping the matched protospacer sites (Quinlan and Hall, 2010). Accessions were collected for matched protospacers occurring within protein coding sequences (CDS) and functionally annotated with cluster of orthologous groups of proteins (COG) domains via the eggnog-mapper (**Figure 1E**) (Tatusov, 2000; Huerta-Cepas et al., 2017). Based on visual inspection of mapped results, 10 additional regions containing arrays missed by CRISPR-Cas finder prediction in certain strains were also masked (Quinlan and Hall, 2010). The plasmid NC_025146.1 was treated as part of the GCF_000829015.1 assembly for all analyses (Ihara et al., 2003; Zhang et al., 2017). Summary statistics for *cas* and spacer analysis and the complete Sp-PS match set is provided in the **Supplementary Files 1, 3**.

Visualization of Spacer–Protospacer Interactions

To estimate the degree of overlap between protospacers present in both G1 *C. botulinum* and *C. sporogenes*, each spacer and protospacer was grouped by species clade. The resulting species-spacer and species-protospacer groups and their associated unique CRISPR spacer IDs were treated as node-edge pairs in a directed force matrix. Node pairs were visualized via Cytoscape and grouped using the yFiles organic layout (Shannon, 2003). The resulting force matrix was colored by spacer clade (G1 *C. botulinum*: green, *C. sporogenes*: blue) and organized such that spacers mapping to protospacers present in both species are centrally located within the diagram, spacers mapping to protospacers present in only one clade at the top and bottom, and unmatched spacers fanning out from the spacer species. Darker edge coloration indicates greater edge density.

Additional Investigation of *Clostridium sporogenes* str. CDC 1632 and Analysis of *ltrA* Distribution

Mauve was used to initially align plasmid pNPD7 (NZ_CP013241.1) and the *C. sporogenes* str. CDC 1632 chromosome (Darling, 2004). Prophage regions were located,

scored, and annotated via PHASTER (Arndt et al., 2016). BlastN, default settings, was used to generate local alignment data between the CDC 1632 putative integrated plasmid (coordinates: 3983434–4238643 bp) and pNPD7 (Altschul et al., 1990). The *ltrA* CDSs identified in *C. sporogenes* str. CDC 1632 were provided as a tblastn query against the 241 strains in the dataset (*E* value 1e-5, word size 2) (**Supplementary File 4**). BlastN, default settings, was used to align a portion of S2 from G1 *C. botulinum* str. 1169 and S6 from G1 *C. botulinum* str. A3 Loch Maree, and matches less than or equal to 1E-50 were visualized via Kablammo and gene cluster diagrams were generated via Gene Graphics (Altschul et al., 1990; Wintersinger and Wasmuth, 2015; Harrison et al., 2018). Representative *ltrA* sequences were aligned via Clustal Omega, default settings (Rice et al., 2000; Sievers et al., 2011). The secondary structure of the catalytic RNA structure flanking *ltrA* (WP_012300946.1) in *C. botulinum* str. A3 Loch Maree was predicted via the MXfold server, and domains were manually annotated through consultation of the group II intron database (Candales et al., 2012; Sato et al., 2021). All phylogenic tree graphics were built via iTOL (Letunic and Bork, 2019). The program phylocorrelate was run in conjunction with the cgSNP tree to investigate correlation between identified group II introns, *bont* genes, and pfam annotated protospacers (**Supplementary File 4**).

Analysis of *bont* and *cas* Integration Sites

Through investigation of the literature and analysis of the predicted *bont* and *cas* gene loci, seven distinct loci encompassing all known sites of *bont* and *cas* gene cluster integration were identified within the subset of complete/closed genomes (n = 43/241). Prophage regions were identified for each chromosome via PHASTER (Arndt et al., 2016). Stable flanking genes were identified for each chromosomal and plasmid site associated with *bont* and/or *cas* gene cluster features. Chromosomal sites were defined as S1: *cysK-brnQ*, S2: *arcA-ytaF*, S3: *efp-cloSI*, and S4: *bglG*-αβ-*hydrolase*; plasmid sites as S5: *dnaX-ATPase*, S6: *viralA-thermonuclase*, and S7: *DUF1292-DUF3854*. Site loci for each genome are provided in **Supplementary File 5**. The closed genomes for *Clostridium botulinum* str. Mfbjulcb8 (genetic *C. sporogenes*) and G1 *C. botulinum* strain 1169 were excluded from analysis due to a unique chromosomal rearrangement that disrupted the insertion sites and a PHASTER prediction error, respectively. Spacers overlapping an annotated site or phage were assigned corresponding codes, and all others were assigned to the chromosome or plasmid. Only chromosomes and *bont*(+) plasmids associated with the 43 closed genomes were included in the analysis.

Statistical Analysis of Protospacer Density Across Closed Genomes

A protospacer density metric was calculated as the number of protospacer loci divided by feature length in base pairs for prophage, plasmid, and chromosomal features in the 43 strains. For determination of chromosomal protospacer density, chromosomal prophage regions and associated protospacers were subtracted from chromosome length and protospacer count. Five plasmids (NCBI Accession: NZ_CP014147.1, NZ_CP013848.1, NZ_CP014218.1, NZ_CP014173.1, NZ_CP031100.1) were classified as phage following observation of numerous structural bacteriophage proteins throughout the length of the plasmid. Protospacer density from plasmids (n = 19), phage (n = 115), and the chromosome (n = 43) were normalized via log transformation [Log10(Protospacer Density) + 7] and protospacer density was assessed across groups via the Welch one-way ANOVA test in the rstatix R package (Wickham, 2011; R Core Team, 2013; Wickham et al., 2019; Kassambara, 2020, 2021). Plasmids and phage with no matched protospacers were excluded from analysis (n = 69). A non-parametric *post hoc* analysis (Games–Howell) was run following the ANOVA to determine statistically significant mean differences between the three feature groups (Kassambara, 2021).

RESULTS

Group I *Clostridium botulinum* and *Clostridium sporogenes* Are Distinct, Closely Related Species

To facilitate analyses of CRISPR-Cas systems throughout G1 *C. botulinum*, we conducted a pan-genomic analysis and constructed a core genome SNP phylogenetic tree of 241 strains of G1 *C. botulinum* and *C. sporogenes* (**Figure 2A**). The analysis revealed 2,003 shared orthologous genes between the two strains (**Supplementary Table 1**), confirming two highly related species (**Figure 2**). This is similar to two previous studies utilizing cgSNP phylogeny approaches, which identified 2,016 and 2,420 shared orthologous genes between G1 *C. botulinum* and *C. sporogenes*, indicating distinct but closely related species (Weigand et al., 2015; Brunt et al., 2020a). The final dataset consisted of 45 closed and 196 contig level assemblies and the phylogeny split into G1 *C. botulinum* and *C. sporogenes* clades (**Figure 2A** and **Supplementary File 1**). Of the 17,472 contigs in the dataset, 1,573 contigs accounting for 2.95% of total nucleotide content were predicted as extrachromosomal (**Supplementary File 2**).

Consistent with prior studies, several non-toxic strains loaded within the G1 *C. botulinum* clade, including *C. sporogenes* str. PA 3679, *C. combesii*, and G1 *C. botulinum* strains SU1575NT and SU1033 (**Figure 2A**; Butler et al., 2017; Dobritsa et al., 2018; Brunt et al., 2020a). Conversely, seven strains previously described and deposited as *C. botulinum* loaded into the *C. sporogenes* clade, four of which possessed no detected *bont* gene cluster (**Figure 2A** and **Supplementary File 1**). Analysis of closed genomes indicated serotypes A/B/F occur as either plasmid-borne or chromosomally integrated clusters in G1 *C. botulinum* (**Figure 2A** and **Supplementary Files 1, 2**). Conversely, only 16 of the 89 *C. sporogenes* assemblies were found to carry a *bont* gene cluster and they exclusively encode BoNT/B1, B2, or B6. Among the three closed *C. sporogenes* genomes encoding *bont* clusters, two occurred on plasmids. Predicted localization of these toxin types in contig level *C. sporogenes* assemblies also favored plasmid localization (**Supplementary File 2**). A single

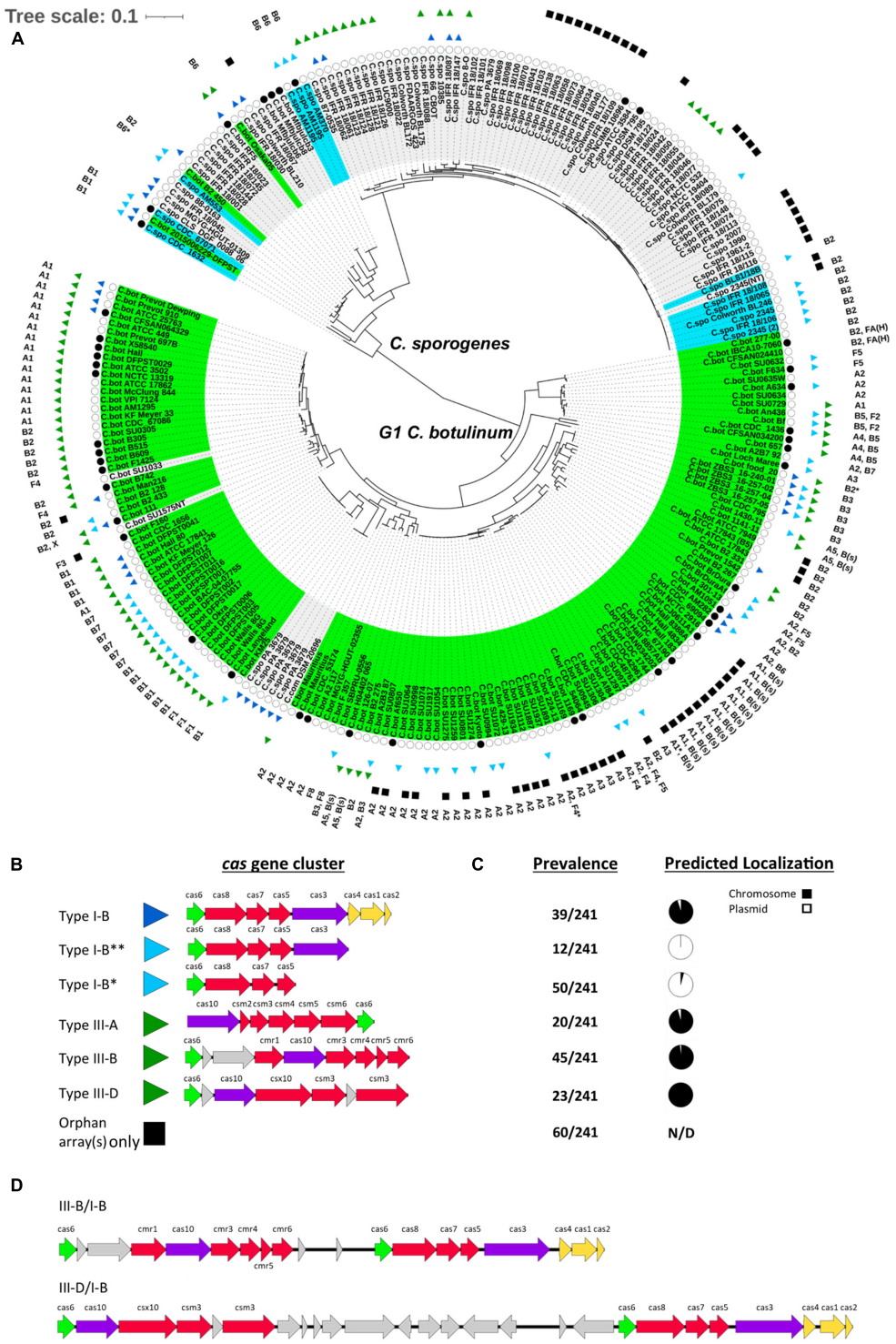


FIGURE 2 | CRISPR-Cas systems in group I *C. botulinum* and *C. sporogenes*. **(A)** A core genome SNP matrix phylogenetic tree of 241 G1 *C. botulinum* and *C. sporogenes* indicating two major clades predominantly demarcated by the respective species. Closed black circles indicate complete genomes (n = 45). Cas gene clusters are represented by colored arrows and correspond to CRISPR-Cas system type in **(B)**. Solid black square indicates strains with predicted CRISPR arrays but no *cas* genes. Predicted BoNT serotype(s) are listed on the outermost band of the tree and strain names shaded by published species and toxin presence: *bont*(+)/G1 *C. botulinum*, *bont*(+)/*C. sporogenes*, bont(−). **(B)** Gene schematic of the major *cas* gene clusters identified in G1 *C. botulinum* and *C. sporogenes*. **(C)** Prevalence of each CRISPR-Cas system type/subtype throughout the assemblies included in the study dataset and distribution by predicted plasmid or chromosomal localization for each contig carrying the CRISPR-Cas system subtype. **(D)** Potential hybrid/coupled type I-B and type III-B or III-D CRISPR-Cas systems generally occur in strains in possession of both.

strain, *C. sporogenes* CDC 1632, the most extreme outlier within the phylogeny, possessed a chromosomally integrated *bont/B1* gene cluster (**Figure 2A**). Taken together, these data indicate that *bont* gene clusters are broadly present throughout G1 *C. botulinum* and frequently chromosomally integrated, while *bont* gene clusters in *C. sporogenes* are limited to *bont/B* and usually plasmid-borne.

