

**Análisis filogenético de las proteínas cristalinas plaguicidas (CRY)**  
**Fundamentos en biología computacional**  
**Universidad EAFIT – 2022-1**  
**Realizado por:**  
**Melissa Montoya Duque**  
**Sebastian Patiño Baena**

**Introducción:**

Las proteínas cristalinas plaguicidas fueron encontradas por primera vez en el organismo *Bacillus thuringiensis*, desde este momento, dichas proteínas han sido de gran utilidad para combatir especies y poblaciones que actúan como peste en diferentes cultivos ya que actúan como toxinas que impiden la alimentación y la reproducción de plagas como las orugas, los escarabajos y nemátodos (Schnepf *et al.*, 1998; Höss *et al.*, 2013). Las proteínas cry son usadas actualmente en conjunto con la técnica molecular de transformación genética para desarrollar plantas resistentes a las plagas, lo que ha facilitado las labores de control en los sectores agrícolas a nivel mundial (Höss *et al.*, 2013).

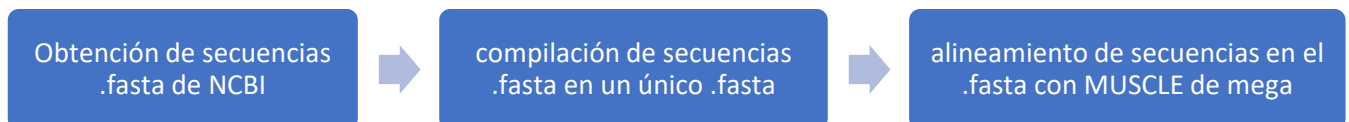
Ahora bien, generalmente dichas proteínas son conocidas y manipuladas a partir del microorganismo mencionado anteriormente, sin embargo, se han encontrado secuencias similares para organismos bastante alejados a escala evolutiva de *Bacillus thuringiensis* como el nemátodo *Priapulius caudatus* o la trucha arcoíris *Oncorhynchus mykiss* (NCBI, 2022), sin embargo se desconoce qué tan cercanas podrían ser las secuencias cry registradas en NCBI que corresponden a organismos filogenéticamente distantes. Los estudios sobre proteínas cry se han enfocado principalmente en dilucidar la evolución de la proteína en *Bacillus thuringiensis* (Bravo *et al.*, 2013; Wu *et al.*, 2007; Pinzon *et al.*, 2007) o sobre su uso para control de plagas (Hernández *et al.*, 2009; Romeis & Meissle, 2011; Khan *et al.*, 2021).

Por lo tanto, el objetivo del presente trabajo es realizar una reconstrucción filogenética de la proteína Cry proveniente del organismo *B. thuringiensis* frente a las proteínas tipo Cry que han sido secuenciadas y publicadas en la base de datos del NCBI para entender así ¿qué tan cercanas están dichas secuencias a escala evolutiva?

**Approach (workflow, programs, pipelines, parameters, steps)**

**1. Alineamiento de secuencias de proteínas cry disponibles en NCBI**

Para este paso se buscaron las secuencias de “pesticidal cristal protein” en NCBI para lo cual se encontraron 13 resultados y realizando curaduría basándonos en si habían repeticiones de secuencias se seleccionaron 8 para el presente trabajo, siguiendo el siguiente workflow para el alineamiento:



Obteniendo lo siguiente:

