



# Arrangement of the *Clostridium baratii* F7 Toxin Gene Cluster with Identification of a $\sigma$ Factor That Recognizes the Botulinum Toxin Gene Cluster Promoters

Nir Dover<sup>1</sup>, Jason R. Barash<sup>1</sup>, Julianne N. Burke<sup>1</sup>, Karen K. Hill<sup>2</sup>, John C. Detter<sup>2</sup>, Stephen S. Arnon<sup>1\*</sup>

**1** Infant Botulism Treatment and Prevention Program, California Department of Public Health, Richmond, California, United States of America, **2** Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America

## Abstract

Botulinum neurotoxin (BoNT) is the most poisonous substances known and its eight toxin types (A to H) are distinguished by the inability of polyclonal antibodies that neutralize one toxin type to neutralize any of the other seven toxin types. Infant botulism, an intestinal toxemia orphan disease, is the most common form of human botulism in the United States. It results from swallowed spores of *Clostridium botulinum* (or rarely, neurotoxicogenic *Clostridium butyricum* or *Clostridium baratii*) that germinate and temporarily colonize the lumen of the large intestine, where, as vegetative cells, they produce botulinum toxin. Botulinum neurotoxin is encoded by the *bont* gene that is part of a toxin gene cluster that includes several accessory genes. We sequenced for the first time the complete botulinum neurotoxin gene cluster of nonproteolytic *C. baratii* type F7. Like the type E and the nonproteolytic type F6 botulinum toxin gene clusters, the *C. baratii* type F7 had an *orfX* toxin gene cluster that lacked the regulatory *botR* gene which is found in proteolytic *C. botulinum* strains and codes for an alternative  $\sigma$  factor. In the absence of *botR*, we identified a putative alternative regulatory gene located upstream of the *C. baratii* type F7 toxin gene cluster. This putative regulatory gene codes for a predicted  $\sigma$  factor that contains DNA-binding-domain homologues to the DNA-binding domains both of BotR and of other members of the TcdR-related group 5 of the  $\sigma^{70}$  family that are involved in the regulation of toxin gene expression in clostridia. We showed that this TcdR-related protein in association with RNA polymerase core enzyme specifically binds to the *C. baratii* type F7 botulinum toxin gene cluster promoters. This TcdR-related protein may therefore be involved in regulating the expression of the genes of the botulinum toxin gene cluster in neurotoxicogenic *C. baratii*.

**Citation:** Dover N, Barash JR, Burke JN, Hill KK, Detter JC, et al. (2014) Arrangement of the *Clostridium baratii* F7 Toxin Gene Cluster with Identification of a  $\sigma$  Factor That Recognizes the Botulinum Toxin Gene Cluster Promoters. PLoS ONE 9(5): e97983. doi:10.1371/journal.pone.0097983

**Editor:** Nancy E Freitag, University of Illinois at Chicago College of Medicine, United States of America

**Received:** September 18, 2012; **Accepted:** April 27, 2014; **Published:** May 22, 2014

**Copyright:** © 2014 Dover et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** These studies were supported by the California Department of Public Health. Submission of the manuscript for publication was approved by the California Department of Public Health and by Los Alamos National Laboratory. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: Stephen.Arn@cdph.ca.gov

## Introduction

Infant botulism caused by nonproteolytic neurotoxicogenic *Clostridium baratii* type F is rare and is notable for its severity and rapidity of onset [1–6]. BoNT/F is also produced by proteolytic *Clostridium botulinum* and by nonproteolytic *C. botulinum*. The type F toxins produced by these organisms have been distinguished into seven subtypes with *C. baratii* toxin categorized as subtype F7 [7].

Botulinum neurotoxin is encoded by the *bont* gene that is part of a gene cluster that includes several nontoxin accessory genes. Two main *bont* gene cluster organizations are known; The hemagglutinin (*ha*) toxin gene cluster is found in *C. botulinum* types A1, A5, B, C, D and G strains, while the *orfX* toxin gene cluster is found in *C. botulinum* types A1–A4, E, F and H, *C. butyricum* type E and *C. baratii* type F strains [8–11]. The toxin gene clusters of proteolytic *C. botulinum* (toxin types A, B, F and H) contain the regulatory gene *botR* in both the *ha* and *orfX* toxin gene clusters that codes for a  $\sigma$  factor that positively controls expression of the structural gene for botulinum toxin as well as of its accessory genes [9,12–14]. Paradoxically, the *botR* regulatory gene is not present in the toxin gene clusters of nonproteolytic *C. botulinum* type E, nonproteolytic

*C. botulinum* type F and nonproteolytic *C. butyricum* type E, yet all of these strains produce botulinum toxin [8,15,16].

The structural genes of the *orfX* toxin gene cluster are transcribed as two divergent polycistronic transcripts. One transcript encodes the *orfX1*, *orfX2* and *orfX3* genes, while the second transcript encodes the *p47*, *ninh* and *bont* genes [17,18]. A similar divergent polycistronic transcription pattern occurs in the botulinum *ha* toxin gene clusters [19–21]. The structural genes are transcribed from conserved promoter sequences located upstream of *orfX1* and *p47* that are recognized and activated by BotR [17,18,22,23].

Bacterial RNA polymerase holoenzyme is a multisubunit protein that consists of a core enzyme and a dissociable  $\sigma$  subunit that is responsible for recognizing DNA promoter sequences [24,25]. Most RNA transcription in growing bacteria begins with a primary  $\sigma$  factor ( $\sigma^{70}$  in *E. coli* and  $\sigma^A$  in Gram-positive bacteria) that associates with the RNA polymerase core enzyme. However, several alternative  $\sigma$  factors can replace the primary  $\sigma$  factor when adaptation to specific stresses or morphological development is required [25,26]. The most well-characterized DNA promoter sequence elements recognized by  $\sigma$  factors are the  $-35$  and  $-10$

elements (TTGACA and TATAAT, respectively, as the *E. coli*  $\sigma^{70}$  consensus recognition sequences), which are designated by their approximate nucleotide distance from the transcription start site.

The botulinum toxin regulatory protein BotR is a  $\sigma$  factor in the TcdR-related group 5 within the  $\sigma^{70}$  family of  $\sigma$  factors. This  $\sigma^{70}$  family group 5 includes regulators of several clostridial toxins; in *Clostridium difficile* TcdR regulates expression of the difficile toxins A and B genes (*tcdA* and *tcdB*), in *C. tetani* TetR regulates expression of the tetanus toxin gene (*telX*) and in *C. perfringens* UviA regulates expression of a gene (*bcn*) that codes for its bacteriocin BCN5 [14,25,27].

We determined by sequencing the complete organization of the botulinum toxin gene cluster of nonproteolytic neurotoxicogenic *C. baratii* type F7, in order to better understand its mechanism(s) of toxin production and to ascertain whether it might contain a *botR* gene. We found that the bacterium contained an *orfX* toxin gene cluster that lacked the regulatory *botR* gene. Unexpectedly, immediately upstream of the type F7 toxin gene cluster, we found a two-gene operon that resembled the *C. perfringens* *wiAB* operon. We further found that *C. baratii* type F7 contains BotR-recognized conserved DNA sequences that are found in all *C. botulinum* strains that carry *botR*. We showed that the *C. baratii* type F7 UviA-like protein, in complex with RNA polymerase core enzyme, specifically recognizes and binds to the *C. baratii* type F7 botulinum toxin gene cluster promoters. This UviA-like, TcdR-related  $\sigma$  factor of neurotoxicogenic *C. baratii* type F7 may participate in regulating production of its botulinum neurotoxin.

## Materials and Methods

### Ethics statement

All animal work was conducted in accord with the standards of the Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC) and was approved by the California Department of Public Health (CDPH) Institutional Animal Care and Use Committee (IACUC) under Animal Use Protocol #12-02.

### Bacterial strains and culture conditions

The neurotoxicogenic *Clostridium baratii* type F7 strain IBCA03-0045 described in this study was isolated from the feces of a California infant botulism patient [6]. Pure cultures grown from isolated single colonies were cultured in CMGS broth (0.5% yeast extract, 1.8% Criterion extract broth, 0.5% glucose and 0.2% soluble starch) and stored at  $-75^{\circ}\text{C}$  in 1% skim milk. The nontoxicogenic *C. baratii* strain IBCA08-0076 was isolated from a fecal specimen submitted for infant botulism testing and was determined to be *C. baratii* based on colony characteristics, Gram stain morphology, an API 20a biochemical profile (Biomérieux, Hazelwood, MO) and 16S rRNA sequencing (GenBank accession number JX847739). *Clostridium botulinum* type E7 strain Detroit was isolated from a Detroit, MI, foodborne botulism outbreak. *Clostridium butyricum* type E4 strain 109 was isolated from an Italian botulism patient [28]. *Clostridium botulinum* type F6 strain IBCA66-5436 was isolated from a California foodborne botulism outbreak [8,29]. Botulinum toxin types were determined using the mouse protection assay [30]. A small amount of frozen ( $-75^{\circ}\text{C}$ ) culture (containing a mixture of CMGS broth and 1% skim milk) was streaked onto 4% Egg Yolk Agar plates (50% Difco egg in brain heart infusion agar) and incubated anaerobically at  $35^{\circ}\text{C}$  for 48 hours. Individual bacterial colonies were removed from the plates, inoculated into 20 ml of pre-reduced TPGY broth (2.5% trypticase peptone, 0.25% protease peptone, 0.2% dextrose, 1% yeast extract, and 0.05% sodium thioglycollate) anaerobically for

24–48 hours at  $35^{\circ}\text{C}$  and then harvested by centrifugation at 3,450 g. The cell pellets were stored at  $-75^{\circ}\text{C}$ .

