Exploring genomic diversity in *Clostridium botulinum* using DNA microarrays

Brian H. Raphael

National Botulism Laboratory Preparedness Team, Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, 1600 Clifton Road, MS G-29, Atlanta, GA 30329, USA

Abstract

Rapid increases in the number of available *Clostridium botulinum* genome sequences have permitted the development of new molecular subtyping methods for this organism. Our laboratory has developed various DNA microarrays in an effort to differentiate strains based on differences in gene content. This review will focus on both high density comparative genomic hybridisation (CGH) microarrays and various focused (low density) oligonucleotide spotted microarrays. Comparison of gene content using DNA microarrays provides investigators with the ability to simultaneously differentiate unrelated strains and to identify strain variable genes. Such genes may play important roles in the pathogenesis, growth, and survival of this organism. Moreover, probes may be optimised as new genome sequences become available leading to improvements in the ability to characterise novel or unusual strains.

Keywords

genomic diversity; microarray; botulism; subtyping; comparative genomic hybridisation; CGH

1 Introduction

Foodborne botulism is considered a public health emergency and laboratory investigations are required to confirm cases and to identify and/or confirm the sources of contamination. From 2000 through 2009, the average number of US foodborne botulism cases reported to CDC was 18.5 (http://www.cdc.gov/nationalsurveillance/botulism_surveillance.html). During that period, the number of multi-case (two or more cases) outbreaks occurring annually was between 2 and 5. Outbreaks involving commercially processed foods are of high concern due to the potential for large numbers of cases spanning multiple states. In support of epidemiological investigations, laboratory testing using the mouse bioassay (CDC, 1998) can demonstrate consistency in the toxin types present in suspect food and in clinical samples of individuals with botulism. However, direct toxin detection in foods and clinical samples is not always possible due to limitations in sample size and quality. Therefore, a significant amount of effort is aimed at isolating toxin producing organisms. Isolated organisms (potentially from multiple sources) can be subjected to additional subtyping analysis to support epidemiological links. Several molecular subtyping

Fax: 404-639-4292, BRaphael@cdc.gov.

methods have been proposed including pulsed-field gel electrophoresis (PFGE) (Hielm et al., 1998; Lin and Johnson, 1995; Nevas et al., 2005), multiple locus sequence typing (MLST) (Jacobson et al., 2008b), multiple locus variable number tandem repeat analysis (VNTR) (Macdonald et al., 2008), and amplified gel electrophoresis (AFLP) (Hill et al., 2007; Keto-Timonen et al., 2005).

Botulinum neurotoxin (BoNT) producing clostridia are a complex group of organisms. *Clostridium botulinum* may produce any of the seven different known serotypes of BoNT (type A-G). Among *C. botulinums* trains, microbiological differences and 16S rRNA sequencing reveal that this species consists of four distinct groups (I to IV) (Collins and East, 1998). Groups I and II strains produce the toxin types typically associated with human botulism. More specifically, Group I strains produce toxin types A, B, and F while Group II strains produce toxin type E and as well as types B and F. It is notable that strains producing dual toxin types (eq. Ab, Ba, Bf, Af) are only found in Group I strains. Similarly, strains which harbour the genes encoding BoNT/A and BoNT/B but only express BoNT/A (designated A(B) strains) are only found within Group I.

Another level of genetic complexity of BoNT producing clostridia is observed upon examination of the neurotoxin genes encoding BoNTs. In addition to differences in the bont genes encoding various toxin serotypes, several investigators have reported the discovery of nucleotide (and amino acid) sequence variation (known as toxin subtype) within a particular serotype (Hill et al., 2007; Smith et al., 2005). The largest difference in nucleotide sequence identity (up to 25%) among toxin subtypes is found among *bont/F1–7* genes (Raphael et al., 2010a).

Genome sequencing efforts have supported the development of DNA microarrays for several pathogenic bacteria including *C. botulinum*. To date, there are 13 complete *C. botulinum* genome sequences and eight incomplete sequences available in the National Center for Biotechnology (NCBI) database (http://www.ncbi.nlm.nih.gov/genome/726). DNA microarrays have been used to analyse gene content among several Group I *C. botulinum* strains (Carter et al., 2009; Lindström et al., 2009; Raphael et al., 2008, 2010b; Sebaihia et al., 2007).

