

# Filogenia y Comparacion de Bacterias pertenecientes a Clostridia Basado en secuencias del 16s RNA Ribosomal

AURA SANCHEZ

Universidad EAFIT  
asanch73@eafit.edu.co

mayo 28 de 2019

## Resumen

*A la hora de realizar una clasificación adecuada de bacterias en general, es necesario conocer una amplia gama de detalles del organismo con el fin de agruparlo de manera adecuada, características tanto fisiológicas, genéticas y metabólicas forman el complejo de aquello que se requiere para dicho fin. Contar con toda esta información para cada una de las especies de bacterias representa un reto significativo empezando por el hecho de que no todas ellas son cultivables en la actualidad, situación que empeora cuanto más se retroceda en la historia y que ha llevado a que varios autores planteen diferentes opciones para su clasificación. En este trabajo se realizara la comparación tomando como especies dentro de clostridium, las mencionadas por (Collins et al., 1994) con el fin de corroborar ,usando datos genéticos, que tan acertada estaba la clasificación propuesta por los autore.*

## I. INTRODUCCION

La hora de realizar estudios de taxonomía y filogenia en bacterias sin duda lo más comúnmente usado como marcador son las secuencias del 16s rRNA debido a varios factores entre ellos el que este presente en la mayoría de bacterias, que cambie relativamente poco en el tiempo y que sea lo suficientemente grande como para almacenar una cantidad representativa de información (Janda Abbott, 2007), al contar con estas secuencias es posible realizar un análisis filogenético de un grupo de interés en específico tal como lo hicieron collins et al., en su artículo publicado en 1994 en el cual proponían una filogenia para clostridium con nuevos géneros y posibles combinaciones de especies, ahora si bien se trata de un articulo detallado, al momento de buscar algunas de las bacterias agrupadas bajo la misma clasificación o grupos relativamente cercanos nos encontramos con una diferencia de funcionalidad drástica, teniendo especies que van desde ocupar un puesto base en la microbiota intestinal, hasta ser causantes de enfermedades.

El genero clostridium esta compuesto por basilios grampositivos anaerobios esporulados distribuido desde el suelo hasta el sistema digestivo de animales y con funciones diversas(Miranda Rojo, n.d.), una de sus especies clostridium difficile es un claro ejemplo ya que se identificó como causante de enfermedad en 1978 y es el agente causal de la colitis pseudomembranosa, enfermedad caracterizada por presentar cuadro diarreico (Rodríguez-pardo, Mirelis, Navarro, 2013). Al tratarse de una categoría diversa es posible que esto ocasione alguna clase de inconsistencia a la hora de clasificarla adecuadamente.

Basados en esta duda principalmente se realizó un análisis de la mayoría de las especies presentadas por los autores y agrupadas como *Clostridium* con el fin de corroborar la validez de la filogenia propuesta y comparar alguna diferencia posible en los resultados obtenidos. Con todo esto en mente se presenta a continuación los métodos usados desde la obtención de las secuencias hasta la construcción final de la filogenia.

## II. METODOLOGÍA

Para comenzar con el análisis filogenético se realizó primero la obtención de las secuencias de cada una de las especies (exactamente las mismas propuestas por Collins et al.,) seguido por su alineamiento, terminando con la selección del modelo y la construcción de la filogenia.

### I. Selección y alineamiento de secuencias:

Se obtuvieron las secuencias correspondientes al marcador 16S rRNA de las especies mostradas en la tabla 1 haciendo uso del el NCBI, una vez almacenadas todas las secuencias se paso a realizar el alineamiento de cada una de estas empleando a Geneious como herramienta principal.

### II. Obtención de un modelo de selección

Una vez se obtuvo el alineamiento era de vital importancia obtener el mejor modelo que estadísticamente soporte de mejor manera mis datos, para esto se usaron los índices AIC y BIC (criterio de información de Akaike y criterio de información bayesiano) (Montesinos et al., 2011), todo esto analizado por medio de jModelTest, todos los valores arrojados por los análisis se modelo bajo los dos índices se muestran en la figura 1y2 en la cual se puede comprobar que basado en ambos valores, el mejor modelo es el GTR+I+G.