### DNA extraction

Genomic DNA for Sanger sequencing was extracted using the MagNA Pure Compact Nucleic Acid Isolation Kit I and the Bacteria Purification Protocol (Roche Applied Science) according to the manufacturer's instructions. Genomic DNA for next-generation sequencing was extracted using the Qiagen DNeasy Blood and Tissue Kit according to the protocol for Gram-positive bacteria.

### RNA extraction

The bacteria were grown as described above until early stationary phase ( $\text{OD}_{600} \sim 1.3$ ), at which point RNA was extracted with the RNeasy Protect Bacteria Mini Kit (Qiagen) that included treatment with DNase.

### Rapid Amplification of 5' Complementary DNA Ends (5' RACE)

The transcription start sites of the *bont* gene cluster mRNA transcripts were mapped with the 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0 (Invitrogen). All primers are listed in Table S1. Primers labeled GSP1 were used for first strand cDNA synthesis; primers labeled GSP2 were used for PCR amplification of the cDNA; primers labeled GSP3 were used for nested PCR amplification and primers labeled seq were used for sequencing of the nested PCR amplicon.

### DNA sequencing

A high-quality draft genome sequence of strain IBCA03-0045 was generated from standard and paired-end libraries that were sequenced on both Illumina and 454 Titanium (Roche Diagnostics) platforms. The 454 data were assembled using Newbler version 2.3, and the Illumina data were assembled with VELVET version 0.7.63. The two assembly results were integrated using parallel Phrap version SPS 4.24 and the integrated assembly was examined with Conseq. A review and BLAST search of the draft genome of *C. baratii* strain IBCA03-0045 identified a contig that contained the *bont/F7* gene within an *orfX* botulinum toxin gene cluster. PCR and sequencing primers were designed based on the draft genome sequence (Table S1). PCR was performed with Phusion Hot Start High-Fidelity DNA Polymerase (New England BioLabs, Ipswich, MA) with thermocycling conditions of  $98^{\circ}\text{C}$  for 1 minute and 32 cycles of:  $98^{\circ}\text{C}$  for 20 seconds,  $57^{\circ}\text{C}$ – $61^{\circ}\text{C}$  for 20 seconds and  $72^{\circ}\text{C}$  for 0.5 to 2.5 minutes. Overlapping PCR amplicons were sequenced using an Applied Biosystems 3730XL DNA Analyzer. Primers used for sequencing of *bont/F7* were previously reported by Raphael et al [7]. Primers for sequencing the type F7 botulinum toxin gene cluster, excluding the *bont/F7* gene, were designed during this study. Universal Primers for sequencing of *C. baratii* 16S rRNA were taken from [http://en.wikipedia.org/wiki/16S\\_ribosomal\\_RNA](http://en.wikipedia.org/wiki/16S_ribosomal_RNA). For the Sanger sequencing of the regulatory sequences (*orfXI-p47* intergenic region) of *C. botulinum* type E7 strain Detroit and *C. butyricum* type E4 strain 109 (GeneBank accession numbers KJ659889 and KJ659890, respectively) genomic DNA was amplified with primers EorfX-GSP2 and Ep47-GSP2 and sequenced with the amplification primers, as well as with primers EorfX-seq and Ep47-GSP3. All primers are listed in Table S1. Sanger DNA sequences were assembled with Sequencher software (Gene Codes, Ann Arbor, MI).

## Sequence analysis and alignments

CLUSTALW multiple alignments of nucleotide sequences, pairwise identities computation and phylogenetic analyses were done with the MEGA4 software [31]. Alignments and comparisons of amino acid sequences were done with the EMBOSS pairwise sequence alignment algorithm and BoxShade ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)).

## Protein expression and purification

The complete coding sequence of the *uviA*-like gene of *C. baratii* type F7 strain IBCA03-0045, optimized for *E. coli* expression using GeneGPS algorithm (GenBank accession number KJ659891), was synthesized (DNA2.0) and cloned into a pET-19b expression vector (Novagen) with a 10 histidine-tag at its N-terminus. The plasmid was transformed into *E. coli* strain BL21DE3 (New England BioLabs), and a single colony was picked and grown overnight in Luria-Bertani medium containing 100 µg/ml ampicillin at 37°C. The overnight culture was diluted 1:100 and isopropyl β-D-1-thiogalactopyranoside (IPTG) (1 mM) was added at an OD<sub>600</sub> of 1.2 to induce the expression of the target protein. The cells were harvested 6 hours post-IPTG induction. The cell pellet from a 500 ml culture was resuspended in 20 ml of binding buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4) and sonicated for 30 seconds on ice. The lysate was centrifuged at 10,322 g for 20 minutes to remove the supernatant. Inclusion bodies were collected and lysed in 10 ml binding buffer containing 6 M Guanidine-HCl by rocking at room temperature for 30 to 60 min until all pellets were dissolved. The lysate was centrifuged at 10,322 g for 20 min and supernatant was collected. A Ni-charged resin column (GenScript) was first equilibrated with 10 volumes binding buffer containing 6 M Guanidine-HCl and the supernatant was loaded on the column. After washing twice with 10 volumes of binding buffer, the protein was eluted with elution buffer (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4, 6 M Guanidine-HCl) and the eluant was collected. The denatured protein was exchanged into refolding buffer (Virovek) and then exchanged to the final buffer (50 mM Tris-HCl, pH 8.5, 150 mM NaCl, 2% Sarkosyl, 10 mM 2-mercaptoethanol).

## Gel mobility shift assay

Two DNA fragments of 250 bp each corresponding to positions 5909–6158 and 6054–6303 (GenBank accession number JX847735) (−166 bp to +84 bp from the *orfX1* and *p47* transcription start sites, respectively) were synthesized, labeled with biotin at the 5′ terminal and gel purified. Two additional DNA fragments of 250 bp each covering the same positions as above were synthesized but with a modification of positions 5947–6037 and 6176–6235 to a repetitive sequence of ACTG nucleotides. The modified positions correspond to the predicted botulinum toxin gene cluster promoters. *Clostridium baratii* UviA-like protein and *E. coli* RNA polymerase core enzyme (Epicentre) were purified by dialysis and dissolved in phosphate buffered saline (PBS) buffer. 1 µl of labeled DNA (0.1 nM), with or without unlabeled DNA (2 nM), was incubated for 30 minutes with either: 1 µl *C. baratii* UviA-like protein (1.5 µM), 1 µl *E. coli* RNA polymerase core enzyme (300 nM), 2 µl pre-incubated (for 30 minutes) complex of both proteins, or with no added proteins. The incubation reactions at room temperature also contained 2 µl binding buffer (Signosis), 1 µl poly d(I-C) and nuclease-free ddH<sub>2</sub>O in a total volume of 10 µl. The reactions were loaded on a 4.5% native polyacrylamide gel prepared in Tris-borate-EDTA (TBE) buffer and electrophoresed for 60 minutes. The gels were transferred to a nylon membrane and the probes were

immobilized with UV cross-linking. The signals in the blots were detected and analyzed using Signosis EMSA assay kit according to the manufacturer's instructions.

## Results

### Characterization of genes of the type F7 botulinum toxin gene cluster

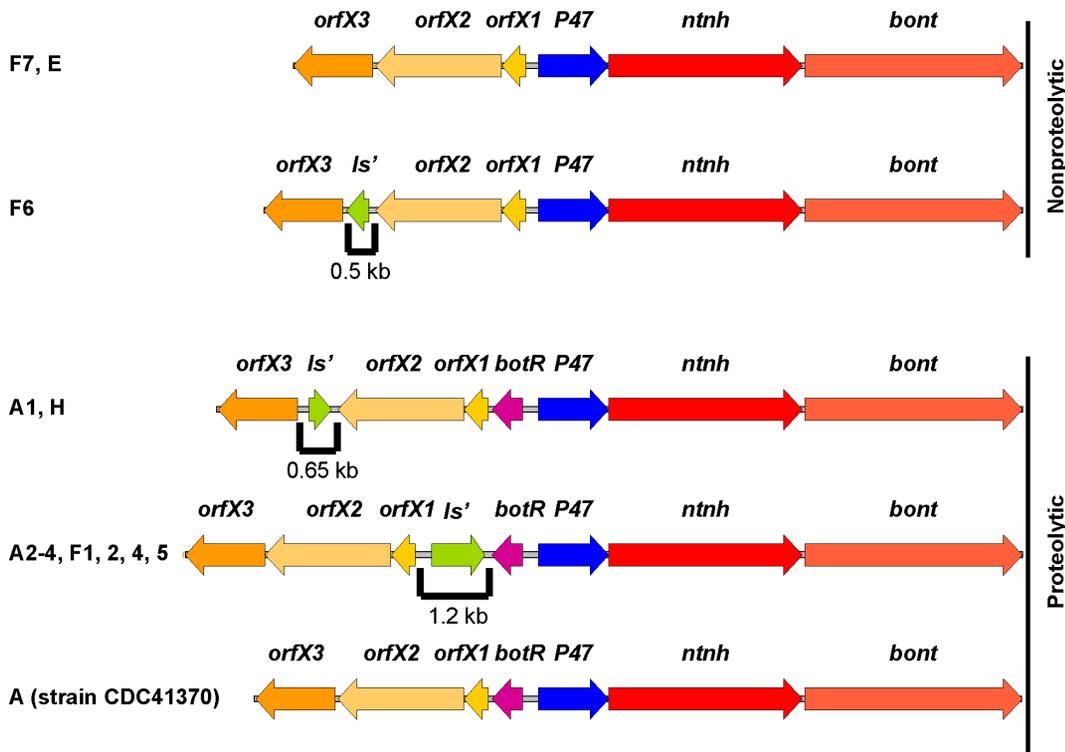
The nonproteolytic *C. baratii* *bont*/F7 gene was contained in an *orfX* botulinum toxin gene cluster (GenBank accession number JX847735) (Fig. 1), as has been found in other type F toxin gene clusters [7]. Also, the type F7 toxin gene cluster lacked a *botR* regulatory gene, as is the case with the nonproteolytic *C. botulinum* type E, nonproteolytic *C. butyricum* type E and nonproteolytic *C. botulinum* type F6 toxin gene clusters [8,16,32].