Comparison of *C. botulinum* strains isolated from outbreak investigations using DNA microarrays has shown that this method may be useful for clustering strains with epidemiological links. We reported that a low density microarray containing probes selected from the type A strain ATCC 3502 genome sequence could be used to cluster outbreak-related type A strains (Raphael et al., 2010b). More recently, we have expanded this platform for Group I *C. botulinum* strains producing other toxin types (including types A, B, and F) utilising probes selected from multiple genome sequences available in public databases. In this review, we describe that these mixed genome microarrays were effective for grouping strains isolated from various foodborne botulism outbreaks. Finally, since Groups I and II *C. botulinum* strains share little DNA homology (Lee and Riemann, 1970), we developed a separate Group II subtyping microarray based mainly on the Alaska E43 genome sequence. This microarray has identified differences in gene content among Group II strains of various serotypes (types B, E, and F).

DNA microarrays may serve an important role in the public health laboratory given the ability to optimise the platform as additional genome sequences become available. In one notable example, researchers developed a prototype microarray containing 1,600 probes representing over 100 viruses (Wang et al., 2002). A larger microarray using probes based on nearly 1,000 viral sequences was used to detect and characterise the novel coronavirus associated with the severe acute respiratory syndrome outbreak in 2003 (Wang et al., 2003).

Other advantages of microarray analysis include limited handling of the viable agent (except for DNA extraction) and relatively short analysis time (overnight hybridisation). Importantly, identification of the specific genes that differ between strains may be important for understanding how genes are acquired and lost by this organism as well as potentially identifying genetic markers associated with properties such as sporulation, toxin production, and growth. This review will focus on the development of specific types of microarrays that our laboratory has used in an effort to subtype and characterise *C. botulinum* strains isolated from botulism outbreak investigations.

2   Types of microarrays

2.1   High density CGH microarrays

Our laboratory developed a comparative genomic hybridisation (CGH) microarray featuring a total of 384,771 overlapping probes (~50–70 bp) representing the ATCC 3502 type A genome sequence in order to evaluate the level of genomic diversity among *C. botulinum* strains (Raphael et al., 2008). These CGH arrays were co-hybridised with genomic DNA from strain ATCC 3502 (labelled with Cy5) and genomic DNA from a test strain (labelled with Cy3). For these arrays, data was expressed as a $\log_2$ ratio of the fluorescence of the reference strain compared to the test strain. Hence, test strains that lack gene content present in the reference strain (ATCC 3502) have high $\log_2$ ratios in the specific location in the genome corresponding to the difference. CGH arrays were initially used to examine type A strains harbouring various toxin subtypes (A1 to A4). With the exception of A(B) strains, these arrays demonstrated that the highest levels of hybridisation was observed with test strains harbouring *bont/A1* (which is the same toxin subtype found in strain ATCC 3502).

Initially, CGH microarrays were used to examine three pairs of outbreak-related subtype A1 isolates (Table 1). Each of these pairs consisted of one strain isolated from an implicated food source and one strain isolated for stool of an individual with botulism. As shown in Figure 1, CGH arrays demonstrated that each pair of isolates could be distinguished from another pair. Moreover, the isolates within each pair were indistinguishable. Notably, the isolate pairs differed with respect to genes encoding a ferrichrome uptake system, a restriction/modification system, and a glycosyltransferase.

The *bont/A* gene is associated with other genes encoding non-toxic and/or regulatory proteins in a region referred to as the toxin gene cluster. Typically, *bont/A1* genes are located in a cluster containing *ha70*, *ha30*, *ha17*, however, some unusual *bont/A1* genes are located in a cluster containing *orfX1–3* termed HA-OrfX1+A1 strains (Jacobson et al., 2008a). Raphael et al. (2008) examined a set of HA-OrfX1+A1 strains isolated from distinct sources.

The *bont/A1* gene associated with these strains harboured five unique single nucleotide polymorphisms and the strains were indistinguishable by CGH microarray analysis.

PFGE using two different enzymes (SmaI and XhoI) was performed with the isolates shown in Table 1 and the strains examined by Raphael et al. (2008). Interestingly, the PFGE profiles of outbreak pair 1 could be distinguished from each other by one or more bands. Also using PFGE, 1 of the 5 isolates examined by Raphael et al. (2008) could be distinguished from the others. It may be hypothesised that in some cases, DNA variation may exist between isolates but such regions are not featured on the microarray.

More recently, CGH arrays were used to assess the relationship of two extensively used laboratory strains; ATCC 3502* from the CDC culture collection and UMass Hall from the University of Massachusetts culture collection (Fang et al., 2010). In this experiment, ATCC 3502* was used as the reference strain and UMASS Hall was used as the test strain. Microarray analysis demonstrated that both ATCC 3502* and UMASS Hall differ with respect to gene content featured on the microarray [representing the strain ATCC 3502 sequenced at the Sanger Center (UK)] and from each other. It is unknown whether such variation is due to differences in these strains' source of origin or if laboratory passage affects gene content. Nonetheless, CGH arrays may provide valuable information to investigators who are interested in determining if their laboratory isolate is equivalent to strains previously reported by others.