CRISPR-Cas Systems (CRISPR-CAS System) and Features Are Broadly Distributed Throughout G1 *Clostridium botulinum* and *Clostridium sporogenes*

Similar to previous observations that type I and type III CRISPR systems are present in *C. botulinum* strains (Carter et al., 2016; Woudstra et al., 2016; Negahdaripour et al., 2017), most of the G1 *C. botulinum* and *C. sporogenes* strains examined in this study (202/241, 84%) contained one or more predicted *cas* gene clusters or an orphan CRISPR array in the absence of a *cas* gene cluster (**Figure 2B**). Type III-A, B, D, and Type I-B CRISPR-Cas systems possessing a defined *cas* gene cluster were found in 141 strains (**Figure 2**). The remaining 100 assemblies possessed no identifiable *cas* gene clusters, although a majority had at least one orphan CRISPR array present. The adaptation module genes *cas1*, *cas2*, and *cas4* (*cas1/2/4*) were only observed in association with type I-B gene clusters and were present in only 39 assemblies, indicating that the capacity to generate novel spacers is relatively rare throughout the population and exclusive to type I-B. While not uncommon for type III systems to lack the adaption module and instead rely on those associated with type I systems (Makarova et al., 2013, 2015), 67 assemblies possessed a type III system with no *cas1/2/4*(+) type I-B system present (**Figure 2A** and **Supplementary File 1**). In addition, two partial variants of the type I-B CRISPR-CAS system lacked adaption (I-B**) and *cas3* nuclease (I-B*) genes (**Figure 2B**). Since *cas6* is the only universally present gene in all investigated assemblies, we analyzed homology of this gene within our data set. Multiple alignment of annotated *cas6* indicated deep divergence at the amino acid level between the partial I-B variants and within complete I-B CRISPR-Cas systems (**Supplementary Figure 1**). These data indicate a diverse range and variable presence of complete and incomplete CRISPR-Cas systems throughout G1 *C. botulinum* and *C. sporogenes*, with no clear species-specific phylogenetic distinction between the observed CRISPR-Cas systems and CRISPR elements.

In Both Species, Complete Type I-B and Type III *cas* Gene Clusters Localize to the Chromosome While Partial I-B *cas* Systems Localize to Plasmids

While primarily utilized by the bacterial host as a means of adaptive immunity, it is increasingly recognized that some MGEs, including bacteriophage, transposases, and plasmids, also possess and utilize CRISPR-Cas systems for regulatory roles and self-preservation (Faure et al., 2019; Klompe et al., 2019; Mcdonald et al., 2019; Varble et al., 2019). Analysis of the genomic localization of the *cas* gene cluster in the 141 assemblies containing them showed that 96% of complete type I-B and type III CRISPR-Cas systems localized to the chromosome or contigs predicted to be chromosomal. The remaining 4% (6 assemblies) of these types with predicted plasmid localization resided on short contigs, which impacts prediction accuracy (Van Bloois et al., 2020). These data indicate that the complete type I-B systems, the sole type identified with the potential to generating novel spacers within both species, are chromosome exclusive. In contrast, the partial type I-B CRISPR-Cas system variants localized exclusively to plasmids (**Figure 2C**). The partial type I-B variant, I-B**, localized to a family of ∼200 kbp, *bont*(-) plasmids, and I-B* to the family of *bont*(+) conjugative plasmids ∼250 kbp (**Figures 2B,C**). Two chromosomally localized type I-B* CRISPR-Cas systems were observed; however, analysis of the genomic regions surrounding the CRISPR-Cas systems indicated that they were localized within chromosomally integrated *bont*(+) plasmids in *C. sporogenes* str. 1632 and *C. botulinum* str. DFPST0006. These chromosomal plasmid integrations have recently been independently reported (Smith et al., 2021a). In strains where both type I-B and III-B or III-D CRISPR-Cas systems were present, the gene clusters encoding the two CRISPR-Cas systems were frequently adjacent to each other within the chromosome (**Figure 2D**). These data show differential localization of *cas* subtypes and in *C. botulinum* and *C. sporogenes*.

Both *cas* and *bont* Gene Clusters Localize to Shared Sites (S1–S4) Within the Chromosome in G1 *Clostridium botulinum* and *Clostridium sporogenes*

CRISPR-Cas systems frequently occur at dynamic sites within the genome that over time can accrue additional genes of related functions such as complementary CRISPR-Cas systems, RM systems, and other genes that may play a defensive role (Makarova et al., 2011; Doron et al., 2018). Having observed some degree of positional overlap between the type III and type I-B systems in the form of adjacent/hybrid systems (**Figure 2D**), we further characterized the regions flanking these CRISPR-Cas systems. Within closed assemblies, type I-B CRISPR-Cas systems are present in two distinct chromosomal genomic regions, while all type III CRISPR-Cas systems localized to only one of the two chromosomal genomic regions (**Figures 3A,B**). Limited examination of contig-level assemblies revealed a subset of III-A CRISPR-Cas systems localized elsewhere, indicating additional CRISPR-Cas system sites may exist (**Supplementary Figure 1**). Analysis of the flanking region showed the *bont* gene clusters predominantly occupy sites near type I-B CRISPR-Cas systems at site 1, which encompasses *brnQ* and *arsC bont* integration sites. Two *bont* gene clusters occurred within fully or partially chromosomally integrated plasmid sequences near the CRISPR-Cas systems integration regions (**Supplementary Figure 2**; Dover et al., 2014; Hill et al., 2015; Smith et al., 2021a). The only *cas* gene clusters observed outside of the chromosome were cas1/2/4(−) I-B** cas gene clusters exclusive to a family of *bont*(−) plasmids and nuclease(−) cas1/2/4/3 type I-B* *cas* gene clusters exclusive

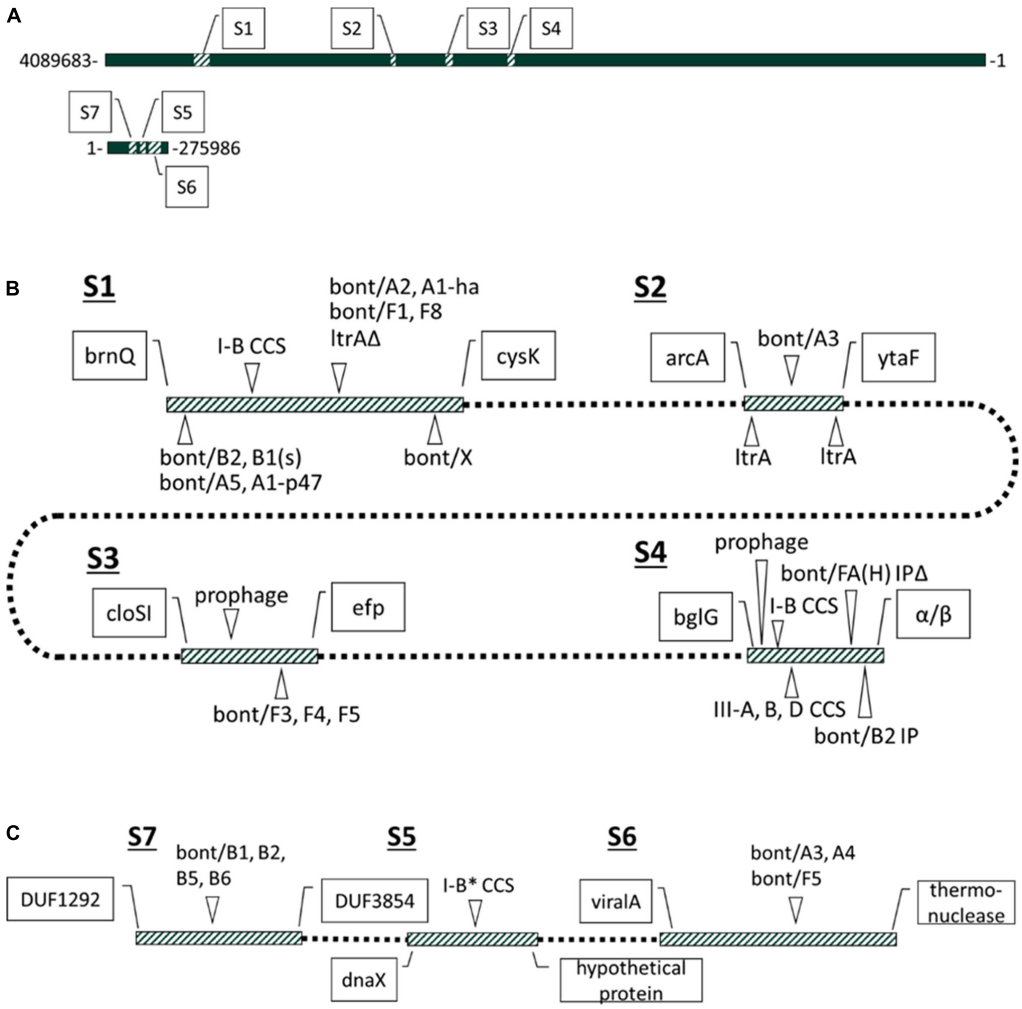


FIGURE 3 | Discrete localization sites of *bont* gene cluster and CRISPR-Cas systems in *C. botulinum* and *C. sporogenes*. **(A)** Genomic locations of seven specified sites associated with *cas* and *bont* gene cluster in G1 *C. botulinum* and *C. sporogenes* host diverse MGEs. Sites annotated within the *C. botulinum* CDC_1436 genome. **(B)** Strain variable *bont*, *cas*, and MGE associated features of the four chromosomal sites known to harbor with *bont* and *cas* gene clusters. At S1, a fragment of the group II intron (*ltrA*Δ) intron-encoded protein co-occurs with *bont* gene clusters integrated at the *arsC* gene. **(C)** Plasmid sites associated with bont and *cas* gene clusters. Boundaries defined for all sites in 43 strains in **Supplementary File 5**.

to *bont*(+) conjugative plasmids. The type I-B** system and the *bont*(−) plasmids were not further characterized beyond that the ∼200-kbp plasmid family is distinct from and unrelated to the ∼250-kbp *bont*(+) conjugative plasmids (**Supplementary Figure 2**). The *bont*(+) conjugative plasmids of G1 *C. botulinum* and *C. sporogenes* were found to share large conserved regions and common elements. This indicates *bont*(+) conjugative plasmids constitute a related plasmid family (**Supplementary Figure 2**), which is consistent with a previous report showing relatedness between *bont/b* bearing plasmids in *C. botulinum* (Hosomi et al., 2014; Orlek et al., 2017). All type I-B* CRISPR-Cas systems localized to one site within the *bont*(+) conjugative plasmid family and were present in all family members with the exception of pCLJ and p1_CDC51232 (**Figure 3C** and **Supplementary Figure 2**). No *bont* gene clusters were detected near the type I-B* plasmid integration site, but the *bont* gene clusters on the *bont*(+) conjugative plasmid family exclusively localized to two distinct plasmid integration sites. Conserved genomic markers were identified within the vicinity of all *bont* and *cas* gene clusters within the 43 closed genomes, defining seven distinct genomic regions. Four chromosomal regions were denoted as sites 1–4 and three plasmid regions as sites 5–7 (**Figures 3A,B**).