Each gene sequence was aligned (CLUSTALW) with its homologs from the *orfX* toxin gene clusters in *C. botulinum* type A, E, and F strains, and the pairwise identities were computed. Most genes of the type F7 cluster shared a low percentage identity with their *orfX* toxin gene clusters homologs in the type A, E, and F strains (Table 1). The pairwise identities of the *bont* genes ranged from 61.0% to 83.7% for nucleotides and from 42.2% to 75.3% for amino acids. The pairwise identities of the non-*bont* genes ranged from 64.2% to 85.4% for nucleotides and from 50.2% to 80.3% for amino acids. Interestingly, the genes of the type F7 toxin gene cluster generally are as similar to their homologs that reside in the proteolytic strains as they are to their homologs that reside in the nonproteolytic strains. Moreover, although both the *C. baratii* type F7 and the *C. botulinum* type F6 toxin gene clusters are nonproteolytic type F clusters that lack the *botR* gene, the *orfX3*, *orfX2*, *orfX1* and *ntnh* genes of the type F7 cluster were the least similar to their type F6 homologs (Table 1).

### Analysis of promoters within the type F7 toxin gene cluster

Because the regulatory *botR* gene is not present in the botulinum toxin gene clusters of nonproteolytic neurotoxic clostridia (Fig. 1), we characterized the promoter sequences of their polycistronic *p47-ntnh-bont* and *orfX1-orfX2-orfX3* transcripts (*Pp47* and *PorfX1* respectively) to search for promoter sequences that might be recognized by a presently unknown regulatory protein(s). The transcription start sites of the *p47-ntnh-bont* and the *orfX1-orfX2-orfX3* transcripts in *C. baratii* type F7 strain IBCA03-0045, in *C. botulinum* type F6 strain IBCA66-5436, in *C. botulinum* type E7 strain Detroit and in *C. butyricum* type E4 strain 109 were found using 5′ RACE promoter mapping, and their predicted −10 and −35 promoter elements were identified (Fig. 2). Unexpectedly, in *C. baratii* type F7 two transcription start sites upstream of *orfX1* were identified. One transcription start site, designated P1*orfX1*, was located 27 bp upstream of the transcript start codon, while the second transcription start site, designated P2*orfX1*, was located 232 bp upstream of the transcript start codon (Figs. 2A and 3). Figure 2 compares the aligned *orfX* neurotoxin gene cluster promoter sequences in neurotoxic clostridia (both proteolytic and nonproteolytic); the predicted −10 and −35 promoter elements are in bold and underlined [17,22].

Sequence analysis of the predicted promoters in all *orfX* toxin gene clusters found that the highly conserved −35 element TTTACA, which is recognized by TcdR-related regulatory proteins, was also present in the *Pp47* promoter of both proteolytic and nonproteolytic strains (Fig. 2, i). In contrast, the conserved *Pp47* −10 element GTTATA was present only in proteolytic *C. botulinum* strains and in nonproteolytic neurotoxic *C. baratii*, while the −10 element in the nonproteolytic type E and F6 toxin



**Figure 1. Gene arrangement of the botulinum *orfX* neurotoxin gene clusters in proteolytic and nonproteolytic clostridia.** The *C. baratii* F7 toxin gene cluster content and arrangement are identical to those of the type E clusters but are different from those of the F1 and F6 toxin gene clusters. Specifically, the nonproteolytic F7, F6 and E toxin gene clusters lack the *botR* regulatory gene that is present in the proteolytic A, F and H *orfX* toxin gene clusters. The nonproteolytic F6 *orfX2-orfX3* intergenic spacing is not homologous to the proteolytic A1 and H *orfX2-orfX3* intergenic spacing, although all contain a degenerated *is* element. All currently known type E clusters have an identical gene arrangement. The complete neurotoxin gene cluster sequences of type F3 is not available in GenBank and therefore could not be included in this Figure. doi:10.1371/journal.pone.0097983.g001

gene clusters consisted of a TATATT sequence. The  $-35$  element TTTACA recognized by TcdR-related regulatory proteins was also present in the *PorfX1* promoters in proteolytic *C. botulinum* strains and in the *C. baratii* F7 P1*orfX1* promoter. In contrast, the nucleotide sequence of the  $-35$  element of *PorfX1* varied in the nonproteolytic type E, F6 and F7 P2 promoters (TTGAAA, TTGTAA and TTCCAT, respectively) (Fig. 2, ii). The *PorfX1*  $-10$  promoter element GTTAAT was conserved in almost all proteolytic strains but varied in the nonproteolytic strains. Interestingly, the predicted  $-10$  promoter element TATAAT of *PorfX1* in *C. botulinum* types E and F6 and in *C. butyricum* type E is identical to the consensus promoter recognition sequence of the primary (housekeeping)  $\sigma$  factors of *E. coli* ( $\sigma^{70}$ ) and *Bacillus subtilis* ( $\sigma^A$ ) [33,34]. This unexpected finding suggests that a regulatory protein closely related to  $\sigma^{70}$  may participate in regulating *PorfX1* in the type E and type F6 toxin gene clusters. The *C. baratii* P1*orfX1*  $-10$  promoter element GTATAT was similar to the *PorfX1*  $-10$  promoter element GTTAAT in proteolytic *C. botulinum* strains, while its P2*orfX1*  $-10$  promoter element TATAAC was similar to the *PorfX1*  $-10$  promoter element TATAAT in the nonproteolytic strains (Fig. 2, ii). Interestingly, in *C. baratii* type F7 the P*p47* promoter and P2*orfX1* promoter share the same predicted  $-10$  element on complementary DNA strands (Fig. 3).

We performed a phylogenetic analysis of the P*p47* and *PorfX1* promoter sequences displayed in Figure 2A. We found that the nonproteolytic *C. baratii* F7 P1*orfX1* promoter is more closely related to the proteolytic *C. botulinum* *PorfX1* promoters in subtypes

A1–A4 and F1, while the *C. baratii* F7 P2*orfX1* promoter is more closely related to the nonproteolytic *C. botulinum* and *C. butyricum* *PorfX1* promoters (Fig. 4). The *C. baratii* F7 P*p47* promoter was more closely related to P*p47* in proteolytic clostridia (Fig. 4).

#### Identification and characterization of a TcdR-related, *Uvia*-like putative $\sigma$ factor

Because no *botR* gene exists within the *C. baratii* F7 toxin gene cluster (Fig. 1) and because the cluster contained predicted  $-35$  promoter elements identical to the sequence recognized by TcdR-related regulatory proteins (Fig. 2), we searched elsewhere for a *tcdR*-like regulatory gene. In the flanking regions of the F7 toxin gene cluster of *C. baratii* strain IBCA03-0045 we identified a putative operon of two open reading frames (ORFs) with an overlap of 20 bp that was located 550 bp upstream of *orfX3* and resembled in size and gene arrangement the *C. perfringens wiaAB* operon (Fig. 5 and [27]). The larger, *wiaA*-like ( $\sim 55\%$  nucleic acid identity) ORF consisted of 555 bp, which is similar in length both to the *botR* gene (which ranges from 537 bp to 546 bp) in proteolytic *C. botulinum* strains and to genes that encode for members of the TcdR-related family of  $\sigma$  factors [12]. The smaller ORF consisted of 222 bp and was *wiaB*-like ( $\sim 58\%$  nucleic acid identity). Sequencing of the equivalent upstream region ( $\sim 1500$  bp) of the F7 toxin gene cluster in six additional toxigenic *C. baratii* strains isolated from six different infant botulism patients found that all of them contained this *wiaAB*-like operon (data not shown).