2.2 Focused type A microarray

While CGH microarrays provide a large amount of genomic data regarding a particular strain, their use involves several drawbacks including large data files, use of specialised software for extracting fluorescence data and data visualisation, and difficulties in comparing data acquired from multiple strains. Therefore, we developed a focused microarray strategy where ~50 bp oligonucleotide probes were designed to target regions whose presence varied from strain-to-strain (similar to those regions highlighted in Figure 1). Based on CGH data from ten type A strains containing different toxin subtypes, we selected 62 strain variable regions for probe design (Raphael et al., 2008). In addition, the focused microarray harboured probes to detect the neurotoxin genes and other selected genes of interest such as those found in toxin gene clusters (e.g., *ha70*, *orfX1*) and *fldB* as a marker for Group I strains (Dahlsten et al., 2008).

Unlike the CGH microarrays in which reference and test strain DNA is co-hybridised, the focused microarrays were hybridised only with test strain genomic DNA labelled with Cy5 and data was expressed as the $\log_{10}$ ratio of the signal fluorescence of a given probe compared to the background signal derived from empty spots (spotted with buffer only). For each strain examined, probes were scored as positive or negative based on the signal ratio and strains with identical probe hybridisation patters were considered to be a cluster. In a comparison of strain pairs from eight outbreaks, strains within an outbreak-pair clustered together supporting epidemiological concordance [i.e., epidemiologically related strains are presumably clonal and should cluster together (van Belkum et al., 2007)]. However, fewer probes hybridised with DNA derived from A(B) strains compared to strains containing *bont/A1* alone.

2.3 Group I subtyping microarray

In order to examine strains of additional serotypes, an expanded microarray was developed for use in subtyping serotype A, B, and F Group I strains (termed Group I subtyping microarray). The Group I subtyping microarray featured probes from the type A focused microarray, those selected from multiple genome sequences, and those targeting additional specific genes for a total of 225 probes (unpublished data).

As shown in Figure 2, we designed 146 subtyping probes targeting regions of the strain ATCC 3502 genome sequence found to be strain variable among type A, B, and F strains based on CGH microarray experiments. An additional ten non-ATCC 3502 probes were selected based on an *in silico* comparison of the type A1 *C. botulinum* strain Hall A, ATCC 19397, and ATCC 3502 genome sequence. Another *in silico* comparison including the OkraB and LangelandF strain genome sequences yielded 24 Okra-specific probes and 20 Langeland specific probes. Among a panel of Group I *C. botulinum* B strains, Lindström et al. (2009) demonstrated that variation in the presence of several arsenate resistance genes (including *arsR* but not *arsC*) was associated with the ability to grow in the presence of sodium arsenite. As a result, probes for both genes were selected for inclusion in the Group I subtyping microarray. Finally, a total of eight probes were designed to target genes found on plasmids in the genome sequences of strains ATCC 3502, Okra B, Loch Maree, and LangelandF.

A total of 55 type A, A(B) and type B *C. botulinum* isolates from 16 separate US foodborne botulism investigations were subjected to microarray analysis. Fluorescence data for each probe was calculated identically to the method used in the type A focused microarray described above. However, for these microarrays, strains were considered to constitute a cluster when the hybridisation results among strains had a Pearson's correlation coefficient 0.95. Each of the strains isolated from a specific outbreak were found to cluster with each other indicating a high level of epidemiological concordance. However, of the 16 outbreaks examined, nine outbreaks contained isolates whose hybridisation profiles were similar enough (correlation coefficient 0.95) to strains of a separate outbreak to cluster together. It is currently unknown if the isolates that cluster with isolates of separate outbreaks share a high level of genetic similarity or if the microarrays used in this study do not contain a sufficient diversity of probes to distinguish among outbreak isolates.

The Group I subtyping microarray was also used to assess a set of seven strains harbouring *bont/F5* isolated from stool and soil in Argentina (Raphael et al., 2010a). These strains produced either BoNT/F5 alone or both BoNT/A2 and BoNT/F5. Although the *bont/F5* nucleotide sequence among these strains was identical, the strains appear to be genetically diverse since only two strains formed a cluster (Table 2).

