



Genomic and Phenotypic Characterization of *Clostridium botulinum* Isolates from an Infant Botulism Case Suggests Adaptation Signatures to the Gut

François P. Douillard,^a Yağmur Derman,^a Cédric Woudstra,^{a*} Katja Selby,^a Tommi Mäklin,^b Martin B. Dorner,^c Harri Saxén,^d Brigitte G. Dorner,^c Hannu Korkeala,^a  Miia Lindström^a

^aDepartment of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland

^bHelsinki Institute for Information Technology HIIT, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

^cBiological Toxins, Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Berlin, Germany

^dNew Children's Hospital, Pediatric Research Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

ABSTRACT In early life, the immature human gut microbiota is prone to colonization by pathogens that are usually outcompeted by mature microbiota in the adult gut. Colonization and neurotoxin production by a vegetative *Clostridium botulinum* culture in the gut of an infant can lead to flaccid paralysis, resulting in a clinical outcome known as infant botulism, a potentially life-threatening condition. Beside host factors, little is known of the ecology, colonization, and adaptation of *C. botulinum* to the gut environment. In our previous report, an infant with intestinal botulism was shown to be colonized by neurotoxicogenic *C. botulinum* culture for 7 months. In an effort to gain ecological and evolutionary insights into this unusually long gut colonization by *C. botulinum*, we analyzed and compared the genomes of *C. botulinum* isolates recovered from the infant feces during the course of intoxication and isolates from the infant household dust. A number of observed mutations and genomic alterations pinpointed at phenotypic traits that may have promoted colonization and adaptation to the gut environment and to the host. These traits include motility, quorum-sensing, sporulation, and carbohydrate metabolism. We provide novel perspectives and suggest a tentative model of the pathogenesis of *C. botulinum* in infant botulism.

IMPORTANCE While the clinical aspects of infant botulism and the mode of action of BoNT have been thoroughly investigated, little is known on the pathogenesis and adaptive mechanisms of *C. botulinum* in the gut. Here, we provide for the first time a comprehensive view on the genomic dynamics and plasticity of *C. botulinum* over time in a case of infant botulism. The genomic and phenotypic analysis of *C. botulinum* isolates collected during the disease course offers an unprecedented view of *C. botulinum* ecology, evolution, and pathogenesis and may be instrumental in developing novel strategies for prevention and treatment of toxicoinfectious botulism.

KEYWORDS *Clostridium botulinum*, botulism, ecology, genomics

In early life, the richness and diversity of the gut microbiota are in a developing stage compared to that in the adult gut microbiota (1), rendering the gut of young infants vulnerable to colonization by pathogens. A major threat is the spores of neurotoxigenic clostridia, which, once ingested, can germinate and outgrow to colonize the infant gut. Vegetative cultures of *Clostridium botulinum* produce the highly potent botulinum neurotoxin (BoNT), resulting in toxicoinfection, with a clinical outcome known as infant botulism. A similar condition is also occasionally reported in adults with underlying factors, such as gastrointestinal disorders, surgery, or antibiotic treatments,

Editor Mark S. Turner, University of Queensland

Copyright © 2022 Douillard et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Miia Lindström, miia.lindstrom@helsinki.fi.

*Present address: Cédric Woudstra, Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg C, Denmark.

The authors declare no conflict of interest.

Received 24 August 2021

Accepted 30 March 2022

Published 2 May 2022

illustrating the important role of the gut microbiota in diseases and health (2–4). Host factors and environmental spore contamination are pivotal in infant botulism, and the etiology and pathogenesis of infant botulism have been well documented (5–7). From a microbial and ecological perspective, the source of contamination remains undetermined in most reported cases. It has been shown that geographical location, consumption of food products contaminated with spores, such as honey or infant milk formula, or potential exposure to contaminated soil, reptile pets, or dust particles may constitute a risk (6–12).

While the clinical course of infant botulism and the mode of action of BoNT have been established (13–15), little is known about the phenotypic and genotypic properties of *C. botulinum* in regard to its ability to colonize and persist in the infant gut, to compete with other gut bacteria for nutrients, and to potentially interact with the host. The long-term persistence of *C. botulinum* in the gut of some infants (16) is likely to trigger the emergence of polymorphic *C. botulinum* populations with altered or enhanced phenotypes favorable for survival and colonization in the gut. Such strategies could allow *C. botulinum* to escape the host immune response and to compete with other gut bacteria for resources. Genomic plasticity and dynamics likely contribute to gut adaptation (17, 18). We hypothesize that *C. botulinum* could adapt to the gut environment over time and therefore favor certain genotypes emerging *in situ* via single nucleotide polymorphisms, insertions, deletions, and chromosomal rearrangements.

Here, we built upon our previous clinical report from Finland, where a 3-month-old infant was diagnosed with infant botulism and, despite clinical recovery in 6 weeks, was colonized with toxinogenic *C. botulinum* culture for as long as 7 months (16). As part of the case follow-up, a large number of fecal samples from the infant as well as environmental samples from the infant household were collected for isolation of *C. botulinum*. Sampling and isolation were performed over a 7-month period, yielding gut isolates from the early onset of the disease (original colonizers) until complete clearance of *C. botulinum* (last colonizers). In the original report, a large set of *C. botulinum* isolates from fecal and dust samples were analyzed by amplified fragment length polymorphism (AFLP) and were shown to be closely related but not identical, suggesting a possible link between an environmental source and the disease (16) and adaptive evolution of *C. botulinum* in the infant gut. The relatively high number of isolates available for this particular infant botulism case provided a unique basis to further elaborate on *C. botulinum* pathophysiology in regard to spore germination, colonization, and persistence in the gut from an ecological and evolutionary perspective.

We analyzed and compared the genomes of *C. botulinum* isolates recovered over time from two distinct niches, infant feces (gut context) and dust samples (environment). This allowed us to examine gene polymorphisms, chromosomal rearrangements, and mobile element exchanges that may have occurred among the isolates within or between the niches. Marked differences between the isolates were found, and mutations as well as other genomic alterations pointed to phenotypic traits that relate to virulence, colonization, and possible adaptation to the gut environment, suggesting that the gut ecosystem may exert a selection pressure on *C. botulinum* over time. Our work brings novel insights into the pathogenesis of *C. botulinum* in the context of intestinal botulism and paves the way for development of prophylactic and therapeutic measures.

RESULTS AND DISCUSSION

Genome and mobilome of *C. botulinum* stool isolate ST7B. *C. botulinum* ST7B (stool isolate, early colonizer) harbored a 3.925-Mb chromosome and a 13.75-kb plasmid with an overall G+C content of 28.16%. The genome of ST7B consisted of a total of 3,512 protein-coding sequences, 80 tRNA genes, and 27 rRNA genes. Two intact prophages and one incomplete prophage were detected. Peculiarly, *C. botulinum* ST7B was closely related to *C. botulinum* CDC_297 (GenBank version number CP006907.1), which is a historical isolate related to a foodborne botulism case (19) and has an

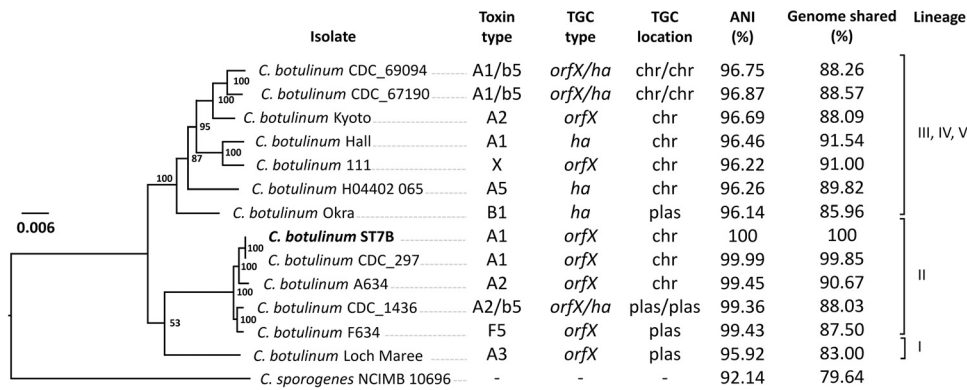


FIG 1 Genetic relatedness of *Clostridium botulinum* ST7B, isolated from infant stool, with a subset of internationally known *C. botulinum* group I strains using the codon tree method as described and generated in the online platform PATRIC (60). The neurotoxin type, neurotoxin gene cluster type, and location of the cluster are indicated. The strains are divided into five lineages (I to V), numbered according to a previous study (70). ANI, average nucleic acid identity; chr, chromosome; ha, hemagglutinin operon; orfX, orfX operon; plas, plasmid; TGC, toxin gene cluster.

average nucleic acid identity (ANI) of 99.99% with ST7B (Fig. 1). In line with the close phylogenetic relatedness of ST7B to CDC_297, the botulinum neurotoxin gene cluster of ST7B was similarly inserted within the *arsC* operon of the chromosome (20) and consisted of *orfX3*, *orfX2*, *orfX1*, *botR*, *p47*, *ntnh*, and *bont* (subtype A1), which is a rare cluster type for A1 strains usually harboring a hemagglutinin operon instead of an *orfX/p47* operon. The entire toxin gene cluster of ST7B is 100% identical to the one present in CDC_297 at the nucleotide level.