**Figura 1:** Valores de AIC (criterio de información de Akaike)

Model	-lnL	K	AIC	delta	weight	cumWeight
GTR+I+G	24483.11815	142	49250.236300	0.000000	1.000000	1.000000
SYM+I+G	24504.51796	139	49287.035920	36.799620	1.02e-008	1.000000
HKY+I+G	24524.66331	138	49325.326620	75.090320	4.95e-017	1.000000
K80+I+G	24565.61108	135	49401.222160	150.985860	1.64e-033	1.000000
GTR+G	24611.93574	141	49505.871480	255.635180	3.09e-056	1.000000
SYM+G	24624.43692	138	49524.873840	274.637540	2.31e-060	1.000000
HKY+G	24648.22065	137	49570.441300	320.205000	2.94e-070	1.000000
K80+G	24699.00938	134	49666.018760	415.782460	5.18e-091	1.000000
F81+I+G	25110.84665	137	50495.693300	1245.457000	0.00e+000	1.000000
JC+I+G	25160.52693	134	50589.053860	1338.817560	0.00e+000	1.000000
F81+G	25197.87831	136	50667.756620	1417.520320	0.00e+000	1.000000
JC+G	25252.51857	133	50771.037140	1520.800840	0.00e+000	1.000000
GTR+I	25869.80053	141	52021.601060	2771.364760	0.00e+000	1.000000
SYM+I	25893.34661	138	52062.693220	2812.456920	0.00e+000	1.000000
HKY+I	25967.79362	137	52209.587240	2959.350940	0.00e+000	1.000000
K80+I	25980.48422	134	52228.968440	2978.732140	0.00e+000	1.000000
F81+I	26462.32829	136	53196.656580	3946.420280	0.00e+000	1.000000
JC+I	26495.73819	133	53257.476380	4007.240080	0.00e+000	1.000000
GTR	28543.74017	140	57367.480340	8117.244040	0.00e+000	1.000000
SYM	28576.06586	137	57426.131720	8175.895420	0.00e+000	1.000000
HKY	28680.30542	136	57632.610840	8382.374540	0.00e+000	1.000000
K80	28686.03251	133	57638.065020	8387.828720	0.00e+000	1.000000
F81	29145.01327	135	58560.026540	9309.790240	0.00e+000	1.000000
JC	29172.06144	132	58608.122880	9357.886580	0.00e+000	1.000000

**Figura 2:** Valores de BIC (criterio de información bayeciana)

Model	-lnL	K	BIC	delta	weight	cumWeight
GTR+I+G	24483.11815	142	50016.079659	0.000000	0.999967	0.999967
SYM+I+G	24504.51796	139	50036.699490	20.619831	3.33e-005	1.000000
HKY+I+G	24524.66331	138	50069.596927	53.517268	2.39e-012	1.000000
K80+I+G	24565.61108	135	50129.312678	113.233018	2.58e-025	1.000000
GTR+G	24611.93574	141	50266.321576	250.241917	4.58e-055	1.000000
SYM+G	24624.43692	138	50269.144147	253.064488	1.12e-055	1.000000
HKY+G	24648.22065	137	50309.318344	293.238685	2.11e-064	1.000000
K80+G	24699.00938	134	50388.716015	372.636355	1.21e-081	1.000000
F81+I+G	25110.84665	137	51234.570344	1218.490685	0.00e+000	1.000000
JC+I+G	25160.52693	134	51311.751115	1295.671455	0.00e+000	1.000000
F81+G	25197.87831	136	51401.240401	1385.160741	0.00e+000	1.000000
JC+G	25252.51857	133	51488.341132	1472.261472	0.00e+000	1.000000
GTR+I	25869.80053	141	52782.051156	2765.971497	0.00e+000	1.000000
SYM+I	25893.34661	138	52806.963527	2790.883868	0.00e+000	1.000000
HKY+I	25967.79362	137	52948.464284	2932.384625	0.00e+000	1.000000
K80+I	25980.48422	134	52951.665695	2935.586035	0.00e+000	1.000000
F81+I	26462.32829	136	53930.140361	3914.060701	0.00e+000	1.000000
JC+I	26495.73819	133	53974.780372	3958.700712	0.00e+000	1.000000
GTR	28543.74017	140	58122.537173	8106.457514	0.00e+000	1.000000
SYM	28576.06586	137	58165.008764	8148.929105	0.00e+000	1.000000
K80	28686.03251	133	58355.369012	8339.289352	0.00e+000	1.000000
HKY	28680.30542	136	58366.094621	8350.014961	0.00e+000	1.000000
F81	29145.01327	135	59288.117058	9272.037398	0.00e+000	1.000000
JC	29172.06144	132	59320.033609	9303.953949	0.00e+000	1.000000

### III. Construcción de la filogenia

Por ultimo se tomaron los archivos del alineamiento ya realizado y se paso a hacer la reconstrucción de la filogenia por medio de Garli, primero fue necesario establecer los parámetros con los que se iba a correr este modelo que en este caso se dividieron en dos, ambos con las mismas secuencias de 16s rRNA pero con valores de Bootstrap (soporte) y replicas diferentes. El primero con 4 replicas y un valor de soporte de 0 y el segundo con una sola replica y valores de soporte de 100. Una vez corridos estos modelos se usa sumtrees para adecuar los valores de soporte, a la mejor de las topologías terminando ya con una filogenia que tras establecer un grupo externo puede pasar a ser analizada.