2.4 Group II subtyping microarray

Between 2001 and 2008, type E toxin was the second most common cause of foodborne botulism in the USA (http://www.cdc.gov/nationalsurveillance/botulism_surveillance.html) suggesting that subtyping of type E Group II strains may have an important role in the public health laboratory. Consistent with another report indicating the lack of gene content

conservation among Group I and II strains examined by microarray analysis (Sebaihia et al., 2007), DNA isolated from Group II strains failed to hybridise to the subtyping probes on the featured on the Group I subtyping microarray. Therefore, we sought to design a separate microarray to characterise Group II strains.

Comparative analysis of the genome sequences of Group II *C. botulinum* strains, 17B (type B) and Alaska E43 (type E) demonstrated that approximately ~80% of the genes encoded by these strains are shared (conserved genes were considered to have   70% nucleotide identity over   70% of the length of the gene). In order to evaluate the genetic diversity among this group of organisms, we designed a 500 probe microarray featuring 495 probes corresponding to approximately every 6th gene in the AlaskaE43 genome sequence. An additional five probes were designed to detect genes encoded by the plasmid (pCLL) found in strain 17B. Interestingly, we observed that three out of the four Group II type B strains examined hybridised to all five of the pCLL-specific probes (unpublished data).

The type E strains examined (n = 15) with these arrays hybridised with 90% of the probes featured on the microarray while type B (n = 4) and type F (n = 3) strains hybridised with 72% of the probes (Figure 3). Compared to the *in silico* analysis of the gene content conservation between types B and E strains, the level of probe hybridisation for types B and F strains to the type E microarray appears to be reduced. Based on our earlier work showing that   82% sequence identity of the target sequence with the oligonucleotide probe is required for hybridisation (Raphael et al., 2010b), it is likely that gene content conservation prediction based on the *in silico* analysis (70%) is less stringent than actual probe hybridisation.

3   Summary and future work

The availability of *C. botulinum* genome sequence data has permitted the development of microarrays with the ability to characterise an increasing number of strains. Microarray examination of a diverse set of strains has shown that gene content may be used to differentiate strains isolated from distinct sources with identical toxin serotypes/subtypes. Moreover, the comparison of outbreak-related strains demonstrated that microarrays can effectively cluster epidemiologically-related strains. Since this methodology can only interrogate genetic information featured on the array, additional subtyping analysis (such as PFGE) remains an important tool for further strain characterisation. However, new probes can be added to existing platforms as improved genome sequence data becomes available. Other advantages of microarray technology include limited handling of viable agents and results within 24 to 48 hours suggesting that microarrays can be used for screening and preliminary analysis especially if a large number of strains need to be examined.

Further work is required to develop new probes based on genome sequences of strains from diverse origins and with different properties (such as toxin subtype). Additionally, platforms with greater presence in public health laboratories, such as suspension microarrays, could be evaluated in order to improve the availability of similar molecular subtyping technology. Moreover, a centralised database of strain-specific gene content would be a significant asset for comparison of strains among public health, academic, and biodefense laboratories.

Notably, such a database could be more easily queried and maintained compared to whole genome sequence data. Finally, microarray analysis of *C. botulinum* has the potential to improve our understanding of the role that genome dynamics play in this organism and to identify genes encoding specific properties of interest such as toxin regulation, sporulation, and growth under specific conditions.

Acknowledgements and disclaimer

Biography

Brian H. Raphael received his PhD in Microbiology and Immunology from the University of Arizona in 2002. He currently serves as the Chief of the Botulism Public Health Research Unit within the National Botulism Laboratory Preparedness Team at CDC. His research interests involve genetic diversity of botulinum toxin producing clostridia.

References

Carter AT, Paul CJ, Mason DR, Twine SM, Alston MJ, Logan SM, Austin JW and Peck MW (2009) 'Independent evolution of neurotoxin and flagellar genetic loci in proteolytic *Clostridium botulinum*', BMC Genomics, Vol. 10, p.115. [PubMed: 19298644]

Centers for Disease Control and Prevention (CDC) (1998) Botulism in the United States (1899–1996). Handbook for Epidemiologists, Clinicians, and Laboratory Workers, US Department of Health and Human Services, CDC, Atlanta, GA.