The plasmid (pST7B) present in *C. botulinum* ST7B was 13,746 bp long and had a G+C content of 27.66%, which is slightly lower than the overall G+C content of the chromosome (28.16%). Nucleic acid BLAST analysis led to the identification of similar plasmids (or contigs) present in other group I *C. botulinum* and *Clostridium sporogenes* genomes (Fig. 2 and Table S1). These include the dual toxin type B2F5 *C. botulinum* strains Bf, An436, CDC69057, H130580885, and H134990001, all related to infant botulism, and also other strains, namely, CDC_297 and F2534/89. The plasmid pST7B was

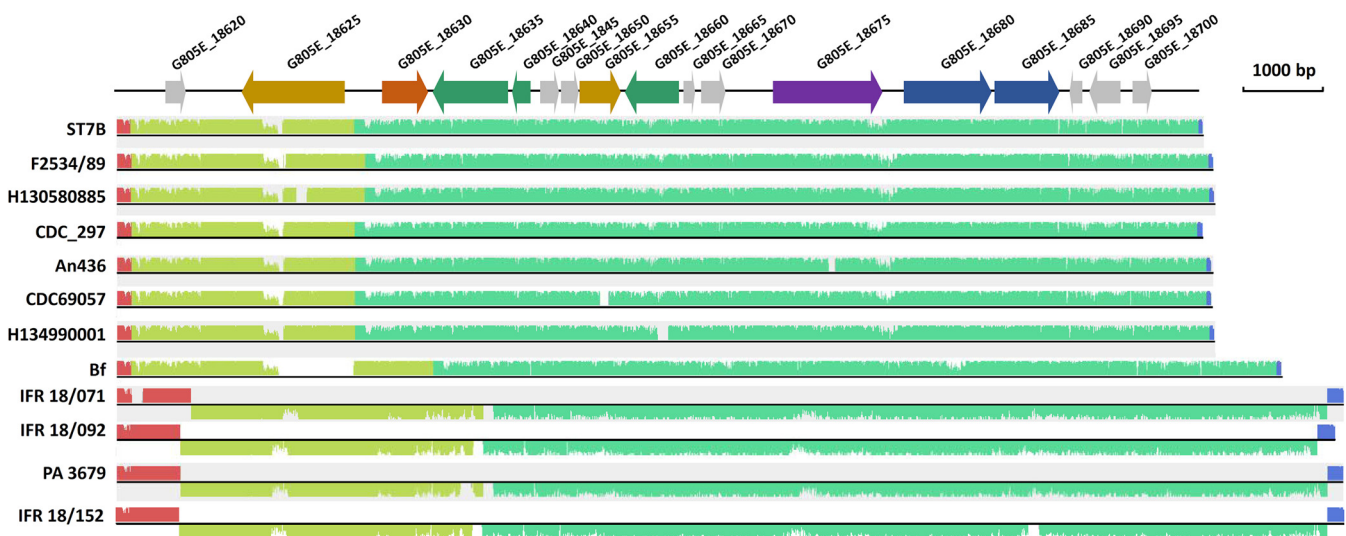


FIG 2 Sequence alignment of *C. botulinum* plasmid pST7B and related plasmids using the Mauve analysis tool. Further information relative to the degree of conservation of pST7B with homologous plasmids/contigs is shown in Table S1. Colored regions correspond to homologous regions between the different plasmid sequences. Blue arrow, gene coding for a restriction-modification system; purple arrow, gene coding for a putative replication protein; green arrow, gene coding for DNA recombination protein; yellow arrow, gene coding for amidase/N-acetylmuramoyl-L-alanine-amidase; orange arrow, gene coding for a putative RNA polymerase sigma factor; gray arrow, genes for coding hypothetical proteins.

TABLE 1 List of sequenced *Clostridium botulinum* group I type A1 isolates recovered from infant stool and environmental sources related to a previously described infant botulism case (16)

| Isolate ^a | Origin | Sampling date | Plasmid ^b |
|----------------------|----------------|-------------------|----------------------|
| H4 | Household dust | 13 May 2010 | + |
| H18 | Household dust | 13 May 2010 | + |
| V1 | Household dust | 13 May 2010 | + |
| V4 | Household dust | 13 May 2010 | - |
| V41 | Household dust | 13 May 2010 | + |
| V62 | Household dust | 13 May 2010 | + |
| V73 | Household dust | 13 May 2010 | + |
| V134 | Household dust | 13 May 2010 | + |
| V174 | Household dust | 13 May 2010 | - |
| V206 | Household dust | 13 May 2010 | + |
| ST25 | Infant stool | 9 April 2010 | + |
| ST7B ^c | Infant stool | 20 April 2010 | + |
| ST19 | Infant stool | 21 April 2010 | + |
| ST4 | Infant stool | 23 April 2010 | + |
| ST7 | Infant stool | 3 May 2010 | + |
| ST21 | Infant stool | 9 May 2010 | + |
| ST29 | Infant stool | 14 May 2010 | + |
| ST31 | Infant stool | 18 May 2010 | + |
| ST32 | Infant stool | 25 May 2010 | + |
| ST33 | Infant stool | 21 June 2010 | + |
| ST34 | Infant stool | 19 August 2010 | + |
| ST39 | Infant stool | 29 September 2010 | + |
| ST40 | Infant stool | 6 October 2010 | + |
| ST41 | Infant stool | 13 October 2010 | + |
| ST43 | Infant stool | 28 October 2010 | + |
| ST44 | Infant stool | 3 November 2010 | + |

^aIsolates are ordered by origin and isolation date. H, environmental isolate from household dust collected from a hand held vacuum cleaner collection bag; V, environmental isolate from household dust collected from a vacuum cleaner collection bag; ST, infant stool isolate.

^b+, plasmid present; -, plasmid absent; based on the present work

^cClosed genome using PacBio and Illumina whole-genome sequencing. All other genomes were only sequenced by Illumina whole-genome sequencing.

also related to plasmids present in *C. sporogenes* and nontoxic group I *C. botulinum* with an identity of ~90% and query coverage of 81% (Table S1). While most genes of pST7B were predicted to encode hypothetical proteins with unknown functions, the few annotated genes provided further insights into the potential role and function of the plasmid. First, pST7B included two genes encoding a type II restriction-modification (R/M) system. These systems are mainly known to act as a defense mechanism against foreign mobile DNA (21) and are not commonly encoded from plasmids (22). The presence of such a system may promote maintenance of pST7B in the bacteria (23) along with protecting the bacterial cells from foreign mobile DNA elements. pST7B also harbored genes encoding an *N*-acetylmuramoyl-L-alanine amidase and a putative amidase domain, whose biological function in the gut remain to be further examined. AmiC, an *N*-acetylmuramoyl-L-alanine amidase, was associated with motility and symbiont association by *Burkholderia* in the insect gut (24). pST7B was present in all our stool isolates (Table 1) and was also conserved to some degree in a number of other infant botulism isolates (Table S1). This leads us to hypothesize that the conserved genetic features present in pST7B and plasmids alike may play a role in colonization and persistence of *C. botulinum* in the infant gut. Further experimental proof of this hypothesis in an intestinal infection model is warranted.

Read mapping analysis singled out the isolate V174. Read mapping of all isolates against ST7B revealed that V174 was distinct from all other isolates. While the percentage of mapped reads for the other isolates ranged from 98.29% to 99.93%, that for V174 was only 87.85% (Data Set S1). This suggested that V174 is unrelated to the other isolates, which is in line with V174 showing an AFLP profile distinct from all the other

isolates related to the infant botulism case (16). Using MiGA, the closest relatives of the *de novo* assembled genome of V174 were *C. botulinum* AM282 (undefined origin, GenBank version number CP013683.1, 98.72% ANI) and strain BrDura (undefined origin, GenBank version number CP014151.1, 98.7% ANI). Remarkably, V174 was present in the same household as the other isolates but was not found in the infant gut, raising interesting hypotheses: (i) the infant did not get exposed to V174 spores, (ii) V174 is lacking genes or traits required for effective germination, colonization, and persistence in the infant gut, and (iii) V174 did not colonize the infant gut through competitive exclusion by the already established *C. botulinum* community. Our present data confirmed that V174 is unrelated to the infant botulism case, as suggested earlier (16). V174 was therefore not included in further comparative analysis.