Whereas S4 is the primary chromosomal CRISPR-Cas system integration site in both species and an atypical *bont* gene cluster integration site, S1 is the primary chromosomal *bont* gene cluster integration site in G1 *C. botulinum* and occasional type I-B CRISPR-Cas systems integration site. Sites 2 and 3 contain *bont* gene clusters but no CRISPR-Cas systems. Similarly, on plasmids *bont* gene cluster associated sites 6 and 7 lack CRISPR-Cas systems, while site 5 hosts the type IB* CRISPR-CAS system and no *bont* gene cluster (**Figure 3C** and **Supplementary Figure 2**).

Further analysis of integration sites within the 43 closed assemblies revealed pseudogenized *cas6* genes and orphan arrays in place of a full *cas* gene cluster in 3 strains at site 1, and in 4 strains at site 4, indicating CRISPR-Cas system degradation and loss in some lineages (**Supplementary File 5**). In addition, these sites are host to a variety of MGEs including ISs (Dineen et al., 2003; Hill et al., 2009) and group II introns (**Figure 3B** and **Supplementary File 5**). Comprehensive analysis of group II introns throughout the study dataset revealed full-length group II introns flanking the *bont/A3* gene cluster at S2 (**Figure 3B** and **Supplementary Figures 3A,B**). In addition, a fragment of the group II intron, intron-encoded protein is present at the site of the disrupted *arsC* gene (site S1) in all G1 *C. botulinum* strains with an *arsC* integrated *bont* gene cluster (**Figure 3B** and **Supplementary Figure 3C**). However, group II introns, like ISs, which have also been observed in the vicinity of and within the boundaries of *bont* gene clusters (Dineen et al., 2003; Dineen, 2004; Smith et al., 2015), can also be found independently of the *bont* gene cluster. Correlation analysis via phylocorrelate supported significant association only between the *ltrA*/*ltrA*-fragment and *bont*/A2, but not other *arsC*-S1 associated *bont* gene clusters (**Supplementary File 4**; Tremblay et al., 2021). This could be due to under/over-representation of certain *bont* subtypes in the study dataset or a true lack of correlation. Despite this, group II introns appear to be one of the more consistently present small MGEs within the vicinity of a diverse group of chromosomal and plasmid localized *bont* gene clusters.

These findings indicate that sites S1 and S4 serve as hypervariable regions that attract and accumulate MGEs and horizontally acquired cargo genes. Sites S2 and S3 lack CRISPR-Cas systems but provide examples of MGEs that occur within the vicinity of *bont* gene clusters. Site S2 was the only chromosomal integration site found for the *bont*/A3 gene cluster flanked by group II introns, and S3 contained *bont*/F3, F4, and F5 gene clusters and bacteriophage have previously been identified in the presence and absence of the *bont*/F gene cluster (**Figure 3B**; Smith et al., 2020). Taken together, we have shown seven chromosomal and plasmid integration sites in G1 *C. botulinum* and *C. sporogenes*, which contain CRISPR gene clusters and/or *bont* gene clusters as well as several other MGEs. This indicates genomic hotspots for integration of both defense islands as well as virulence genes and other MGEs. While the association between the *bont* gene cluster and some MGEs occurs sporadically, it is also necessary to account for CRISPR-Cas system targeting of associated MGEs which could, by proximity, limit the horizontal transfer of *bont* gene clusters.

CRSIPR-Cas Systems of G1 *Clostridium botulinum* and *Clostridium sporogenes* Predominantly Target Plasmids and Bacteriophages but Not *bont* Gene Clusters

To investigate whether predicted CRISPR-Cas systems in G1 *C. botulinum* and *C. sporogenes* could potentially modulate the range of *bont* gene transfer through immunity by either directly targeting the *bont* gene cluster or associated MGEs, we investigated the CRISPR array encoded spacers and identified their predicted cognate targets (protospacers). We first examined the global protospacer matches identified via spacer mapping against all 241 strains and the RefSeq plasmid database. Across all assemblies, a pool of 6,208 spacers was identified. Of those, 60.4% mapped to protospacer targets present in the strain and/or RefSeq plasmid dataset (**Supplementary Table 2A**; Brooks et al., 2019), with a total of 55,729 spacer–protospacer matches identified. The high percentage of matched spacers reflects the stringent quality control applied to spacers, strain redundancy, and high prevalence of protospacers within the study data set. Both G1 *C. botulinum* and *C. sporogenes* genomes were heavily targeted (**Supplementary Table 2A**). Of all spacer matches, 26% mapped to proteins with detectable conserved COG domains (**Supplementary Table 2A**; Tatusov, 2000; Huerta-Cepas et al., 2017). Categorization of remaining hits by RefSeq annotation revealed 20% mapped to phage associated proteins, 16% to intergenic loci, 9% to proteins with an annotated putative function but no COG match, 4% to proteins with a domain of unknown function (DUF), and 25% to hypothetical proteins (**Supplementary Table 2A**). Consistent with other studies, a small fraction of spacers (n=130), were self-matches within the same genome (**Supplementary File 6**). This low prevalence is largely consistent with the expectation that self-matches are deleterious to the host (Shmakov et al., 2020). Among protospacer matches belonging to known COG categories, the most heavily targeted proteins in both *C. sporogenes* and G1 *C. botulinum* were those relating to: (1) replication, recombination, and repair; (2) transcription; and (3) cell wall/membrane/envelop biogenesis (**Supplementary Table 2B**). Overall, these results are consistent with general expectations that CRISPR-Cas systems primarily target gene products that bacteriophage and plasmids require to replicate and propagate (Bhaya et al., 2011).

Investigation of protospacer annotations indicated the *bont* gene cluster is not a direct target of CRISPR-Cas systems in either species. No protospacers were identified within any of the genes of the primary *bont* gene cluster genes: *bont*, *ntnh*, *ha17*, *ha33*, *ha70*, *p-47*, *orfX1–3*, or *botR*. Very few annotated matches to IS elements were identified, namely, spacers from G1 *C. botulinum* str. B2 331 and the *C. sporogenes* PA3679 strains matched protospacers in a *tnpB* sequence of IS200/IS605-like ISs present exclusively on several predicted *bont*(+) conjugative plasmids (**Supplementary File 6**). No protospacers within annotated group II introns (*ltrA*) were identified. A single redundant spacer present in both *C. sporogenes* strains IFR 18/061 and IFR 18/062 was observed to match a putative peptidoglycan binding gene, which may frequently co-occur with *p47*/*orfX*(+) *bont* gene clusters (Smith et al., 2021b; **Supplementary File 6**). However, these same spacers were also observed to match other copies of this gene that occurred independently of the *bont* gene clusters. These findings suggest that, in aggregate at the species wide-scale, CRISPR-Cas systems do not represent a direct barrier to the trafficking of *bont* gene cluster genes throughout G1 *C. botulinum* and *C. sporogenes*.

To determine whether plasmid-borne *bont* gene clusters may be indirectly targeted through targeting of the plasmid

vehicle, spacer–protospacer matches to verified plasmids were assessed. Both the *bont*(+) ∼250 kbp and *bont*(−) ∼200-kbp plasmids are targeted by spacers from G1. *C. botulinum* and *C. sporogenes*, and a higher density of matches is observed relative to those in plasmids from other species (**Figures 4A,B** and **Table 1**). Despite being an outlier in the *bont*(+) plasmid family, due to the absence of a ∼100-kbp region including the predicted type IV secretion system, pCLD possesses equivalent protospacer density to the rest of the family (**Figures 4A,B**). These results indicate that plasmid protospacers are broadly distributed across the length of the plasmid, and protospacer profiles can be used to further characterize plasmid families. Matches involved a range of functions including replication and toxin–antitoxin matches, potentially representing examples of spacers both with generalized anti-plasmid targeting and with specific targeting of gene products representing a direct threat to host survival (**Table 2**). In addition, spacers with matches to pCLD come from a broad range of CRISPR-Cas system(+) assemblies including those harboring hybrid CRISPR Cas systems type I-B/III, I-B only, III only, and plasmid-borne

