



UNIVERSIDAD DE  
**COSTA RICA**

# Construcción de árboles filogenéticos

---

Bradd Mendoza Guido

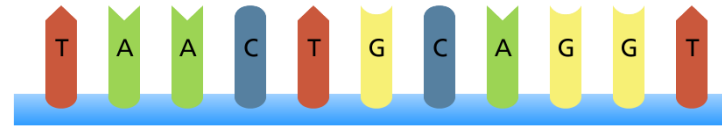
Genómica de procariontes B0634-SP8221



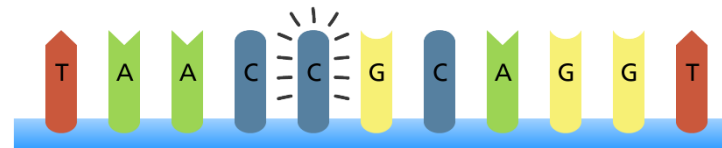
# Evolución

- Cambio en las frecuencias alélicas de una población
- **Mutación**
- Flujo genético
- Deriva genética
- Selección natural.

Original sequence



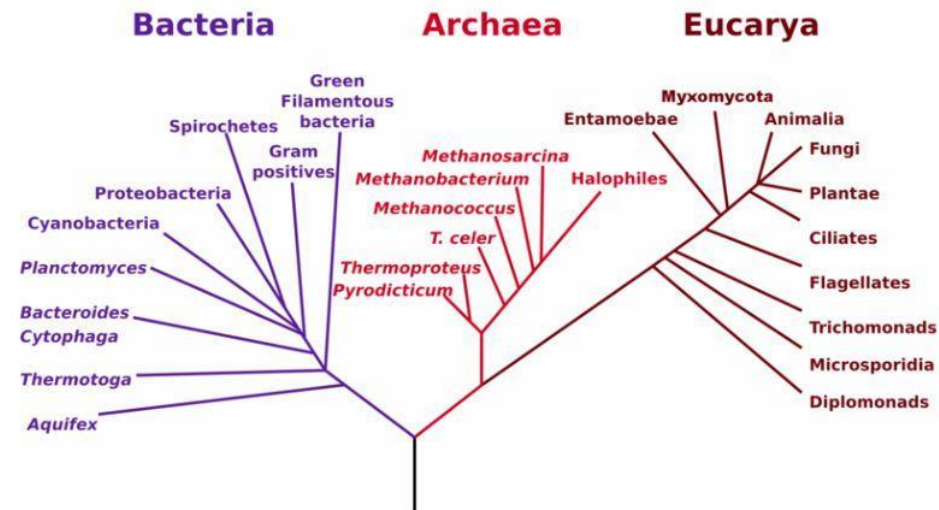
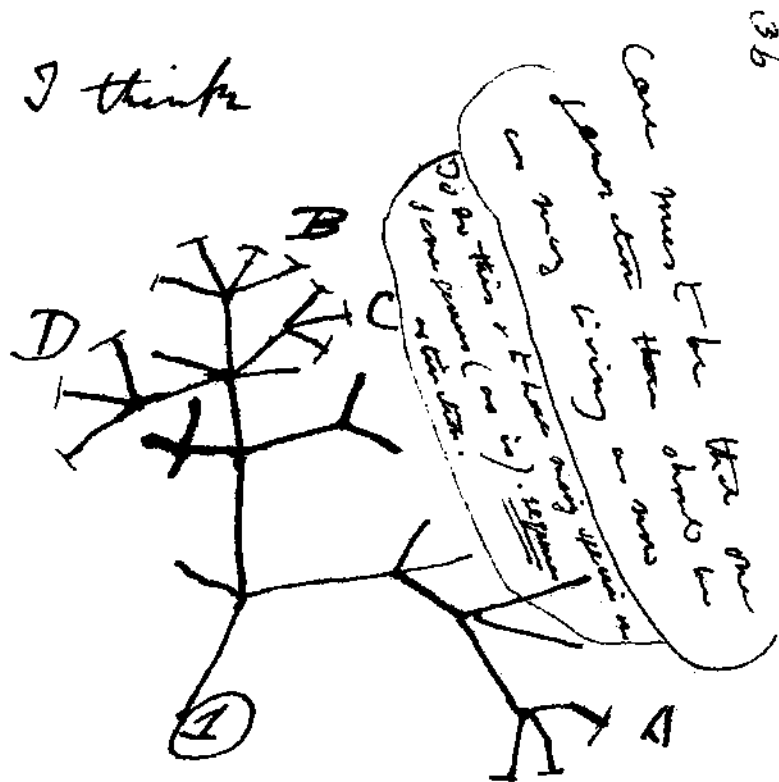
Point mutation





