Identification of a botulinum neurotoxin-like gene cluster in *Bacillus toyonensis*

Xin Wei[1], Briallen Lobb[1], Kang Wang[2], Min Dong[2], Andrew C. Doxey[1*]

[1]Department of Biology, University of Waterloo, Waterloo, Ontario, Canada

[2]Department of Urology, Boston Children's Hospital, Department of Surgery and Department of Microbiology, Harvard Medical School, Boston, MA, USA

*Corresponding author. Email: acdoxey@uwaterloo.ca

Abstract

Clostridial neurotoxins, which include botulinum neurotoxins (BoNTs) and tetanus neurotoxin (TeNT) are the most potent toxins known, and are the causative agents of the neuroparalytic diseases, botulism and tetanus. Until recently, the clostridial neurotoxin family was restricted to the genus *Clostridium*, but members of this protein family have been found in a growing number of non-*Clostridium* species including *Weissella*, *Enterococcus*, and *Paraclostridium*. Here, we report the bioinformatic identification and analysis of a novel clostridial neurotoxin homolog in a *Bacillus toyonensis* genome recently deposited into the NCBI Genbank database. This putative toxin shares 26-29% identity with its closest BoNT relatives, suggesting that it is likely a novel BoNT-like toxin. It possesses key functional motifs (e.g., HExxH) indicative of toxin protease activity, contains the four characteristic BoNT domains, and is located in a BoNT-like genomic neighborhood containing the upstream non-toxic non-hemagglutinin (NTNH) gene as well as several P47-related genes. Phylogenetically, the toxin clusters as a divergent member of the recently discovered lineage of BoNT-like toxins that includes BoNT/X, BoNT/En, and the insecticidal PMP1. Genomic analysis of the *B. toyonensis* isolate CH177 revealed additional virulence factors and toxin genes indicative of potential pathogenicity targeting an unknown host species. The *Bacillus toyonensis* BoNT-like protein (BTNT) adds to a growing number of non-clostridial BoNT-like toxins, adding further information on the intriguing phylogenetic distribution and evolutionary history of the most potent toxins known.

Introduction

Botulinum neurotoxins (BoNTs) and tetanus neurotoxin (TeNT) are members of a broader protein family known as clostridial neurotoxins (CNTs), and are the two most potent toxins known [1]. Their sophisticated molecular mechanism of action, role in disease, intriguing phylogenetic distribution and evolutionary history, and use as therapeutics or as scaffolds for design of novel toxin-based therapeutics, make them of considerable interest to the scientific and biomedical community [2–11].

The BoNT mechanism of action can be summarized as follows: after entering host motor neurons by receptor-mediated endocytosis, the BoNT protease is released into the cytosol where it cleaves host SNARE proteins (soluble N-ethylmaleimide-sensitive factor attachment protein receptor). Cleavage of host SNARE proteins inhibits the release of neurotransmitters, resulting in flaccid paralysis [1–3]. The TeNT mechanism of action is similar but it undergoes a unique retrograde transport mechanism to reach inhibitory neurons, resulting in a spastic paralytic phenotype [12].

CNT genes encode a single polypeptide chain that is proteolytically cleaved into two distinct protein chains (light chain and heavy chain), which are connected by an inter-chain disulfide bond. The light chain (LC) includes the N-terminal metalloprotease domain responsible for proteolytic cleavage of host SNARE proteins, and the heavy chain (HC) contains the translocation domain (HN) responsible for translocation of the LC across the host endosome to the cytosol, and the C-terminal receptor binding domains (HCn and HCc) that mediates binding to a variety of host receptors such as polysialogangliosides, synaptotagmin I/II and glycosylated SV2 [13].

Historically, BoNTs were only known to exist within a variety of *Clostridium* species, and were phylogenetically grouped into seven subtypes (A-G). BoNT subtypes, and TeNT which falls within the BoNT phylogenetic tree, are highly divergent from one another, exhibiting amino acid sequence identities as low as ~30% [14]. Despite their sequence divergence, they are structurally and functionally similar, and retain their overall mechanism of action with the notable exception of TeNT as described above. In addition, different BoNT serotypes exhibit unique host-specificities, with BoNT/A, B, E, and F known to cause human botulism, and subtypes C, D and hybrid CD associated with wildlife botulism outbreaks.