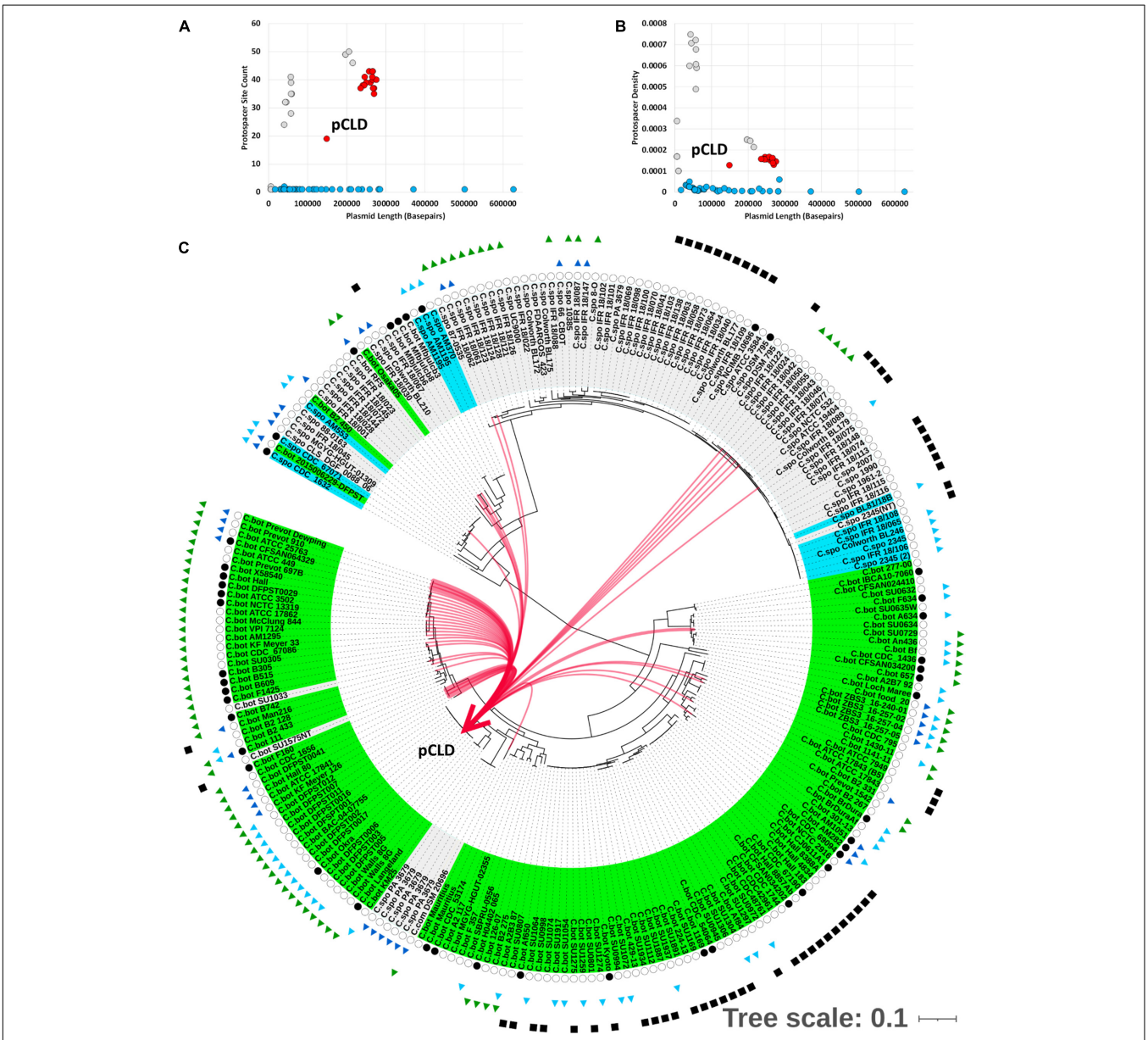


FIGURE 4 | *bont*(+) conjugative plasmids targeted by CRISPR-Cas systems in G1 *C. botulinum* and *C. sporogenes*. **(A)** Plasmid length (*x*-axis) by number of (redundant) spacer–protospacer (Sp-PS) matches (*y*-axis) for Refseq plasmids; BoNT(+) plasmids in red, bont(−) plasmids from *C. sporogenes* and G1. *C. botulinum* in gray, matches to plasmids from other species in blue. **(B)** Transformation of **(A)** to plasmid length (*x*-axis) by Protospacer Density (Sp-PS matches/plasmid length) (*y*-axis) by plasmid length. **(C)** Visual representation of matched spacer containing assemblies to plasmid pCLD.

TABLE 1 | Count of RefSeq plasmids and protospacer loci by species matched to spacers from the 241 strain G1 *C. botulinum* and *C. sporogenes* dataset.

Species	Matched protospacer positions	Matched plasmid count
Acinetobacter baumannii	1	1
Bacillus cereus	5	5
Bacillus thuringiensis	5	5
Chroococcidiopsis thermalis	2	1
Clostridioides difficile	2	2
G1 Clostridium botulinum	991	31
G2 Clostridium botulinum	2	1
Clostridium perfringens	1	1
Clostridium sporogenes	126	4
Enterococcus avium	1	1
Enterococcus faecium	12	12
Enterococcus mundtii	1	1
Eubacterium eligens	1	1
Lactiplantibacillus plantarum	1	1
Methanomethylovorans hollandica	1	1

TABLE 2 | Annotations associated with pCLD protospacers matched to spacers derived from 44 G1 *C. botulinum* and *C. sporogenes* strains (**Figure 4C**).

pCLD protospacer annotations	Spacer–protospacer matches
AAA family ATPase	7
Bacitracin ABC transporter ATP-binding protein	2
DNA polymerase III subunit delta	1
Helix-turn-helix transcriptional regulator	1
Hypothetical protein	26
Intergenic	14
Methyltransferase	1
Phosphoadenosine phosphosulfate reductase family protein	1
Type II toxin-antitoxin system death-on-curing family toxin	1
Viral A-type inclusion protein	4

partial I-B variants (**Figure 4C**). Overall, these results indicate that the *bont*(+) plasmids are targeted by a range of CRISPR-Cas systems present throughout both species. However, questions remain as to how impactful this is at the species level, considering the prevalence of *bont*(+) plasmids and CRISPR-Cas features in both species.

The horizontal mobility of the *bont* gene cluster could also be affected through targeting of nearby and interceding genomic features. To investigate the presence of protospacers in intergenic matches within and near the *bont* gene cluster and potentially associated smaller MGEs, we globally categorized protospacers in sites S1–S7 in a subset of strains with closed assemblies. Intergenic protospacers accounted for 15 and 17% of *C. sporogenes* and G1 *C. botulinum* hits (**Supplementary Table 2A**). However, there are limitations in the accurate identification of promiscuous and often pseudogenized MGEs, such as group II introns and ISs, by annotation alone. To comprehensively examine potential intergenic protospacers and

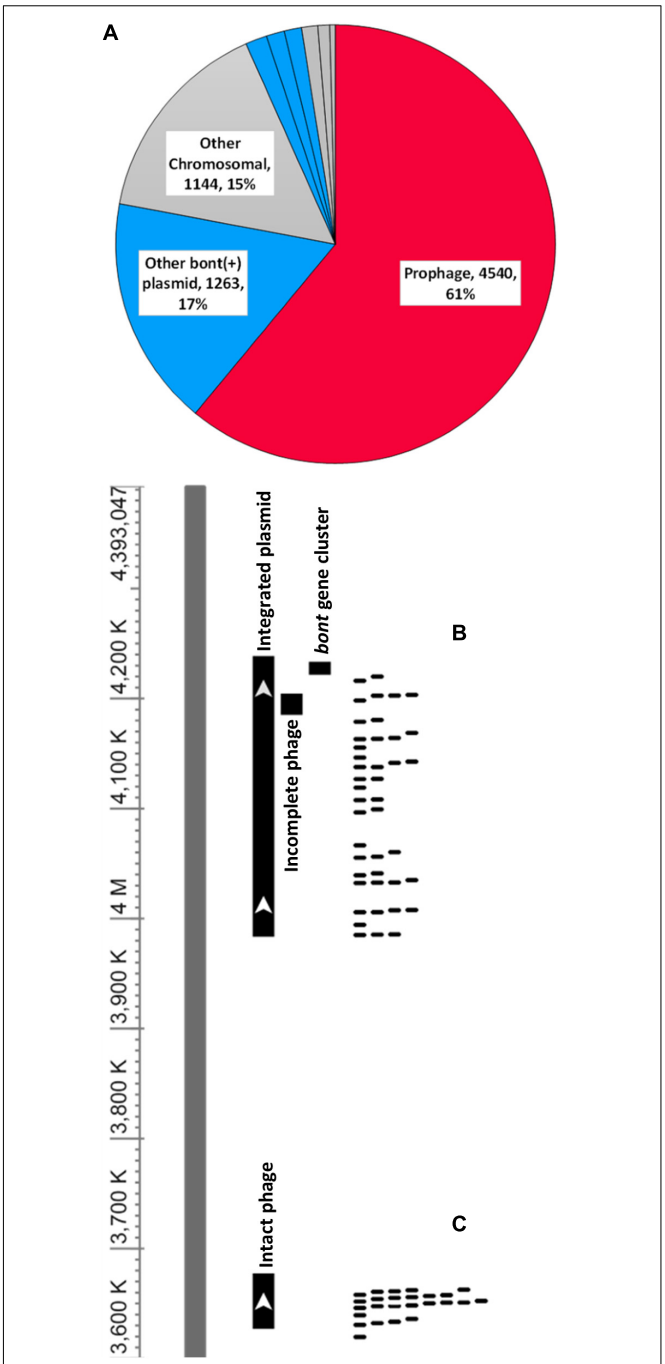


FIGURE 5 | Botulinum neurotoxins are not targets of CRISPR spacers. **(A)** Pie chart showing 7,005 spacer–protospacer pairs from 43 closed genomes categorized by genomic localization. Individual sites are detailed in **Table 3**. Genomic sites 1–7 are defined as in **Figure 3**. Boundaries for each genome are provided in the **Supplementary File 5**. Prophage refers to any protospacer present within a chromosomal prophage region. Other *bont*(+) plasmid to any *bont*(+) plasmid protospacers falling outside the defined sites S5–7 and other chromosomal to any protospacers outside of a predicted phage or defined sites S1–4 (**Figures 3B,C**). **(B,C)** Protospacers occur at a higher density in prophage and plasmids relative to the chromosome. Mapping spacers to the closed genome of *C. sporogenes* CDC 1632 revealed through protospacer matches, both prophage associated regions and an integrated *bont*(+) plasmid.

TABLE 3 | Proportion of protospacer hits falling within the defined *bont* and *cas* associated gene clusters.

Site	# Spacer–protospacer matches	% of total matches
S6	118	1.59
S7	99	1.33
S5	95	1.28
S4	89	1.20
S1	64	0.86
S3	31	0.42
S0	0	0.00

*No protospacers were identified with S2 (associated with **Figure 5A**).*

MGEs within and around *bont* gene clusters in an annotation agnostic way, we examined protospacer matches within the sites established in **Figure 3**. Of the 7,443 spacer-protospacer matches present in the closed assemblies surveyed, which included 43 chromosomes and 11 *bont*(+) plasmids, 61% were found within predicted prophage (**Figure 5A**). Within the defined genomic sites S1–7, protospacer matches accounted for 7% of all matches. However, no matches to *bont* gene cluster genes, group II introns, or ISs were detected (**Figure 5A**, **Table 3**, and **Supplementary File 7**). At site S2, which harbors the *bont*/A3 gene cluster, no protospacers were identified. Of the remainder of protospacer matches, outside of annotated bacteriophage and the seven defined sites S1–S7, 17% were found in the *bont*(+) plasmid and 15% in the chromosome. A visual pile-up representation of these dynamics in *C. sporogenes* str. CDC 1632 demonstrates the utility of this approach in the investigation of integrated plasmids and prophage (**Figures 5B,C**). Taken together, these results indicate that the spacer arrays in G1 *C. botulinum* and *C. sporogenes* are predominantly composed of anti-phage and anti-plasmid spacers and do not directly target *bont* gene clusters, nearby insertional elements, or group II introns.

Initially, in terms of protospacer density, the *bont*(+) conjugative plasmids appeared to be among the least targeted plasmids present in G1 *C. botulinum* and *C. sporogenes* (**Figure 4B**). Investigation of the ∼40–50 kbp plasmids with ∼3–5 × greater protospacer density than the *bont*(+) plasmids suggested through the presence of numerous structural phage genes that these sequences represent circularized phage genomes, not plasmids. To more fully understand the degree to which *bont*(+) plasmids are targeted by CRISPR-Cas, we determined the protospacer density for each plasmid, phage, and the chromosome outside of prophage regions. Within the 43 strains investigated, 170 phage regions and 33 plasmids were identified (**Figure 6A**). No protospacers were identified in 55 phage and 14 plasmids (**Figure 6B**). All *bont*(+) plasmids possessed between 19 and 43 protospacer loci. A Welch one-way ANOVA of log-transformed protospacer density in chromosomal (n = 43), plasmid (n = 19), and phage (n = 115) revealed a statistically significant difference in protospacer density between groups: F = 1,270 (2,66.05), $p \leq 0.0001$. A Ganes–Howell *post hoc* analysis revealed statistically significant differences ($p \leq 0.0001$) in mean protospacer density between all groups (**Figure 6C** and **Supplementary File 8**). These findings indicate that plasmids, including all *bont*(+) plasmids, and phage are enriched in protospacers relative to the chromosome, indicating that uptake of the *bont* gene cluster may be impacted by CRISPR-Cas system targeting of the plasmid.