Collins MD and East AK (1998) 'Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins', Journal of Applied Microbiology, Vol. 84, No. 1, pp.5–17. [PubMed: 15244052]

Dahlsten E, Korkeala H, Somervuo P and Lindström M (2008) 'PCR assay for differentiating between Group I (proteolytic) and Group II (nonproteolytic) strains of Clostridium botulinum', International Journal of Food Microbiology, Vol. 124, No. 1, pp.108–111. [PubMed: 18374440]

Fang P, Raphael BH, Maslanka SE, Cai S and Singh BR (2010) 'Analysis of genomic differences among Clostridium botulinumtype A1 strains', BMC Genomics, Vol. 11, p.725. [PubMed: 21182778]

Hielm S, Björkroth J, Hyytiä E and Korkeala H (1998) 'Genomic analysis of *Clostridium botulinum* group II by pulsed-field gel electrophoresis', Applied and Environmental Microbiology, Vol. 64, No. 2, pp.703–708. [PubMed: 9464411]

Hill KK, Smith TJ, Helma CH, Ticknor LO, Foley BT, Svensson RT, Brown JL, Johnson EA, Smith LA, Okinaka RT, Jackson PJ and Marks JD (2007) 'Genetic diversity among botulinum neurotoxin-producing clostridial strains', Journal of Bacteriology, Vol. 189, No. 3, pp.818–832. [PubMed: 17114256]

Jacobson MJ, Lin G, Raphael B, Andreadis J and Johnson EA (2008a) 'Analysis of neurotoxin cluster genes in *Clostridium botulinum* strains producing botulinum neurotoxin serotype A subtypes', Applied and Environmental Microbiology, Vol. 74, No. 9, pp.2778–27869. [PubMed: 18326685]

Jacobson MJ, Lin G, Whittam TS and Johnson EA (2008b) 'Phylogenetic analysis of *Clostridium botulinum* type A by multi-locus sequence typing', Microbiology, Vol. 154, No. 8, pp.2408–2415. [PubMed: 18667573]

Keto-Timonen R, Nevas M and Korkeala H (2005) 'Efficient DNA fingerprinting of *Clostridium botulinum* types A, B, E, and F by amplified fragment length polymorphism analysis', Applied and Environmental Microbiology, Vol. 71, No. 3, pp.1148–1154. [PubMed: 15746312]

Lee WH and Riemann H (1970) 'The genetic relatedness of proteolytic *Clostridium botulinum* strains', Journal of General Microbiology, Vol. 64, No. 1, pp.85–90. [PubMed: 4930571]

Lin W and Johnson EA (1995) 'Genome analysis of *Clostridium botulinum* type A by pulsed-field gel electrophoresis', Applied and Environmental Microbiology, Vol. 61, No. 12, pp.4441–4447. [PubMed: 8534108]

Lindström M, Hinderink K, Somervuo P, Kiviniemi K, Nevas M, Chen Y, Auvinen P, Carter AT, Mason DR, Peck MW and Korkeala H (2009) 'Comparative genomic hybridization analysis of two predominant Nordic group I (proteolytic) Clostridium botulinum type B clusters', Applied and Environmental Microbiology, Vol. 75, No. 9, pp.2643–2651. [PubMed: 19270141]

Macdonald TE, Helma CH, Ticknor LO, Jackson PJ, Okinaka RT, Smith LA, Smith TJ and Hill KK (2008) 'Differentiation of *Clostridium botulinum* serotype A strains using multiple-locus variable number tandem repeat analysis', Applied and Environmental Microbiology, Vol. 74, No. 3, pp.875–882. [PubMed: 18083878]

Nevas M, Lindstrom M, Hielm S, Bjorkroth KJ, Peck MW and Korkeala H (2005) 'Diversity of proteolytic *Clostridium botulinum* strains, determined by a pulsed-field gel electrophoresis approach', Applied and Environmental Microbiology, Vol. 71, No. 3, pp.1311–1317. [PubMed: 15746333]

Raphael BH, Choudoir MJ, Lúquez C, Fernández R and Maslanka SE (2010a) 'Sequence diversity of genes encoding botulinum neurotoxin type F', Applied and Environmental Microbiology, Vol. 76, No. 14, pp.4805–4812. [PubMed: 20511432]

Raphael BH, Joseph LA, McCroskey LM, Lúquez C and Maslanka SE (2010b) 'Detection and differentiation of *Clostridium botulinum* type A strains using a focused DNA microarray', Molecular and Cellular Probes, Vol. 24, No. 3, pp.146–153. [PubMed: 20056143]

Raphael BH, Lúquez C, McCroskey LM, Joseph LA, Jacobson MJ, Johnson EA, Maslanka SE and Andreadis JD (2008) 'Genetic homogeneity of *Clostridium botulinum* type A1 strains with unique toxin gene clusters', Applied and Environmental Microbiology, Vol. 74, No. 14, pp.4390–4397. [PubMed: 18502928]