**Tabla 1:** *Especies seleccionadas*

Especies	
Anaerococcus hydrogenalis	Eubacterium cellulosolvens
Anaerococcus prevotii	Eubacterium tenue
Acetitomaculum ruminis	Faecalicatena orotica
Anaerococcus tetradius	Filifactor villosus
Anaerocolumna aminovalerica	Finegoldia magna
Asaccharospora irregularis	Gottschalkia acidurici
Blautia coccoides	Gottschalkia purinilytica
Blautia hansenii	Hathewayia histolytica
Blautia producta	Hathewayia limosa
Caldicellulosiruptor bescii	Hathewayia proteolytica
Caldicellulosiruptor naganoensis	Helcococcus kunzii
Clostridioides difficile	Hungateiclostridium aldrichii
Clostridioides mangenotii	Hungateiclostridium thermocellum
C. aerotolerans	Maledivibacter halophilus
C. aminobutyricum	PaeniC. ghonii
C. aminophilum	Paeniclostridium sordellii
C. bifermentans	Parvimonas micra
C. celerecrescens	Peptoniphilus asaccharolyticus
C. cellulosi	Peptoclostridium litorale
C. clostridioforme	Peptoniphilus lacrimalis
C. felsineum	Peptostreptococcus anaerobius
C. filamentosum	Roseburia cecicola
C. leptum	Ruminiclostridium cellobioparum
C. paradoxum	Ruminiclostridium cellulolyticum
C. polysaccharolyticum	Ruminiclostridium papyrosolvens
C. populeti	Ruminococcus torques
C. sphenoides	Sedimentibacter hydroxybenzoicus
C. sporosphaeroides	Terrisporobacter glycolicus
C. sticklandii	Terrisporobacter mayombeii
C. symbiosum	Thermoclostridium stercorarium
C. thermoalcaliphilum	Tissierella praeacuta
C. xylanolyticum	Tyzzera nexilis
Clostridium symbiosum	Romboutsia lituseburensis
Coprococcus eutactus	

### III. RESULTADOS

La topológica obtenida al finalizar el análisis de máxima verosimilitud con Garli y la posterior suma de los arboles y obtención de valores de Bootstrap con sumtrees fue enraizada usando a *clostridium polysaccharolyticum* como grupo externo hasta contar con la filogenia que se muestra en la figura 3. Ahora si bien la filogenia en si cuenta con valores de soporte altos esto no confirma que las relaciones genéticas del grupo graficadas ahí sean correctas ya que es solo un reflejo del soporte que le dan, mis datos específicamente, a la construcción de la topología.