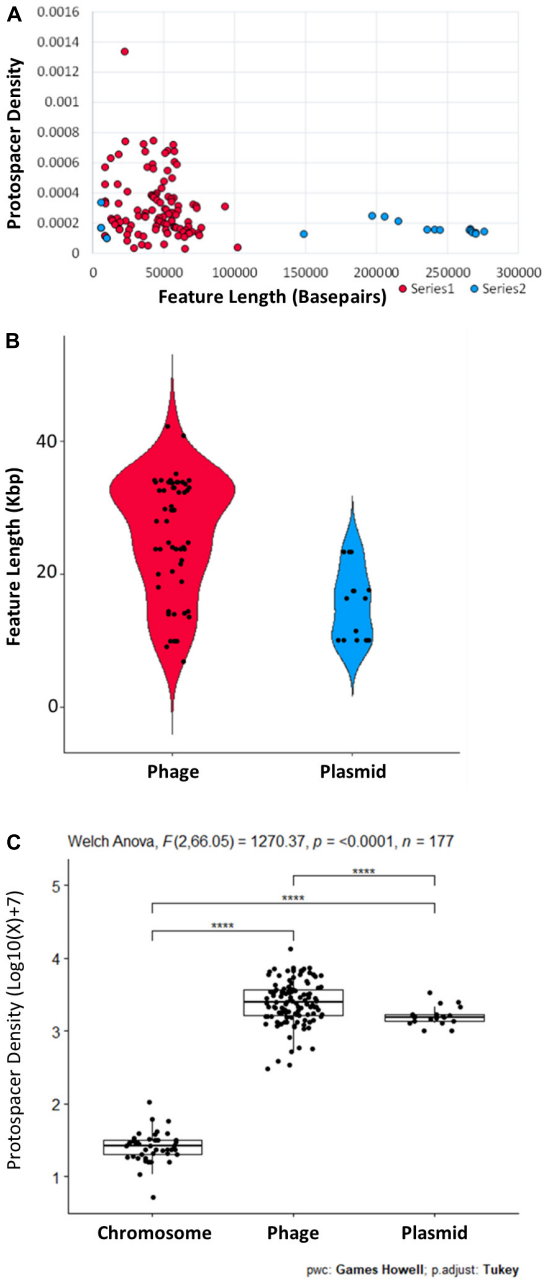


FIGURE 6 | Phage and plasmids vary in length and protospacer density. **(A)** Protospacer density in phage (n = 115) and plasmids (n = 19) from 43 closed G1 *C. botulinum* and *C. sporogenes* genomes plotted relative to size in base pairs. **(B)** Range in length of phage (n = 55) and plasmids (n = 14) with no protospacer matches. **(C)** Phage and plasmids are significantly enriched in protospacers relative to the chromosome. Additional data and parameters are provided in **Supplementary File 8**. **** indicate statistical significance.

G1 *Clostridium botulinum* and *Clostridium sporogenes* Possess a Large Shared Mobilome

To better understand the potential relevance of these data at the population scale, we investigated the overlap of protospacers in G1 *C. botulinum* and *C. sporogenes*. Species exclusive spacer–protospacer (Sp-PS) matches (e.g., *C. sporogenes* spacers that exclusively match *C. sporogenes* protospacers) accounted for only 15.2% of total matches (**Figure 7** and **Table 4**). Hits from one species exclusively against protospacers present in the other represented 6.8% of total hits (**Figure 7** and **Table 4**). The vast majority, 77.6%, of matched spacers were predicted to target protospacers present in both species. These results, in conjunction with the finding that protospacer targets are predominantly prophage and plasmid associated, indicate that G1 *C. botulinum* and *C. sporogenes* possess a large shared mobilome. Analysis of viral protospacers was limited to prophage present within closed G1 *C. botulinum* and *C. sporogenes* genomes, limiting insight into the broader host range targeted bacteriophage. However, matched RefSeq plasmids from other soil-dwelling Gram-positive genera included *Paenibacillus* and *Enterococcus*, which are known to possess genes homologous to *bont* gene cluster genes (Zhang et al., 2018; Nowakowska et al., 2019). With further development, these data might enable enhanced risk assessment through the quantification of the normal range of horizontal gene transfer between G1 *C. botulinum* and other species.

DISCUSSION

In this study, we investigated whether CRISPR-Cas systems present in the closely related species G1 *C. botulinum* and

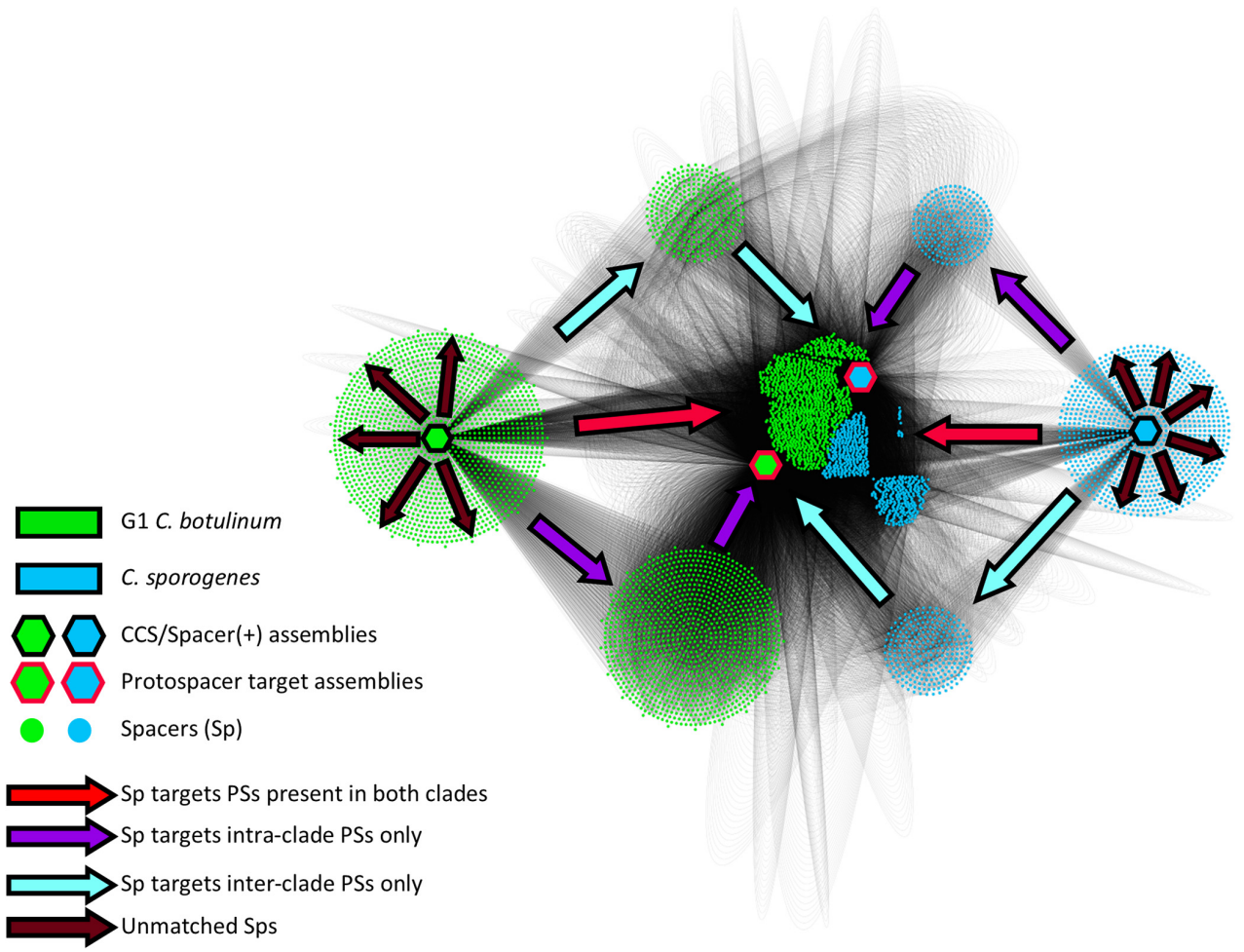


FIGURE 7 | Most targeted protospacers are present in both G1 *C. botulinum* and *C. sporogenes*. Figure shows a directed matrix made in Cytoscape (Shannon, 2003) connecting CRISPR-Cas(+) assemblies (black outlined hexagonal node) to protospacers(+) (red outlined hexagonal node) through a shared intermediate node representing a unique spacer (circular node). For all spacer–protospacer interactions, each strain was collapsed into one of four clade-specific groups (G1 *C. botulinum*/*C. sporogenes* + spacer/protospacer). Both species are represented as hexagonal nodes connected to each spacer ID by edges that are 1:1 between spacer and species-spacer, and 1:X matched protospacers between spacers and protospacer-species. The arrow legend indicates spacer targets. Red arrows indicate directionality of spacers that target protospacers present in both species and represent 78% of all spacer–protospacer pairs (**Table 4**).

TABLE 4 | Directed force matrix.

Protospacer species	Spacer species	Spacer species	Grand totals
	G1 C. botulinum	***C. sporogenes***	
G1 C. botulinum	7,202 (G1–G1 Only)	1,840 (Cspo-G1 Only)	
C. sporogenes	1,975 (G1-Cspo Only)	1,309 (Cspo-Cspo Only)	
G1 [Share of G1 + *C. spo*]	17,265	9,499	
C. spo [Share of G1 + *C. spo*]	10,118	6,203	
Unmatched spacers	1,502	954	
Total shared	27,383	15,702	43,085
Total in-clade only	7,202	1,309	8,511
Total cross-clade only	1,975	1,840	3,815

C. sporogenes represent a potential barrier to the acquisition of the botulinum neurotoxin virulence factor via horizontal gene transfer. Through application of a spacer mapping and reference masking approach to predict spacers and identify cognate protospacer targets within a diverse sample of 145 G1 *C. botulinum* and 96 *C. sporogenes* strains (**Figure 1**), we uncovered evidence of a targeting of bacteriophage and plasmids. None of the annotated protospacers within the ∼56,000 spacer–protospacer matches occurred within the constituent *bont* gene cluster genes: *bont*, *ntnh*, *botR*, *ha17/33/70*, *p47*, and *orfX1/2/3* present in toxigenic strains in the study dataset (n = 154/241). Investigation of all protospacer hits within closed genomes (n = 43) indicated no protospacers present between *bont* gene cluster genes and none targeting nearby group II introns and ISs. As of today, no IS elements near the *bont* gene clusters have been shown to be functional. Emerging research will further elucidate their potential role in the transfer of *bont* gene clusters and their (non)targeting by restriction systems such as CRISPR-Cas. Intriguingly, the IS200/IS605 IS elements, the only IS family with matched spacers by CRISPR-Cas systems in our study, have themselves been recently demonstrated to be programmable RNA-guided nucleases related to the type V CRISPR-Cas system effector (Cas12) (Altae-Tran et al., 2021). Functional characterization of these ISs alongside the endogenous CRISPR-Cas systems is required to determine whether this targeting represents competition or collaboration in these two species. The absence of targeting of other IS elements supports the hypothesis that IS elements play a beneficial role to the bacterial host in acquiring and retaining potentially beneficial gene(s) such as the *bont* gene cluster. However, targeting of IS element activity by an alternate mechanism such as other restrictive processes, such as RM systems, cannot be excluded based on our data.

Approximately 80% of the protospacer targets of CRISPR-Cas systems in G1 *C. botulinum* and *C. sporogenes* were present across both species, indicating these species possess a large shared mobilome (**Figure 7**). Through protospacer inference, the CRISPR-Cas system targeted mobilome predominantly consists of bacteriophage and to a lesser extent plasmids (**Figures 5**, **6**). These findings run contrary to a previous report of more frequent matches between spacers and plasmid associated protospacers than phage associated protospacers in *C. botulinum* (Biswas et al., 2013; Negahdaripour et al., 2017; Brooks et al., 2019). These differences are explained by differing methodology, as GenBank-Phage used in that study to search for protospacers only identified sequenced bacteriophage present in GenBank-Phage while the mapping approach utilized in this study enabled identification of all annotated bacteriophage within the chromosome (Rohwer, 2003; Hatfull and Hendrix, 2011; Biswas et al., 2013; Arndt et al., 2016). Our observed association of 61% of protospacers with bacteriophage approaches the 70–90% rate reported in most bacterial species (Shmakov et al., 2017). This likely still represents an undercount of phage associated protospacers, as examination of protospacer annotations (e.g., phage tail family protein) at other chromosomal sites (15%) suggests the presence of phage remnants or non-annotated phage (**Supplementary File 7**).