Mobile elements. As for chromosomal regions or mobile elements present in ST7B and missing in others, the dust isolate V4 was the only one devoid of the plasmid pST7B. As hypothesized above, the absence of pST7B may be detrimental for gut persistence or colonization. No other chromosomal or mobile elements present in ST7B were missing in other sequenced isolates.

To first determine the diversity among the gut or dust isolates, we also searched for additional mobile genetic elements or chromosomal regions that were not found in our reference isolate ST7B. Reads unmapped against *C. botulinum* ST7B were *de novo* assembled but did not yield sequence assemblies larger than 1 kb, except in one case. In isolate V73 (dust isolate), assembly of the 1.71% unmapped reads yielded one contig of 51 kb, predicted to encode a prophage. *De novo* genome assembly showed that this 51-kb prophage-like segment was inserted within *yabG*, which encodes the sporulation-specific protease YabG (25–27). Regulation of sporulation genes by prophage disruption has been reported in other spore-formers; in *Clostridioides difficile* and *Bacillus subtilis*, cryptic prophage-encoding sequences (also called “skin,” for *sigK* intervening sequence) are inserted within *sigK* and coordinate the biosynthesis of the late sporulation sigma factor SigK during sporulation (28–30). The spore counts produced by V73 and ST7B were similar (Fig. S1). Further work on the biological function of this insertion element in *yabG* will reveal if and how it contributes to the sporulation cascade and/or the spore germination properties. It will also bring insights into the ecological relevance of prophage-bacterium interactions and prophage-mediated DNA rearrangement elements.

***C. botulinum* population polymorphism may be shaped by the environment and host.** Read mapping against ST7B of 25 *C. botulinum* isolates related to the infant botulism case identified a number of single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) (Fig. 3 and 4, Data Set S1). SNPs present in all isolates except the reference genome ST7B were further compared with all deposited sequences. The results suggested that these SNPs emerged in the reference genome (ST7B) and not in all the other isolates. We therefore reassigned these SNPs to ST7B (Fig. 3 and 4). Interestingly, most dust isolates (H4, H8, V4, V134, V206) and two stool isolates (ST4, ST31) were identical based on SNP calling (Fig. 3). Their position within a neighbor-joining tree (Fig. 3) suggests that these isolates were the most closely related ones to the common ancestor, from which all other related isolates derived. However, we cannot conclude whether the dust isolates were the source of the infant botulism case or whether they were shed during the early phase of infection. Neither can we exclude the possibility of external sources of the infection.

The dust isolates harbored at most 2 SNPs/InDels (median $\mu_{\text{dust}} = 0$), and stool isolates had up to 9 SNPs/InDels (median $\mu_{\text{stool}} = 3$) (Fig. 3). Some of the isolates extracted from late stool samples (ST34, ST39, ST40, ST43, and ST44) also showed a slightly higher number of SNPs and InDels than earlier stool isolates (Fig. 4), illustrating the possible adaptive pressure exerted by the gut niche to the *C. botulinum* population. Most SNPs and InDels present among all isolates were found in genes relevant for pathogenesis and colonization, as further detailed below (Fig. 4 and Table S2). Specifically, three chromosomal regions were particularly prone to polymorphisms and were related to (i) quorum-sensing (*agr-2* operon), (ii) motility (flagellar apparatus), and

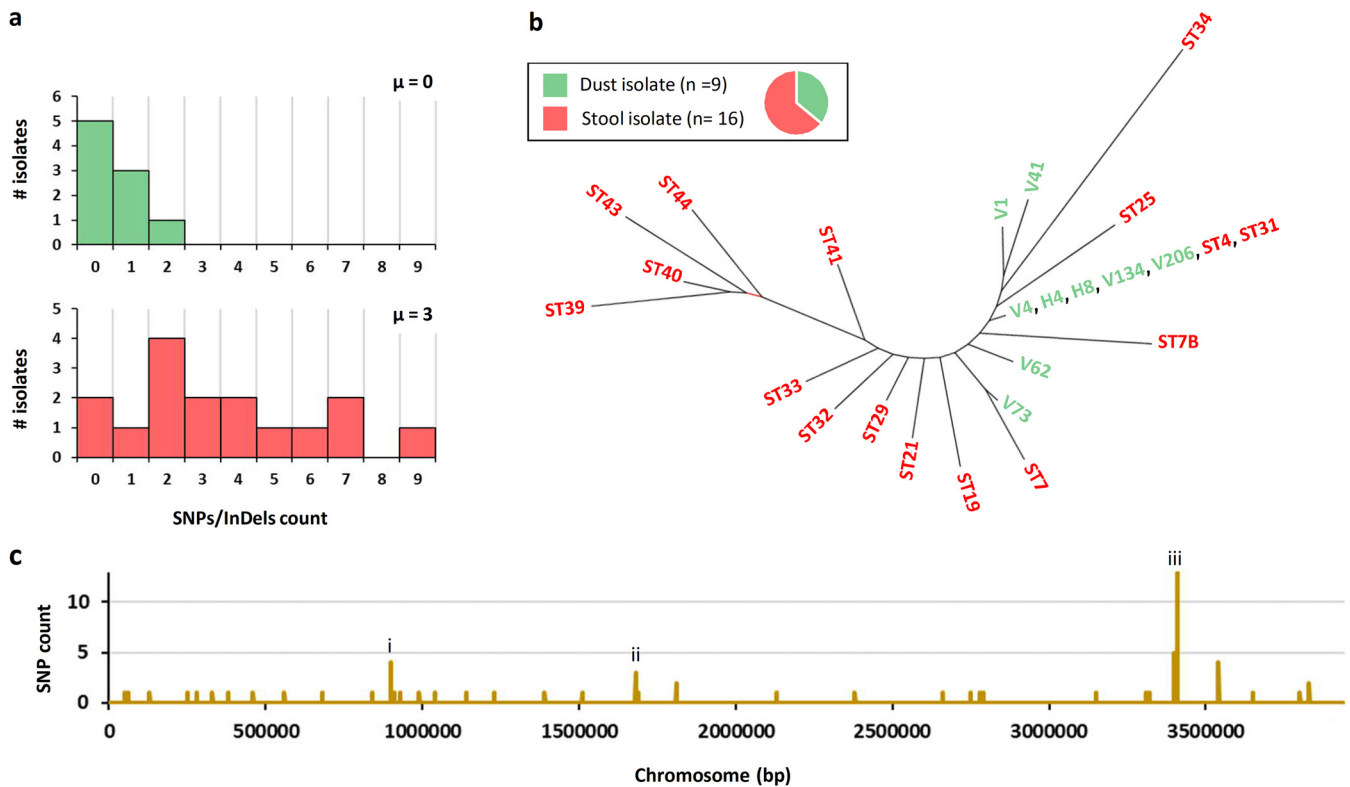


FIG 3 SNP analysis of all environmental (green) and stool isolates (red) of *Clostridium botulinum*. (a) Distribution of the isolates based on their SNP/InDels count. μ , median. (b) Neighbor-joining tree of all isolates based on Hamming distance (Saitou-Nei criterion) using PHYLOViZ software (64). (c) Chromosomal distribution of SNPs and identification of three hot spot mutation regions, labeled (i) motility (flagellar apparatus), (ii) metabolism and transport (trehalose-specific phosphotransferase system [PTS]), and (iii) quorum-sensing (*agr-2* operon).

(iii) metabolism and transport (trehalose-specific phosphotransferase system [PTS]). These systems are further discussed below. Other genes with predicted roles in gut persistence, stress resistance, or antimicrobial resistance are further detailed in Table S2. Also, a SNP present in *mutS* (DNA mismatch repair) in a late fecal isolate (ST39) may suggest the emergence of a novel lineage, potentially yielding to an even higher geno-phenotypic heterogeneity and subpopulations with enhanced adaptation or resistance to prevailing conditions (Table S2).

