CbMT Sequence Typing for Identification and Tracking of Foodborne *Clostridium botulinum* Outbreaks

Tom Edlind, MicrobiType LLC, Plymouth Meeting, PA

MicrobiType
Microbial Typing Services
(www.microbitype.com)

ABSTRACT

Introduction: *Clostridium botulinum* sporadically contaminates various foods, where it may propagate anaerobically and produce a potent neurotoxin. To identify outbreaks and track down their source, rapid epidemiological analysis is required, which in turn reqires strain typing. Multiple methods have been developed for *C. botulinum* typing, including multilocus variable number of tandem repeats analysis (MLVA) and, most recently, whole genome sequencing. To varying extents these methods have limitations relating to strain resolution, turnaround time, data portability, technical complexity, and cost. Furthermore, these methods typically depend on pure cultures and are hence incompatible with crude samples such as foods and stool; this is problematic since clinical laboratories increasingly rely on culture-independent diagnostic tests.

Purpose: A promising alternative is polymorphic locus sequence typing (PLST), which analyzes through conventional, inexpensive, and technically robust PCR and Sanger sequencing the one or two most phylogenetically informative loci within the genome of a bacterial species.

Methods: To extend PLST to *C. botulinum*, genome sequences of representative strains were bioinformatically analyzed for tandem repeat regions that combined extensive polymorphism (insertions/deletions and single nucleotide variants) with presence in all or nearly all strains.

Results: The most promising loci, CbMT1 and CbMT2, include TCTATAC and AGTTCT repeats within intergenic and membrane protein-coding regions, respectively. Both repeats were previously reported to be the most informative loci in MLVA studies, yielding diversity indexes (DI) of 0.95 and 0.92 based on length alone. CbMT1 and CbMT2 loci from the ca. 120 *C. botulinum* strains represented in GenBank databases were downloaded and phylogenetically analyzed. This resolved 69 and 87 distinct alleles, respectively, with several clusters representing strains known or likely to be epidemiologically related. DIs following removal of epidemiological replicates were 0.97 and >0.99.

Significance: CbMT typing, facilitated by safe and convenient outsourcing, represents a promising new tool for *C. botulinum* epidemiology.

METHODS

Bioinformatics. Tandem repeats were identified in the genome of strain ATCC 3502 using Tandem Repeat Database (https://tandem.bu.edu). Repeat regions plus 500 nuc flanks were used as queries in BLASTN searches of GenBank Nucleotide/nr database. Loci that combined high strain resolution with presence in all or nearly all strains and lack of interuption (e.g., by transposable elements) were further evaluated by BLASTN searches of the Refseq genomes database, Clustal-Omega sequence alignments, and DNA parsimony analysis (http://evolution.genetics.washington.edu/phylip.html). For selected loci, primers were designed for optimal specificity by exhaustive analyses of database sequences.

PLST. Genomic DNA from *C. botulinum* strain Walls8G (VPI 4404) was obtained from BEI Resources (item NR-2713). DNA (0.3 μl) was used as PCR template in 20 μl reactions with locus-specific primers (proprietary sequences) and Taq polymerase (New England BioLabs) as recommended by the manufacturer. Amplification was for 30 cycles of 94ºC 20 sec, 55ºC 1 min, and 70ºC 1 min, followed by 70ºC 3 min. Aliquots were checked by agarose gel electrophoresis, and submitted for dideoxynucleotide sequencing (Genewiz) following ExoSAP-IT (Affymetrix) treatment and addition of sequencing primer. Sequences were edited as needed by inspection of chromatograms, and trimmed to common termini for clustal alignment and dnapars analysis.

REFERENCES

Edlind T, Brewster J, Paoli G (2016). Enrichment, amplification, and sequence-based typing of *Salmonella enterica* and other foodborne pathogens. J Food Prot 80:15-24.

Edlind T, Liu Y (2015). Development and evaluation of a commercial sequence-based strain typing service for *Listeria monocytogenes*. J Microb Biochem Technol 7:351-362.