Sebaihia M, Peck MW, Minton NP, Thomson NR, Holden MT, Mitchell WJ, Carter AT, Bentley SD, Mason DR, Crossman L, Paul CJ, Ivens A, Wells-Bennik MH, Davis IJ, Cerdeño-Tárraga AM, Churcher C, Quail MA, Chillingworth T, Feltwell T, Fraser A, Goodhead I, Hance Z, Jagels K, Larke N, Maddison M, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, White B, Whithead S and Parkhill J (2007) 'Genome sequence of a proteolytic (Group I) Clostridium botulinum strain Hall A and comparative analysis of the clostridial genomes', Genome Research, Vol. 17, No. 7, pp.1082–1092. [PubMed: 17519437]

Smith TJ, Lou J, Geren IN, Forsyth CM, Tsai R, Laporte SL, Tepp WH, Bradshaw M, Johnson EA, Smith LA and Marks JD (2005) 'Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization', Infection and Immunity, Vol. 73, No. 9, pp.5450–5457. [PubMed: 16113261]

van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S and Struelens MJ (2007) 'Guidelines for the validation and application of typing methods for use in bacterial epidemiology', Clinical Microbiology and Infectious Diseases, Vol. 13, Suppl. 3, pp.1–46.

Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D and DeRisi JL (2002) 'Microarray-based detection and genotyping of viral pathogens', Proceedings of the National Academy of Sciences USA, Vol. 99, No. 24, pp.15687–15692.

Wang D, Urisman A, Yu-Tsueng L, Springer M, Ksiazek TG, Erdman DD, Mardis ER, Hickenbotham M, Magrini V, Eldred J, Latreille P, Wilson RK, Ganem D and DeRisi JL (2003) 'Viral discovery and sequence recovery using DNA microarrys', PLoS Biology, Vol. 1, No. 2, p.e2. [PubMed: 14624234]
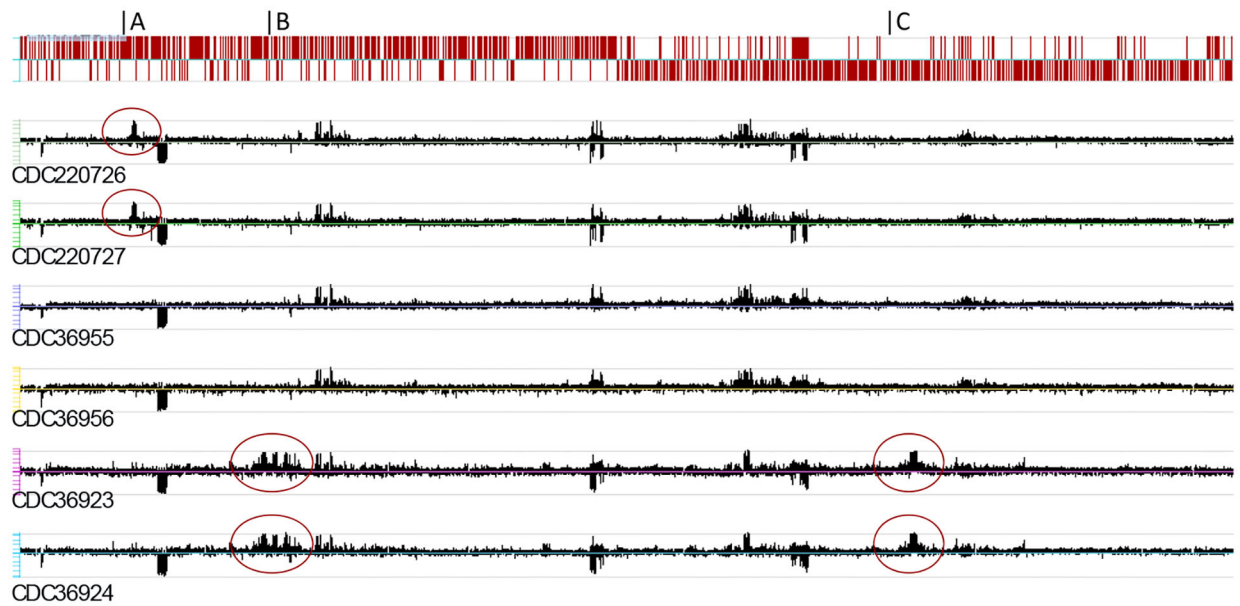
Figure 1.