Agr quorum sensing systems regulate virulence in numerous pathogens, such as *C. perfringens*, *Enterococcus faecalis*, and *Staphylococcus aureus* (31–33). *C. botulinum* ATCC 3502 has two *Agr* quorum sensing systems, of which the *agr-1* operon appears to have a role in sporulation control and the *agr-2* operon in positive regulation of neurotoxin production (34). Similar to ATCC 3502, ST7B and related isolates encode conserved *agr-1* and *agr-2* operons. Of all chromosomal regions, the *agr-2* operon was most prone to mutations (14 distinct locations in total), particularly in the genomes of later stool isolates (Fig. 3). The presence of nonsense mutations or frameshift mutations in the *agr-2* operon may suggest a defective or impaired *Agr-2* quorum sensing system in a number of isolates (Fig. 5). In an effort to determine if these *agr-2* mutated isolates had defective toxin production, we quantified BoNT production in a subset of isolates harboring mutations in the *agr-2* operon (Fig. S2). Large differences in toxin production were observed among the isolates. However, differences in toxin production could not be solely explained by mutations located within the *agr-2* operon, suggesting that other mutations also impact neurotoxinogenesis directly or indirectly. Epigenetic factors are also likely to modulate toxinogenesis, as exemplified in ST40, where the low BoNT production could not be attributed to the presence of any particular SNPs (Fig. S2). In addition, toxin quantification of the isolates did not fit the clinical course of the infection. Indeed, the long-term colonization of the infant gut by

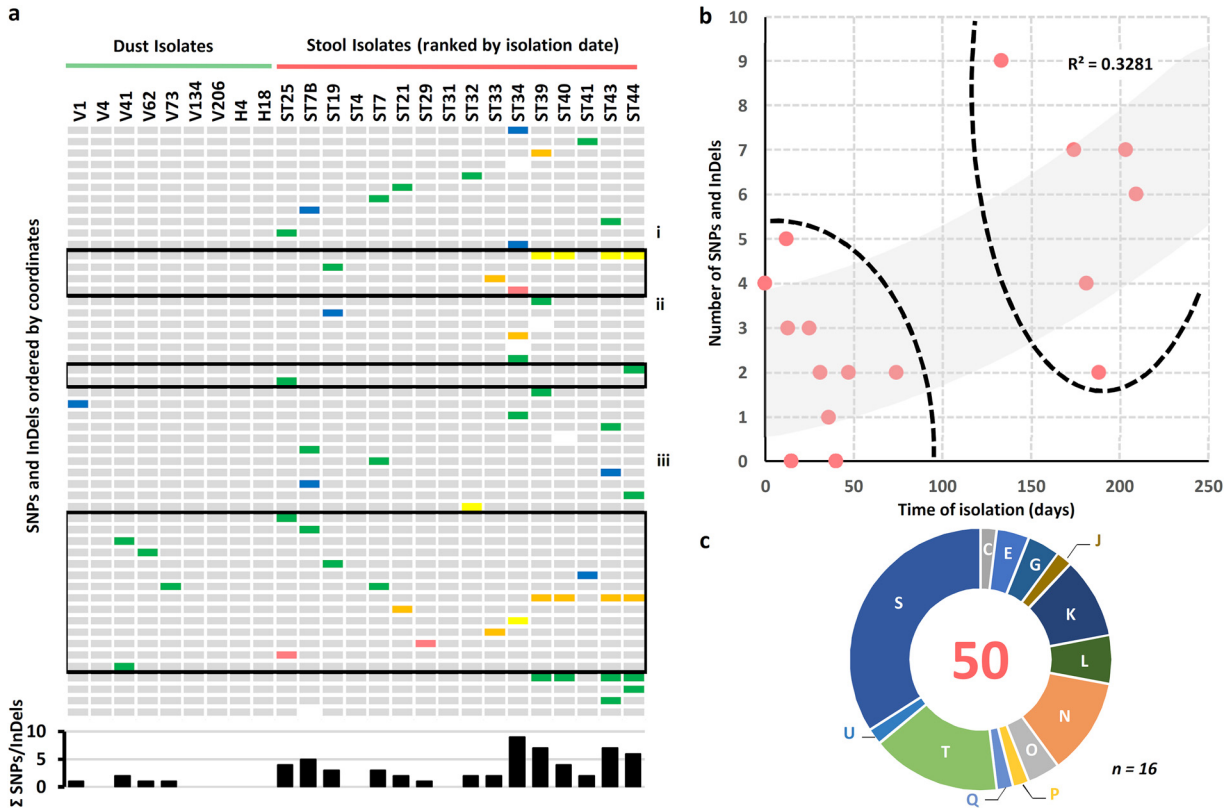


FIG 4 Spatial, temporal, and functional distribution of SNPs identified in isolates. (a) Heat map showing the SNPs and InDels reported for each strain. Isolates are ordered according to their isolation origin and date. White, silent mutation; blue, intergenic region; purple, sense mutation located in rRNA; green, missense mutation; orange, frameshift deletion; yellow, frameshift insertion; red, nonsense mutation. The histogram shows the total of SNPs and Indels per isolate. (b) Number of SNPs and InDels reported for each stool isolate according to the time (in days) from the isolation of the first *Clostridium botulinum* isolate. The two distinct clusters of stool samples are highlighted by dashed lines. The trendline (gray) and its corresponding R^2 value are also indicated. (c) Repartition of the SNPs and InDels in stool isolates per clusters of orthologous groups of proteins (COGs) as determined using eggNOG v5.0 (71). The central number corresponds to the total count of mutated coordinates. The letter identification system describes the different COGs, as previously defined (72). *n*, number of stool isolates.

C. botulinum was not associated with any clinical symptoms in the late time points, in line with the negative mouse bio-assays obtained from stool samples positive for *C. botulinum* (16). These findings suggest that toxin production of isolates grown under laboratory conditions may not truly reflect their toxinogenic phenotype in the gut environment, since exogenous or host factors may intervene in toxin regulation.

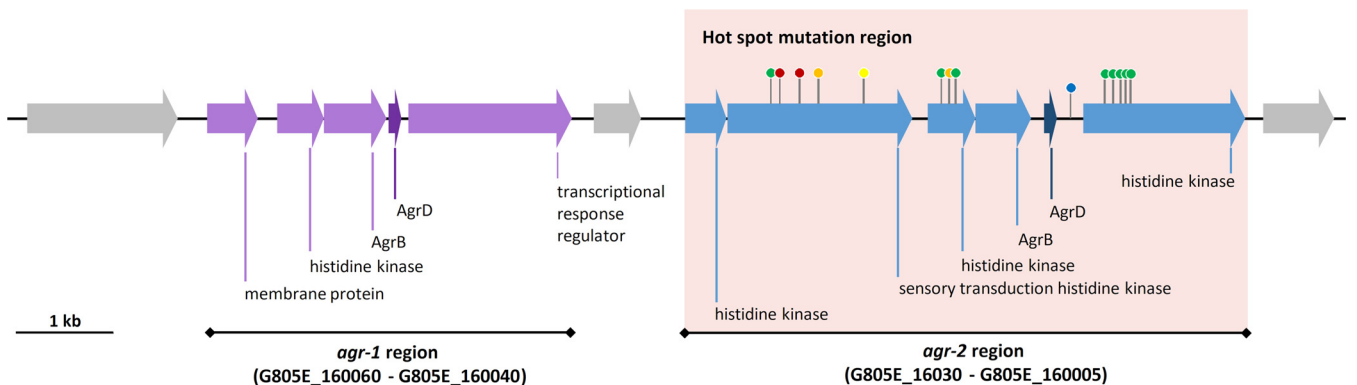


FIG 5 Detailed view of the *agr-1* (purple) and *agr-2* (blue) regions as predicted in the genome of *Clostridium botulinum* strain ST7B. Mutations detected in other isolates compared to ST7B are indicated with color-coded pins. Green, missense mutation; yellow, frameshift insertion; orange, frameshift deletion; red, nonsense mutation; blue, mutation in intergenic region.

Remarkably, isolates from late fecal samples also harbored mutations in genes associated with the flagellar apparatus (*flaA*) or its biosynthesis (*flhA* and *flgN*) (Fig. 3). Flagella confer motility, known to play a role in host colonization of gut pathogens, such as *C. difficile* (35), and to be associated with pleiotropic effects, such as toxin production in *C. difficile* (36, 37). Flagella are an important colonization factor, but on the other hand, immunogenicity of the flagella may be detrimental for the bacteria in regard to the host defense system (38). Thus, bacteria have developed strategies to escape the immune response, such as direct or indirect downregulation of flagellar production (39–41). Alternatively, this might also happen by mutational emergence of isolates with flagellar defects or atypical flagellar filaments or assembly. The frameshift mutations in *flhA* and *flgN* identified in the late-colonizing *C. botulinum* isolates (Fig. 4 and Data Set S1) impacted motility to different degrees (Fig. S3), whereas the mutation (substitution) observed in *flaA* encoding flagellin in isolate ST19 did not alter motility. In ST19, the findings suggest that while the flagellar assembly remained intact and functional, putative alterations in the flagellin structure may have provided ecological benefits, such as escape from the host immune system or enhanced motility and chemotaxis for competing with gut microbes. The motility phenotype was associated with colony morphotype. Motile strains formed smooth-edged colonies, while nonmotile strains formed rough-edged colonies (Fig. S3 and Table S3). Similar findings relating to colony morphotype and motility behaviors have been reported in *C. difficile* (42).