We observed a clear bifurcation in protospacer density within G1 *C. botulinum* and *C. sporogenes* plasmids relative to those observed in other species (**Figure 4B**). Relative to a previous study that generated spacer–protospacer matches utilizing blastN within the CRISPR-Target program and RefSeq plasmids (Biswas et al., 2013; Negahdaripour et al., 2017; Brooks et al., 2019), we observed lower diversity in the protospacer containing plasmids sequenced in other genera. While the reference mapping settings utilized in our study allowed identification of protospacers with up to two mismatches and local alignment allowed some flexibility at the ends of the spacer-protospacer alignment, this approach is conservative compared with an 80% identity blastN threshold (Negahdaripour et al., 2017) and may have excluded more distant matches in other genera. Interestingly, nearly all matched plasmids were Gram-positive bacteria, and hits to plasmids in genera including *Enterococcus* and *Paenibacillus* are consistent with rare, but documented horizontal gene flows of *bont* gene cluster constituents (Zhang et al., 2018; Nowakowska et al., 2019). CRISPR-Cas system spacers will generally target MGEs that are most often encountered by the host strains (Shmakov et al., 2017), indicating that the plasmids present in the G1 *C. botulinum* and *C. sporogenes* mobilomes have been present long enough for CRISPR-CAS system mediated immunity to develop in certain strains. This also indicates plasmid targeting occurs, and CRISPR-Cas systems may constitute a barrier to uptake in the minority of analyzed strains possessing a functional CRISPR-Cas system and the appropriate spacer(s) (**Figure 4C**). For example, *C. sporogenes* PA 3679 (genetic G1 *C. botulinum*) possess spacers against several *bont*(+) plasmids, which could limit conjugation of *bont*(+) plasmids into this strain, and by extension acquisition of toxicity. However, this would be achieved through targeting of the vehicle, not the toxin. The cross-strain variation in CRISPR-Cas systems also suggests they are unlikely to play a major restrictive role regulating toxin transfer at the species or bi-species level.

Type I-B, III-B, and III-D CRISPR-Cas systems have been identified in (G1–G3) *C. botulinum* by prior studies utilizing closed genomes, with type III-B CRISPR-Cas systems reported as the most prevalent (Hatoum-Aslan and Marraffini, 2014; Negahdaripour et al., 2017; Puigbò et al., 2017). Our study utilized both closed and contig-level genomes extended these

findings to neighbor species *C. sporogenes* and additionally identified the presence of III-A CRISPR-Cas systems in both species. Through identification of conserved genomic markers and utilization of verified and predicted plasmids, we were able to determine exclusive chromosomal localization of the type III systems and identified key differences in the *cas* gene composition of chromosomal and plasmid type I-B CRISPR-Cas systems (**Figures 2B,C**). These additional findings are the result of study design, scope, and time elapsed since the previous studies were conducted (Hatoum-Aslan and Marraffini, 2014; Negahdaripour et al., 2017; Puigbò et al., 2017). For example, utilization of the RFPlasmid program (Van Bloois et al., 2020) enabled the identification of several non-canonical variants of the type I-B CRISPR-CAS system. In contrast to the observed presence/absence and pseudogenization of individual type III genes across strains, potentially indicating a loss of function, chromosomal type I-B systems were rarely pseudogenized. The exclusive and mostly conserved presence of a nuclease(−) type I-B* on *bont*(+) conjugative plasmids may indicate degraded or functionally atypical CRISPR-Cas system. Recently, a CRISPR-Cas mediated toxin–antitoxin system was linked to the retention of effector gene function in type I-B systems in archaea and some bacterial species (Li et al., 2021). A similar mechanism could explain the persistence of the plasmid-borne type I-B* systems consisting of only *cas6* and the *cas5/7/8* effectors. In addition, the type I-B* CRISPR-Cas system also appears to have the *cas* gene set necessary to perform a CRISPRi type function, which have been proposed to potentially play a regulatory role (Vial and Hommais, 2020; Wimmer and Beisel, 2020). There is also some similarity between the type I-B* CRISPR-Cas system and a distinct group of type I-F CRISPR-Cas systems in *Vibrio* spp. transposons that lack a *cas3* nuclease gene and possess *cas6*, *cas7*, and a *cas5*/8 fusion genes and are utilized to achieve CRISPR-mediated site-specific transposition within the genome (Peters et al., 2017; Klompe et al., 2019; Mcdonald et al., 2019). The integrated *bont*(+) plasmids (Dover et al., 2014; Smith et al., 2021a) observed at integration sites 1 and 4 are intriguing within that context and suggest that much remains to be learned about these still cryptic plasmids (**Figure 3B**). Direct functional investigation and characterization of this unique, *bont*(+) plasmid exclusive, type I-B system will provide additional insight into its function and potential relationship to the *bont* gene cluster.

CRISPR-Cas systems are dynamic recombination sites, which makes accurate identification of CRISPR Cas systems and *cas* types challenging. *C. botulinum* B1 Okra was reported to possess a chromosomal type III-B and plasmid-borne I-B CRISPR-Cas systems (Negahdaripour et al., 2017), while another study from the same year identified that same strain as possessing a III-D CRISPR-Cas system and highlighting it as an example of a recombination event whereby the III-B CRISPR-Cas system was supplanted by a III-D CRISPR-Cas system with *cas6* remaining unaffected (Kristensen et al., 2017). Our data also showed phylogenetic co-clustering of the type III-B and III-D associated Cas6 proteins, which is consistent with potential recombination between type III-B and III-D CRISPR-Cas systems. However, our study also discovered the presence of type III-A CRISPR-Cas systems in several strains, with the associated Cas6 not phylogenetically grouping with type III-B and III-D associated Cas6 proteins (**Supplementary Figure 1**). Detailed examination of gene gain/loss falls beyond the scope of this study; however, the data collected potentially lends itself to such analysis in future studies. For example, we observed evidence that the type I-B system within site 1 occurs in a minority of both *C. botulinum* and *C. sporogenes* strains (**Supplementary Figure 1** and **Supplementary File 5**). A blast search shows that *C. tepidum*, the nearest neighbor species to G1 *C. botulinum* and *C. sporogenes* (Dobritsa et al., 2017), also possesses a homologous type I-B CRISPR-Cas system at the same location (NZ_JADPGM010000006.1). This could indicate either that the common ancestor of all three species possessed a I-B CRISPR-Cas system at site 1 (vertical heritage) or that the I-B integration at site 1 has occurred independently multiple times (horizontal acquisition). The presence of pseudogenized *cas6* genes and broad presence of orphan CRISPR features throughout both species would support vertical heritage while the relative scarcity of the type I-B CRISPR-Cas system at site 1 would support horizontal acquisition. Understanding the dynamics of additional acquisition of genes at sites beyond the BoNT gene clusters will enable deeper investigation of how these hypervariable genomic regions are governed.

The results of our study have revealed broad similarities between G1 *C. botulinum* and *C. sporogenes* in both the types of CRISPR-Cas systems present and the mobile targets that they defend against. A recent pan-genomic analysis by Brunt and colleagues found that unique genes to G1 *C. botulinum* and *C. sporogenes* map regularly throughout the length of the genome with no identifiable hotspots (Brunt et al., 2020a). This is consistent with findings in this study that (1) the prophage and plasmids that make up the bulk of the CRISPR-Cas system targeted mobilome are predominantly shared between G1 *C. botulinum* and *C. sporogenes*, and (2) the chromosomal locations of hypervariable sites/hotspots seem to be shared across the two species. In future comparative genomic studies of recombination and integration at hypervariable sites, it may prove beneficial to consider both species in the context of additional species outgroups such as *C. tepidum*. In particular, additional closed genomes from a broader and more diverse range of strains will provide further insight into the regulation of these sites and the selective pressures that enable these sites to acquire, host, and eliminate sophisticated genomic defense modules, the most potent known biological toxin, or nothing at all.

Our systematic investigation of CRISPR-Cas systems in G1 *C. botulinum* and *C. sporogenes* revealed a predominantly shared mobilome between these neighboring species and widespread (∼83%) presence of CRISPR-Cas system features across strains of both. However, the capacity to utilize the adaptive immune component of CRISPR-Cas systems was present in only 16% of strains with chromosomally localized type I-B CRISPR-Cas systems, and the plasmid exclusive presence of partial type I-B systems presents the possibility that plasmids, including the family that carries the *bont* gene cluster, are utilizing CRISPR-Cas with some degree of autonomy from the host. Inclusion of contig level genomes did present analytical challenges. For example, we

did not systematically predict PAM sequences associated with the type I-B systems as it was often challenging to associate CRISPR arrays and *cas* gene clusters in contig-level assemblies. PAM determination would be best achieved on closed subsets and ideally in the context of functional characterization of the CRISPR-Cas systems. However, inclusion of contig level genomes ultimately led to a larger, more diverse spacer set than would have been obtainable through closed genomes alone.

Taken together, our data show that despite being the most prominent horizontally trafficked gene cluster in *Clostridium*, the *bont* gene cluster was not directly targeted by the endogenous CRISPR-Cas systems of G1 *C. botulinum* and *C. sporogenes*. However, these systems do appear to target the conjugative plasmids that traffic the *bont* gene clusters in certain G1 *C. botulinum* and *C. sporogenes* strains. Future functional investigation of the diverse endogenous CRISPR-Cas systems in both species will provide further insight into the regulation of these shared dynamic genomic regions host to both complementary genomic defense systems and the most potent known bacterial toxin.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

TW, AD, J-DS, and SP conceived and designed the research. TW acquired and interpreted the data. TW, BT, and MB analyzed the data. TW wrote the manuscript. TW, BT, MB, AD, SS, J-DS, and SP contributed to article revision, read and approved the submitted version, and provided approval for the publication of the content. All the authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by PHS, DHHS-Food and Drug Administration U01FD001418F and by National Institutes of Health, NIAID R01AI139306.

ACKNOWLEDGMENTS

The authors thank J. P. van Pijkeren and Jason Peters for their critical review of the data and article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.787726/full#supplementary-material

Supplementary Figure 1 | Cas6 protein alignment and phylogeny. Multiple alignment (Clustal Omega Default) (Sievers et al., 2011) and phylogenetic analysis (raxml -PROTGAMMAAUTO) (Stamatakis, 2014) of all Cas6 proteins identified within the study dataset indicates divergence between the type I-B systems present at genomic site 1, site 4, and plasmid-borne variants (**Supplementary File 5**). Type III-B and III-D Cas6 form a single, high identity branch, while type III-A Cas6 proteins load on separate branches. All Cas6 loci are listed in **Supplementary File 6**.

Supplementary Figure 2 | Plasmid-borne *bont* gene clusters occur on a family of related conjugative plasmids. Plasmid pNPD7 and the putative integrated plasmid from *C. sporogenes* CDC 67071 share the conserved conjugation region present in most G1 *C. botulinum bont*(+) conjugative plasmids. Local alignment (blastN) of select *bont*(+) plasmids against pCBG; with a type I-B* CRISPR-Cas system present at S5, a *bont/A* gene cluster at S6, and *bont/B* gene cluster at S7. A region of chromosome from *C. botulinum* IBCA10-7060 from 2339765–2389948 and a *bont*(–) conjugative plasmid pRSJ2_2 (NZ_CP013709.1) were additionally included. BRIG was run with pCBG set as the reference with the following settings: blastN, 70% upper identity threshold, 50% lower identity threshold, and annotated with site information (Alikhan et al., 2011). The *bont* genes and the type I-B* gene clusters from pCBG are highlighted (red) within plasmid sites 5, 6, and 7 (gray) as defined in **Figure 3**. **(A)** The type I-B* CRISPR-Cas systems are present on all family plasmids except a subset of *bont/A, B* producing plasmids such as pCLJ. **(B)** The plasmid fragment adjacent to chromosomal *bont/FA(H)* at site 4 in strain (**Figure 3**) is part of an insert unique to *bont/A, B* plasmids. **(C)** The *bont/B1* plasmid is ∼100 kbp shorter than other family members and is missing the region including the putative conjugal type IV secretion system. **(D)** The *bont*(–) ∼200-kbp plasmids carrying the type I-B** CRISPR-Cas system are unrelated to *bont*(+) 250-kbp plasmids.