The extremely limited taxonomic distribution, unique structure, and remarkable potency of CNTs, has made them intriguing targets of study for researchers interested in the evolution of toxins and proteins in general [7, 8, 15, 16]. The first indication that BoNT-related proteins might exist outside of the *Clostridium* genus came in 2015 with a study by Mansfield et al. that reported the identification of a BoNT homolog in the organism, *Weissella oryzae* [16]. This highly divergent BoNT-like protein retained several characteristics of BoNTs and was reported to retain the ability to cleave SNARE substrates [16, 17], but lacked several other characteristics of BoNT gene clusters including the presence of nearby accessory genes. The host-specificity of this "BoNT/Wo" protein is still unclear.

Since then, several additional non-clostridial BoNTs have been detected through genome analysis and database searching [18], including BoNT/En from a strain of *Enterococcus faecium* [19, 20] and PMP1 from *Paraclostridium bifermentans* [21]. Phylogenetically, BoNT/En and PMP1 fall into a distinct lineage of BoNT-like toxins, and form their own branch outside of the subfamily of CNTs targeting vertebrates including BoNTs A-G and TeNT. These BoNT-like toxins therefore appear to be a sister clade to the vertebrate BoNTs. Growing evidence suggests that this sister clade of BoNT-like toxins, which also includes the recently discovered BoNT/X from *Clostridium botulinum* strain 111, may have an association with insect hosts [7, 21, 22].

In this work, we report the identification and initial bioinformatic analysis of a novel putative BoNT-like toxin from a particular strain of *Bacillus toyonensis* ("isolate CH177"), whose genome was recently added to the NCBI database on Apr 12, 2023. Our analysis of *B. toyonensis* BoNT-like toxin suggests that it is related to the BoNT/X/En/PMP1 lineage and that it likely forms a novel BoNT-like toxin with unique properties.

Results and Discussion

BLAST-based identification of BoNT-related genes in B. toyonensis

On July 6, 2023, we performed a BLASTp search against the NCBI nr database with default parameters using BoNT/A1 (NCBI accession # WP_011948511) as a query. A previously unreported homolog annotated as "hypothetical protein" (accession # HDR7951433) from *Bacillus toyonensis* was detected with an *E*-value of 3 x 10^{-53}, and amino acid sequence identity of 26.87%. A second protein (HDR7951432), also from *B. toyonensis,* was detected with an *E*-value of 4 x 10^{-33} and 24.6% amino acid sequence identity. Inspection of both sequences and subsequent BLAST analysis revealed that the first protein (HDR7951433) is a homolog of BoNT, while the second (HDR7951432) is a homolog of the NTNH protein, whose gene is typically encoded upstream of *bont* in *bont* gene clusters. Consistent with this, a zinc-metalloprotease HExxH motif unique to BoNTs was found in HDR7951433 but was absent in HDR7951432. We therefore designated HDR7951433 as BTNT for putative *Bacillus toyonensis neurotoxin* and HDR7951432 as BT-NTNH.

Next, we aligned the putative *B. toyonensis* neurotoxin (BTNT) with other members of the BoNT family (**Figure 1**). It is most similar to PMP1 from *Paraclostridium bifermentans*, but is still highly divergent at only ~29% identity. This degree of sequence divergence is expected for different serotypes within the BoNT family as shown in **Figure 1**.

	PMP1	BoNT/A1	BoNT/X	BoNT/En	TeNT
BTNT	28.99	24.15	26.97	25.75	26.22
PMP1		30.27	36.07	34.78	29.35
BoNT/A1			30.51	28.92	34.23
BoNT/X				38.62	29.13
BoNT/En					29.05

Figure 1. Pairwise sequence similarity of BoNT toxins including the newly identified BoNT-like toxin from *B. toyonensis* (HDR7951433.1).

Gene neighborhood analysis

To examine the gene neighborhood surrounding the newly identified *B. toyonensis* BoNT-like toxin gene, we retrieved the associated contig (DAOJGQ010000093.1) and gene annotations from the NCBI Assembly (GCA_029709895.1), and visualized it using AnnoView (annoview.uwaterloo.ca). We also manually analyzed and annotated the nearby genes using BLAST. *btnt* has a gene neighborhood structure that is characteristic of other *bont* genes with the *bt-ntnh* gene located immediately upstream as expected, as well as three P47/ORFX-like genes (**Figure 2**). BLAST analysis revealed that the P47/ORFX containing proteins putatively match the proteins ORFX3, P47, and ORFX2 in *P. bifermentans*, *E. faecium*, and *C. botulinum*. This general gene arrangement is conserved among other members of the X/En/PMP1 clade of BoNT-like toxins, and also some other BoNT genes (e.g., BoNT/A4). It is noticeable that *p47* is located within the *orfX* locus in the genomic neighborhood of BNTN, whereas it is usually located upstream or downstream of the *orfX* locus in the genomic neighborhood of other BoNT toxins. As the available contig is short, the identity of genes further upstream and downstream are currently unknown.

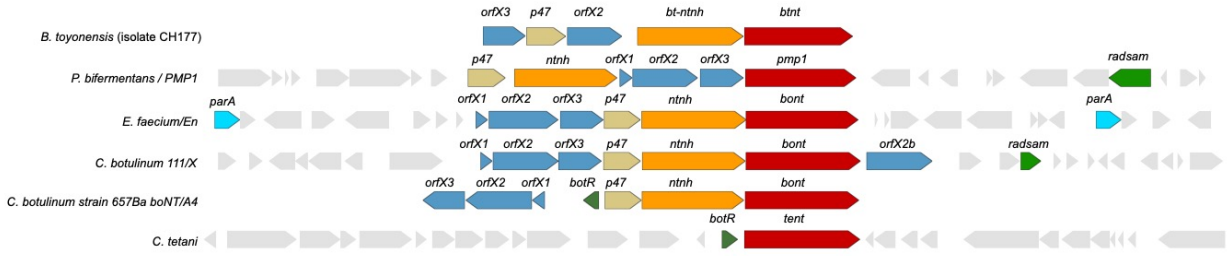

Figure 2. Visualization of a newly identified *bont*-like gene cluster in *B. toyonensis* containing homologs of *bont, ntnh, and* P47-domain containing proteins. Also shown for comparison are *bont* gene clusters from other species.

Sequence analysis

Next, we aligned BTNT with all representatives of all other BoNT subtypes, and examined the alignment for presence/absence of key functional residues/motifs (**Figure 3**). Although BTNT contains the conserved HExxH motif as described earlier, it appears to lack the double Cys residues within the disulfide linker region at the canonical positions conserved in other BoNTs. This conserved disulfide bond connects the heavy and light chain, and is essential for the toxicity of neurotoxins [23]. However, other Cys residues may perform this role, such as the Cys-590 located eight residues further downstream. BTNT also possesses the conserved PxxG motif identified in the translocation domain of BoNTs and more distantly related BoNT homologs [24]. It lacks the SxWY ganglioside cell receptor-binding motif [25] but possesses a similar SSEY motif at this site, which could indicate an ability to bind a similar receptor.

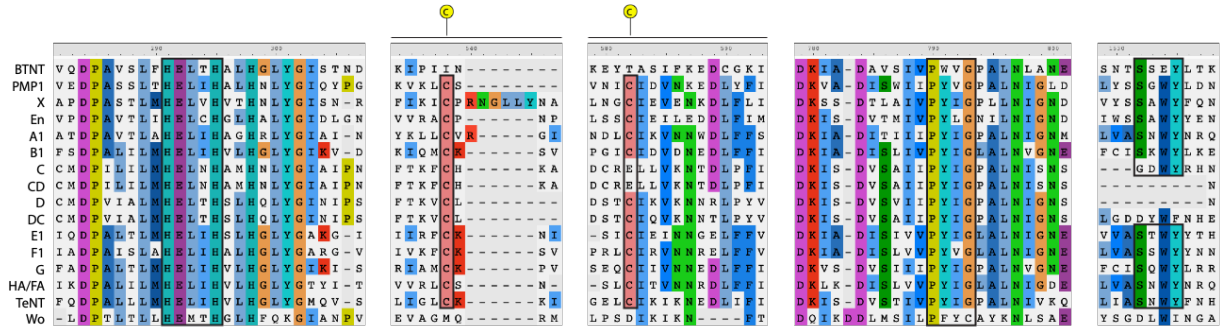

Figure 3. Alignment of *B. toyonensis* BoNT-like toxin with other subtypes and presence/absence of key functional motifs.

Phylogenetic analysis

To investigate the evolutionary relationships between BTNT and other BoNTs, we constructed a maximum-likelihood tree using IQ-TREE (**Figure 4**). BTNT clustered with the X/En/PMP1 lineage as the outermost branch. Its closer evolutionary relationship to this BoNT lineage is consistent with its gene neighborhood structure (**Figure 2**).

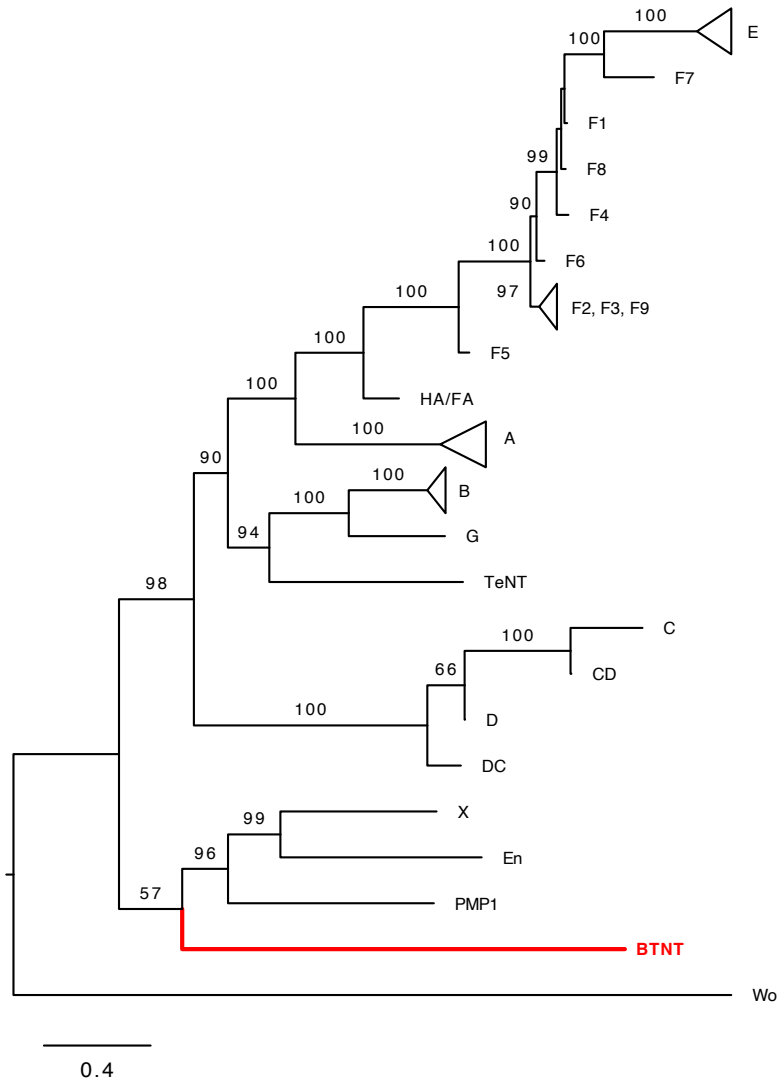

Figure 4. Maximum-likelihood phylogenetic tree of BoNTs and related toxins including the newly identified *B. toyonensis* BoNT-like toxin. Bootstrap values > 50 are indicated above the nodes.

Genomic analysis B. toyonensis isolate CH177

Based on available information in the NCBI, the BTNT-containing *B. toyonensis* strain was recently added to the NCBI database under the accession # GCA_029709895.1 and the isolate name “CH177”. The assembly has 296 contigs, and the draft WGS was generated using the SKESA v2.2. de novo assembler [26].

B. toyonensis isolate CH177 was found to be phylogenetically and genomically similar to other *B. toyonensis* strains (data not shown). All currently assembled *B. toyonensis* strains lack detectable homologs of BoNT genes, suggesting that the BTNT gene cluster is a recent acquisition into *B. toyonensis* isolate CH177. Further sequencing and assembly will be required to establish whether it is chromosomally encoded or whether it resides on a plasmid that has been uniquely acquired by this strain.

Next, we examined the CH177 assembly for additional evidence of pathogenicity-related genes using VFanalyzer within the VFDB database [27]. VFanalyzer detected hemolytic enterotoxin genes (*hblA*, *hblC*, and *hblD*), and a non-hemolytic enterotoxin (*nheC*), all of which have been detected in other related *Bacillus* species. Notably, in a commonly used probiotic *B. toyonensis* strain, these toxin genes are not detectably expressed or are expressed at much lower levels than *B. cereus* [28]. Also detected was the PlcR-PapR quorum sensing system, a hemolysin III family protein, haemolysin *xhlA* gene, and sphingomyelinase. The *islA* gene was also detected, which is a virulence factor in *Bacillus cereus* and has been shown to be expressed in the insect hemocoel and under iron-depleted conditions [29, 30]. Ultimately, these predictions suggest a pathogenicity-related function for CH177, but the host is still unknown.

Methods

Sequence and phylogenetic analysis of BoNT-like toxin in B. toyonensis

The BoNT-like toxin in *B. toyonensis*, PMP1 (WP_150887772.1) as well as representative BoNT sequences for each subtype (downloaded from bontbase.org) were used to create a multiple protein sequence alignment. A multiple alignment was then constructed using MAFFT v7.407 with the L-INS-i algorithm [31]. Jalview was used for alignment visualization. IQ-TREE version 2.0.3 was used to generate a phylogenetic tree of all BoNT-toxins [32]. We then midpoint rooted the tree, collapsed some clades that include toxins from the same subtype, and removed bootstrap values lower than 50%.

Gene neighborhood analysis

The neighboring genes in the ±20kb surrounding the neurotoxin genes were obtained from the NCBI database in .gbk format. These .gbk format files were uploaded to AnnoView (annoview.uwaterloo.ca) for gene neighborhood visualization. Genes in *B. toyonensis* were annotated based on BLAST searches against the NCBI-nr protein database, while all other gene annotations were derived from Wei et al. [22]

Genome analysis of B. toyonensis strain CH177

A dataset of 310 *B. toyonensis* genomes was downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=155322). A whole genome SNP-based phylogeny was created using Parsnp v1.2 [33] with default parameters and the *B. toyonensis* P18 strain (accession # GCF_016605985.1) selected as a reference genome. The predicted proteome (.fasta file) was submitted to the VFanalyzer tool in the VFDB database [27] with the *Bacillus* genus chosen as a reference taxon. Default parameters were used.

References

1. **Rossetto O, Montecucco C**. Tables of Toxicity of Botulinum and Tetanus Neurotoxins. *Toxins (Basel)*;11. Epub ahead of print 22 November 2019. DOI: 10.3390/TOXINS11120686.
2. **Williamson CHD, Sahl JW, Smith TJ, Xie G, Foley BT, *et al.*** Comparative genomic analyses reveal broad diversity in botulinum-toxin-producing Clostridia. *BMC Genomics* 2016;17:180.
3. **Collins MD, East AK**. Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *J Appl Microbiol* 1998;84:5–17.
4. **Dong M, Masuyer G, Stenmark P**. Botulinum and Tetanus Neurotoxins. *Annu Rev Biochem* 2019;88:811–837.
5. **Poulain B, Popoff MR**. Why Are Botulinum Neurotoxin-Producing Bacteria So Diverse and Botulinum Neurotoxins So Toxic? *Toxins (Basel)*;11. Epub ahead of print 1 January 2019. DOI: 10.3390/TOXINS11010034.
6. **Tehran DA, Pirazzini M**. Novel Botulinum Neurotoxins: Exploring Underneath the Iceberg Tip. *Toxins (Basel)*;10. Epub ahead of print 1 May 2018. DOI: 10.3390/TOXINS10050190.
7. **Mansfield MJ, Doxey AC**. Genomic insights into the evolution and ecology of botulinum neurotoxins. *Pathog Dis*;76. Epub ahead of print 1 June 2018. DOI: 10.1093/femspd/fty040.
8. **DasGupta BR**. Botulinum neurotoxins: perspective on their existence and as polyproteins harboring viral proteases. *J Gen Appl Microbiol* 2006;52:1–8.
9. **Roy D, Sadick NS**. Therapeutic uses of botulinum toxin. *N Engl J Med* 1991;324:561–568.
10. **Keith F, John C**. Targeted secretion inhibitors-innovative protein therapeutics. *Toxins (Basel)* 2010;2:2795–2815.
11. **Pellizzari R, Rossetto O, Schiavo G, Montecucco C**. Tetanus and botulinum neurotoxins: mechanism of action and therapeutic uses. *Philos Trans R Soc Lond B Biol Sci* 1999;354:259–68.
12. **Surana S, Tosolini AP, Meyer IFG, Fellows AD, Novoselov SS, *et al.*** The travel diaries of tetanus and botulinum neurotoxins. *Toxicon* 2018;147:58–67.
13. **Pirazzini M, Rossetto O, Eleopra R, Montecucco C**. Botulinum Neurotoxins: Biology, Pharmacology, and Toxicology. *Pharmacol Rev* 2017;69:200–235.
14. **Peck MW, Smith TJ, Anniballi F, Austin JW, Bano L, *et al.*** Historical perspectives and guidelines for botulinum neurotoxin subtype nomenclature. *Toxins*;9. Epub ahead of print 2017. DOI: 10.3390/toxins9010038.
15. **Doxey AC, Lynch MDJ, Müller KM, Meiering EM, McConkey BJ**. Insights into the evolutionary origins of clostridial neurotoxins from analysis of the Clostridium botulinum strain A neurotoxin gene cluster. *BMC Evol Biol* 2008;8:316.
16. **Mansfield MJ, Adams JB, Doxey AC**. Botulinum neurotoxin homologs in non-Clostridium species. *FEBS Lett* 2015;589:342–348.

17. **Zornetta I, Azarnia Tehran D, Arrigoni G, Anniballi F, Bano L, *et al.*** The first non Clostridial botulinum-like toxin cleaves VAMP within the juxtamembrane domain. *Sci Rep* 2016;6:30257.
18. **Doxey AC, Mansfield MJ, Montecucco C**. Discovery of novel bacterial toxins by genomics and computational biology. *Toxicon* 2018;147:2–12.
19. **Brunt J, Carter AT, Stringer SC, Peck MW**. Identification of a novel botulinum neurotoxin gene cluster in Enterococcus. *FEBS Lett* 2018;592:310–317.
20. **Zhang S, Lebreton F, Mansfield MJ, Miyashita S-I, Zhang J, *et al.*** Identification of a Botulinum Neurotoxin-like Toxin in a Commensal Strain of Enterococcus faecium. *Cell Host Microbe* 2018;23:169-176.e6.
21. **Contreras E, Masuyer G, Qureshi N, Chawla S, Dhillon HS, *et al.*** A neurotoxin that specifically targets Anopheles mosquitoes. *Nat Commun*;10. Epub ahead of print 1 December 2019. DOI: 10.1038/S41467-019-10732-W.
22. **Wei X, Wentz T, Lobb B, Mansfield M, Zhen W, *et al.*** Identification of divergent botulinum neurotoxin homologs in Paeniclostridium ghonii. *bioRxiv* 2022;2022.08.17.504336.
23. **Pirazzini M, Azarnia Tehran D, Zanetti G, Megighian A, Scorzeto M, *et al.*** Thioredoxin and Its Reductase Are Present on Synaptic Vesicles, and Their Inhibition Prevents the Paralysis Induced by Botulinum Neurotoxins. *Cell Rep* 2014;8:1870–1878.
24. **Mansfield MJ, Wentz TG, Zhang S, Lee EJ, Dong M, *et al.*** Bioinformatic discovery of a toxin family in Chryseobacterium piperi with sequence similarity to botulinum neurotoxins. *Sci Rep* 2019;9:1634.
25. **Stenmark P, Dong M, Dupuy J, Chapman ER, Stevens RC**. Crystal structure of the botulinum neurotoxin type G binding domain: insight into cell surface binding. *J Mol Biol* 2010;397:1287–1297.
26. **Souvorov A, Agarwala R, Lipman DJ**. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol*;19. Epub ahead of print 4 October 2018. DOI: 10.1186/S13059-018-1540-Z.
27. **Liu B, Zheng D, Jin Q, Chen L, Yang J**. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 2019;47:D687–D692.
28. **Abdulmawjood A, Herrmann J, Riede S, Jimenez G, Becker A, *et al.*** Evaluation of enterotoxin gene expression and enterotoxin production capacity of the probiotic strain Bacillus toyonensis BCT-7112T. *PLoS One*;14. Epub ahead of print 1 April 2019. DOI: 10.1371/JOURNAL.PONE.0214536.
29. **Fedhila S, Daou N, Lereclus D, Nielsen-LeRoux C**. Identification of Bacillus cereus internalin and other candidate virulence genes specifically induced during oral infection in insects. *Mol Microbiol* 2006;62:339–355.
30. **Daou N, Buisson C, Gohar M, Vidic J, Bierne H, *et al.*** IlsA, a unique surface protein of Bacillus cereus required for iron acquisition from heme, hemoglobin and ferritin. *PLoS Pathog*;5. Epub ahead of print November 2009. DOI: 10.1371/JOURNAL.PPAT.1000675.
31. **Katoh K, Standley DM**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–80.

32. **Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ**. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
33. **Treangen TJ, Ondov BD, Koren S, Phillippy AM**. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol*;15. Epub ahead of print 2014. DOI: 10.1186/S13059-014-0524-X.