Trehalose transport and metabolism have been subject to intense research in regard to potential roles in *C. difficile* hyper-virulence (43–45), stress tolerance and antibiotic (fluoroquinolone) resistance in *C. perfringens* (46), and drug tolerance and resistance in *Mycobacterium tuberculosis* (47). In both dust and stool isolates, *treP* encoding the trehalose-specific PTS EIBC component (transporter) was mutated at position 270 (Fig. 3). Thus, *treP* in ST25 has a threonine residue at position 270, as opposed to proline in ST7B and leucine in ST44 (most common variant). We hypothesized that this may favorably alter the phenotype of *C. botulinum*. While antibiotic susceptibility assays were inconclusive (Table S4), growth of ST25 and ST44 in semi-defined medium supplemented with D-trehalose suggests that these isolates utilize D-trehalose, as opposed to ST7B (Fig. S4). Interestingly, the SNP present in *treP* (position 270) of ST25 and ST44 was not isogenic based on SNP calling data (Data Set S1). We hypothesized that this may have resulted from multiple passaging of bacterial clones exogenously or from differences in selective pressure. Indeed, consecutive subculturing of isolates ST25 and ST44 in rich medium under laboratory conditions led to the changes in the *treP* SNP frequency in ST25 and ST44, suggesting that the pheno-genotype of wild isolates may be prone to further polymorphisms upon maintenance under laboratory conditions (Data Set S1). Further work with directed mutagenesis will investigate whether the different *treP* variants are associated with altered D-trehalose transport efficiency and if *treP* polymorphism is dictated by the gut environmental pressure, by providing a competitive edge.

Along with neurotoxin production, sporulation is another key phenotypic trait in *C. botulinum*. In *Bacilli* and *Clostridia*, sporulation is regulated by the sporulation master switch Spo0A and involves a tightly controlled cascade of multiple sigma factors and their regulons in the mother cell and in the forespore (48). Spo0A also controls metabolic pathways and biofilm formation, and, importantly, regulates virulence in spore formers (49, 50). *C. botulinum* V1 and ST34 harbored mutations either upstream of *spo0A* or in its CDS, respectively. Interestingly, both strains demonstrated particularly early spore formation in two different media (Fig. S1), suggesting that these particular mutations may also impact the general physiology of the cells.

Concluding remarks. Here, we provide a comprehensive view of the genomic dynamics and plasticity of *C. botulinum* during the course of infant botulism, to further comprehend the adaptation strategies developed by *C. botulinum* to colonize and persist in the gut environment. While the isolates from this infant botulism case evidently shared a recent and common ancestor, marked genomic alterations were observed

among the isolates. These included plasmid loss, prophage-mediated insertion element, and point mutations in genes with predicted functions in pathogenesis as well as host colonization and persistence in the gut. We also reported large variations in the phenotype of the isolates in terms of sporulation, toxin production, trehalose utilization, and motility, illustrating possible selection or adaptation to the gut conditions. While we were able to correlate genotypes and phenotypes to some extent, phenotypic heterogeneity among the isolates could not be fully associated with corresponding genotypes, suggesting that some SNPs may have remained undetected or that some phenotypic traits might result from specific SNP combinations or epigenetic factors. Nevertheless, our findings shed light on the pathogenesis of *C. botulinum* and led us to speculate on a two-stage model of infant botulism; upon ingestion, environmental spores with adequate phenotypic traits are first selected for their ability to survive and colonize the infant gut (stage I). This *C. botulinum* subpopulation is then able to germinate and outgrow, resulting in the onset of the disease. Subsequently, the population is driven toward adapting to the host gut conditions (stage II), until it reaches a tipping point where host factors and other gut microbes lead to clearance of *C. botulinum*. It is important to bear in mind that the phenotypic diversity observed *in vitro* may not truly reflect the phenotypic diversity of the same pool of isolates in the infant gut. Environmental and ecological conditions, i.e., pH, nutrient availability, competitive exclusion, host-bacterium cross-interactions, bacterium-bacterium interactions, and the host immune system, are likely to impact the phenotype and fitness of each isolate. In addition to DNA-level alterations, phenotypic variation is also prone to arise from epigenetic factors including gene regulation, translation, and protein-protein interactions.

This work provides novel perspectives on the pathogenesis and adaptive mechanisms of *C. botulinum* in toxicoinfectious botulism that may be instrumental in developing novel strategies for prevention and treatment. Comprehensive analysis of the infant gut microbiota will further refine the bacterial traits and host factors involved in the different stages of intestinal botulism in infants.

MATERIALS AND METHODS

Bacterial strains, genomic DNA isolation, and genome sequencing. All *C. botulinum* isolates were derived from an infant botulism case previously reported by Derman et al. (16) (Table 1). Routinely, *C. botulinum* cultures were anaerobically grown in Trypticase-peptone-glucose-yeast extract (TPGY) broth at 37°C for 16 h, unless specified. Total genomic DNA was isolated using a Wizard genomic DNA purification kit (Promega, Wisconsin, USA) and quantified with a Qubit fluorometric assay (Thermo Scientific, Massachusetts, USA).

Isolated in the early onset of the disease (early colonizer), *C. botulinum* ST7B was selected as the reference genome for this work. The closed genome of ST7B was obtained using the PacBio sequencing platform (Pacific Biosciences, California, USA) at the DNA Sequencing and Genomics Laboratory (Institute of Biotechnology, University of Helsinki, Helsinki, Finland). The PacBio DNA library was constructed using the DNA template prep kit v2.0 provided by Pacific Biosciences. Genomic DNA was shredded (11-kp-long fragments) with Megaruptor (Diagenode, New Jersey, USA). The SMRTbell libraries were prepared by ligating blunt-end hairpin adapters to the fragmented DNA samples. Sequencing primers were annealed, and polymerase was bound to the SMRTbell templates using the PacBio DNA/polymerase binding kit P6. MagBeads were used to add SMRTbell libraries to SMRTcells. The PacBio RS II instrument was run for a total of 360 min.

In addition, DNA from all isolates was sent for short-read whole-genome sequencing at the Institute for Molecular Medicine Finland (FIMM), University of Helsinki (Helsinki, Finland). The Illumina sequencing library was prepared using a Nextera DNA Flex library prep kit (Illumina, California, USA), and DNA fragments of 500 to 850 bp were selected using BluePippin (Sage Science, Massachusetts, USA). Genome sequencing was performed using Illumina HiSeq 2500 or NovaSeq 6000 (Illumina) and yielded paired-end reads (2×101 bp).

Genome assembly of *C. botulinum* ST7B. We assembled the complete genome of ST7B using both Illumina and PacBio sequencing reads. We sketched concatenated paired-end Illumina reads with Mash (51) v2.0 (with *k*-mer size of 32 and minimum *k*-mer copy number of 3) to obtain an initial estimate for the genome size. The initial estimate was used as the target genome size parameter for Flye (52) v2.6 (with the plasmid recovery option and zero polishing iterations) to assemble the concatenated FASTQ-format PacBio subreads once. We then used the sum of the lengths of the contigs in the resulting assembly to update the target genome size parameter and reran Flye on the PacBio reads using the updated target genome size parameter, the plasmid recovery option, and 10 polishing iterations to produce the long-read-only Flye assembly. Illumina reads were aligned to the Flye assembly using the Burrows-Wheeler Aligner MEM algorithm (BWA-MEM v0.7.17-r1198-dirty) (53), and the alignment was

sorted and indexed with SAMtools v1.9-63-g649f04f (54). The alignment was used to correct errors in the Flye assembly with Pilon v1.23 (55), producing the final, error-corrected assembly.

Read mapping, SNP calling, and variant filtering. In an effort to accurately identify SNPs and InDels, we defined a number of analytical parameters and filters most suitable for our work based on the review by Olson et al. (56). We did not expect a large number of variants, SNPs, and InDels among the isolates, which justified the need to carefully assess the quality of the variants in order to report reliable phylogenetic observations. First, rather than using a publicly available genome as a reference for read mapping, the genome of an early colonizer, namely, ST7B, was sequenced and closed (present work) to reduce any possible SNP calling bias. We suspected sequencing errors in our ST7B reference genome sequence. Therefore, we verified them by Sanger sequencing and, if confirmed, removed them from subsequent analysis (Data Set S1). Second, to mitigate errors related to sample processing, sequencing, and read mapping, all isolates described in the study were sequenced with a high sequencing depth (average mapping coverage higher than 150×) using paired-end reads (Data Set S1). Sufficient sequencing depth and the use of paired-end sequencing reads are known to reduce such errors (56). Illumina paired-end raw reads from all isolates were imported in CLC Genomics Workbench v11.0.1 (Qiagen GmbH, Germany). Reads were trimmed and mapped to the ST7B genome (Data Set S1). SNP, InDel, and structural variant calling and filtering were then carried out in CLC Genomics Workbench v11.0.1 (Data Set S1). Variants located at positions with ambiguous nonspecific read matches to the reference sequence were filtered out. We subsequently inspected variants using the CLC Genomic Workbench v11.0.1 viewer tool. The SNPs and InDels located within genes with biological relevance for this work were also confirmed by Sanger sequencing. Such genes were associated with motility, quorum sensing, sporulation, and carbohydrate transport (Data Set S1). Finally, sequencing reads with no matches to the ST7B genome were *de novo* assembled in CLC Genomics Workbench v11.0.1 with the following parameters: mapping mode = map reads back to contigs (slow), update contigs = yes, automatic bubble size = yes, minimum contig length = 1,000, automatic word size = yes, perform scaffolding = yes, auto-detect paired distances = yes, mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.5, similarity fraction = 0.8, create list of unmapped reads = no, colorspace alignment = no, guidance only reads = no, min distance = 1, max distance = 1,000, in order to identify possible extra genetic elements present in these isolates.

Genome assembly of V174. In line with previous results, read mapping of the isolate V174 against ST7B revealed V174 to be distinct from the other isolates (Data Set S1). The genome of V174 was therefore assembled solely based on Illumina paired-end sequencing reads. *De novo* assembly was carried out in CLC Genomics Workbench v11.0.1 using the same parameters as listed above.

Comparative genomic analyses. The MiGA (Microbial Genomes Atlas) Web interface (57) was used to reveal the taxonomic identification, average nucleotide identity (ANI), and shared genome percentage of the isolates. PHASTER analysis (58, 59) was carried out to search for prophages. The genome sequence of isolate ST7B (stool) was further analyzed using the comprehensive genome analysis tool available at PATRIC (60), the RAST annotation tool kit (RASTtk) (61), the genome annotation tool Prokka (62), and the multiple-genome alignment tool Mauve v2.4.0 (63). Dendrograms and a minimum spanning tree were generated based on SNPs in each isolate using PHYLOVIZ (64).

Motility assay. From an overnight culture in TPGY medium, *C. botulinum* isolates ST19, ST34, ST39, ST40, ST43, ST44, and ST7B were inoculated by stabbing wells containing 3 mL of anaerobic semisolid TPGY agar medium (0.3% wt/vol agar content) in 6-well plates. The plates were incubated anaerobically at 37°C and observed for motility after 24 h. Three biological replicate series of each isolate were included.

Sporulation assay. Spore counts of *C. botulinum* V1, ST34, V73, and ST7B cultures were determined (65) in TPGY and Trypticase-peptone-yeast extract (TPY) media after anaerobic incubation at 37°C for 24 h, 48 h, and 120 h postinoculation. For each time point, a culture aliquot was sampled and heat-treated at 80°C for 15 min. Spore counts were determined using the most-probable-number (MPN) technique as previously described (66). Three biological replicates of each isolate were included.

Toxin quantification. *C. botulinum* isolates with SNPs located within the *agr-2* operon (Data Set S1) were grown at 37°C anaerobically in either Trypticase-yeast extract D-glucose (TYG) medium, Trypticase-peptone-yeast extract (TPY) medium, or TPY medium supplemented with 0.4% wt/vol D-glucose. Each isolate was inoculated 1/100 into 10 mL of fresh medium. Samples were collected after 24 h and stored at −80°C. Toxins present in the cultures were then quantified using a sandwich enzyme-linked immunosorbent assay (ELISA) for detection of BoNT/A as previously described (67, 68). Three biological replicates of each isolate were included.

Antibiotic resistance assay. *C. botulinum* isolates ST7B (reference early isolate), ST41 (late stool isolate), ST25 (early stool isolate with *treP* SNP), and ST44 (late stool isolate with *treP* SNP) were anaerobically grown in 5 mL of TPGY medium overnight at 37°C, and 100 μL of culture was streaked onto prewarmed anaerobic bovine blood agar plates, to which Etest strips (bioMérieux, France) were applied to determine the MIC for the fluoroquinolones moxifloxacin (MX), levofloxacin (LE), and ciprofloxacin (CI), known to be linked with trehalose metabolism (46, 47). Three biological replicates were included. The MIC values were determined after 24-h incubation according to the manufacturer's protocol.

Growth curve assay. *C. botulinum* isolates ST7B (reference isolate), ST25, and ST44 were first anaerobically grown in 5 mL of TPY medium overnight. Then, we used semi-defined medium (called CDM-I), which is a medium essentially based on previous work by Whitmer and Johnson (69) where the essential amino acid mixture was replaced by Casamino Acids (10 g/L) and L-tryptophan (0.1 g/L) and also supplemented with CoCl₂ (1,189.7 mg/200 mL) and Na₂SeO₃ (17.3 mg/200 mL). Each isolate was then inoculated 1/100 into CDM-I or CDM-I plus 0.4% wt/vol D-trehalose. Then, 200 μL per well was dispensed into

a 96-well plate. The plate was placed into the Hidex Sense plate reader (Hidex Oy, Finland) at 37°C for 24 h under anaerobic conditions. The optical density at 600 nm was measured every 15 min. The growth curve assay was performed in triplicate (three biological replicates consisting of four technical replicates) for each condition and isolate.

Data availability. The sequencing reads and all associated metadata were deposited in the NCBI Sequence Read Archive (SRA) under the project number [PRJNA610151](https://www.ncbi.nlm.nih.gov/sra/PRJNA610151) (Data Set S1). The closed genome of *C. botulinum* ST7B was deposited under the accession numbers [CP050251](https://www.ncbi.nlm.nih.gov/nuclink/CP050251) (chromosome) and [CP050252](https://www.ncbi.nlm.nih.gov/nuclink/CP050252) (plasmid).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, PDF file, 0.5 MB.

FIG S1, PDF file, 0.1 MB.

FIG S2, PDF file, 0.02 MB.

FIG S3, PDF file, 0.2 MB.

FIG S4, PDF file, 0.1 MB.

TABLE S1, PDF file, 0.01 MB.

TABLE S2, DOCX file, 0.04 MB.

TABLE S3, PDF file, 0.01 MB.

TABLE S4, PDF file, 0.01 MB.

ACKNOWLEDGMENTS

The study was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-CoG whyBOther, grant agreement number 683099). The work was also financed by the HiLife Fellows Program, University of Helsinki, the Academy of Finland (grants number 299700 and 310261).

We thank Hanna Korpunen for her technical assistance. We also thank FIMM Genomics (Institute for Molecular Medicine Finland, Biocenter Finland), and the DNA Sequencing and Genomics Laboratory of the Institute of Biotechnology (University of Helsinki).

We declare that we have no conflict of interest.