Supplementary Figure 3 | Full-length and/or fragmented group II introns near select subtypes of *bont/A*, *B*, and *F* gene clusters. **(A)** Local alignment (blastN default) of the plasmid-borne *bont*/A3 gene cluster in *C. botulinum* A3 Loch Maree against the chromosomal *bont*/A3 gene cluster in *C. botulinum* 1169. The CDS for the putative group II intron encoded protein (IEP) is annotated as *ltrA*. **(B)** Predicted secondary structure of the group II intron surrounding CDS ltrA in strain A3 Loch Maree with conserved domains (I–VI) labeled. **(C)** Location of a *ltrA* fragment present at the *arsC* disruption site in a subset of chromosomally integrated *bont*/A and F gene clusters. **(D)** Multiple alignment of full-length *ltrA* genes and *ltrA* fragments from representative strains (Clustal omega, default settings). **(E)** Phylogenetic distribution of *ltrA* fragments (filled blue circle) and full-length *ltrA* genes (filled red square) across the study strains. *Primary hit: *E* value 0.0, 86% coverage, 99.08% identity, hits ≤ *E* value 1e-50 displayed.

REFERENCES

Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Altae-Tran, H., Kannan, S., Demircioglu, F. E., Oshiro, R., Nety, S. P., McKay, L. J., et al. (2021). The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* 374, 57–65.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/s0022-2836(05)80360-2

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Arnon, S. S., Midura, T. F., Damus, K., Thompson, B., Wood, R. M., and Chin, J. (1979). Honey and other environmental risk factors for infant botulism. *J. Pediatr.* 94, 331–336. doi: 10.1016/s0022-3476(79)80863-x

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: versatile Small RNAs for Adaptive Defense and Regulation. *Annu. Rev. Genet.* 45, 273–297. doi: 10.1146/annurev-genet-110410-132430

Biswas, A., Gagnon, J. N., Brouns, S. J. J., Fineran, P. C., and Brown, C. M. (2013). CRISPRTarget. *RNA Biol.* 10, 817–827. doi: 10.4161/rna.24046

Brooks, L., Kaze, M., and Sistrom, M. (2019). A Curated, Comprehensive Database of Plasmid Sequences. *Microbiol. Resour. Announc.* 8, e01325–18. doi: 10.1128/mra.01325-18

Brunt, J., Carter, A. T., Stringer, S. C., and Peck, M. W. (2018). Identification of a novel botulinum neurotoxin gene cluster in Enterococcus. *FEBS Lett.* 592, 310–317. doi: 10.1002/1873-3468.12969

Brunt, J., Van Vliet, A. H. M., Stringer, S. C., Carter, A. T., Lindström, M., and Peck, M. W. (2020b). Pan-Genomic Analysis of *Clostridium botulinum* Group II (Non-Proteolytic *C. botulinum*) Associated with Foodborne Botulism and Isolated from the Environment. *Toxins* 12:306. doi: 10.3390/toxins12050306

Brunt, J., Van Vliet, A. H. M., Carter, A. T., Stringer, S. C., Amar, C., Grant, K. A., et al. (2020a). Diversity of the Genomes and Neurotoxins of Strains of *Clostridium botulinum* Group I and Clostridium sporogenes Associated with Foodborne, Infant and Wound Botulism. *Toxins* 12:586. doi: 10.3390/toxins12090586

Butler, R. R., Schill, K. M., Wang, Y., and Pombert, J.-F. (2017). Genetic Characterization of the Exceptionally High Heat Resistance of the Non-toxic Surrogate Clostridium sporogenes PA 3679. *Front. Microbiol.* 8:545. doi: 10.3389/fmicb.2017.00545

Candales, M. A., Duong, A., Hood, K. S., Li, T., Neufeld, R. A. E., Sun, R., et al. (2012). Database for bacterial group II introns. *Nucleic Acids Res.* 40, D187–D190. doi: 10.1093/nar/gkr1043

Carte, J., Wang, R., Li, H., Terns, R. M., and Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 22, 3489–3496. doi: 10.1101/gad.1742908

Carter, A. T., Austin, J. W., Weedmark, K. A., and Peck, M. W. (2016). Evolution of Chromosomal*Clostridium botulinum*Type E Neurotoxin Gene Clusters: evidence Provided by Their Rare Plasmid-Borne Counterparts. *Genome Biol. Evol.* 8, 540–555. doi: 10.1093/gbe/evw017

Chandler, M., and Mahillon, J. (2002). "Insertion sequences revisited," in *Mobile DNA II*, eds N. L. Craig, R. Craigie, M. Gellert and A. Lambowitz (Washington, DC: American Society of Microbiology).

Che, Y., Yang, Y., Xu, X., Břinda, K., Polz, M. F., Hanage, W. P., et al. (2021). Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2008731118. doi: 10.1073/pnas.2008731118

Collins, M. D., and East, A. K. (1998). Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *J. Appl. Microbiol.* 84, 5–17. doi: 10.1046/j.1365-2672.1997.00313.x

Contreras, E., Masuyer, G., Qureshi, N., Chawla, S., Dhillon, H. S., Lee, H. L., et al. (2019). A neurotoxin that specifically targets Anopheles mosquitoes. *Nat. Commun.* 10:2869. doi: 10.1038/s41467-019-10732-w

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

Darling, A. C. E. (2004). Mauve: multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Dineen, S. (2004). Nucleotide sequence and transcriptional analysis of the type A2 neurotoxin gene cluster in *Clostridium botulinum*. *FEMS Microbiol. Lett.* 235, 9–16. doi: 10.1016/j.femsle.2004.04.002

Dineen, S. S., Bradshaw, M., and Johnson, E. A. (2003). Neurotoxin Gene Clusters in *Clostridium botulinum* Type A Strains: sequence Comparison and Evolutionary Implications. *Curr. Microbiol.* 46, 345–352. doi: 10.1007/s00284-002-3851-1

Ding, W., Baumdicker, F., and Neher, R. A. (2018). panX: pan-genome analysis and exploration. *Nucleic Acids Res.* 46:e5. doi: 10.1093/nar/gkx977

Dobritsa, A. P., Kutumbaka, K. K., and Samadpour, M. (2018). Reclassification of Eubacterium combesii and discrepancies in the nomenclature of botulinum neurotoxin-producing clostridia: challenging Opinion 69. Request for an Opinion. *Int. J. Syst. Evol. Microbiol.* 68, 3068–3075. doi: 10.1099/ijsem.0.002942

Dobritsa, A. P., Kutumbaka, K. K., Werner, K., Wiedmann, M., Asmus, A., and Samadpour, M. (2017). Clostridium tepidum sp. nov., a close relative of Clostridium sporogenes and *Clostridium botulinum* Group I. *Int. J. Syst. Evol. Microbiol.* 67, 2317–2322. doi: 10.1099/ijsem.0.001948

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., et al. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359:eaar4120. doi: 10.1126/science.aar4120

Dover, N., Barash, J. R., Hill, K. K., Davenport, K. W., Teshima, H., Xie, G., et al. (2013). *Clostridium botulinum* Strain Af84 Contains Three Neurotoxin Gene Clusters: bont/A2, bont/F4 and bont/F5. *PLoS One* 8:e61205. doi: 10.1371/journal.pone.0061205

Dover, N., Barash, J. R., Hill, K. K., Xie, G., and Arnon, S. S. (2014). Molecular Characterization of a Novel Botulinum Neurotoxin Type H Gene. *J. Infect. Dis.* 209, 192–202. doi: 10.1093/infdis/jit450

Eklund, M., and Poysky, F. (1974). Interconversion of type C and D strains of *Clostridium botulinum* by specific bacteriophages. *Appl. Microbiol.* 27, 251–258.

Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., et al. (2019). CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* 17, 513–525. doi: 10.1038/s41579-019-0204-7

Gonzalez-Escalona, N., Thirunavukkarasu, N., Singh, A., Toro, M., Brown, E. W., Zink, D., et al. (2014). Draft Genome Sequence of Bivalent *Clostridium botulinum* Strain IBCA10-7060, Encoding Botulinum Neurotoxin B and a New FA Mosaic Type. *Genome Announc.* 2, e01275–14. doi: 10.1128/genomea.01275-14

Harrison, K. J., Crécy-Lagard, V. D., and Zallot, R. (2018). Gene Graphics: a genomic neighborhood data visualization web application. *Bioinformatics* 34, 1406–1408. doi: 10.1093/bioinformatics/btx793

Hatfull, G. F., and Hendrix, R. W. (2011). Bacteriophages and their genomes. *Curr. Opin. Virol.* 1, 298–303. doi: 10.1016/j.coviro.2011.06.009

Hatoum-Aslan, A., and Marraffini, L. A. (2014). Impact of CRISPR immunity on the emergence and virulence of bacterial pathogens. *Curr. Opin. Microbiol.* 17, 82–90. doi: 10.1016/j.mib.2013.12.001

Henderson, I. (1996). Genetic characterisation of the botulinum toxin complex of *Clostridium botulinum* strain NCTC 2916. *FEMS Microbiol. Lett.* 140, 151–158. doi: 10.1016/0378-1097(96)00172-3

Hill, K. K., Xie, G., Foley, B. T., and Smith, T. J. (2015). Genetic diversity within the botulinum neurotoxin-producing bacteria and their neurotoxins. *Toxicon* 107, 2–8. doi: 10.1016/j.toxicon.2015.09.011

Hill, K. K., Xie, G., Foley, B. T., Smith, T. J., Munk, A. C., Bruce, D., et al. (2009). Recombination and insertion events involving the botulinum neurotoxin complex genes in *Clostridium botulinum* types A, B, E and F and Clostridium butyricumtype E strains. *BMC Biol.* 7:66. doi: 10.1186/1741-7007-7-66

Hosomi, K., Sakaguchi, Y., Kohda, T., Gotoh, K., Motooka, D., Nakamura, S., et al. (2014). Complete nucleotide sequence of a plasmid containing the botulinum neurotoxin gene in *Clostridium botulinum* type B strain 111 isolated from an infant patient in Japan. *Mol. Genet. Genomics* 289, 1267–1274. doi: 10.1007/s00438-014-0887-4

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., et al. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Ihara, H., Kohda, T., Morimoto, F., Tsukamoto, K., Karasawa, T., Nakamura, S., et al. (2003). Sequence of the gene for *Clostridium botulinum* type B neurotoxin associated with infant botulism, expression of the C-terminal half of heavy chain and its binding activity. *Biochim. Biophys. Acta Gene Struct. Expr.* 1625, 19–26. doi: 10.1016/s0167-4781(02)00537-7

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

Johnson, E. A., and Montecucco, C. (2008). Botulism. *Handb. Clin. Neurol.* 91, 333–368.

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1

Jolley, K. A., and Maiden, M. C. (2010). BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. doi: 10.1186/1471-2105-11-595

Kassambara, A. (2020). *ggpubr: 'ggplot2' Based Publication Ready Plots*. Available online at: https://CRAN.R-project.org/package=ggpubr (accessed September 30, 2021).

Kassambara, A. (2021). *rstatix: pipe-Friendly Framework for Basic Statistical Tests*. Available online at: https://CRAN.R-project.org/package=rstatix (accessed September 30, 2021).