Figura 3: filogenia completa

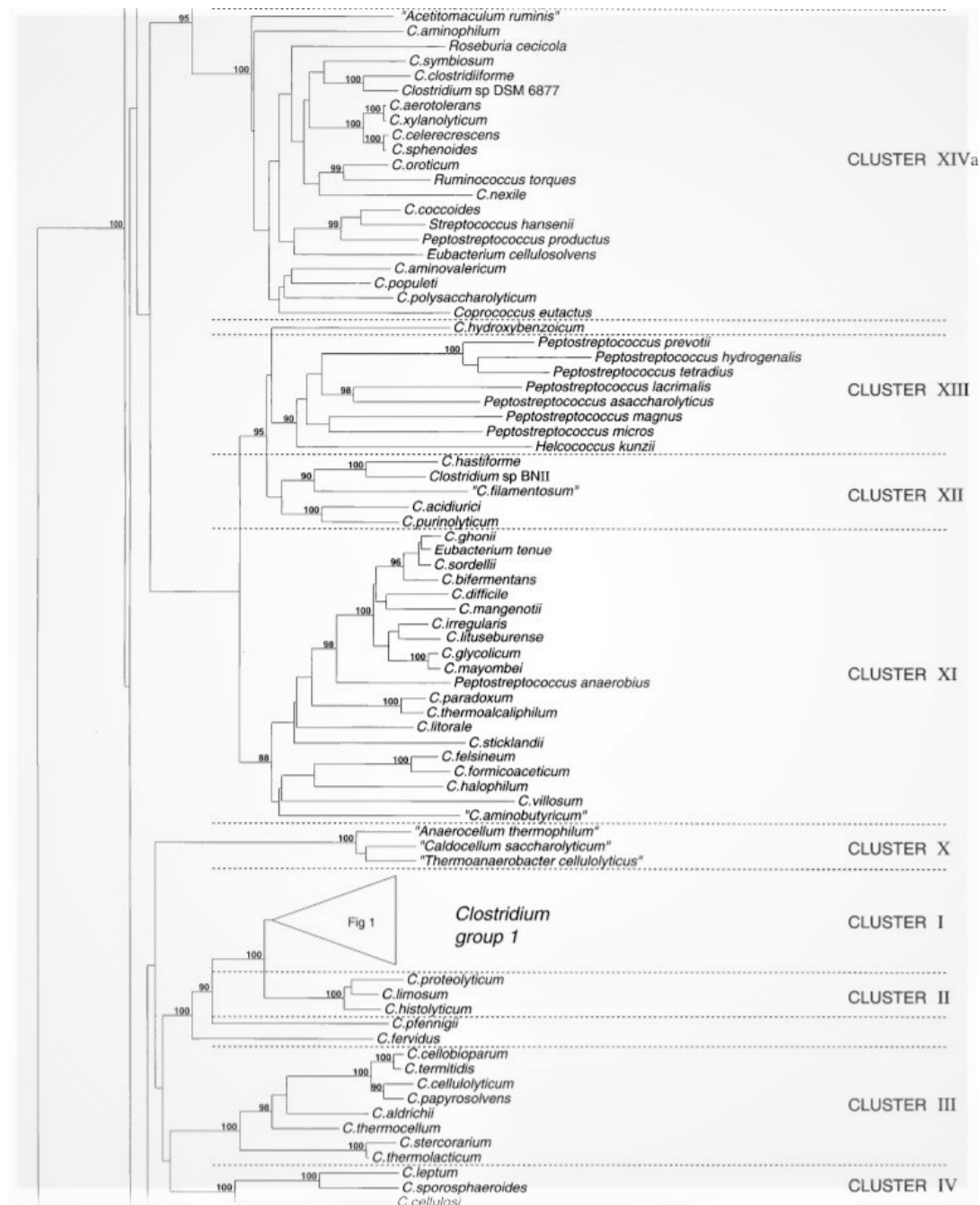


#### IV. ANÁLISIS

a la hora de comparar los resultados de ambas filogenias es posible observar que aunque algunas de las especies se encuentran ubicadas de forma diferente, la estructura de los nodos y la mayoría de grupos están ubicados igual. todo eso indicando que los datos de este marcador genético si soportan la clasificación de clostridium como aquel grupo que abarca todas estas

especies propuestas por collins et al., pero no limitado solo a esto, ya que algunas especies se han adicionado a este genero desde eso junto con otras que en un principio se creía pertenecían al genero pero que terminaron siendo parte de otras familias dentro de la misma clase, lo que explica la aun existente relación genética pero el cambio de la clasificación.

Figura 4: filogenia propuesta por(Collins et al., 1994)



## V. CONCLUSIÓN

tras realizar las filogenias desde cero son varias las conclusiones a las que se puede llegar, primero que construir topologías aunque es basado en marcadores genéticos específicos que cumplan características de interés, sigue estando en gran parte en manos de la probabilidad, en tomar el supuesto mas correcto y el mas cercano a mis datos para responder una pregunta especifica y darle orden a las relaciones entre organismos, lo que también nos permite concluir que los marcadores elegidos fueron los adecuados para el tipo de organismo seleccionado arrojando valores de soporte altos.

## VI. REFERENCIAS

Collins, M. D., Lawzon, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayz, J., Garcia, P., ... Farrow, J. A. E. (1994). The Phylogeny of the Genus *Clostridium*: Proposal of Five New Genera and Eleven New Species Combinations. *International Journal of Systematic Bacteriology*, 44(4), 812–826. <https://doi.org/10.1099/00207713-44-4-812>

Janda, J. M., Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, 45(9), 2761–2764. <https://doi.org/10.1128/JCM.01228-07>

Miranda, C., Rojo, D. (n.d.). *Clostridium perfringens*: INFECCIONES DE PIEL Y TEJIDOS BLANDOS. 1–10.

Montesinos, A., Directores De Tesina, L., Raúl, E., Diharce, V., Escobar, L. A., Guanajuato, R., Agosto De 201, G. (2011). ESTUDIO DEL AIC Y BIC EN LA SELECCIÓN DE MODELOS DE VIDA CON DATOS CENSURADOS que para obtener el grado de Maestro en Ciencias con Especialidad en Probabilidad y Estadística

Rodríguez-pardo, D., Mirelis, B., Navarro, F. (2013). Infecciones producidas por *Clostridium difficile*. *Medicina Clinica*, 31(4), 254–263. <https://doi.org/10.1016/j.eimc.2012.12.010>