CGH results of outbreak-related type A1 strains

Notes: The top track depicts each coding region sequence (CDS) in the *C. botulinum* strain ATCC 3502 genome sequence. Subsequent tracks depict $\log_2$ signal ratios of the reference strain (ATCC3502) compared to each test strain (indicated below each track). Regions whose presence varies among outbreak-pairs are circled and their genomic location relative to ATCC 3502 genome is depicted by a letter above the top track. Region A corresponds to a ferrichrome uptake system, region B corresponds to a restriction/modification system, and region C corresponds to a glycosyltransferase.
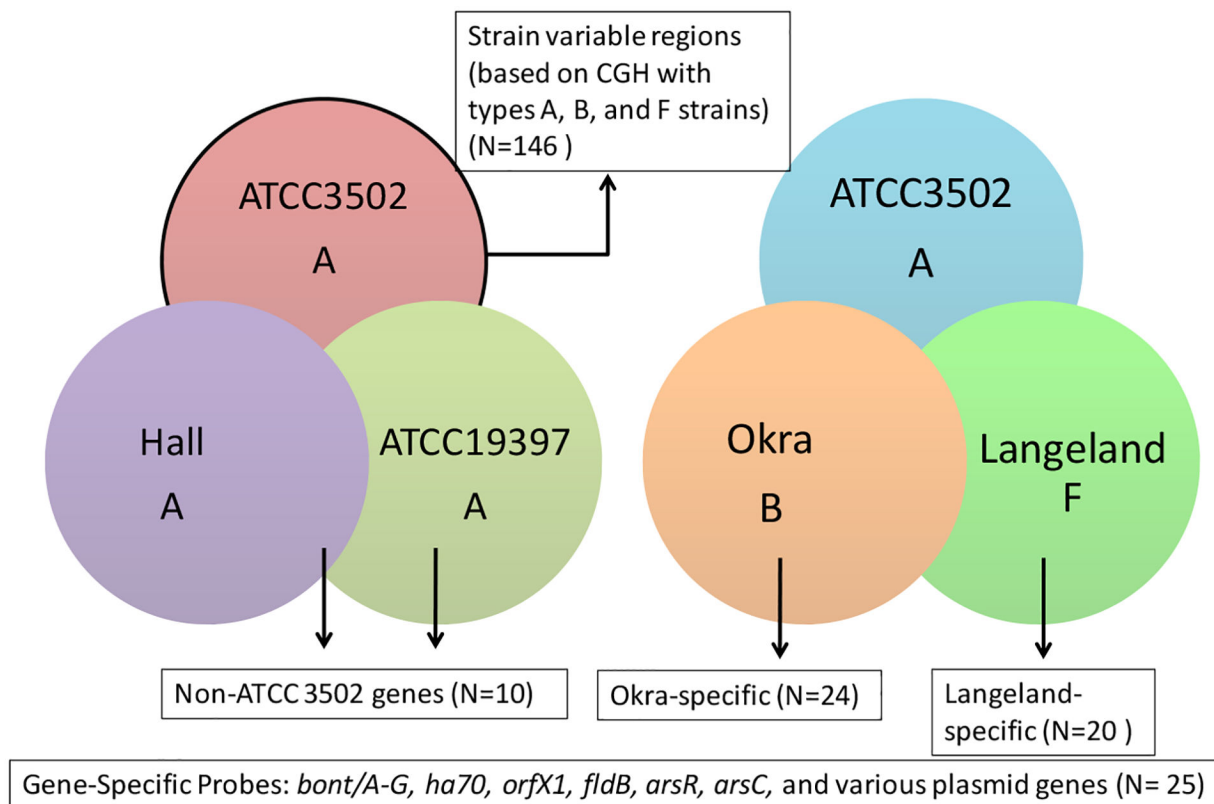
Figure 2.
Scheme used to select probes for the Group I subtyping microarray