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., and Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* 571, 219–225. doi: 10.1038/s41586-019-1323-z

Kristensen, D. M., Wolf, Y. I., and Koonin, E. V. (2017). ATGC database and ATGC-COGs: an updated resource for micro- and macro-evolutionary studies of prokaryotic genomes and protein family annotation. *Nucleic Acids Res.* 45, D210–D218. doi: 10.1093/nar/gkw934

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Lee, H., Dhingra, Y., and Sashital, D. G. (2019). The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *eLife* 8:e44248. doi: 10.7554/elife.44248

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, M., Gong, L., Cheng, F., Yu, H., Zhao, D., Wang, R., et al. (2021). Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. *Science* 372:eabe5601. doi: 10.1126/science.abe5601

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013). The basic building blocks and evolution of CRISPR–Cas systems. *Biochem. Soc. Trans.* 41, 1392–1400. doi: 10.1042/bst20130038

Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2018). Classification and Nomenclature of CRISPR-Cas Systems: where from Here? *CRISPR J.* 1, 325–336. doi: 10.1089/crispr.2018.0033

Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011). Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *J. Bacteriol.* 193, 6039–6056. doi: 10.1128/jb.05535-11

Mansfield, M. J., Adams, J. B., and Doxey, A. C. (2015). Botulinum neurotoxin homologs in non-Clostridium species. *FEBS Lett.* 589, 342–348. doi: 10.1016/j.febslet.2014.12.018

Mansfield, M. J., Wentz, T. G., Zhang, S., Lee, E. J., Dong, M., Sharma, S. K., et al. (2019). Bioinformatic discovery of a toxin family in *Chryseobacterium piperi* with sequence similarity to botulinum neurotoxins. *Sci. Rep.* 9:1634.

Marshall, K. M., Bradshaw, M., and Johnson, E. A. (2010). Conjugative Botulinum Neurotoxin-Encoding Plasmids in *Clostridium botulinum*. *PLoS One* 5:e11087. doi: 10.1371/journal.pone.0011087

Mcdonald, N. D., Regmi, A., Morreale, D. P., Borowski, J. D., and Boyd, E. F. (2019). CRISPR-Cas systems are present predominantly on mobile genetic elements in Vibrio species. *BMC Genomics* 20:105. doi: 10.1186/s12864-019-5439-1

Nawrocki, E. M., Bradshaw, M., and Johnson, E. A. (2018). Botulinum neurotoxin–encoding plasmids can be conjugatively transferred to diverse clostridial strains. *Sci. Rep.* 8:3100. doi: 10.1038/s41598-018-21342-9

Negahdaripour, M., Nezafat, N., Hajighahramani, N., Rahmatabadi, S. S., and Ghasemi, Y. (2017). Investigating CRISPR-Cas systems in *Clostridium botulinum* via bioinformatics tools. *Infect. Genet. Evol.* 54, 355–373. doi: 10.1016/j.meegid.2017.06.027

Nevas, M., Lindström, M., Hielm, S., BjöRkroth, K. J., Peck, M. W., and Korkeala, H. (2005). Diversity of Proteolytic *Clostridium botulinum* Strains, Determined by a Pulsed-Field Gel Electrophoresis Approach. *Appl. Environ. Microbiol.* 71, 1311–1317. doi: 10.1128/aem.71.3.1311-1317.2005

Nowakowska, M. B., Douillard, F. P., and Lindström, M. (2019). Looking for the X Factor in Bacterial Pathogenesis: association of orfX-p47 Gene Clusters with Toxin Genes in Clostridial and Non-Clostridial Bacterial Species. *Toxins* 12:19. doi: 10.3390/toxins12010019

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., and Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* 21, 528–534. doi: 10.1038/nsmb.2820

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., Mcveigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Orlek, A., Stoesser, N., Anjum, M. F., Doumith, M., Ellington, M. J., Peto, T., et al. (2017). Plasmid classification in an era of whole-genome sequencing: application in studies of antibiotic resistance epidemiology. *Front. Microbiol.* 8:182. doi: 10.3389/fmicb.2017.00182

Peters, J. E., Makarova, K. S., Shmakov, S., and Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U. S. A.* 114, E7358–E7366. doi: 10.1073/pnas.1709035114

Plagens, A., Richter, H., Charpentier, E., and Randau, L. (2015). DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiol. Rev.* 39, 442–463. doi: 10.1093/femsre/fuv019

Puigbò, P., Makarova, K. S., Kristensen, D. M., Wolf, Y. I., and Koonin, E. V. (2017). Reconstruction of the evolution of microbial defense systems. *BMC Evol. Biol.* 17:94. doi: 10.1186/s12862-017-0942-y

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

R Core Team (2013). *R: a Language And Environment For Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277.

Rohwer, F. (2003). Global Phage Diversity. *Cell* 113:141. doi: 10.1016/s0092-8674(03)00276-9

Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., and Marraffini, L. A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* 161, 1164–1174. doi: 10.1016/j.cell.2015.04.027

Sato, K., Akiyama, M., and Sakakibara, Y. (2021). RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat. Commun.* 12:941. doi: 10.1038/s41467-021-21194-4

Shannon, P. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2017). The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *mBio* 8, e01397–17. doi: 10.1128/mbio.01397-17

Shmakov, S. A., Wolf, Y. I., Savitskaya, E., Severinov, K. V., and Koonin, E. V. (2020). Mapping CRISPR spaceromes reveals vast host-specific viromes of prokaryotes. *Commun. Biol.* 3:321. doi: 10.1038/s42003-020-1014-1

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Siguier, P., Gourbeyre, E., and Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* 38, 865–891. doi: 10.1111/1574-6976.12067

Silas, S., Lucas-Elio, P., Jackson, S. A., Aroca-Crevillén, A., Hansen, L. L., Fineran, P. C., et al. (2017). Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *eLife* 6:e27601. doi: 10.7554/elife.27601

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* 30, 1335–1342. doi: 10.1038/emboj.2011.41

Smith, T. J., Hill, K. K., Foley, B. T., Detter, J. C., Munk, A. C., Bruce, D. C., et al. (2007). Analysis of the Neurotoxin Complex Genes in *Clostridium botulinum* A1-A4 and B1 Strains: BoNT/A3, /Ba4 and /B1 Clusters Are Located within Plasmids. *PLoS One* 2:e1271. doi: 10.1371/journal.pone.0001271

Smith, T. J., Hill, K. K., Xie, G., Foley, B. T., Williamson, C. H. D., Foster, J. T., et al. (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. doi: 10.1016/j.meegid.2014.12.002

Smith, T. J., Williamson, C. H. D., Hill, K. K., Johnson, S. L., Xie, G., Anniballi, F., et al. (2021b). The Distinctive Evolution of orfX Clostridium parabotulinum Strains and Their Botulinum Neurotoxin Type A and F Gene Clusters Is Influenced by Environmental Factors and Gene Interactions via Mobile Genetic Elements. *Front. Microbiol.* 12:566908. doi: 10.3389/fmicb.2021.566908

Smith, T. J., Tian, R., Imanian, B., Williamson, C. H. D., Johnson, S. L., Daligault, H. E., et al. (2021a). Integration of Complete Plasmids Containing Bont Genes into Chromosomes of Clostridium parabotulinum, Clostridium sporogenes, and Clostridium argentinense. *Toxins* 13:473. doi: 10.3390/toxins13070473

Smith, T. J., Xie, G., Williamson, C. H. D., Hill, K. K., Fernández, R. A., Sahl, J. W., et al. (2020). Genomic Characterization of Newly Completed Genomes of Botulinum Neurotoxin-Producing Species from Argentina, Australia, and Africa. *Genome Biol. Evol.* 12, 229–242. doi: 10.1093/gbe/evaa043

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tatusov, R. L. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33

Tremblay, B. J.-M., Lobb, B., and Doxey, A. C. (2021). PhyloCorrelate: inferring bacterial gene–gene functional associations through large-scale phylogenetic profiling. *Bioinformatics* 37, 17–22. doi: 10.1093/bioinformatics/btaa1105

Tsuzuki, K., Kimura, K., Fujii, N., Yokosawa, N., Indoh, T., Murakami, T., et al. (1990). Cloning and complete nucleotide sequence of the gene for the main component of hemagglutinin produced by *Clostridium botulinum* type C. *Infect. Immun.* 58, 3173–3177. doi: 10.1128/iai.58.10.3173-3177.1990

Van Bloois, L. V. D. G., Wagenaar, J. A., and Zomer, A. L. (2020). RFPlasmid: predicting plasmid sequences from short read assembly data using machine learning. *bioRxiv* [Preprint] doi: 10.1101/2020.07.31.230631

Van Der Oost, J., Westra, E. R., Jackson, R. N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 12, 479–492. doi: 10.1038/nrmicro3279

Vandecraen, J., Chandler, M., Aertsen, A., and Van Houdt, R. (2017). The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* 43, 709–730. doi: 10.1080/1040841x.2017.1303661

Varble, A., Meaden, S., Barrangou, R., Westra, E. R., and Marraffini, L. A. (2019). Recombination between phages and CRISPR-cas loci facilitates horizontal gene transfer in staphylococci. *Nat. Microbiol.* 4, 956–963. doi: 10.1038/s41564-019-0400-2

Vial, L., and Hommais, F. (2020). Plasmid-chromosome cross-talks. *Environ. Microbiol.* 22, 540–556. doi: 10.1111/1462-2920.14880

Watanabe, T., Nozawa, T., Aikawa, C., Amano, A., Maruyama, F., and Nakagawa, I. (2013). CRISPR Regulation of Intraspecies Diversification by Limiting IS Transposition and Intercellular Recombination. *Genome Biol. Evol.* 5, 1099–1114. doi: 10.1093/gbe/evt075

Weigand, M. R., Pena-Gonzalez, A., Shirey, T. B., Broeker, R. G., Ishaq, M. K., Konstantinidis, K. T., et al. (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

Wickham, H. (2011). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686.

Willems, A., East, A. K., Lawson, P. A., and Collins, M. D. (1993). Sequence of the gene coding for the neurotoxin of *Clostridium botulinum* type A associated with infant botulism: comparison with other clostridial neurotoxins. *Res. Microbiol.* 144, 547–556. doi: 10.1016/0923-2508(93)90004-l

Wimmer, F., and Beisel, C. L. (2020). CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. *Front. Microbiol.* 10:3078. doi: 10.3389/fmicb.2019.03078

Wintersinger, J. A., and Wasmuth, J. D. (2015). Kablammo: an interactive, web-based BLAST results visualizer. *Bioinformatics* 31, 1305–1306. doi: 10.1093/bioinformatics/btu808

Woudstra, C., Le Maréchal, C., Souillard, R., Bayon-Auboyer, M.-H., Mermoud, I., Desoutter, D., et al. (2016). New Insights into the Genetic Diversity of *Clostridium botulinum* Group III through Extensive Genome Exploration. *Front. Microbiol.* 7:757. doi: 10.3389/fmicb.2016.00757

Zhang, J., Kasciukovic, T., and White, M. F. (2012). The CRISPR Associated Protein Cas4 Is a 5′ to 3′ DNA Exonuclease with an Iron-Sulfur Cluster. *PLoS One* 7:e47232. doi: 10.1371/journal.pone.0047232

Zhang, S., Lebreton, F., Mansfield, M. J., Miyashita, S. I., Zhang, J., Schwartzman, J. A., et al. (2018). Identification of a Botulinum Neurotoxin-like Toxin in a Commensal Strain of Enterococcus faecium. *Cell Host Microbe* 23, 169–176.e6. doi: 10.1016/j.chom.2017.12.018

Zhang, S., Masuyer, G., Zhang, J., Shen, Y., Lundin, D., Henriksson, L., et al. (2017). Identification and characterization of a novel botulinum neurotoxin. *Nat. Commun.* 8:14130. doi: 10.1038/ncomms14130

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact