

Evaluación de la variación para el gen de la neurotoxina tipo B2 en *Clostridium botulinum*

Nicolas Montoya Leon¹, John Anderson Acosta² y Raquel Bautista Escobar¹.

¹ Estudiante de Microbiología, facultad de ciencias, Universidad de los Andes, rbaustista@uniandes.edu.co, n.montoyal2@uniandes.edu.co

² Estudiante de Ingeniería Biomédica,, facultad de Ingeniería, Universidad de los Andes, ja.acostar1@uniandes.edu.co,

Resumen

Introducción

Las bacterias *Clostridium botulinum* producen neurotoxinas botulínicas, que provocan botulismo (Dong, M. et al, 2019), una enfermedad donde la persona afectada experimenta cólicos abdominales, insuficiencia respiratoria, dificultad al tragar y al hablar, visión doble, vómito y debilidad con parálisis (Medlineplus, 2023).

En este sentido, dichas bacterias producen 8 exotoxinas, es decir, 8 variaciones, de las cuales se evaluó la neurotoxina tipo B2. La cual es de interés científico por su impacto en el sistema nervioso, esta bloquea la liberación de acetilcolina, un neurotransmisor que controla el movimiento muscular(Naumann, M. et al, 2008). Asimismo, tiene uso médico, por ejemplo, en procedimientos estéticos conocidos como el botox.

Objetivos

Analizar diferentes alineamientos de bacterias de la especie *Clostridium botulinum*, específicamente al gen encargado de la producción de la neurotoxina tipo B2, con el fin de descifrar si existe o no variación en la producción y respuesta fenotípica de dicha molécula.

Hipótesis nula: No existe variación dentro de las diferentes cepas de *C. botulinum* respecto a la producción de la neurotoxina B2.

Hipótesis alternativa: Dependiendo de la cepa, existe variación significativa para los genes

encargados de la producción de la neurotoxina B2.

Parte experimental y datos

II-A Descarga y Alineamientos múltiples en EMBL-EBI a través del algoritmo de MUSCLE

Inicialmente, se descarga el conjunto de secuencias del portal de GenBank haciendo uso de la herramienta PoPset en el servidor del NCBI: número de acceso [2557445913](#). Seguidamente, se emplea un alineamiento múltiple por medio del algoritmo de MUSCLE en la página de EMBL-EBI, estos resultados se pueden observar en la **Fig.1** al igual que en con otros resultados en [Results for Job ID: muscle-I20240510-173150-0175-30095129-p1m](#) con otros resultados en EMBL-EBI.

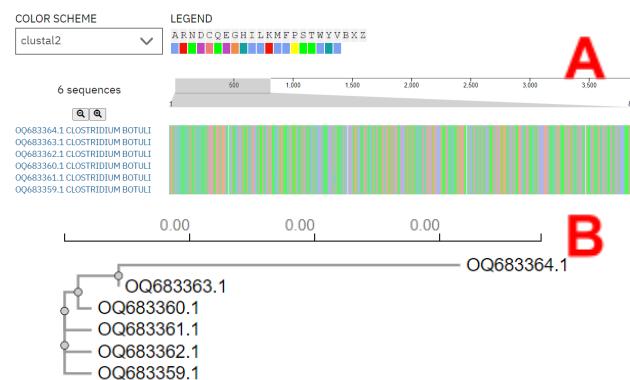


Fig. 1. Resultados del alineamiento Múltiple con el algoritmo de MUSCLE, a) indica los 6 alineamientos de las secuencias de nucleótido de los organismos de la especie *C. botulinum* b) Representa el árbol filogenético del alineamiento.

II-B Visualización en Jalview y revisión de las lecturas desde la suite de paquetes de Seqtk y EMBOSS en el cluster

Para analizar nuestros resultados, se empleó el programa de JalView para comparar los resultados de MUSCLE obtenidos en la sección anterior **Fig.2**. Asimismo, los atributos que más resaltan es la identificación de 3876 de bases en todas las secuencias y la alta conservación de la mayoría de las bases. Sumado a esto, se generó un árbol de distancia media y un árbol de filogenia de las secuencias.

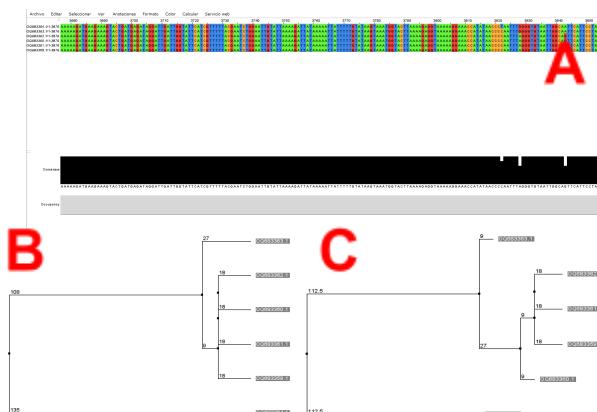


Fig. 2. Resultados obtenidos del alineamiento múltiple en jalview a) alineamiento múltiple de las secuencias asignando un color a cada nucleótido b) Árbol de filogenia c) Árbol de la distancia media de cada secuencia

Ahora bien, para obtener más datos o propiedades del alineamiento se decidió hacer uso del cluster y de los paquetes de Seqtk y EMBOSS. En este sentido, los comandos de `comp` para caracterizar el contenido de cada nucleótido, para seqtk, y el comando `Sixpack` e `infoseq` para representar la información básica sobre las secuencias y de los posibles ORFs **Fig.3**.

```
A module load seqtk/1.3 sixpack toxinaB2.fasta -auto
module load emboss/6.6.0 infoseq toxinaB2.fasta
```

```
B [biol229%hypatia ja.acostar] $ seqtk comp toxinaB2.fasta
Q0683359.1 3876 1622 375 601 1278 0 0 0 64 0 0
Q0683360.1 3876 1620 375 602 1279 0 0 0 64 0 0
Q0683361.1 3876 1622 374 601 1279 0 0 0 64 0 0
Q0683362.1 3876 1620 375 601 1280 0 0 0 64 0 0
Q0683363.1 3876 1621 375 601 1279 0 0 0 64 0 0
Q0683364.1 3876 1619 377 604 1276 0 0 0 70 0 0
```

ISA	Database Name	Accession	Type	Length	%GC	Organism	Description
Fasta::toxinaB2.fasta::00683359.1 -		Q0683359.1	-	3876	25.18	Clostridium botulinum strain Bo65 neurotoxin type B2 gene, complete cds	
Fasta::toxinaB2.fasta::00683360.1 -		Q0683360.1	-	3876	25.21	Clostridium botulinum strain Bo65 neurotoxin type B2 gene, complete cds	
Fasta::toxinaB2.fasta::00683361.1 -		Q0683361.1	-	3876	25.15	Clostridium botulinum strain Bo59 neurotoxin type B2 gene, complete cds	
Fasta::toxinaB2.fasta::00683362.1 -		Q0683362.1	-	3876	25.18	Clostridium botulinum strain Bo16 neurotoxin type B2 gene, complete cds	
Fasta::toxinaB2.fasta::00683363.1 -		Q0683363.1	-	3876	25.18	Clostridium botulinum strain Bo289 neurotoxin type B2 gene, complete cds	
Fasta::toxinaB2.fasta::00683364.1 -		Q0683364.1	-	3876	25.31	Clostridium botulinum strain Bo267 neurotoxin type B2 gene, complete cds	

C

```

C Q F I P K D E G W H T E *
S S F L K M K D G L N X
V H S * R * R M D * T X
-----|-----|-----|-----|
3841 CAGTCATCTCTAAAGATGAGGATGGATTGAATAA 3876
-----|-----|-----|-----|
3841 GTCAGTAAAGATTTCTACTCTACCTTAATT 3876
C N M G L S S P H I S Y
A T * E * L H L T S Q I
L E N R F I F S P N F L
F1 F2 F3
F6 F5 F4
#####
Minimum size of ORFs : 1
Total ORFs in Frame 1 : 1
Total ORFs in Frame 2 : 124
Total ORFs in Frame 3 : 198
Total ORFs in Frame 4 : 98
Total ORFs in Frame 5 : 110
Total ORFs in Frame 6 : 46
Total ORFs : 557
#####
```

Fig. 3. Resultados del análisis de las lecturas en el cluster a) representa todos los comandos y módulos usados para generar el análisis b) resultados del comando `-comp` y de `-infoseq` c) resultado obtenidos de `-sixpack`

Análisis de datos

III-A Construcción de la matriz, un árbol de similitud por Neighbor Joining y de máxima verosimilitud.

Al observar los datos **Fig. 4**. Se determina que al construir la matriz estos tienen valores numéricos muy bajos, que suelen oscilar entre 0,000516 y 0,003366. Estos valores bajos indican un alto grado de similitud entre las secuencias del gen. Como lo menciona Roozgard et al. (2013). Los valores más bajos significan una distancia genética menor, lo que significa que las secuencias son muy similares o casi idénticas. Del mismo modo, el árbol indica una alta conservación del gen entre las especies del estudio. Sumado a esto, el árbol ilustra las relaciones evolutivas y las distancias genéticas en las que se observa que la secuencia más alejada es la Q0683364.1. Finalmente, la imagen 3 muestra una tabla con las puntuaciones de la máxima verosimilitud para varios modelos

de sustitución de nucleótidos utilizados para analizar los datos de la secuencia.

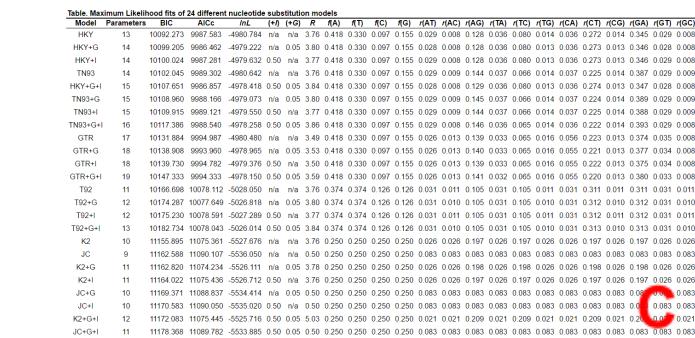
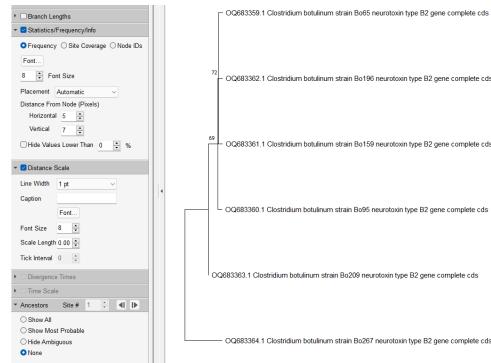
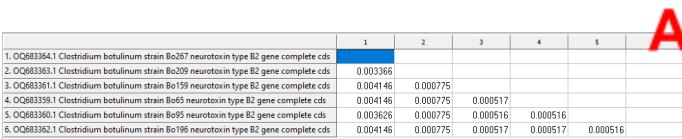


Fig. 4. Resultados del análisis a través de MEGA del alineamiento multiple a) describe la construcción de la matriz b) el árbol de similitud por Neighbor Joining c) la matriz de máxima verosimilitud

III-B Construcción del HMMs y Visualización del HMMs en Skylign

Posteriormente, luego del uso de estas herramientas se construyeron los HMMs de los alineamientos **Fig. 5.** a través del cluster y del comando observado en las imágenes. El resultado que más destaca es el hecho de la alta fidelidad o conservación de los nucleótidos en los alineamientos, esto podría indicar o liderar el planteamiento de que este gen en la especie *C.*

botulinum es sumamente conservado entre las especies, teniendo en total 15 variaciones en 3 secuencias de 3876 nucleótidos analizados

hmmbuild perfil.hmm toxinaB2.fasta

3865	2.19028	2.02190	2.27728	0.42595	3865	t	-	-
	1.38629	1.38629	1.38629	1.38629				
3866	2.27098	2.66446	0.29847	2.46273	3866	g	-	-
	0.85270	3.66261	3.66261	1.46634	0.26236	1.09861	0.40547	
3867	2.27098	2.66446	0.29847	2.46273	3867	g	-	-
	1.38629	1.38629	1.38629	1.38629				
3868	0.35040	2.44657	2.25350	2.26361	3868	a	-	-
	0.85270	3.66261	3.66261	1.46634	0.26236	1.09861	0.40547	
3869	2.19028	2.02190	2.27728	0.42595	3869	t	-	-
	1.38629	1.38629	1.38629	1.38629				
3870	2.19028	2.02190	2.27728	0.42595	3870	t	-	-
	1.38629	1.38629	1.38629	1.38629				
3871	2.27098	2.66446	0.29847	2.46273	3871	g	-	-
	0.85270	3.66261	3.66261	1.46634	0.26236	1.09861	0.40547	
3872	0.35040	2.44657	2.25350	2.26361	3872	a	-	-
	0.85270	3.66261	3.66261	1.46634	0.26236	1.09861	0.40547	
3873	0.35040	2.44657	2.25350	2.26361	3873	a	-	-
	1.38629	1.38629	1.38629	1.38629				
3874	2.19028	2.02190	2.27728	0.42595	3874	t	-	-
	1.38629	1.38629	1.38629	1.38629				
3875	0.85270	3.66261	3.66261	1.46634	0.26236	1.09861	0.40547	
	1.38629	1.38629	1.38629	1.38629				
3876	0.35040	2.44657	2.25350	2.26361	3876	a	-	-
	1.38629	1.38629	1.38629	1.38629				
	0.02669	3.63661	*	1.46634	0.26236	0.00000	*	A

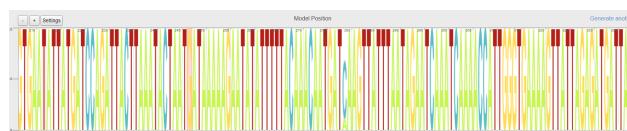


Fig. 5. Construcción del HMMs desde el cluster de Hypatia, a) comando y resultado obtenido del cluster b) Grafica resultante del alineamiento en la web de Skylign

III-C Realización de un blast estándar y árbol filogenético

Para finalizar, se realizó un blast estándar de los datos, y un análisis filogenético con otras secuencias guardadas en las bases de BLAST **Fig. 6.** Como se observa en la imagen se tiene una identidad del 100% y 0 gaps presente en el

alineamiento, esto como se mencionó en las anteriores secciones, indica que el gen de la neurotoxina tipo B2 está muy altamente conservado entre las especies. Por otro lado, El árbol relaciona el alineamiento con alineamientos de la especie *C. botulinum* y firmicutes. El autor Jacobson et al. (2008), indica que en su análisis de las secuencias del gen de la neurotoxina tipo B2 en diferentes cepas de *Clostridium botulinum*, observaron un grado muy alto de conservación de la secuencia, con identidades de secuencia por pares que oscilaban entre el 97,8% y el 100%.

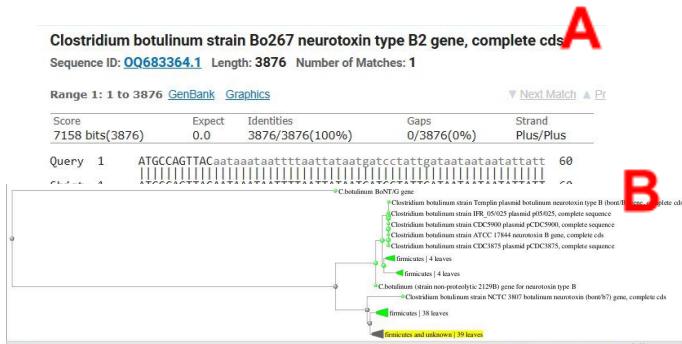


Fig. 6. Resultados del análisis en BLAT a) Indica los resultados del BLAST b) Resultados filogenéticos

Conclusiones

La neurotoxina tipo B2 producida por *Clostridium botulinum* es de gran interés científico debido a su impacto en el sistema nervioso y sus aplicaciones médicas, como el procedimiento estético conocido como Botox.

Los alineamientos múltiples de las secuencias del gen que codifica para la neurotoxina B2 en diferentes cepas de *C. botulinum* muestran un alto grado de conservación entre las especies.

La matriz de distancias genéticas construida presenta valores extremadamente bajos, cercanos a cero, lo que indica una similitud casi idéntica entre las secuencias analizadas.

Los árboles filogenéticos generados a partir de los alineamientos también reflejan esta alta conservación del gen, con pocas variaciones entre las cepas estudiadas.

El análisis de máxima verosimilitud de los modelos de sustitución de nucleótidos respalda la observación de una secuencia altamente conservada para este gen.

La construcción de Modelos Ocultos de Markov (HMM) y su visualización en Skylign ratifican la alta fidelidad y conservación de los nucleótidos en los alineamientos.

El análisis BLAST y la comparación con otras secuencias almacenadas en las bases de datos muestran una identidad del 100% y 0 gaps, lo que refuerza la idea de que el gen de la neurotoxina B2 está altamente conservado entre las especies de *C. botulinum* y bacterias relacionadas.

Bibliografía

- Dong, M., Masuyer, G., y Stenmark, P. (2019). Neurotoxinas botulínicas y tetánicas. *Revisión anual de bioquímica*, 88, 811-837.
<https://doi.org/10.1146/annurev-biochem-013118-111654>
 - Naumann, M., So, Y., Argoff, C. E., Childers, M. K., Dykstra, D. D.,
 - Gronseth, G. S., Jabbari, B., Kaufmann, H. C., Schurch, B., Silberstein, S. D., Simpson, D. M., & Subcomité de Evaluación de Terapéutica y Tecnología de la Academia Americana de Neurología (2008). Evaluación: Neurotoxina botulínica en el tratamiento de los trastornos autónomos y el dolor (una revisión basada en la evidencia): informe del Subcomité de Evaluación de Terapéutica y Tecnología de la

- Academia Americana de Neurología.
Neurología, 70(19), 1707–1714.
<https://doi.org/10.1212/01.wnl.0000311390.87642.d8>
- Jacobson, M. J., Lin, G., Raphael, B., Andreadis, J. D., & Johnson, E. A. (2008). Analysis of Neurotoxin Cluster Genes in Clostridium botulinum Strains Producing Botulinum Neurotoxin Serotype A Subtypes. *Applied And Environmental Microbiology*, 74(9), 2778-2786.
<https://doi.org/10.1128/aem.02828-07>

- Roozgard, A., Barzigar, N., Wang, S., Jiang, X., Ohno-Machado, L., & Cheng, S. (2013). Nucleotide sequence alignment using sparse coding and belief propagation. *IEEE*.
<https://doi.org/10.1109/embc.2013.6609568>
- U.S. National Library of Medicine. (2023, August 23). *Botulismo: Medlineplus Enciclopedia Médica*. MedlinePlus.
<https://medlineplus.gov/spanish/ency/article/000598.htm>