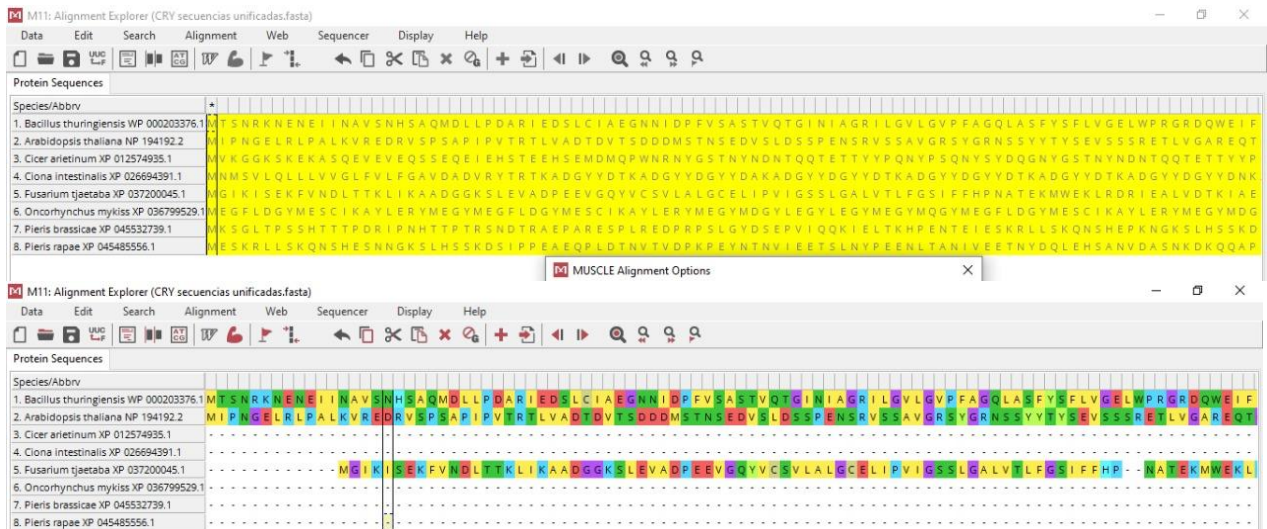
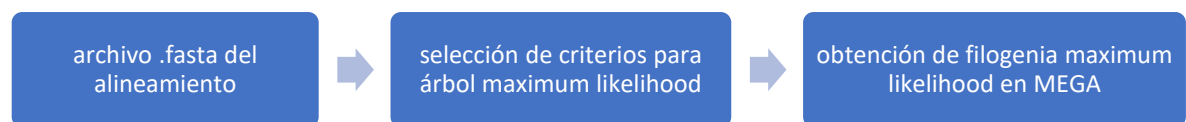


Figura 1. Alineamiento de proteínas cry usando muscle en el programa MEGA

## 2. Obtención de la filogenia máximo likelihood

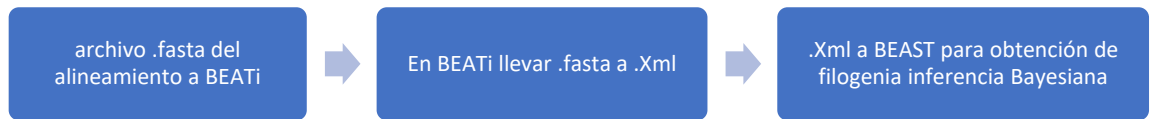
Una vez realizado el alineamiento, se descargó el archivo .fasta de este y se llevó nuevamente a MEGA para obtener la filogenia por el método máximo likelihood con los criterios por defecto de MEGA, eligiendo un Bootstrap de 100. El Bootstrap es necesario ya que es el número de veces que se va a repetir el proceso para saber qué tanto se repiten los nodos que se están formado y así asignar un valor más cercano a la realidad, así el Bootstrap asigna medidas de precisión (sesgo, varianza, intervalos de confianza, error de predicción, etc.) a las estimaciones de la muestra (Efron & Tibshirani, 1994).

Así, el enfoque de Bootstrap con un valor de arranque de 100 es un enfoque usado para evaluar la fiabilidad de la filogenia basada en secuencias. Los valores de Bootstrap de un árbol filogenético indica que, de 100, cuántas veces se observa la misma rama cuando se repite la generación de un árbol filogenético en un conjunto de datos remuestreado (Ojha *et al.*, 2022).



## 3. Obtención de la filogenia inferencia bayesiana

Nuevamente, se usa el archivo .fasta obtenido del alineamiento en MEGA, el cual se llevo a BEATi para ser convertido a archivo .Xml que pudiera ser leído por BEAST para la obtención de la filogenia por inferencia Bayesiana. En este caso se realizó un Bootstrap de 10.000.000 ya que para este tipo de métodos es necesario un mayor número de repeticiones por el método que maneja basado en densidades de probabilidad. En este caso (BI) en lugar de simular la distribución de muestreo de una estadística que estima un parámetro, el Bootstrap bayesiano simula la distribución posterior del parámetro.



## Results & Discussion

### 1. Filogenia maximum likelihood

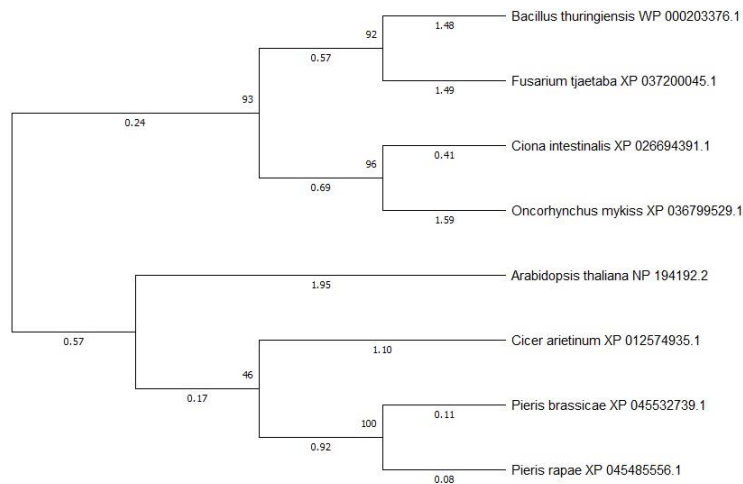


Figura 2. Filogenia obtenida por el método Máximo Likelihood en MEGA

### 2. Filogenia inferencia bayesiana

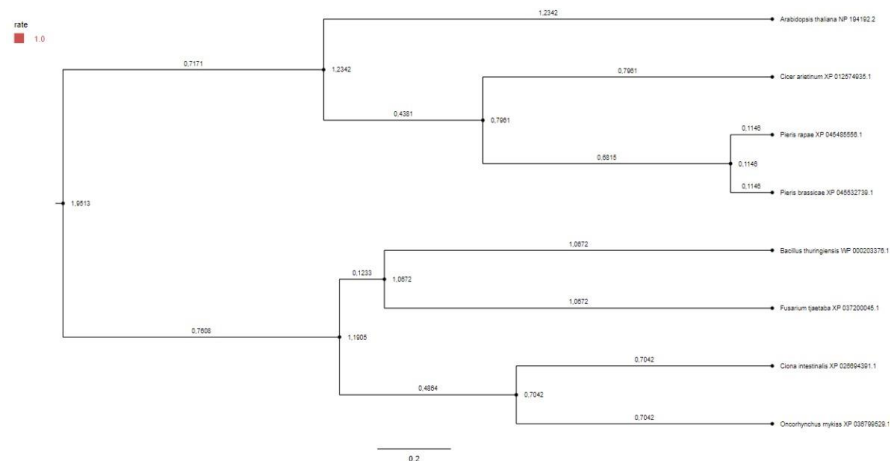
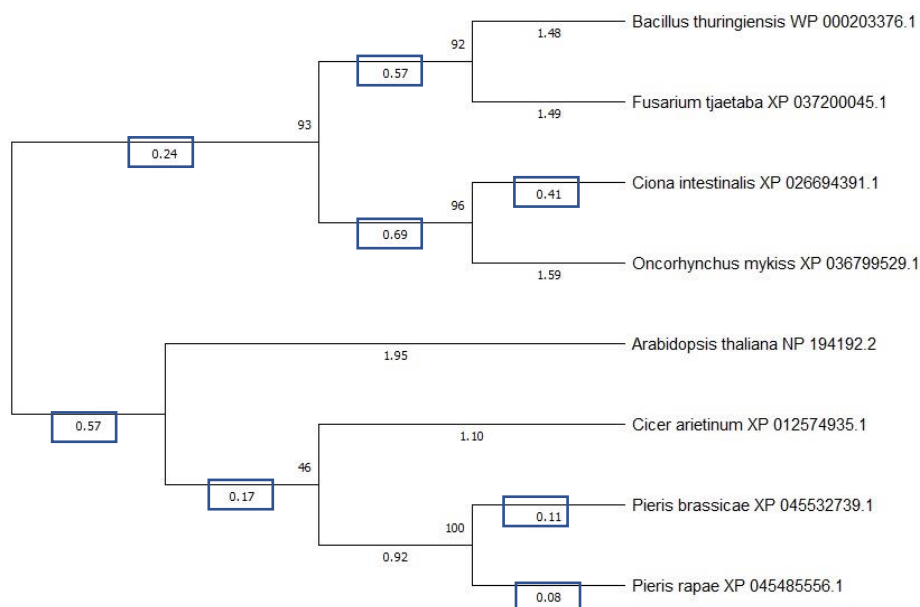


Figura 3. Filogenia obtenida por el método inferencia Bayesiana en BEAST

Se observa que independientemente del método usado las relaciones filogenéticas observadas son las mismas, con los mismos grupos hermanos y agrupación de taxones. Sin embargo, es de observar que los valores que soportan las ramas son diferentes para cada una de las filogenias y esto es porque, en esencia, son métodos estadísticos con una aproximación diferente para obtener los resultados. En general, el método de Maximum likelihood devuelve un valor fijo y usa parámetros fijos desconocidos, mientras que el de inferencia Bayesiana

Y para el caso las filogenias realizadas por Maximum Likelihood los valores entre 0.75 y 0.95 indican que está bien soportada dicha rama, si es menor a 0.75 no lo está, y vemos que nuevamente en la filogenia aquí obtenida no se tienen valores aceptables para muchas de las ramas como se señala a continuación:



*Figura 5. Filogenia Maximum Likelihood con valores menores a 0.75 resaltados*

Al analizarlas en conjunto resulta curioso que varias de las ramas con valores que no se ajustan a los valores aceptables coinciden, y esto puede deberse a que al ser secuencias de organismos tan distantes genéticamente donde quizá la proteína cry surgió por evolución convergente, al realizar las filogenias será más complicado que se repitan de 100 veces una cantidad superior a 75 (al pertenecer a organismos poco emparentados); sin embargo es en estos casos, en que no se trabaja con marcadores moleculares que permiten rastrear de manera más precisa las relaciones filogenéticas entre organismos, donde se hace necesario trabajar con más de un método para dilucidar la filogenia, siendo que si bien los valores de ninguna de ellas (ML o BI) se ajustan a lo deseable, se soportan entre sí al evidenciar que las relaciones filogenéticas encontradas son las mismas.

### **Conclusión:**

Los resultados encontrados permiten evidenciar que al realizar filogenias de organismos poco emparentados con genes que no corresponden a marcadores moleculares no es suficiente con un solo método, pues los valores de las ramas si bien en un método por si solo pueden no ajustarse a lo deseable, es posible encontrar que los métodos se soporten entre sí. Adicionalmente, se encontró que las secuencias de las proteínas cry son cercanas entre organismos cercanamente emparentados (e.g *Pieris*) y lejanas entre organismos lejanamente emparentados (e.g *Bacillus* y *Arabidopsis*).

### **Referencias:**

Beerli, P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, 22(3), 341-345.

Bravo, A., Gómez, I., Porta, H., García-Gómez, B. I., Rodriguez-Almazan, C., Pardo, L., & Soberón, M. (2013). Evolution of *Bacillus thuringiensis* Cry toxins insecticidal activity. *Microbial biotechnology*, 6(1), 17-26.

Douady, C. J., Delsuc, F., Boucher, Y., Doolittle, W. F., & Douzery, E. J. (2003). Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular biology and evolution*, 20(2), 248-254.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Hernández-Martínez, P., Ferré, J., & Escriche, B. (2009). Broad-spectrum cross-resistance in *Spodoptera exigua* from selection with a marginally toxic Cry protein. *Pest Management Science: formerly Pesticide Science*, 65(6), 645-650.

Kahn, T. W., Duck, N. B., McCarville, M. T., Schouten, L. C., Schweri, K., Zaitseva, J., & Daum, J. (2021). A *Bacillus thuringiensis* Cry protein controls soybean cyst nematode in transgenic soybean plants. *Nature communications*, 12(1), 1-12.

National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2022 Abril 11]. Available from: <https://www.ncbi.nlm.nih.gov/>

Ojha, K. K., Mishra, S., & Singh, V. K. (2022). Computational molecular phylogeny: concepts and applications. In *Bioinformatics* (pp. 67-89). Academic Press.

Pinzon, E. H., Sierra, D. A., Suarez, M. O., Orduz, S., & Florez, A. M. (2017). DNA secondary structure formation by DNA shuffling of the conserved domains of the Cry protein of *Bacillus thuringiensis*. *BMC biophysics*, 10(1), 1-10.

Romeis, J., & Meissle, M. (2011). Non-target risk assessment of Bt crops–Cry protein uptake by aphids. *Journal of Applied Entomology*, 135(1-2), 1-6.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D. R., & Dean, D. H. (1998). *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiology and molecular biology reviews: MMBR*, 62(3), 775–806. <https://doi.org/10.1128/MMBR.62.3.775-806.1998>

Wu, J. Y., Zhao, F. Q., Bai, J., Deng, G., Qin, S., & Bao, Q. Y. (2007). Adaptive evolution of cry genes in *Bacillus thuringiensis*: implications for their specificity determination. *Genomics, Proteomics & Bioinformatics*, 5(2), 102-110.