Genomic epidemiology of *Clostridium botulinum* isolates from temporally related cases of infant botulism in New South Wales, Australia.

Nadine McCallum²³#, Timothy J Gray¹, Qinning Wang¹², Jimmy Ng¹, Leanne Hicks¹, Trang Nguyen¹, Marion Yuen¹, Grant A. Hill-Cawthorne³⁴, Vitali Sintchenko¹²³

¹Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, NSW, Australia.²Centre for Infectious Diseases and Microbiology – Public Health, Westmead Hospital, Westmead, NSW, Australia.³The Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, the University of Sydney, Sydney, NSW, Australia.⁴School of Public Health, the University of Sydney, Sydney, NSW, Australia.

#Corresponding author:
Dr Nadine McCallum
Centre for Infectious Diseases and Microbiology, level 3, ICPMR
Westmead Hospital, Westmead, NSW2145
AUSTRALIA
Phone: + 61 2 9845 6255 | F: + 61 2 9891 5317
Email: nadine.holmes@sydney.edu.au
Abstract

Infant botulism is a potentially life-threatening paralytic disease that can be associated with prolonged morbidity if not rapidly diagnosed and treated. Four infants were diagnosed and treated for infant botulism in NSW, Australia, between May 2011 and August 2013. Despite the temporal relationship between cases, there was no close geographical clustering or other epidemiological links. *Clostridium botulinum* isolates, three of which produced botulism neurotoxin serotype A (BoNT/A) and one BoNT serotype B (BoNT/B), were characterised using whole-genome sequencing (WGS). *In silico* multi-locus sequence typing (MLST) found that two of the BoNT/A-producing isolates shared an identical novel sequence type, ST84. The other two isolates were single locus variants of this sequence type (ST85 and ST86). All BoNT/A-producing isolates contained the same chromosomally-integrated BoNT/A2 neurotoxin gene cluster. The BoNT/B-producing isolate carried a single plasmid-borne bont/B gene cluster, encoding BoNT subtype B6. Single nucleotide polymorphism (SNP)-based typing results corresponded well with MLST, however, the extra resolution provided by whole-genome SNP comparisons showed that the isolates differed from each other by over 3,500 SNPs. WGS analyses indicated that the four infant botulism cases were caused by genomically distinct strains of *C. botulinum* that were unlikely to have originated from a common environmental source. Isolates did however cluster together, when compared with international isolates, suggesting that *C. botulinum* from environmental reservoirs throughout NSW have descended from a common ancestor. Analyses showed that the high resolution of WGS provided important phylogenetic information that would not be captured by standard seven-loci multi-locus sequence typing (MLST).
Introduction

Genomic epidemiology has provided novel insights into the genetic characteristics and phylogenetic diversity of botulinum neurotoxin (BoNT)-producing *Clostridium* species (1-8).

BoNT are responsible for causing the serious paralytic disease botulism and their potent neuro-paralytic activity make them one of the top, Tier 1, agents considered to pose a significant threat to public health if used for bioterrorism (Electronic Code of Federal Regulations – Title 42; Part 73: http://www.ecfr.gov/cgi-bin/retrieveECFR?r=PART&Tn=42y1.0.1.6.61) (9, 10). The same properties also make them a powerful tool for both medical therapeutic and cosmetic applications (11).

Botulism is a very rare disease in Australia (National Notifiable Diseases Surveillance System: NNDSS- http://www9.health.gov.au/cda/source/cda-index.cfm), with only twenty cases reported since 1991. However, a global survey found that Australia had one of the highest numbers of notified cases of infant botulism in the world (12, 13). Infant botulism results from the ingestion of *Clostridium botulinum* spores which germinate and temporarily colonise the infant’s colon and vegetative cells begin to grow and produce BoNT (14-16).

This form of the disease only occurs in infants, generally under one year old, as they have relatively low levels of gastric acid and poorly developed normal gut microflora (17). It is epidemiologically distinct from foodborne botulism which results from the ingestion of preformed BoNT from contaminated food (18).

The first clinical signs in infant botulism are usually constipation and lethargy, with progression to poor feeding, muscular weakness, abnormal eye movements and, in the absence of intervention, progressive symmetrical flaccid paralysis and respiratory arrest due to neuromuscular junction blockage (19, 20). A number of risk factors have been identified,
including the consumption of honey (21-24), herbal infusions (25), contamination of infant food such as powdered formula (26), exposure to household dust (27) and disrupted soil (20, 28). Contact with pet terrapins has also recently been identified as a risk factor for infant botulism caused by \textit{Clostridium butyricum} producing BoNT/E (29). However, sources of infection or specific risk factors have generally not been identified for either sporadic or clustered cases (30). Infant botulism has a low mortality rate but can be associated with significant morbidity. However, the timely administration of Human Botulism Immune Globulin Intravenous (BIG-IV) has been shown to reduce morbidity, as evidenced by a significant reduction in the mean length of hospital stay (19, 20, 28, 31, 32).

The diagnosis of infant botulism is confirmed by isolation of \textit{C. botulinum} from the stool, followed by confirmation of toxin production by the mouse neutralization assay (33). Emerging evidence suggests that the genotyping of strains can help to link clustered cases and identify potential shared sources of infections (34-38). Unfortunately the resolution of standard typing methods and the frequent absence of environmental samples often make it difficult to conclusively link cases or identify environmental reservoirs. The use of genome sequencing (WGS) to compare entire bacterial genomes at the single nucleotide level does however have the capacity to greatly improve clustering of bacterial pathogens (39).

There are seven well characterised BoNT serotypes (A to G) and a potential eighth type was recently reported (H) (40, 41) but still to be confirmed. The neurotoxins are produced by four genetically and physiologically distinct groups of \textit{C. botulinum} (Groups I to IV) (42, 43) or less frequently by \textit{Clostridium butyricum} or \textit{Clostridium baratii} (44-47). Different serotypes of toxin can cause neuromuscular paralysis of significantly different durations and are further classified into subtypes based on \textit{bont} gene sequence variation. The majority of human botulism cases are caused by \textit{C. botulinum} Group II, non-proteolytic strains that
produce toxin serotypes B, E or F; or by *C. botulinum* Group I, proteolytic strains that
produce toxins A, B or F. Group I strains are closely related to non-toxogenic *Clostridium*
*sporogenes* isolates (7, 37, 48). Some strains are bivalent and express more than one type of
BoNT (e.g. serotype BoNT/Ab), whereby the uppercase letter denotes the dominant toxin
serotype (2). Some BoNT/A strains also carry a silent serotype B toxin, with the presence of
the inactive toxin indicated in parentheses (e.g. BoNT/A(B)) (2). BoNT is produced by
neurotoxin gene complexes that are located on mobile genetic elements (MGE) which can
move horizontally between strains with diverse genetic backgrounds, making the *bont*-gene
or BoNT-complex sequences an unreliable target for phylogenetic comparisons (35, 49, 50).
The largest epidemiological studies of infant botulism have been carried out in Japan, where
31 cases have been reported since 1986 (4), and California in the US, where 978 cases have
been reported since 1976, with 20-50 cases reported each year (34). Dabritz, et al., applied
amplified fragment-length polymorphism (AFLP) analysis to genotype isolates and BoNT-
specific real-time PCR to characterise BoNT subtypes. Comparative analyses identified
several distinct clades associated with spatiotemporal clusters or clusters containing both
infant and environmental (i.e. honey) isolates, implicating common-source exposure (34).
Considerable diversity was documented in the genomic backgrounds of isolates expressing
the same toxin subtypes, indicating extensive horizontal transfer of neurotoxin gene clusters
(34). The most in-depth genomic characterisation of *C. botulinum* isolates causing infant
botulism was carried out on all isolates from reported cases of infant botulism in Japan
between 2006 and 2011 (4). Comparative genomic analysis of the ten sequenced isolates
and 13 reference sequences available in GenBank revealed that only ~33% of the *C.
botulinum* genomes represented a core-genome, reflecting the large phylogenetic distances
separating lineages of *C. botulinum* (49). Core-genome SNP analysis was shown to improve
cluster resolution, separating the largest group of serotype A(B) isolates into two distinct lineages (4).

Currently there is no information about the genomic characteristics of *C. botulinum* in Australia. Therefore, whole genome sequencing (WGS) was used to perform comparative genomic analyses on *C. botulinum* isolates from a potential cluster of four infant botulism cases that occurred in NSW within a 16-month period.
Materials and methods

Case reports and isolate characterisation. In Australia clinical cases of infant botulism are rare with only six cases notified in NSW in the last 15 years. Four of these cases were reported between September 2011 and August 2013. These four cases were documented in infants born at term after uncomplicated pregnancies. All were admitted to the intensive care units of paediatric referral hospitals in Sydney and required tracheal intubation. Infant botulism was suspected following characteristic electromyography features. Only two cases were treated with Botulism Immune Globulin Intravenous (Human) (BIG-IV) (19) and all recovered. Public health investigations were not able to identify epidemiologic links between the cases and no risk factors or potential sources of infection were identified. The clinical course of the cases are summarised in table 1, and clinical reports on cases 1 and 2 have been previously published (31).

C. botulinum was isolated from stool samples of all cases. Isolates were cultured on Botulism Selective Media and identified using API 20A (Biomerieux, Hazelwood, MO), Rapid ID 32A (Remel, Lenexa, KS) and 16S rDNA sequencing. Mouse bioassay was used to confirm the presence of C. botulinum neurotoxin and to identify the toxin serotype (33). Monovalent neutralising anti-toxins were supplied by the Centers for Disease Control and Prevention, Atlanta, GA.

DNA extraction and WGS. Genomic DNA was extracted from pure cultures using the DNeasy Blood & Tissue Kit (QIAGEN) and 200-bp fragment Libraries were prepared using the IonXpress Fragment Library Kit and IonXpress Barcode Adapters, following the protocol for 100 ng-DNA input library preparation (Life Technologies, USA). Samples were then pooled and sequenced together on an Ion 318 Chip in an Ion Torrent PGM, using the Ion PGM Template OT2 200 Kit and Ion PGM 200 Sequencing Kit (Life Technologies, USA).
Bioinformatic analyses. Sequence data were processed and analysed using the Torrent Suite 4.0.2. Sequencing reads were mapped to the reference genomes: *C. botulinum* ATCC 3502 A1 (accession: AM412317.1), *C. botulinum* A2 str. Kyoto (accession: NC_012563.1) and *C. botulinum* B1 str. Okra (accession: CP000939.1) using the Torrent Alignment plugin v4.0-r77189 and variants detected with the variantCaller plugin v4.0-r76860, using default settings: positions used to call single nucleotide polymorphisms (SNPs) had to have a minimum Phred-scaled call quality $\geq 10$, a minimum read fold-coverage $\geq 6$ and a maximum strand bias of 0.95. De novo assembly was also performed on sequencing reads using the Assembler plugin v3.4.2.0 plugin, which uses MIRA Version 3.9.9-Development to assemble contigs. Contigs were then further assembled into scaffolds using Mauve v2.3.1 (51). WebACT (http://www.webact.org/WebACT/home) was used to produce BLAST comparisons of sequenced isolates against each other and the reference genome and plasmid sequences. Seven-loci multiple locus sequence typing (MLST) (35) was performed *in silico*, by uploading de novo-assembled contig files to the PubMLST *Clostridium botulinum* query page (http://http://pubmlst.org/cbotulinum/). The MLST allele sequences obtained from NSW1_A2 were then downloaded and concatenated to create an MLST pseudomolecule (supplementary table 1). Reads from all four strains were then aligned to this pseudomolecule to confirm the allele sequences, which were then submitted to the PubMLST *Clostridium botulinum* sequence query page to confirm the allele profiles. MLST allele profiles were then compared to the list of allele profiles available for download from PubMLST and a phylogenetic tree was created using the PubMLST Treedrawing tool (http://pubmlst.org/cgi-bin/mlstanalyse/mlstanalyse.pl?site=pubmlst&page=treedraw&referer=pubmlst.or). De novo-assembled contigs were also uploaded to CSI Phylogeny 1.0a (https://cge.cbs.dtu.dk/services/CSIPhylogeny/), a webserver which identifies SNPs from
whole genome sequencing data, filters and validates the SNP positions and then infers phylogeny based on concatenated SNP profiles (52). *C. botulinum* ATCC 3502 was used as the reference strain. SNPs were excluded if they were in regions with a minimum fold coverage of <10, within 10-bp of another SNP or <15-bp from the end of a contig. The analysis also included isolates belonging to each of the five lineages of *C. botulinum* group I strains, recently described by Gonzalez-Escalona et al. (3).

BoNT-complex sequences were extracted from *de novo*-assembled contigs and compared to publically available BoNT cluster sequences using BLASTn. Phylogenetic analyses of BoNT gene sequences were performed using MEGA6. Multiple sequence alignments were performed using MUSCLE and phylogenetic analyses performed using the maximum likelihood method based on the Kimura 2-parameter model (53, 54). The accession numbers of BoNT gene sequences used to construct phylogenetic trees are listed in supplementary table 2.

**Genomic Data.** The genomic data have been deposited in the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/Traces/sra/) under the accession number: PRJNA273428.
Results and Discussion

Three of the *C. botulinum* isolates were confirmed to produce toxin A (NSW1_A2, NSW2_A2 and NSW3_A2) and one (NSW4_B6) to produce toxin B in the mouse bioassay. Sequence reads from NSW1_A2, NSW2_A2 and NSW3_A2 were mapped to reference genome sequences of *C. botulinum* ATCC 3502 A1 and *C. botulinum* A2 str. Kyoto and reads from NSW4_B6 were mapped to *C. botulinum* B1 str. Okra and *C. botulinum* A2 str. Kyoto. Read mapping and single nucleotide variant (SNV) detection metrics, supplementary table 3, showed that all four NSW isolates were highly diverse in comparison to the reference strains. All isolates were more similar to *C. botulinum* A2 str. Kyoto than their other respective reference strain, but there were still high levels of sequence divergence, with reads from BoNT/A-carrying strains covering only approximately 88-90% of the *C. botulinum* A2 str. Kyoto genome and containing over 40,000 variants and reads from the BoNT/B-carrying isolate covering approximately 89% of the genome and containing over 37,000 variants.

*In silico* MLST analysis revealed that two of the BoNT/A-carrying isolates shared the same MLST profile (*aroE*-11, *mdh*-5, *aceK*-11, *oppB*-9, *rpoB*-8, *recA*-6, *hsp*-8), which represented a new sequence type (ST) that was a 6-locus match to both ST10 and ST53 allele profiles. This profile has been submitted to the PubMLST *C. botulinum* database and designated ST84. NSW2_A2 contained a new *mdh* allele sequence which was submitted to the PubMLST *C. botulinum* database and named *mdh*-30, the resulting ST profile (*aroE*-11, *mdh*-30, *aceK*-11, *oppB*-9, *rpoB*-8, *recA*-6, *hsp*-8) was designated ST85. The BoNT/B carrying isolate also had a new ST (*aroE*-13, *mdh*-5, *aceK*-11, *oppB*-9, *rpoB*-8, *recA*-6, *hsp*-8), which was a 6-locus match to ST84 and a 5-locus match to ST10, ST8 and ST15 and ST53 (figure 1). Because all four of
Isolate relatedness was further examined by whole-genome SNP comparisons. The CSI Phylogeny webserver was used to identify SNPs separating the four NSW infant botulism isolates from one another and from selected published genome sequences available in NCBI GenBank. The SNP distance matrix comparison genomes are included in supplementary table 5. The topology of this tree closely resembled that of the MLST-based phylogenetic tree (supplementary figure 1), with the four NSW isolates clustering together. The SNP-difference matrix, calculated by CSI Phylogeny, showed that \( \textit{bont}\)/A-carrying isolates harboured over 3,500 SNP differences when compared to one another and over 8,000 SNP differences when compared to the \( \textit{bont}\)/B-carrying isolate. These numbers are not likely to reflect exact numbers of SNPs between isolates as some may belong to repetitive or poorly aligned regions that can interfere with SNP detection and there will be a certain amount of bias introduced by the use of a reference genome for variant detection; as regions found in our isolates, but not present in \textit{C. botulinum} A2 Str. Kyoto, will be not be included in the analysis. However, a second variant comparison, using CLC Genomics Workbench, confirmed that all isolates were separated from one another by at least 4,000 SNPs and that the SNPs were distributed across the entire genome (supplementary figure 2) and. This suggests that even though these strains are likely to have descended from a common clonal ancestor, they are not sufficiently related to one another to indicate that any of the cases had been infected from the same environmental source of \textit{C. botulinum} spores. All \( \textit{bont}\)/A-carrying isolates had a single chromosomally-encoded BoNT/A2 neurotoxin cluster integrated into the \( \textit{arsC}\)-operon integration site (49) (supplementary figure 3).

Multiple sequence alignments showed that the 13,786-bp sequences of all three BoNT/A2
neurotoxin clusters were identical, apart from one single nucleotide difference in strain NSW1_A2 (data not shown). The synonymous SNP was in orfX1 at nucleotide position 105 (C>T). The high degree of homology suggested that the neurotoxin gene cluster was horizontally acquired by all three isolates from a common donor source. Extensive horizontal gene transfer of toxin clusters amongst strains from diverse clonal lineages is thought to be responsible for the lack of phylogenetic correlation between toxin subtype and genomic background (38, 49, 56-58). This BoNT/A2 neurotoxin cluster lacked a 1.2-kb insertion between botR and orfX1 that is typically present in A2 toxin gene clusters. The absence of this insertion was previously described in the BoNT/A2 neurotoxin cluster of C. botulinum type A str. Mascarpone (58). Other regions of the cluster shared higher nucleotide identity levels with other diverse bont/A2 cluster gene sequences from the BLASTn database (figure 3). Supplementary table 6 shows the locations of SNPs detected when mapping reads from NSW1_A2, NSW2_A2 and NSW3_A2 to the bont cluster region of C. botulinum type A str. Mascarpone (accession number: DQ310546.1). The BoNT/B-producing isolate contained a single plasmid-encoded neurotoxin gene cluster that shared 99% identity with the bont/B6 cluster from C. botulinum B str. Osaka05 (59) (figure 3 and supplementary figure 3). The plasmid containing the bont/B cluster in NSW4_B6 (pNSW4_B6) appears to be similar in size and structure to plasmid pCB111 (accession: AB855771.1) from C. botulinum B str. 111 (60) (supplementary figures 3 and 4), which encodes a BoNT/B2 subtype toxin. bont toxin gene phylogenies showed that the identical bont/A gene sequences from the BoNT/A-producing strains fell within the A2 subtype group, which contains the neurotoxin from C. botulinum A2 str. Kyoto, which was isolated from a case of infant botulism in Kyoto, Japan in 1978 (61) (figure 4A). The bont/B gene sequence fell within the B6 subtype cluster
together with BoNT/B-encoding gene from *C. botulinum* B str. Osaka05, isolated from a case of infant botulism in Osaka, Japan in 2005 (59) and the BoNT/B-encoding gene from *C. botulinum* B str. Okayama2011 from a case of infant botulism in Okayama, Japan in 2011 (4) (figure 4B). Isolate NSW4_B6 from NSW is the first strain identified outside Japan to carry a BoNT subtype B6. There were four SNP differences between Osaka05 and Okayama2011 (4) and eight and six SNP differences between NSW4_B6 and Osaka05 and Okayama2011, respectively. Supplementary table 7 shows the locations of SNPs detected when mapping reads from NSW4_B6 to the *bont* cluster region of *C. botulinum* B str. Osaka05. Strains Osaka05 and Okayama2011 were previously shown to belong to a phylogenetically distinct *C. botulinum* group I clade, based on approx. 50% of their core-genome SNP markers not matching any of the other analysed strains (4). In figure 2, these isolates were more closely related to *C. sporogenes* PA3679 (48) than other group I *C. botulinum* isolates and were distantly removed from NSW4_B6, which clustered with the *bont*/A2-carrying NSW isolates. These observations indicated that *bont*/B6-carrying plasmids may be acquired by phylogenetically diverse *C. botulinum* isolates from different geographical areas. High levels of similarity between *bont* gene sequences of isolates from infant botulism cases in NSW and those from cases of infant botulism in Japan, give further weight to evidence indicating that different BoNT subtypes have differing toxicogenic and/or immunogenic properties that could predispose those subtypes to different disease etiologies (43, 56). *C. botulinum* spores are ubiquitous in the environment and have been isolated from soil, marine and fresh water samples from most parts of the world (13). Genomic analyses suggest that diverse populations of clonally-related non-toxigenic Clostridia are present throughout different regions of NSW. The acquisition of MGE-containing specific BoNT clusters could then predispose these isolates to causing infant botulism. However, further
analyses of environmental and both infant botulism and food-borne botulism isolates would need to be carried out to confirm this.

Conclusion

This study reports the first genomic characterisation of *C. botulism* isolates from Australia. Whole-genome SNP-based analysis revealed that isolates from four temporally-related cases of infant botulism belonged to a separate WGS clade of *C. botulinum* group I isolates that did not cluster with strains from any of five previously defined lineages (3) or core-genome SNP-based clades (1) of published isolates. By standard MLST and *bont* gene sequencing, two of the three *bont/A*-carrying isolates would have appeared identical, indicating that they could have belonged to an epidemiologically-related cluster. However, the high resolution power of WGS revealed that all strains were separated from one another by at least 3,500 SNPs. There have been no in-depth studies investigating rates of SNP accumulation over time in *C. botulinum*. However, based on estimated molecular clock rates for other bacterial pathogens the numbers of SNPs separating the four NSW isolates indicated that they had probably diverged from their most recent common ancestor at least hundreds of years ago (62). Conversely, the presence of only one SNP amongst the three BoNT/A2 cluster sequences indicated that this genomic region had been more recently acquired by all three isolates, potentially from the same donor source.

Acknowledgments

The authors thank Basel Suliman and Vishal Ahuja for assistance in bioassay testing. The study was supported by capacity development grant from the NSW Ministry of Health.
References


evidence for the presence of more than one strain of Clostridium botulinum in clinical specimens and food. J Med Microbiol 54:769-776.


Table 1. Clinical information relating to the four cases of infant botulism

<table>
<thead>
<tr>
<th>Case (isolate name)</th>
<th>Toxin</th>
<th>Background</th>
<th>Month/Year</th>
<th>Location in NSW</th>
<th>Age at presentation</th>
<th>Duration of intubation</th>
<th>Duration of ICU admission</th>
<th>Duration of Hospital admission</th>
<th>BIG-IV*</th>
</tr>
</thead>
</table>
| 1 (NSW1_A2)         | A     | • Exclusive breastfeeding  
|                     |       | • Semi-rural living  
|                     |       | • Recent renovation of residence | May 2011 | Bathurst | 10 weeks | 45 days | 53 days | 72 days | Not given |
| 2 (NSW2_A2)         | A     | • Formula feeds  
|                     |       | • Urban living | Septemb er 2011 | Sydney | 8 weeks | 7 days | 7 days | 23 days | Not given |
| 3 (NSW3_A2)         | A     | • Formula feeds  
|                     |       | • Urban living – had visited farm | March 2013 | Sydney | 19 weeks | 13 days | 23 days | 25 days | Given on day 5 of illness |
| 4 (NSW4_B6)         | B     | • Exclusive breastfeeding  
|                     |       | • Urban living | August 2013 | Sydney | 27 weeks | 14 days | 17 days | 30 days | Given on day 6 of illness |

*Human Botulism Immune Globulin Intravenous (BIG-IV)
Figure Legends

**Figure 1.** MLST-based phylogenetic tree of NSW infant botulism isolates. Positions of the new ST84 represented by NSW1_A2 and NSW3_A2 (aroE-11, mdh-5, aceK-11, oppB-8, rpoB-8, recA-6, hsp-8) the new ST85 represented by NSW3_A2 (aroE-11, mdh-30, aceK-11, oppB-9, rpoB-8, recA-6, hsp-8) and the new ST86 represented by NSW4_B6 (aroE-13, mdh-5, aceK-11, oppB-9, rpoB-8, recA-6, hsp-8) are indicated. The corresponding 7-locus allele profiles used to create the tree are shown on the right. Branch lengths indicate the linkage distance between isolates calculated PubMLST Tree drawing tool.

**Figure 2.** SNP-based phylogenetic tree of NSW infant botulism isolates and selected *C. botulinum* genome sequences from GenBank. SNP detection and phylogenetic inference were performed by the CSI Phylogeny webserver (52). Branch lengths correspond to numbers of nucleotide substitutions per site. The clustering of published isolates corresponded well with the lineage groupings 1-5 of *C. botulinum* Group I strains reported by Gonzalez-Escalona et. al. (3), which are indicated in gray. Isolates sequenced in this study and the Japanese infant botulism isolates Osaka05 and Okayama2011 did not cluster with any of these previously described lineages.

**Figure 3.** Structure of *bont* gene clusters from NSW infant botulism isolates.

A) Comparison of the *bont/A2* cluster from *C. botulinum* A2 str. Kyoto and the three *bont/A2*-carrying NSW infant botulism isolates. All isolates had identical *bont* complex sequences apart from NSW1_A2, which had a single nucleotide difference in *orfX1*. All isolates had the 1.2-kb deletion between *orfX1* and *botR* that is also present in *C. botulinum* strain Mascarpone (accession: DQ310546.1) and *C. botulinum* strain CDC41370 (accession: FJ981696.1). B) Comparison of the *bont/B* cluster from *C. botulinum* strain Osaka05 and the *bont/B*-carrying NSW infant botulism isolate. The closest BLASTn percent-identity hit for the...
whole gene cluster is indicated above the cluster schematic and the number of identical nucleotide hits to the closest BLASTn match for each gene is shown beneath, for ORFs where the closest match was not the reference \textit{bont} cluster, the corresponding strain name is given.

\textbf{Figure 4.} \textit{bont} gene phylogenetic analyses. A) Dendrogram of selected \textit{bont}/A nucleotide sequences. The \textit{bont}/A sequence from NSW1\_A2, NSW2\_A2 and NSW3\_A2 clustered with \textit{bont}/A2 gene sequences. B) Dendrogram of selected \textit{bont}/B genes, showing the NSW4\_B6 gene clustered with \textit{bont}/B6 sequences. Neurotoxin subtypes are indicated to the right of their corresponding colour-coded branches. Branch lengths represent the number of substitutions per site as indicated by the scale bars.
Figure 2

H04402_065
Okra
F_str_230613
Langeland
NSW4_B6
NSW1_A2
NSW2_A2
NSW3_A2
A2_Kyoto
NCTC_2916

Loch_Maree
Ba4_str_657
CDC_297

ATCC_3502
ATCC_19397
Hall

C. sporogenes_PA3679
Okayama2011
Osaka05

0.05
Figure 3

A

*C. botulinum* A2 Kyoto – BoNT/A2 neurotoxin gene cluster (15 kb)

NSW1_A2, NSW2_A2 and NSW3_A2 – BoNT/A2 neurotoxin gene cluster (13.8 kb) – 98% ID to *C. botulinum* A2 Kyoto

B

*C. botulinum* B6 Osaka05 – BoNT/B6 neurotoxin gene cluster (11.7 kb)

NSW4_B6 – BoNT/B6 neurotoxin gene cluster (11.7 kb) – 99% ID (11607/11704) to *C. botulinum* B Osaka05
Figure 4