REFERENCES

- Wopereis H, Oozeer R, Knipping K, Belzer C, Knol J. 2014. The first thousand days - intestinal microbiology of early life: establishing a symbiosis. *Pediatr Allergy Immunol* 25:428–438. <https://doi.org/10.1111/pai.12232>.
- Guru PK, Becker TL, Stephens A, Cannistraro RJ, Eidelman BH, Hata DJ, Brumble L. 2018. Adult intestinal botulism: a rare presentation in an immunocompromised patient with short bowel syndrome. *Mayo Clin Proc Innov Qual Outcomes* 2:291–296. <https://doi.org/10.1016/j.mayocpiqo.2018.06.005>.
- Harris RA, Anniballi F, Austin JW. 2020. Adult intestinal toxemia botulism. *Toxins* (Basel 12:81.). <https://doi.org/10.3390/toxins12020081>.
- Sheppard YD, Middleton D, Whitfield Y, Tyndel F, Haider S, Spiegelman J, Swartz RH, Nelder MP, Baker SL, Landry L, Maceachern R, Deamond S, Ross L, Peters G, Baird M, Rose D, Sanders G, Austin JW. 2012. Intestinal toxemia botulism in 3 adults, Ontario, Canada, 2006–2008. *Emerg Infect Dis* 18:1–6. <https://doi.org/10.3201/eid1801.110533>.
- Rosow LK, Strober JB. 2015. Infant botulism: review and clinical update. *Pediatr Neurol* 52:487–492. <https://doi.org/10.1016/j.pediatrneurol.2015.01.006>.
- Arnon SS, Midura TF, Damus K, Thompson B, Wood RM, Chin J. 1979. Honey and other environmental risk factors for infant botulism. *J Pediatr* 94:331–336. [https://doi.org/10.1016/S0022-3476\(79\)80863-X](https://doi.org/10.1016/S0022-3476(79)80863-X).
- Arnon SS, Midura TF, Clay SA, Wood RM, Chin J. 1977. Infant botulism: epidemiological, clinical, and laboratory aspects. *JAMA* 237:1946–1951. <https://doi.org/10.1001/jama.1977.03270450036016>.
- Thompson JA, Glasgow LA, Warpinski JR, Olson C. 1980. Infant botulism: clinical spectrum and epidemiology. *Pediatrics* 66:936–942. <https://doi.org/10.1542/peds.66.6.936>.
- Nevas M, Lindström M, Virtanen A, Hiemi S, Kuusi M, Arnon SS, Vuori E, Korkeala H. 2005. Infant botulism acquired from household dust presenting as sudden infant death syndrome. *J Clin Microbiol* 43:511–513. <https://doi.org/10.1128/JCM.43.1.511-513.2005>.
- Brett MM, McLauchlin J, Harris A, O'Brien S, Black N, Forsyth RJ, Roberts D, Bolton FJ. 2005. A case of infant botulism with a possible link to infant formula milk powder: evidence for the presence of more than one strain of *Clostridium botulinum* in clinical specimens and food. *J Med Microbiol* 54:769–776. <https://doi.org/10.1099/jmm.0.46000-0>.
- Grant KA, Nwarfor I, Mpmugo O, Mithani V, Lister P, Dixon G, Nixon G, Planche T, Courtney M, Morgan J, McLauchlin J. 2009. Report of two unlinked cases of infant botulism in the UK in October 2007. *J Med Microbiol* 58:1601–1606. <https://doi.org/10.1099/jmm.0.011510-0>.
- Shelley EB, O'Rourke D, Grant K, McArdle E, Capra L, Clarke A, McNamara E, Cunney R, McKeown P, Amar CFL, Cosgrove C, Fitzgerald M, Harrington P, Garvey P, Grainger F, Griffin J, Lynch BJ, McGrane G, Murphy J, Ni Shuibhne N, Prosser J. 2015. Infant botulism due to *C. butyricum* type E toxin: a novel environmental association with pet terrapins. *Epidemiol Infect* 143:461–469. <https://doi.org/10.1017/S0950268814002672>.
- Pellizzari R, Rossetto O, Schiavo G, Montecucco C. 1999. Tetanus and botulinum neurotoxins: mechanism of action and therapeutic uses. *Philos Trans R Soc Lond B Biol Sci* 354:259–268. <https://doi.org/10.1098/rstb.1999.0377>.
- Dolly JO, Aoki KR. 2006. The structure and mode of action of different botulinum toxins. *Eur J Neurol* 13:1–9. <https://doi.org/10.1111/j.1468-1331.2006.01648.x>.
- Dressler D, Adib Saberi F. 2005. Botulinum toxin: mechanisms of action. *Eur Neurol* 53:3–9. <https://doi.org/10.1159/000083259>.
- Derman Y, Korkeala H, Salo E, Lonnqvist T, Saxén H, Lindström M. 2014. Infant botulism with prolonged faecal excretion of botulinum neurotoxin and *Clostridium botulinum* for 7 months. *Epidemiol Infect* 142:335–339. <https://doi.org/10.1017/S0950268813001258>.
- Porse A, Gumpert H, Kubicek-Sutherland JZ, Karami N, Adlerberth I, Wold AE, Andersson DI, Sommer MOA. 2017. Genome dynamics of *Escherichia*