Fillo S, Giordani F, Anniballi F, Gorgé O, Ramisse V, Vergnaud G et al. (2011). *Clostridium botulinum* group I strain genotyping by 15-locus multilocus variable-number tandem-repeat analysis. J Clin Microbiol 49(:4252-63.

Raphael BH, Shirey TB, Luquez C, Maslanka SE (2014). Distinguishing highly-related outbreak-associated *Clostridium botulinum* type A(B) strains. BMC Microbiology 14:192.

Williamson CHD, Sahl JW, Smith TJ, Xie G, Foley BT, Smith LA, Fernandez RA et al. (2016). Comparative genomic analyses reveal broad diversity in botulinum-toxin-producing *Clostridia*. BMC Genomics 17:180.

CORRESPONDENCE

Mail: MicrobiType LLC, 5110 Campus Drive, Plymouth Meeting, PA 19462.
Email: tom@microbitype.com. Phone: 215-850-2975. Website: www.microbitype.com.

RESULTS

Fig. 1. Typing methods commonly employed by the CDC and public health labs (www.cdc.gov/pulsenet).

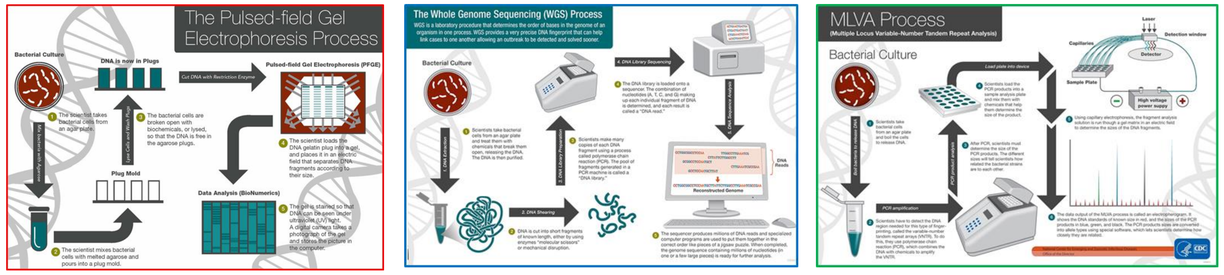

Fig. 2. Diagram of PLST-CbMT method for *C. botulinum* typing, with alternative sample types.

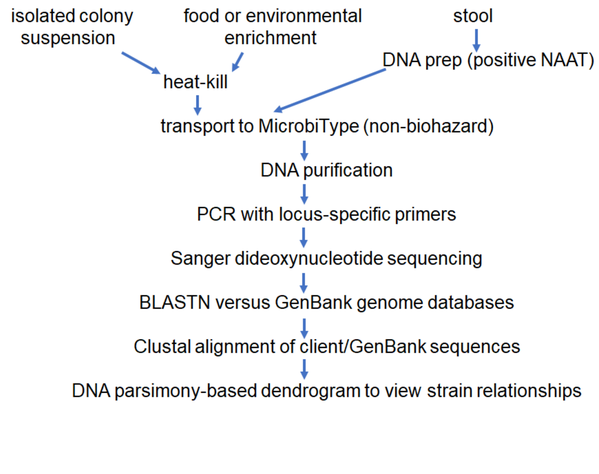

Fig. 4. Dendrogram of CbMT2 sequences from all *C. botulinum* Group I strains in GenBank databases (as of October 2017). For space reasons, replicates combined and labelled "/#". Neurotoxin serotypes (e.g., A1) roughly correlate with CbMT2 clusters; exceptions reflect horizontal gene transfer (Williamson et al., 2016).

Alleles/strains: 87/116
Simpson's DI: 0.99
(combining known or likely epidemiological replicates)

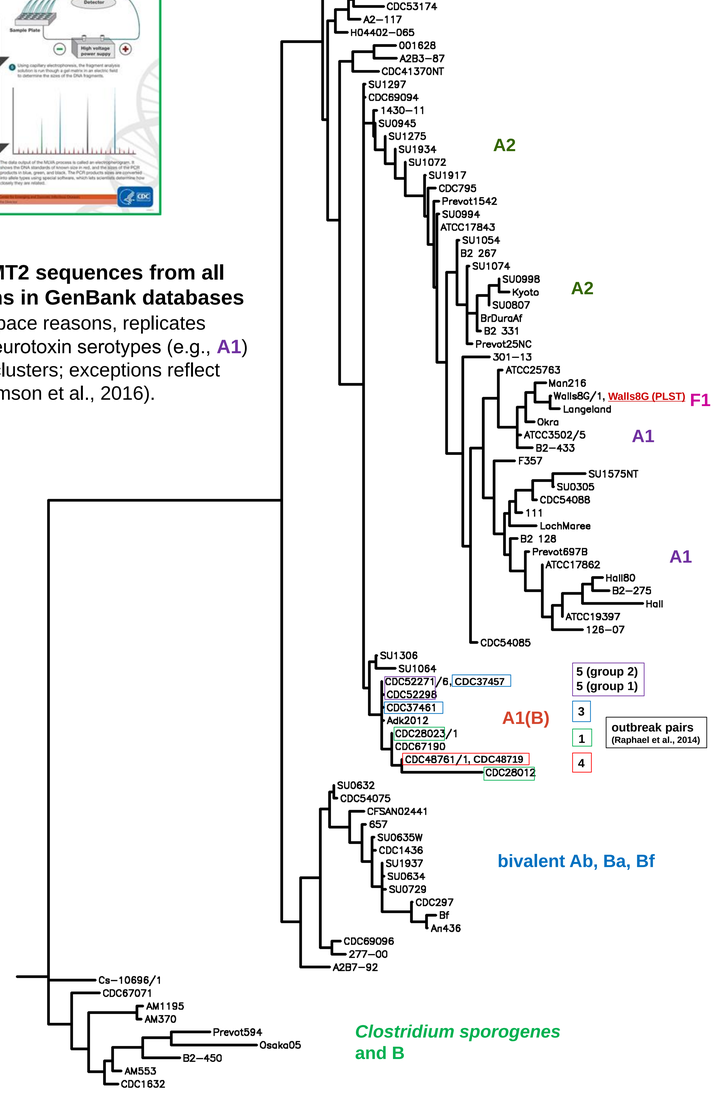

Table 1. Tandem repeat loci in *C. botulinum* strain ATCC 3502 that represent the most promising PLST targets, based on BLASTN analysis of GenBank Nucleotide/nr database.

PLST locus	repeat location[a]	length	number	sequence	alleles/strains[b]	MLVA (DI)[c]
CbMT1	435456-435531	7	8	TCTATAC	30/35	cbms03/VNTR2 (0.95)
CbMT2	589115-589195	6	13	AGTTCT	35/36	cbms04/VNTR3 (0.92)
CbMT3	8760-8858	15	6	GAAGAAAATTTAAAT	31/36	cbms01/VNTR1 (0.87)

[a] Location of the repeat region in *C. botulinum* strain ATCC 3502 (GenBank accession AM412317).
[b] Number of distinct alleles among the *C. botulinum* Group I strains represented in GenBank Nucleotide (nr) database (as of February 2018).
[c] MLVA data from Fillo et al., 2011. DI, diversity index

Fig. 3. Sequence alignment of CbMT2 loci from representative *C. botulinum* strains, illustrating polymorphism in the form of both insertions/deletions of the tandem repeat and single nucleotide polymorphisms (SNPs).

```
CDC67071    tcttatatagattatgcagattttaaagaagaaaacctagaaatattattaaattggactaatattaatataaaaaataaaactgataagaatctttagtaaattgtttatatagaaat
CDC28012    tcttatatagactatgtagatttaaaaaaagaaaacataaaaatattattaaattggattaacattaatataaaaaatgaaactaatgaagaatctttagtaaattatttctatacaaat
CDC52271/6  tcttatatagactatgtagatttaaaaaaagaaaacataaaaatattattaaattggattaacattaatataaaaaatgaaactaatgaagaatctttagtaaattatttctatacaaat
ATCC3502/5  tcttatatagactatgtagatttaaaaaaagaaaacataaaaatattattaaattggattaacattaatataaaaaataaaactgatgaagaatctttagtaaattatttctatacaaat
CDC69096    tcttatatagactatgtagattttaaaaaagaaaacataaaaatattattaaattggattaacattaatataaaaaataaaactgatgaagaatgtttagtaaattatttctatacaaat
            ********** **** ****** *** ******** ** ****************** *** *************** ***** ** ****** ********** *** **** ****

CDC67071    aatttactaaaatatgatgaaaaaaatactaaaattaaagatacttctagttctagttct----------agttctagttctagttctagttctagttctagttctagttct
CDC28012    aatttactaaaaaatcatgaaaaaa-tggtgaaattaaaactagttct----------acttctagttctagttctagttctagttctagttctagttct
CDC52271/6  aatttactaaaaaatcatgaaaaaaatggtgaaattaaaactagttctactttctagttc----------tagttctagttctagttctagttct
ATCC3502/5  agtttactaaaaaatcatgaaaaaaatggtgaaattaaaactagttctagttctagttctagttctagttctagttctagttctagttctagttctagttctagttct
CDC69096    aatttattaaaaaatcatgggaaaaatggtgaaattaaaactagttctagttctag----------ttctagttctagttctagttct
            * **** ***** ** ***  **** *  * ********  ** ****                                      **********************

CDC67071    aaagattctagcactaagctctcaaatgaacaacctgatataaaaattaaaaaggctgattctatttatctacaaaatatccataaagaatttaataataatgttattaatattgaacaa
CDC28012    aaagattctagcaccaaaatcacagatgaacaaccctatataaaaattaaaaaagctgattctatttatctacaaaacgcccatacagaatctaataataatttttattaatatcaaacag
CDC52271/6  aaagattctagcaccaaaatcacagatgaacaaccctatataaaaattaaaaaagctgattctatttatctacaaaacgcccatacagaatctaataataatttttattaatatcaaacag
ATCC3502/5  aaagattctagcaccaaaatcgcagatgaacaaccctatataaaaattaaaaaggttgattctatttatctacaaaacgcccatacagaatctaataataatttttattaatatcaaacag
CDC69096    aaaaattctagcaccaaaatcacagatgaacaaccctatataaaaattaaaaaggctgattctatttatctacaaaacgcccatacagaatctaataataatttttattaatattaagcag
            *** ********** **  ** ** **********  ********************* * ********************   ***** ***** ********** **********  * **
```

CONCLUSIONS

PLST represents a promising approach to epidemiological analysis of *C. botulinum* environmental, foodborne, and clinical isolates:

1. CbMT2-PLST provides excellent strain resolution that should be sufficient for routine epidemiological analysis. Additional loci can be analyzed as needed to confirm or extend CbMT2 results.

2. Crude inactivated lysates can be used as PLST templates; propagation of live cultures are not required to prepare DNA template. Colonies growing vegetatively can be safely shipped for typing following heat inactivation.

3. PLST should be compatible with any molecular-based method for *C. botulinum* detection; i.e., it can employ the same DNA template, saving time and reducing the need for further biohazardous workup.

4. Turnaround time is relatively rapid at 2-3 days.

5. Costs are minimal for consumables (~1$ for PCR reagents), equipment (conventional thermal cycler and electrophoresis apparatus), and dideoxynucleotide sequencing ($6).

6. User-friendly interpretation of results, employing standard web-based tools to discern strain relationships.

7. Fully portable sequence-based data; i.e., in contrast to length-based PFGE or MLVA data, PLST data can be confidently compared over time and between labs.