Notes: A total of 225 probes were selected. 146 probes corresponding to strain variable regions of the ATCC 3502 genome sequence were selected based on CGH analysis using Group I types A, B, and F test strains. *In silico* analysis was used to select the remaining probes representing genes not present in ATCC 3502, specific for either OkraB or LanglelandF strains, or other specific genes of interest.
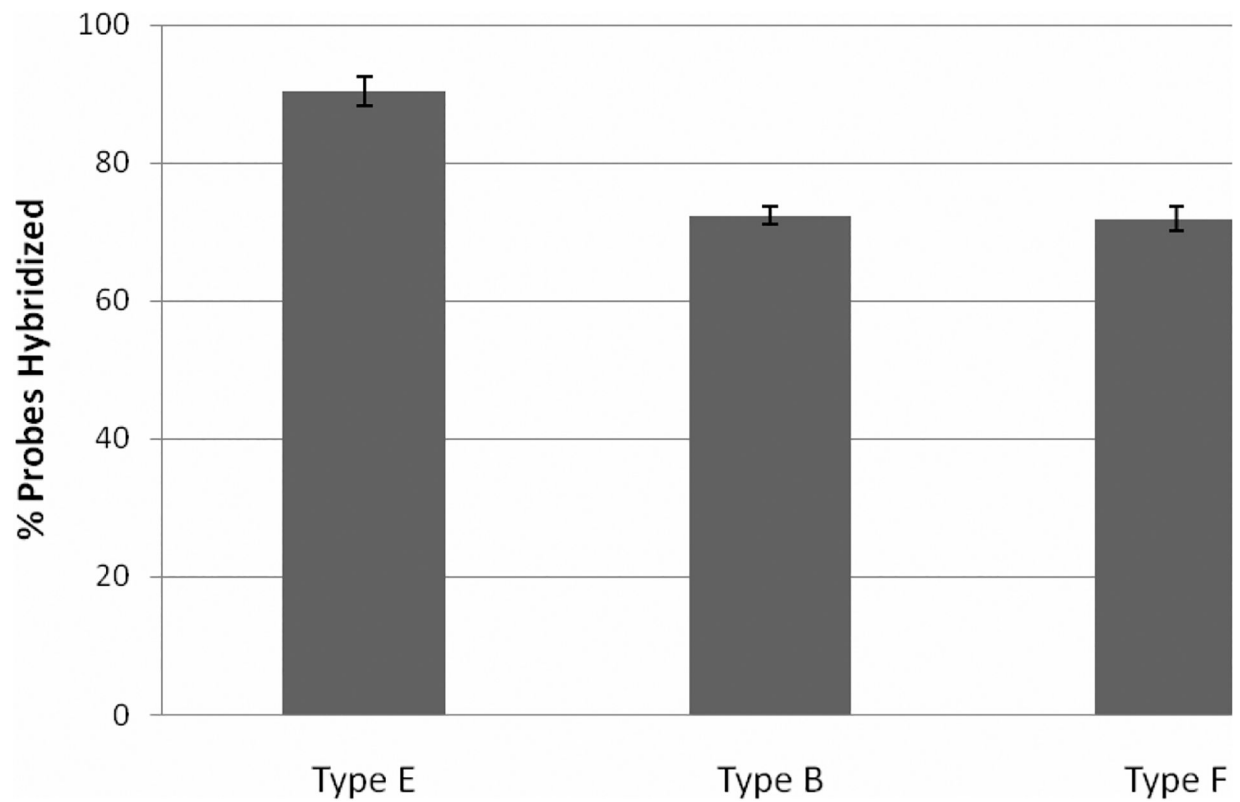
Figure 3.

Hybridisation of the Group II subtyping microarray by type E (n = 15), type B (n = 4) and type F (n = 3) *C. botulinum* strains

Notes: Shown is the average % of probes hybridised (i.e., where $\log_{10}$ of the ratio of probe fluorescent compared to background   1.0) for each toxin serotype indicated. Error bars depict standard deviations.

Table 1

Outbreak-related *C. botulinum* type A strains examined using CGH microarrays

Outbreak pair	Isolate	State	Source
I	CDC220726	ID	Stool
	CDC220727	ID	Home canned peppers
II	CDC36923	ID	Stool
	CDC36924	ID	Home canned trout
III	CDC36955	TX	Infant stool
	CDC36956	TX	Honey

Table 2

Matrix of correlation coefficients of the hybridisation results for *bont/F5* encoding strains using the Group I subtyping microarray (see online version for colours)

Source	Soil	Soil	Soil	Stool	Soil	Soil	Stool
Year	1978	1978	1987	1981	1978	1978	1982
Type	F	F	F	F	Af	Af	Af
Strain	*CDC54074*	*CDC54075*	*CDC54085*	*CDC54090*	*CDC54079*	*CDC54084*	*CDC54096*
CDC54074	-						
CDC54075	0.87993	-					
CDC54085	0.46607	0.50220	-				
CDC54090	0.87797	0.96182	0.49452	-			
CDC54079	0.72697	0.78066	0.54211	0.78138	-		
CDC54084	0.48753	0.50817	0.74646	0.50907	0.62223	-	
CDC54096	0.53524	0.55109	0.77836	0.55955	0.64965	0.68921	-

Note: Grey highlight indicates a correlation coefficient ≥ 0.95.