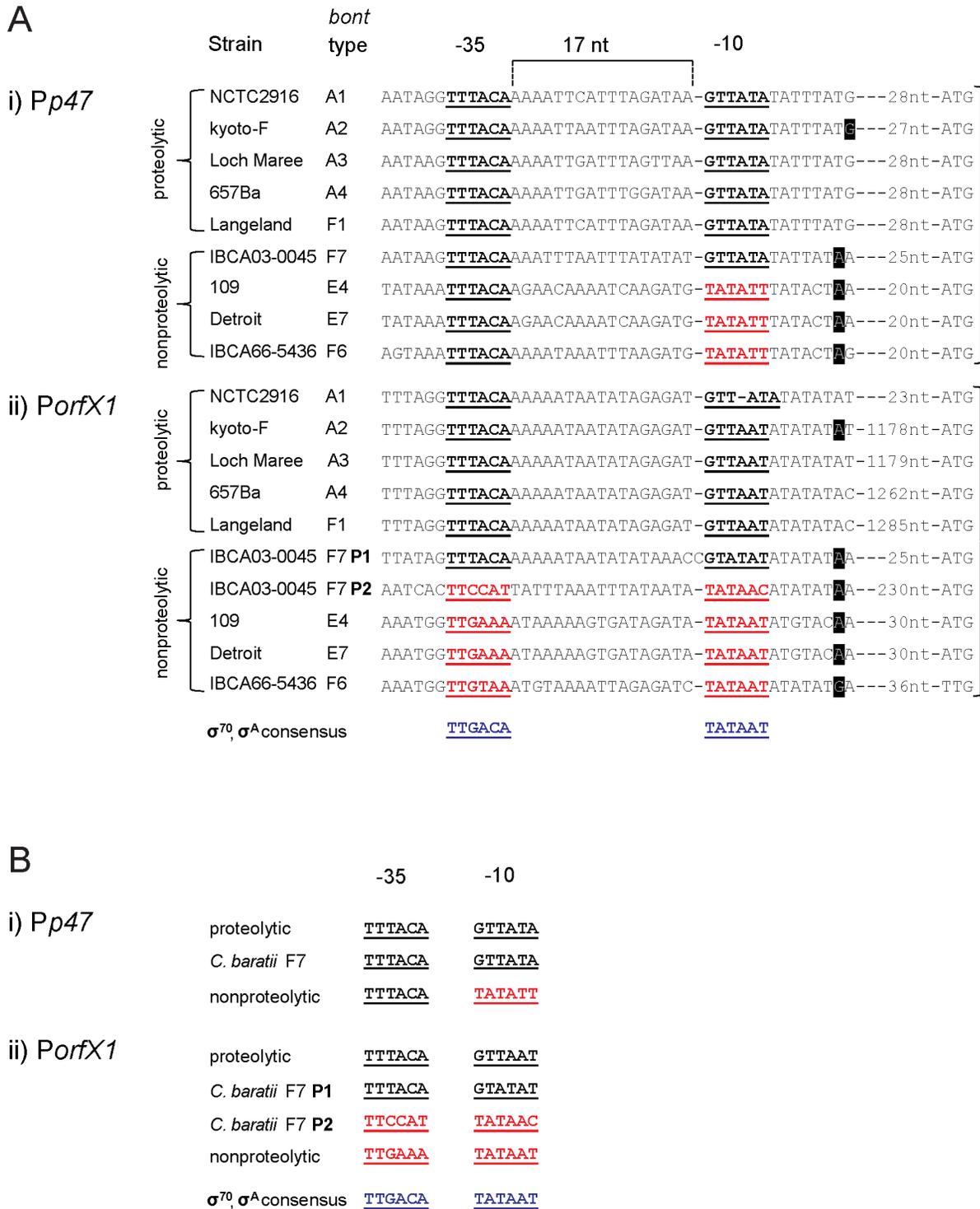
**Table 1.** Comparison of nucleotide and amino acid identities of the genes of the *C. baratii* type F7 *orfX* cluster to the genes of the *C. botulinum* types A, E, and F *orfX* clusters.

Strain	GenBank Accession no.	bont Subtype	% nucleotide identity (% amino acid identity)					
			<i>orfX3</i>	<i>orfX2</i>	<i>orfX1</i>	<i>p47</i>	<i>nrth</i>	<i>bont</i>
NCTC2916 A(B)	Y14238, AY497357, X52066	A1	82.9 (77.4)	65.8 (50.3)	77.1 (64.3)	78.8 (71.8)	84.0 (76.6)	61.8 (42.7)
Kyoto-F	CP001581	A2	82.3 (76.3)	66.1 (50.6)	78.3 (67.1)	80.0 (71.8)	84.0 (76.0)	61.1 (42.2)
CDC41370 Ab	FJ981696	A	82.3 (76.5)	65.7 (50.2)	77.1 (65.0)	80.5 (72.0)	84.0 (76.2)	61.6 (42.7)
Loch Maree	CP000963	A3	82.2 (75.6)	65.6 (50.5)	77.8 (65.7)	80.4 (72.0)	83.9 (75.9)	61.0 (42.9)
657Ba	CP001081	A4	82.2 (76.1)	64.6 (50.9)	79.2 (65.7)	84.2 (77.6)	83.3 (75.5)	61.0 (42.9)
Alaska E43	CP001078	E3	85.4 (80.3)	76.0 (65.6)	74.7 (61.4)	78.4 (67.4)	84.4 (75.6)	76.4 (65.1)
Langeland	CP000728	F1	82.8 (76.7)	66.0 (50.7)	76.1 (62.9)	78.7 (71.5)	84.0 (75.6)	83.7 (75.3)
Bf	ABDP01000023	F2	82.3 (76.0)	66.1 (50.8)	78.4 (65.2)	84.2 (77.4)	84.6 (76.6)	80.7 (70.5)
CDC54086	GU213218	F3	NA	NA	NA	NA	NA	80.8 (70.9)
AF84	AOSX01000018	F4	83.9 (77.8)	82.7 (76.1)	77.7 (65.2)	80.7 (72.7)	83.9 (73.9)	82.7 (73.7)
AF84	AOSX01000021	F5	82.3 (75.6)	66.2 (50.8)	78.4 (65.2)	80.0 (71.8)	84.7 (74.7)	76.4 (65.5)
IBCA66-5436	HQ441176	F6	79.5 (73.8)	64.2 (50.3)	74.5 (62.1)	83.3 (75.4)	79.5 (69.8)	81.5 (71.9)

NA, not available in GenBank.

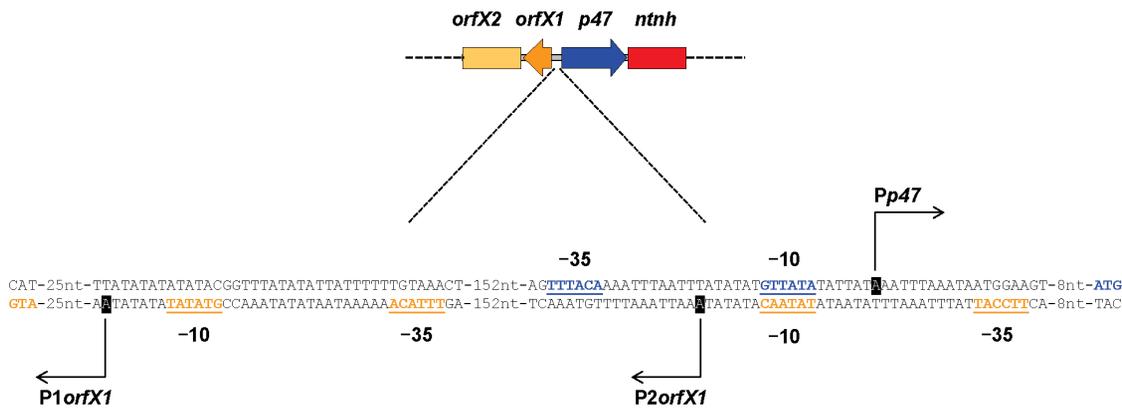
Note that even the most similar homologous genes were approximately 15% different (nucleotides) and 25% different (amino acids). Also, the most dissimilar homologous genes were approximately 60%–80% different (nucleotides) and 43%–75% different (amino acids). Only one representative of the several known *bont/E* gene clusters (*bont/E3* of strain Alaska E43) was included in Table 1 because the *bont/E* subtype gene cluster sequences are conserved. Both Alaska E43 and IBCA66-5463 are nonproteolytic (Group II), while the other strains in the Table are proteolytic (Group I).

doi:10.1371/journal.pone.0097983.t001



**Figure 2. Promoter sequences of *orfX* botulinum toxin gene clusters.** A) Aligned sequences of the (i) *Pp47* and (ii) *PorfX1* promoters in botulinum *orfX* toxin gene clusters. B) Comparison of the consensus -10 and -35 promoter elements in botulinum *orfX* toxin gene clusters. The sequences of the -10 and -35 promoter elements are bolded and underlined. The BotR-recognized conserved neurotoxin gene cluster promoter elements are bolded and in black. The unique promoter elements of *orfX* toxin gene clusters that reside in nonproteolytic strains are bolded and in red. The  $\sigma^{70}$  and  $\sigma^A$  consensus recognition sequences are bolded and in blue. Transcription start sites, where data are available, are blocked in black. Note that the *C. baratii* type F7 *Pp47* and P1 *orfX1* promoters both contain the conserved, and BotR-recognized, -10 and -35 elements (or a similar sequence). The *PorfX1* -10 and -35 promoter elements in types E and F6 are similar to the consensus sequences recognized by *E. coli*  $\sigma^{70}$  and *Bacillus subtilis* ( $\sigma^A$ ) (bottom row of Figure). F7 P1 and F7 P2 refer to the two putative promoters of the *orfX* operon in *C. baratii* type F7. The two type E strains listed in the Figure represent the two known bacterial species that express type E botulinum toxin, *C. botulinum* (E7) and neurotoxicogenic *C. butyricum* (E4).

doi:10.1371/journal.pone.0097983.g002



**Figure 3. The *orfX1*-*p47* intergenic region containing the *p47* and *orfX1* promoters in *C. baratii* type F7.** The *orfX1*-*p47* intergenic region is magnified. The  $-10$  and  $-35$  promoter elements are underlined and colored according to their gene affiliation (*p47* in blue and *orfX1* in gold). The transcription start sites are blocked in black. The translation start codons (ATG) are indicated and colored. Note that the *Pp47* and *P2orfX1* promoters share the same  $-10$  element on complementary DNA strands.  
 doi:10.1371/journal.pone.0097983.g003

Conserved Domain [35] and CDART [36] queries on the NCBI website of the predicted translated amino acid sequence of the *C. baratii* *wiaA*-like upstream ORF identified domains that matched the botulinum-toxin-associated regulatory protein BotR, as well as domains that matched sequences within the RNA polymerase  $\sigma^{70}$  subunit regions 2 and 4. The  $\sigma^{70}$  regions 2 and 4 recognize the  $-10$  and  $-35$  promoter elements, respectively [37]. Dupuy et al. reported that UviA is a TcdR-related  $\sigma$  factor that regulates the expression of bacteriocin BCN5 of *C. perfringens* [27]. The role of UviB is unknown [27].

We also identified similar *wiaA*-like genes in *wiaAB*-like operons in the genome sequences of nonproteolytic *C. botulinum* type E strain Alaska (GenBank accession number CP001078 region: 866476-867021), in nonproteolytic *C. botulinum* type B strain Eklund 17B (GenBank accession number CP001056 region: 887914-888459) and in nonproteolytic *C. botulinum* type F6 strain IBCA66-5436 (GenBank accession number JX847736). These *wiaA*-like genes are not within close proximity to the botulinum toxin gene clusters (e.g., 303,631 bp downstream of *bont/E* in *C. botulinum* type E strain Alaska). The *wiaA*-like genes of nonproteolytic *C. botulinum* types B, E and F6 were 95%–99% identical to each other and were ~50% identical to the *wiaA*-like gene in *C. baratii* type F7. However, we could not find a similar *wiaA*-like gene in the draft genomes of the nonproteolytic neurotoxicogenic *C. butyricum* type E strains BL5262 and 5521 (GenBank accession numbers NZ\_A-COM00000000 and NZ\_ABDT00000000, respectively).

Members of the TcdR-related  $\sigma$  factors share the conserved region 4.2, which is the region that interacts with the  $-35$  promoter element, while region 2.4 that interacts with the  $-10$  promoter element is more variable [12,27]. We aligned the four nonproteolytic *C. botulinum* and the nonproteolytic *C. baratii* UviA-like regions 2.4 and 4.2 predicted amino acid sequences with those of TcdR, UviA, TetR and BotR. Remarkably, all eight sequences shared the same conserved region 4.2 (Fig. 6 i and ii). All eight sequences also contained the highly conserved Serine-Arginine-Glutamine motif (SRQ) that is absent in non-TcdR-related alternative  $\sigma$  factors also belonging to the  $\sigma^{70}$  family ([12] and Fig. 6iii). As reported before, the amino acids sequence of region 2.4 varied among the TcdR-related  $\sigma$  factors [12].

### Nontoxicogenic *C. baratii*

To determine whether the *C. baratii* *wiaA*-like gene and *bont/F7* were genetically linked, DNA extracted from neurotoxicogenic *C.*

*baratii* strain IBCA03-0045 type F7 and from nontoxicogenic *C. baratii* strain IBCA08-0076 were tested by PCR for the presence of both genes. As expected, the neurotoxicogenic *C. baratii* strain was PCR-positive for both the *bont/F7* and the novel *wiaA*-like genes. In contrast, the nontoxicogenic *C. baratii* strain was PCR-negative for both genes (Fig. 7). Identical results were obtained for an additional six toxicogenic and two nontoxicogenic *C. baratii* strains (data not shown).

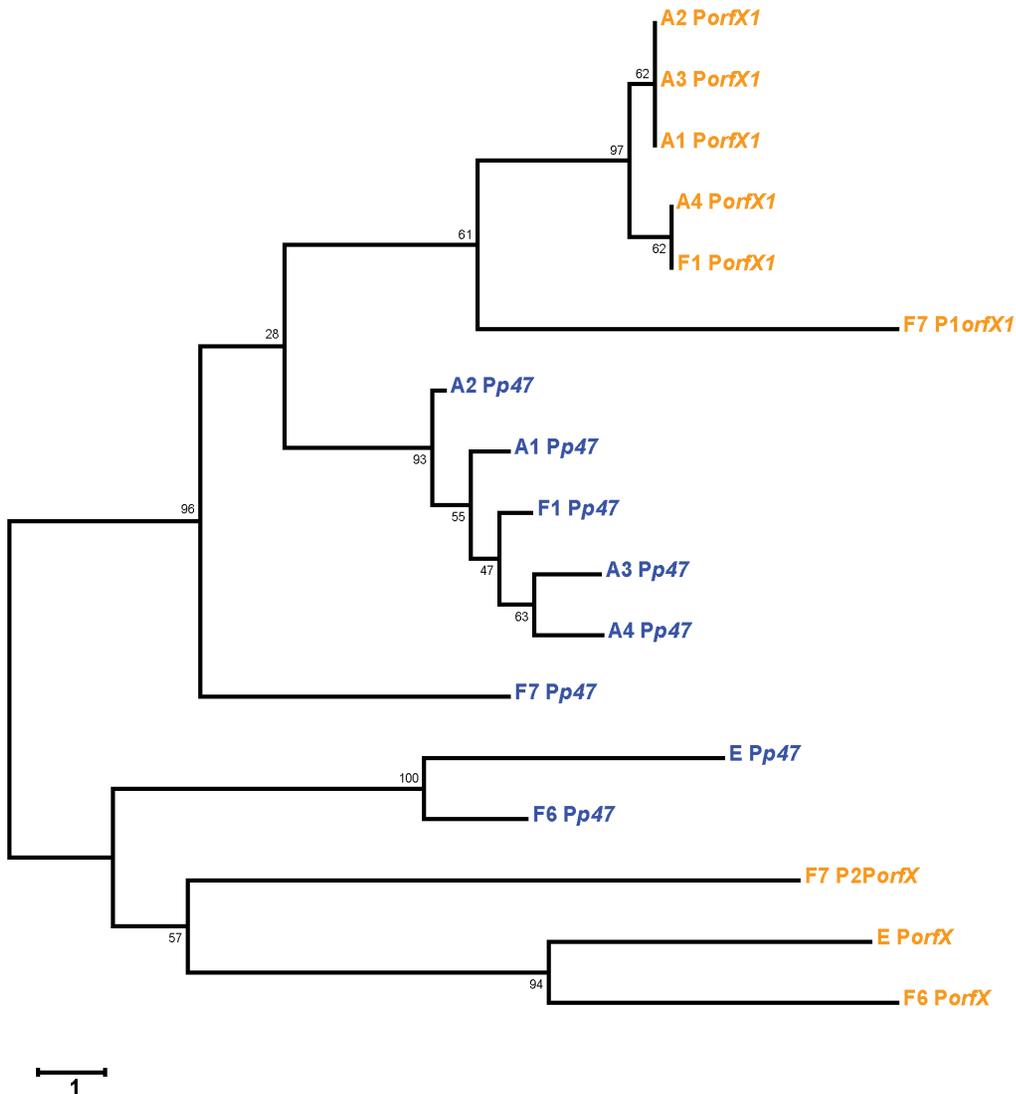
### UviA-like protein functions like a $\sigma$ factor that enables RNA polymerase core enzyme binding to the F7 botulinum toxin gene cluster promoters

We overexpressed and purified the *C. baratii* type F7 UviA-like protein in order to study its association with the botulinum toxin gene cluster promoters in gel mobility shift assays. Fig. 8 displays the electrophoretic mobility of 250 bp DNA fragments that contained the overlapping *Pp47* and *P2orfX1* promoters (Fig. 8A) or the *P1orfX1* promoter (Fig. 8B). RNA polymerase core enzyme and the purified UviA-like protein by themselves did not shift the migration of DNA fragments that contained the type F7 botulinum toxin gene cluster promoters (Fig. 8A and B, lanes 2 and 3). In contrast, following preincubation of the purified UviA-like protein and RNA polymerase core, a shift in migration was observed (Fig. 8A and B, lane 4), indicating that the UviA-like-RNA polymerase core complex bound to both promoter-containing DNA fragments. An excess of unlabeled homologous DNA fragments prevented the binding of the labeled fragments, indicating specific interactions (Fig. 8A and B, lane 5).

We repeated the gel mobility shift assays described above with similar DNA fragments that did not contain any promoter regions (Fig. 8C and D). In these DNA fragments the promoter sequences were replaced with a random nucleotide sequence. No shift in migration of either DNA fragment was observed, further establishing the specificity of the binding of the UviA-like-RNA polymerase core complex to the type F7 botulinum toxin gene cluster promoters.

### Discussion

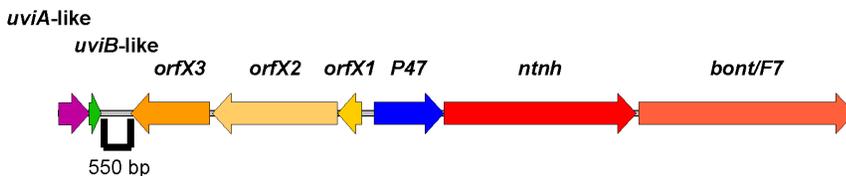
Here we report the sequence and arrangement of the entire botulinum toxin gene cluster of nonproteolytic *C. baratii* type F7. Its *orfX* toxin gene cluster does not contain a *botR* gene, similarly to nonproteolytic *C. botulinum* type F6 [8], nonproteolytic *C. botulinum*



**Figure 4. Phylogenetic analysis of the botulinum neurotoxin gene cluster promoters.** The *Pp47* and *PorfX1* promoter sequences presented in Fig. 2 (upstream of the *p47* and *orfX1* transcripts, respectively) were aligned and a phylogenetic tree was drawn to scale using the Neighbor-Joining method, with branch lengths in the same units as the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). Phylogenetic analyses were conducted in MEGA4 software. Bootstrap values and number of base differences per site are shown. Note that the F7 *P1orfX1* promoter sequence is phylogenetically more related to its type A and F1 homologs, while the F7 *P2orfX1* promoter is more related to its type E and F6 homologues. This finding implies that the F7 *orfX1* transcript may be regulated by a BotR-like protein through its P1 promoter and by a type E or type F6-like regulator through its P2 promoter. doi:10.1371/journal.pone.0097983.g004

type E and nonproteolytic *C. butyricum* type E toxin gene clusters (Fig. 1) [16]. We unexpectedly found that some of the *C. baratii* type F7 toxin gene cluster promoters are similar to the conserved BotR-recognized promoters (Figs. 2 and 4) [14]. We identified a

gene coding for a putative UviA-like regulatory protein immediately upstream of the *C. baratii* type F7 toxin gene cluster that shared a similar DNA-binding-domain with BotR and the TcdR-related group 5 of the  $\sigma^{70}$  family (Fig. 5 and 6) [12,27]. The



**Figure 5. Arrangement of the *C. baratii* type F7 neurotoxin gene cluster and its associated *uviAB*-like operon.** The type F7 toxin gene cluster arrangement with its upstream *uviAB*-like operon are shown. doi:10.1371/journal.pone.0097983.g005

Strain	Regulator	Region 2.4	Region 4.2	Accession	Regulated proteins
i) <i>C. perfringens</i>	UviA	RSLLNSRKDYINRNKIKY	LNDKSEVEIAEIMGTSRQWLNPIRNTALKKLKE	BAD90626	Bacteriocin
<i>C. difficile</i>	TcdR	RSLLNKTFEFSRRYSRMEF	LKGYSDRETSVKLISRQAVNKARNRAFKKTKK	ABI93806	Difficile toxins
<i>C. tetani</i>	TetR	RSLKRYCDDCNKKNRDEK	FEDYKECQIAKLLHMSRQAVYKKNVLAALKLEP	AAO37455	Tetanus toxin
<i>C. botulinum</i> type A1	BotR/A	RCLKRYCSTCMKKNRDEK	RKDYKDEIAKLLISRQSVYKKNVLAALKLEP	EDT82570	Botulinum toxin
ii) <i>C. baratii</i> type F7	TcdR - related proteins	RSLINRAIDAIENKKEEA	KYQESIAIADKDGISRQFVNKIENNALIKLKK	AGR53833	Botulinum toxins?
<i>C. botulinum</i> type B4		RCLNKSINHYKINSYEN	YLQSDVEIAERLISRQAVNKSRRQALRGLKN	ACD23823	
<i>C. botulinum</i> type F6		RCLNKSINHYKINSYEN	YLQSDVEIAERLISRQAVNKSRRQALRGLKN	AGR53843	
<i>C. botulinum</i> type E3		RCLNKSINYYKINSYEN	YLQSDVEIAERLISRQAVNKSRRQALEBLKN	ACD51683	
iii) <i>C. botulinum</i> type E3	Sigma D	IRIKGAIIDQIRNRPTTK	YEKTLKREISILSVSESRCQLHARSHSNLRE	ACD52641	motility proteins
<i>C. botulinum</i> type E3	Sigma E	RCLPNEIMYLRENSKIFA	TREKIQKEVAMLCISQSYISLEKKILRRLKK	ACD51156	sporulation stage II
<i>C. botulinum</i> type E3	Sigma G	PMIIGEIRRYLRDMNSLIV	IRGRITQMEVAEELCISQAQVSELEKTLKHMFK	ACD51817	sporulation stage III
<i>C. botulinum</i> type E3	Sigma H	LCVTRQITAIKTATROFH	LDGKSYQEIACDLDRHSKSDNALQRVKREKLEK	ACD51421	sporulation stage 0

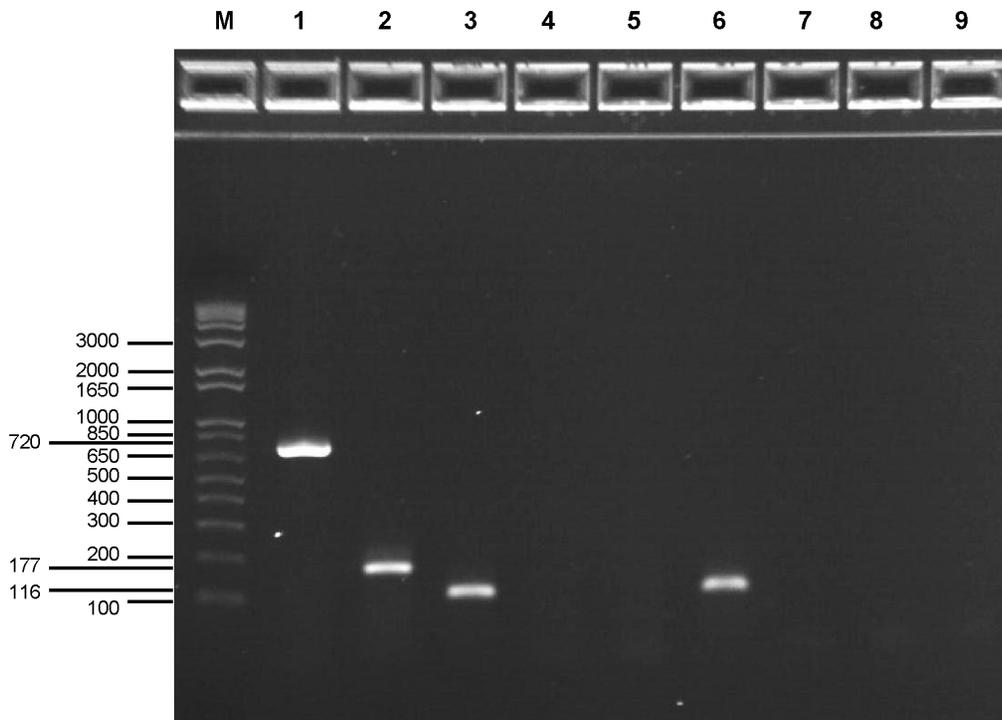
**Figure 6. Alignment of the 2.4 and 4.2 regions of TcdR-related  $\sigma$  factors of selected toxigenic clostridia.** Note the highly conserved Serine-Arginine-Glutamine (SRQ) motif in all 4.2 regions of the eight TcdR-related  $\sigma$  factors (i and ii). The alignment includes the 2.4 and 4.2 regions of: i) the TcdR-related  $\sigma$  factors UviA (*C. perfringens*), TcdR (*C. difficile*), TetR (*C. tetani*), BotR (*C. botulinum* type A) (adapted from [27]) and ii) the putative TcdR-related  $\sigma$  factors of *C. baratii* type F7, nonproteolytic *C. botulinum* type B4, *C. botulinum* type E3 and *C. botulinum* type F6. For comparison, (iii) four non-TcdR-related members of the  $\sigma^{70}$  family found in *C. botulinum* type E3 strain Alaska ( $\sigma^E$ ,  $\sigma^G$ ,  $\sigma^H$  and  $\sigma^D$ ) were also included in the alignment. BoxShade ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)) with a threshold of 0.5 was used for shading of identical amino acids (black) and similar amino acids (grey). Accession numbers are from GenBank. The zigzag sign represents the non-continuity of the 2.4 and 4.2 regions.

doi:10.1371/journal.pone.0097983.g006

purified UviA-like regulatory protein in complex with RNA polymerase core enzyme specifically recognized the *C. baratii* type F7 botulinum toxin gene cluster promoters (Fig. 8). Therefore, the UviA-like protein of *C. baratii* type F7 may serve as an alternative  $\sigma$

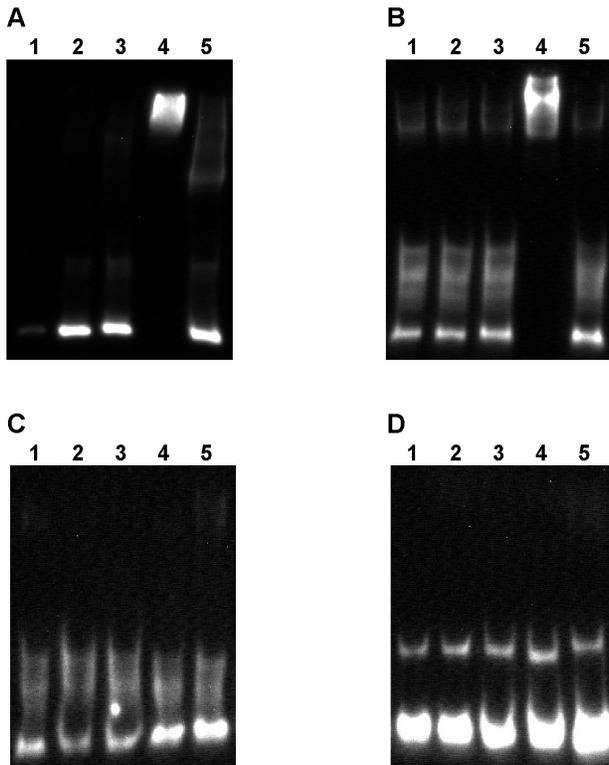
factor that recognizes the botulinum toxin gene cluster promoters of this organism.

In contrast to the types E and F6 *otfX* toxin gene clusters in nonproteolytic clostridia, the type B4 *ha* toxin gene cluster in the nonproteolytic strain Eklund 17B and the type G *ha* toxin gene



**Figure 7. PCR amplification comparison of toxigenic and nontoxigenic *C. baratii* strains.** Genomic DNA isolated from toxigenic (strain IBCA03-0045) (lanes 1–3) and nontoxigenic (strain IBCA08-0076) (lanes 4–6) *C. baratii* was used as templates. No template PCR (lanes 7–9) was used as a negative control. The PCR was performed with primer pairs specific for *bont/F7* (primers baratii58F and baratii60R, lanes 1, 4 and 7), the *uviA*-like gene (primers baratii1F and baratii2R, lanes 2, 5 and 8), and the 23S rRNA gene (primers baratii23sF1 and baratii23sR1, lanes 3, 6 and 9) of *C. baratii*. The toxigenic *C. baratii* strain was PCR-positive for both the *bont/F7* and the *uviA*-like genes (lanes 1 and 2), while the nontoxigenic strain was PCR-negative for both these genes (lanes 4 and 5). This finding suggests that the *C. baratii bont/F7* neurotoxin gene and the *uviA*-like gene are genetically linked. Note that both *C. baratii* strains were PCR-positive for the 23S rRNA gene that served as a positive control (lanes 3 and 6). M = Molecular weight markers (1 Kb Plus DNA Ladder, Invitrogen). The sizes (in base pairs) of the Molecular weight markers and the amplicons are presented at the left.

doi:10.1371/journal.pone.0097983.g007



**Figure 8. Gel mobility shift assays of the F7 botulinum gene cluster promoters with the UviA-like protein.** Labeled DNA fragments containing (A) or lacking (C) the *Pp47* and the *P2orfX1* promoters and containing (B) or lacking (D) the *P1orfX1* promoter were tested in gel mobility shift assays. The labeled DNA (lane 1) was incubated with RNA polymerase core (lane 2), UviA-like protein (lane 3), UviA-like protein preincubated with RNA polymerase core (lane 4) and UviA-like protein preincubated with RNA polymerase core and an excess of unlabeled DNA fragment (lane 5). Only the UviA-like-RNA polymerase core complex bound the botulinum toxin gene cluster promoters (lanes A4 and B4) and induced a shift in the labeled DNA migration. An excess of unlabeled DNA (lanes A5 and B5) or elimination of the promoter sequences (lanes C4 and D4) prevented the binding. doi:10.1371/journal.pone.0097983.g008

cluster in the nonproteolytic strain ATCC 27322 both contain the *botR* gene [38]. Hence, the absence of the *botR* regulatory gene within the toxin gene cluster may be a general characteristic of the *orfX* toxin gene clusters that reside in nonproteolytic neurotoxic clostridia. If this generalization is correct, then regulation of *bont* expression in *orfX* toxin gene clusters may differ between the nonproteolytic and the proteolytic neurotoxic clostridia.

However, the  $-35$  promoter element TTTACA that is associated with  $\sigma$  factor recognition and transcription of the polycistronic mRNA that codes for BoNT is conserved in all analyzed *bont* toxin gene clusters (Fig. 2i and [17]), as well as in other clostridial toxin gene promoters [22]. This promoter sequence conservation suggests that a similar transcription regulator participates in the expression of all *bont* toxin gene clusters (and possibly of other clostridial toxin genes), regardless of the presence or absence of the *botR* regulatory gene.

The nonproteolytic *C. baratii* type F7 *Pp47* and *P1orfX1* toxin gene cluster promoter sequences differed from their *orfX* toxin gene cluster promoters in nonproteolytic *C. botulinum* types E and F6 and were more homologous to the conserved promoters in proteolytic *C. botulinum* strains that are recognized by BotR (Fig. 2 and Fig. 4). This homology implies that the *C. baratii* *Pp47* and

*P1orfX1* promoters might be recognized by a BotR-like regulatory protein belonging to the TcdR-related group 5 subfamily of  $\sigma$  factors.

Interestingly, the predicted *PorfX1*  $-35$  promoter elements in the nonproteolytic type E and F6 strains (TTGAAA and TTGTAA, respectively) differed from the highly conserved  $-35$  element (TTTACA) recognized by the TcdR-related group 5 subfamily of  $\sigma$  factors (Fig. 2). Additionally, the *PorfX1*  $-10$  promoter element in nonproteolytic type E and F6 strains was identical to the primary  $\sigma$  factor (*E. coli*  $\sigma^{70}$  and *B. subtilis*  $\sigma^A$ ) consensus recognition sequence TATAAT. Promoters that contain the  $-35$  TTGAAA and  $-10$  TATAAT motifs are known to be primary  $\sigma^A$ -dependent promoters in clostridia [33]. The *C. baratii* *P2orfX1* promoter sequence is more similar to the *PorfX1* promoters in nonproteolytic type E and F6 strains than to the *C. baratii* *P1orfX1* and the *PorfX1* promoters in the proteolytic strains (Fig. 4).

Although the *Pp47*  $-35$  promoter motif in nonproteolytic type E and F6 strains was identical to the conserved  $-35$  motif recognized by TcdR-related  $\sigma$  factors, their  $-10$  promoter motif (TATATT) was highly homologous to the  $\sigma^A$  consensus recognition sequence (TATAAT) (Fig. 2). It is intriguing to speculate that the *orfX1* transcript, and perhaps even the *p47* transcript in nonproteolytic type E and F6 strains, are transcribed by an RNA polymerase that utilizes the primary  $\sigma$  factor ( $\sigma^A$ ), rather than the alternative  $\sigma$  factor (BotR) that is used by all proteolytic *C. botulinum* [14].

In *C. baratii* type F7 we identified a putative regulatory gene immediately upstream of the toxin gene cluster, as also occurs with the toxin gene clusters of *C. botulinum* types C and D [20,39]. But unlike *C. botulinum* types C and D, the *C. baratii* upstream putative regulatory gene is part of an operon with a similar size, structure and sequence to the *uviAB* operon of *C. perfringens* that codes for UviA and UviB [27]. Like BotR, UviA is a member of the TcdR-related group 5 of the  $\sigma^{70}$  family and regulates the expression of a toxin (bacteriocin BCN5), while the role of UviB is unknown [25,27]. The presence of similar regulatory genes that are involved in the regulation of unrelated toxins in different clostridium species may represent horizontal gene transfer and subsequent independent evolution [40]. Predicted amino acid sequence analysis found that the UviA-like protein of *C. baratii* type F7 is also a member of the  $\sigma^{70}$  family. Moreover, the UviA-like protein of *C. baratii* type F7 contained the same conserved region 4.2 of the TcdR-related group 5 subfamily of  $\sigma$  factors that recognizes the conserved  $-35$  promoter element TTTACA (Fig. 6). The highly conserved and unique Serine-Arginine-Glutamine (SRQ) motif in region 4.2 of the TcdR-related group 5 subfamily of  $\sigma$  factors may be a useful marker for recognizing members of this group.

However, the predicted amino acid sequence of region 2.4 of the UviA-like protein of *C. baratii* type F7 is quite different from the amino acid sequence of region 2.4 of the other TcdR-related  $\sigma$  factors (Fig. 6). Dupuy et al. reported that region 2.4 of the TcdR-related  $\sigma$  factors, the region responsible for the recognition of the  $-10$  promoter element, is not conserved. This observation explains the sequence variability of the  $-10$  promoter elements of the promoters that are recognized by TcdR-related  $\sigma$  factors [12].

We used gel mobility shift assays to show that the UviA-like protein of *C. baratii* type F7, in association with RNA polymerase core, binds specifically a DNA fragment that contains both the *Pp47* and *P2orfX1* promoters and to a DNA fragment that contains the *P1orfX1* promoter. The binding of UviA-like protein in complex with RNA polymerase core (Fig. 8A and B lane 4), but not by itself (Fig. 8A and B lane 3), to the DNA fragments that

contained the promoters demonstrates the recognition of the promoter sequences by the UviA-like protein in playing the role of a  $\sigma$  factor.

The *wiA*-like putative regulatory gene of *C. baratii* could not be detected by PCR in nontoxicogenic strains that lack *bont*/F7 (Fig. 7). This finding indicates that the two genes are genetically linked and suggests that they may also be functionally related. However, gene sequence variability may also have contributed to the negative PCR results. Genomic sequence comparison of toxicogenic and nontoxicogenic *C. baratii* strains is needed to confirm the absence of the *wiA*-like gene in nontoxicogenic *C. baratii* strains.

We located similar *wiAB*-like operons that encode putative TcdR-like  $\sigma$  factors in other nonproteolytic neurotoxicogenic clostridia that, like nonproteolytic *C. baratii*, lack the *botR* gene (e.g., *C. botulinum* type E and *C. botulinum* type F6). However, the *wiAB*-like operon resides in close proximity to the *bont* toxin gene cluster only in the genome of *C. baratii*. Also, *C. baratii* was the only nonproteolytic neurotoxicogenic *Clostridium* that contained *orfX* toxin

gene cluster promoters similar to the toxin gene cluster promoters in proteolytic neurotoxicogenic clostridia (Figs. 2 and 4). Therefore, it appears that *C. baratii* type F7 may have a unique molecular mechanism that controls the expression of its botulinum toxin gene cluster. If so, this unique mechanism may help explain the special clinical features observed in type F7 infant botulism patients.

## Supporting Information

**Table S1 Primers used for amplification, sequencing and mapping of various genes.**  
(DOCX)

## Author Contributions

Conceived and designed the experiments: ND SSA. Performed the experiments: ND JRB JNB. Analyzed the data: ND. Contributed reagents/materials/analysis tools: KKH JCD. Wrote the paper: ND SSA.

## References

- Hoffman RE, Pincomb BJ, Skeels MR (1982) Type F infant botulism. *Am J Dis Child* 136: 370–371.
- Hall JD, McCroskey LM, Pincomb BJ, Hatheway CL (1985) Isolation of an organism resembling *Clostridium botulinum* type F which produces type F botulinum toxin from an infant with botulism. *J Clin Microbiol* 21: 654–655.
- Suen JC, Hatheway CL, Steigerwalt AG, Brenner DJ (1988) Genetic confirmation of identities of neurotoxicogenic *Clostridium botulinum* and *Clostridium butyricum* implicated as agents of infant botulism. *J Clin Microbiol* 26: 2191–2192.
- Gimenez JA, Gimenez MA, DasGupta BR (1992) Characterization of the neurotoxin isolated from a *Clostridium botulinum* strain implicated in infant botulism. *Infect Immun* 60: 518–522.
- Paisley JW, Lauer BA, Arnon SS (1995) A second case of infant botulism type F caused by *Clostridium botulinum*. *Pediatr Infect Dis J* 14: 912–914.
- Barash JR, Tang TWH, Arnon SS (2005) First case of infant botulism caused by *Clostridium botulinum* type F in California. *J Clin Microbiol* 43: 4280–4282.
- Raphael BH, Choudoir MJ, Luquez C, Fernandez R, Maslanka SE (2010) Sequence diversity of genes encoding botulinum neurotoxin type F. *Appl Environ Microbiol* 76: 4805–4812.
- Dover N, Barash JR, Hill KK, Dettler JC, Arnon SS (2011) Novel structural elements within the non-proteolytic *Clostridium botulinum* type F toxin gene cluster. *Appl Environ Microbiol* 77: 1904–1906.
- Jacobson MJ, Lin G, Raphael B, Andreadis J, Johnson EA (2008) Analysis of neurotoxin cluster genes in *Clostridium botulinum* strains producing botulinum neurotoxin serotype A subtypes. *Appl Environ Microbiol* 74: 2778–2786.
- Smith TJ, Hill KK, Foley BT, Dettler JC, Munk AC, et al. (2007) Analysis of the neurotoxin complex genes in *Clostridium botulinum* A1–A4 and B1 strains: BoNT/A3, /Ba4 and /B1 clusters are located within plasmids. *PLoS One* 2: e1271.
- Dover N, Barash JR, Hill KK, Xie G, Arnon SS (2014) Molecular characterization of a novel botulinum neurotoxin type H gene. *J Infect Dis* 209: 192–202.
- Dupuy B, Raffestin S, Matamouros S, Mani N, Popoff MR, et al. (2006) Regulation of toxin and bacteriocin gene expression in *Clostridium* by interchangeable RNA polymerase sigma factors. *Mol Microbiol* 60: 1044–1057.
- Marvaud JC, Gibert M, Inoue K, Fujinaga Y, Oguma K, et al. (1998) botR/A is a positive regulator of botulinum neurotoxin and associated non-toxin protein genes in *Clostridium botulinum* A. *Mol Cell Probes* 29: 1009–1018.
- Raffestin S, Dupuy B, Marvaud JC, Popoff MR (2005) BotR/A and TetR are alternative RNA polymerase sigma factors controlling the expression of the neurotoxin and associated protein genes in *Clostridium botulinum* type A and *Clostridium tetani*. *Mol Microbiol* 55: 235–249.
- East AK, Bhandari M, Stacey JM, Campbell KD, Collins MD (1996) Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic *Clostridium botulinum* types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component. *Int J Syst Bacteriol* 46: 1105–1112.
- Kubota T, Yonekura N, Hariya Y, Isogai E, Isogai H, et al. (1998) Gene arrangement in the upstream region of *Clostridium botulinum* type E and *Clostridium butyricum* BL6340 progenitor toxin genes is different from that of other types. *FEMS Microbiol Lett* 158: 215–221.
- Dineen SS, Bradshaw M, Karasek CE, Johnson EA (2004) Nucleotide sequence and transcriptional analysis of the type A2 neurotoxin gene cluster in *Clostridium botulinum*. *FEMS Microbiol Lett* 235: 9–16.
- Henderson I, Whelan SM, Davis TO, Minton NP (1996) Genetic characterisation of the botulinum toxin complex of *Clostridium botulinum* strain NCTC 2916. *FEMS Microbiol Lett* 140: 151–158.
- Bradshaw M, Dineen SS, Maks ND, Johnson EA (2004) Regulation of neurotoxin complex expression in *Clostridium botulinum* strains 62A, Hall A-hyper, and NCTC 2916. *Anaerobe* 10: 321–333.
- Hauser D, Eklund MW, Boquet P, Popoff MR (1994) Organization of the botulinum neurotoxin C1 gene and its associated non-toxic protein genes in *Clostridium botulinum* C 468. *Mol Gen Genet* 243: 631–640.
- Hauser D, Gibert M, Marvaud JC, Eklund MW, Popoff MR (1995) Botulinum neurotoxin C1 complex genes, clostridial neurotoxin homology and genetic transfer in *Clostridium botulinum*. *Toxicon* 33: 515–526.
- Raffestin S, Marvaud JC, Cerrato R, Dupuy B, Popoff MR (2004) Organization and regulation of the neurotoxin genes in *Clostridium botulinum* and *Clostridium tetani*. *Anaerobe* 10: 93–100.
- Raffestin S, Couesnon A, Pereira Y, Mazuet C, Popoff MR (2009) Botulinum and tetanus neurotoxins: molecular biology, toxin gene regulation and mode of action. In: Bruggemann H, Gotschalk G, editors. *Clostridia: molecular biology in the post-genomic era*. Norfolk, United Kingdom: Caister Academic Press. pp. 1–23.
- Travers AA, Burgess RR (1969) Cyclic re-use of the RNA polymerase sigma factor. *Nature* 222: 537–540.
- Helmann JD (2002) The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* 46: 47–110.
- Ishihama A (2000) Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* 54: 499–518.
- Dupuy B, Mani N, Katayama S, Sonenshein AL (2005) Transcription activation of a UV-inducible *Clostridium perfringens* bacteriocin gene by a novel sigma factor. *Mol Microbiol* 55: 1196–1206.
- Fencia L, Franciosa G, Pourshaban M, Aureli P (1999) Intestinal toxemia botulism in two young people, caused by *Clostridium butyricum* type E. *Clin Infect Dis* 29: 1381–1387.
- Midura TF, Nygaard GS, Wood RM, Bodily HL (1972) *Clostridium botulinum* type F: isolation from venison jerky. *Appl Microbiol* 24: 165–167.
- Hatheway CL, Ferreira JL (1996) Detection and identification of *Clostridium botulinum* neurotoxins. *Adv Exp Med Biol* 391: 481–498.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
- Chen Y, Korkeala H, Aarnikunnas J, Lindstrom M (2007) Sequencing the botulinum neurotoxin gene and related genes in *Clostridium botulinum* type E strains reveals *orfX3* and a novel Type E neurotoxin subtype. *J Bacteriol* 189: 8643–8650.
- Han SO, Yukawa H, Inui M, Doi RH (2005) Molecular cloning and transcriptional and expression analysis of *engO*, encoding a new noncellulosomal family 9 enzyme, from *Clostridium cellulovorans*. *J Bacteriol* 187: 4884–4889.
- Saujet L, Monot M, Dupuy B, Soutourina O, Martin-Verstraete I (2011) The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. *J Bacteriol* 193: 3186–3196.
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39: D225–D229.
- Geer LY, Domrachev M, Lipman DJ, Bryant SH (2002) CDART: protein homology by domain architecture. *Genome Res* 12: 1619–1623.
- Paget MS, Helmann JD (2003) The sigma70 family of sigma factors. *Genome Biol* 4: 203.
- Bhandari M, Campbell KD, Collins MD, East AK (1997) Molecular characterization of the clusters of genes encoding the botulinum neurotoxin complex in *Clostridium botulinum* (*Clostridium argentinense*) type G and nonproteolytic *Clostridium botulinum* type B. *Curr Microbiol* 35: 207–214.

39. Nakajima H, Inoue K, Ikeda T, Fujinaga Y, Sunagawa H, et al. (1998) Molecular composition of the 16S toxin produced by a *Clostridium botulinum* type D strain, 1873. *Microbiol Immunol* 42: 599–605.
40. Popoff MR, Bouvet P (2013) Genetic characteristics of toxigenic Clostridia and toxin gene evolution. *Toxicon* 75: 63–89.