- coli* during antibiotic treatment: transfer, loss, and persistence of genetic elements *in situ* of the infant gut. *Front Cell Infect Microbiol* 7:126.
18. Suzuki R, Satou K, Shiroma A, Shimoji M, Teruya K, Matsumoto T, Akada J, Hirano T, Yamaoka Y. 2019. Genome-wide mutation analysis of *Helicobacter pylori* after inoculation to Mongolian gerbils. *Gut Pathog* 11:45. <https://doi.org/10.1186/s13099-019-0326-5>.
 19. Smith TJ, Hill KK, Xie G, Foley BT, Williamson CHD, Foster JT, Johnson SL, Chertkov O, Teshima H, Gibbons HS, Johnsky LA, Karavis MA, Smith LA. 2015. Genomic sequences of six botulinum neurotoxin-producing strains representing three clostridial species illustrate the mobility and diversity of botulinum neurotoxin genes. *Infect Genet Evol* 30:102–113. <https://doi.org/10.1016/j.meegid.2014.12.002>.
 20. Hill KK, Xie G, Foley BT, Smith TJ, Munk AC, Bruce D, Smith LA, Brettin TS, Detter JC. 2009. Recombination and insertion events involving the botulinum neurotoxin complex genes in *Clostridium botulinum* types A, B, E and F and *Clostridium butyricum* type E strains. *BMC Biol* 7:66. <https://doi.org/10.1186/1741-7007-7-66>.
 21. Vasu K, Nagaraja V. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* 77:53–72. <https://doi.org/10.1128/MMBR.00044-12>.
 22. Oliveira PH, Touchon M, Rocha EPC. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res* 42:10618–10631. <https://doi.org/10.1093/nar/gku734>.
 23. Kulakauskas S, Lubys A, Ehrlich SD. 1995. DNA restriction-modification systems mediate plasmid maintenance. *J Bacteriol* 177:3451–3454. <https://doi.org/10.1128/jb.177.12.3451-3454.1995>.
 24. Lee JB, Byeon JH, Jang HA, Kim JK, Yoo JW, Kikuchi Y, Lee BL. 2015. Bacterial cell motility of *Burkholderia* gut symbiont is required to colonize the insect gut. *FEBS Lett* 589:2784–2790. <https://doi.org/10.1016/j.febslet.2015.08.022>.
 25. Shrestha R, Cochran AM, Sorg JA. 2019. The requirement for co-germinants during *Clostridium difficile* spore germination is influenced by mutations in *yabG* and *cspA*. *PLoS Pathog* 15:e1007681. <https://doi.org/10.1371/journal.ppat.1007681>.
 26. Takamatsu H, Kodama T, Imamura A, Asai K, Kobayashi K, Nakayama T, Ogasawara N, Watabe K. 2000. The *Bacillus subtilis yabG* gene is transcribed by SigK RNA polymerase during sporulation, and *yabG* mutant spores have altered coat protein composition. *J Bacteriol* 182:1883–1888. <https://doi.org/10.1128/JB.182.7.1883-1888.2000>.
 27. Takamatsu H, Imamura A, Kodama T, Asai K, Ogasawara N, Watabe K. 2000. The *yabG* gene of *Bacillus subtilis* encodes a sporulation specific protease which is involved in the processing of several spore coat proteins. *FEMS Microbiol Lett* 192:33–38. <https://doi.org/10.1111/j.1574-6968.2000.tb09355.x>.
 28. Krogh S, O'Reilly M, Nolan N, Devine KM. 1996. The phage-like element PB5X and part of the *skin* element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous. *Microbiology* 142:2031–2040. <https://doi.org/10.1099/13500872-142-8-2031>.
 29. Haraldsen JD, Sonenshein AL. 2003. Efficient sporulation in *Clostridium difficile* requires disruption of the *sigmaK* gene. *Mol Microbiol* 48:811–821. <https://doi.org/10.1046/j.1365-2958.2003.03471.x>.
 30. Pereira FC, Saujet L, Tomé AR, Serrano M, Monot M, Couture-Tosi E, Martin-Verstraete I, Dupuy B, Henriques AO. 2013. The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. *PLoS Genet* 9:e1003782. <https://doi.org/10.1371/journal.pgen.1003782>.
 31. Thoendel M, Kavanaugh JS, Flack CE, Horswill AR. 2011. Peptide signaling in the staphylococci. *Chem Rev* 111:117–151. <https://doi.org/10.1021/cr100370n>.
 32. Ohtani K, Yuan Y, Hassan S, Wang R, Wang Y, Shimizu T. 2009. Virulence gene regulation by the *agr* system in *Clostridium perfringens*. *J Bacteriol* 191:3919–3927. <https://doi.org/10.1128/JB.01455-08>.
 33. Nakayama J, Chen S, Oyama N, Nishiguchi K, Azab EA, Tanaka E, Kariyama R, Sonomoto K. 2006. Revised model for *Enterococcus faecalis* *fsr* quorum-sensing system: the small open reading frame *fsrD* encodes the gelatinase biosynthesis-activating pheromone propeptide corresponding to staphylococcal *agrd*. *J Bacteriol* 188:8321–8326. <https://doi.org/10.1128/JB.00865-06>.
 34. Cooksley CM, Davis IJ, Winzer K, Chan WC, Peck MW, Minton NP. 2010. Regulation of neurotoxin production and sporulation by a putative *agrBD* signaling system in proteolytic *Clostridium botulinum*. *Appl Environ Microbiol* 76:4448–4460. <https://doi.org/10.1128/AEM.03038-09>.
 35. Stevenson E, Minton NP, Kuehne SA. 2015. The role of flagella in *Clostridium difficile* pathogenicity. *Trends Microbiol* 23:275–282. <https://doi.org/10.1016/j.tim.2015.01.004>.
 36. Aubry A, Hussack G, Chen W, KuoLee R, Twine SM, Fulton KM, Foote S, Carrillo CD, Tanha J, Logan SM. 2012. Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. *Infect Immun* 80:3521–3532. <https://doi.org/10.1128/IAI.00224-12>.
 37. Barketi-Klai A, Monot M, Hoys S, Lambert-Bordes S, Kuehne SA, Minton N, Collignon A, Dupuy B, Kansau I. 2014. The flagellin FliC of *Clostridium difficile* is responsible for pleiotropic gene regulation during *in vivo* infection. *PLoS One* 9:e96876. <https://doi.org/10.1371/journal.pone.0096876>.
 38. Liaudet L, Deb A, Pacher P, Mabley JG, Murthy KGK, Salzman AL, Szabo C. 2002. The flagellin-TLR5 axis: therapeutic opportunities. *Drug News Perspect* 15:397–409. <https://doi.org/10.1358/dnp.2002.15.7.840075>.
 39. Lai MA, Quarles EK, Lopez-Yglesias AH, Zhao X, Hajjar AM, Smith KD. 2013. Innate immune detection of flagellin positively and negatively regulates salmonella infection. *PLoS One* 8:e72047. <https://doi.org/10.1371/journal.pone.0072047>.
 40. Akerley BJ, Cotter PA, Miller JF. 1995. Ectopic expression of the flagellar regulon alters development of the *Bordetella*-host interaction. *Cell* 80:611–620. [https://doi.org/10.1016/0092-8674\(95\)90515-4](https://doi.org/10.1016/0092-8674(95)90515-4).
 41. Rossez Y, Wolfson EB, Holmes A, Gally DL, Holden NJ. 2015. Bacterial flagella: twist and stick, or dodge across the kingdoms. *PLoS Pathog* 11:e1004483. <https://doi.org/10.1371/journal.ppat.1004483>.
 42. Garrett EM, Sekulovic O, Wetzel D, Jones JB, Edwards AN, Vargas-Cuevas G, McBride SM, Tamayo R. 2019. Phase variation of a signal transduction system controls *Clostridioides difficile* colony morphology, motility, and virulence. *PLoS Biol* 17:e3000379. <https://doi.org/10.1371/journal.pbio.3000379>.
 43. Collins J, Robinson C, Danhof H, Knetsch CW, van Leeuwen HC, Lawley TD, Auchtung JM, Britton RA. 2018. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature* 553:291–294. <https://doi.org/10.1038/nature25178>.
 44. Saund K, Rao K, Young VB, Snitkin ES. 2020. Genetic determinants of trehalose utilization are not associated with severe *Clostridium difficile* infection outcome. *Open Forum Infect Dis* 7:ofz548. <https://doi.org/10.1093/ofid/ofz548>.
 45. Eyre DW, Didelot X, Buckley AM, Freeman J, Moura IB, Crook DW, Peto TEA, Walker AS, Wilcox MH, Dingle KE. 2019. *Clostridium difficile* trehalose metabolism variants are common and not associated with adverse patient outcomes when variably present in the same lineage. *EBioMedicine* 43:347–355. <https://doi.org/10.1016/j.ebiom.2019.04.038>.
 46. Park M, Mitchell WJ, Rafii F. 2016. Effect of trehalose and trehalose transport on the tolerance of *Clostridium perfringens* to environmental stress in a wild type strain and its fluoroquinolone-resistant mutant. *Int J Microbiol* 2016:4829716. <https://doi.org/10.1155/2016/4829716>.
 47. Lee JJ, Lee S-K, Song N, Nathan TO, Swarts BM, Eum S-Y, Ehart S, Cho S-N, Eoh H, et al. 2019. Transient drug-tolerance and permanent drug-resistance rely on the trehalose-catalytic shift in *Mycobacterium tuberculosis*. *Nat Comm* 10:2928. <https://doi.org/10.1038/s41467-019-10975-7>.
 48. Paredes CJ, Alsaker KV, Papoutsakis ET. 2005. A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol* 3:969–978. <https://doi.org/10.1038/nrmicro1288>.
 49. Molle V, Fujita M, Jensen ST, Eichenberger P, González-Pastor JE, Liu JS, Losick R. 2003. The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol* 50:1683–1701. <https://doi.org/10.1046/j.1365-2958.2003.03818.x>.
 50. Pettit LJ, Browne HP, Yu L, Smits WK, Fagan RP, Barquist L, Martin MJ, Goulding D, Duncan SH, Flint HJ, Dougan G, Choudhary JS, Lawley TD. 2014. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC Genom* 15:160. <https://doi.org/10.1186/1471-2164-15-160>.
 51. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 17:132. <https://doi.org/10.1186/s13059-016-0997-x>.
 52. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
 53. Li HW. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* <https://doi.org/10.48550/arXiv.1303.3997>.
 54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.

55. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
56. Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, Keim P, Morrow JB, Salit ML, Zook JM. 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Front Genet* 6:235.
57. Rodríguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) web-server: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res* 46:W282–w8. <https://doi.org/10.1093/nar/gky467>.
58. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkw387>.
59. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res* 39:W347–W352. <https://doi.org/10.1093/nar/gkr485>.
60. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 42:D581–D91. <https://doi.org/10.1093/nar/gkt1099>.
61. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
62. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
63. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>.
64. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* 13:87. <https://doi.org/10.1186/1471-2105-13-87>.
65. Derman Y, Isokallio M, Lindstrom M, Korkeala H. 2013. The two-component system CBO2306/CBO2307 is important for cold adaptation of *Clostridium botulinum* ATCC 3502. *Int J Food Microbiol* 167:87–91. <https://doi.org/10.1016/j.ijfoodmicro.2013.06.004>.
66. Food and Drug Administration. 1998. Bacteriological analytical manual, 8th ed. AOAC International, Gaithersburg, MD.
67. Simon S, Fiebig U, Liu Y, Tierney R, Dano J, Worbs S, Endermann T, Nevers M-C, Volland H, Sesardic D, Dorner MB. 2015. Recommended immunological strategies to screen for botulinum neurotoxin-containing samples. *Toxins (Basel)* 7:5011–5034. <https://doi.org/10.3390/toxins7124860>.
68. Pauly D, Kirchner S, Stoermann B, Schreiber T, Kaulfuss S, Schade R, Zbinden R, Avondet M-A, Dorner MB, Dorner BG. 2009. Simultaneous quantification of five bacterial and plant toxins from complex matrices using a multiplexed fluorescent magnetic suspension assay. *Analyst* 134: 2028–2039. <https://doi.org/10.1039/b911525k>.
69. Whitmer ME, Johnson EA. 1988. Development of improved defined media for *Clostridium botulinum* serotypes A, B, and E. *Appl Environ Microbiol* 54:753–759. <https://doi.org/10.1128/aem.54.3.753-759.1988>.
70. Gonzalez-Escalona N, Timme R, Raphael BH, Zink D, Sharma SK. 2014. Whole-genome single-nucleotide-polymorphism analysis for discrimination of *Clostridium botulinum* group I strains. *Appl Environ Microbiol* 80: 2125–2132. <https://doi.org/10.1128/AEM.03934-13>.
71. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
72. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <https://doi.org/10.1186/1471-2105-4-41>.