# Árbol filogenético

- Representación gráfica que muestra las relaciones evolutivas estimadas de un grupo de organismos.





## Por qué crear árboles filogenéticos?

- Identificar especies
- Estudiar la relación entre dos o más especies
- Discriminar cepas, o poblaciones de una misma especie
- Descubrir nuevas variantes de genes y/o proteínas



# Árboles genéticos vs protéicos

- Genes

- Mayor número de caracteres
- Cambios en sus secuencias pueden no verse en las proteínas
- Genes bajo poca selección natural ayudan a entender mejor la historia evolutiva de los organismos asociada a tiempos biológicos

- Proteínas

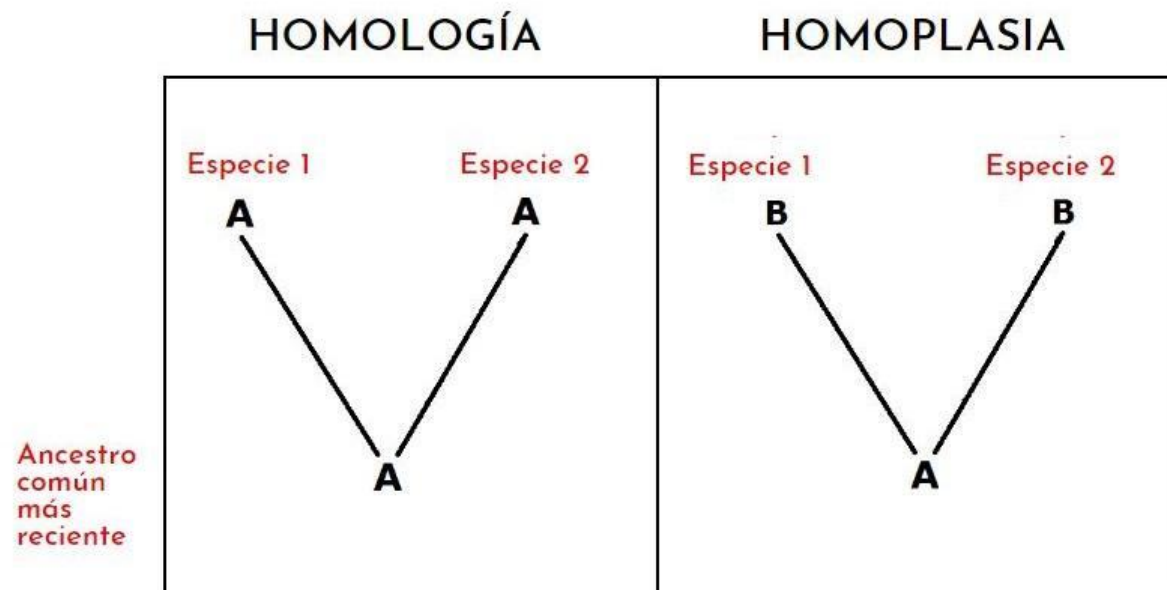
- Son las moléculas que realmente están bajo el efecto de la selección
- Ayudan a descifrar la derivación de nuevas moléculas
- Estudiar las relaciones evolutivas a nivel fenotípico

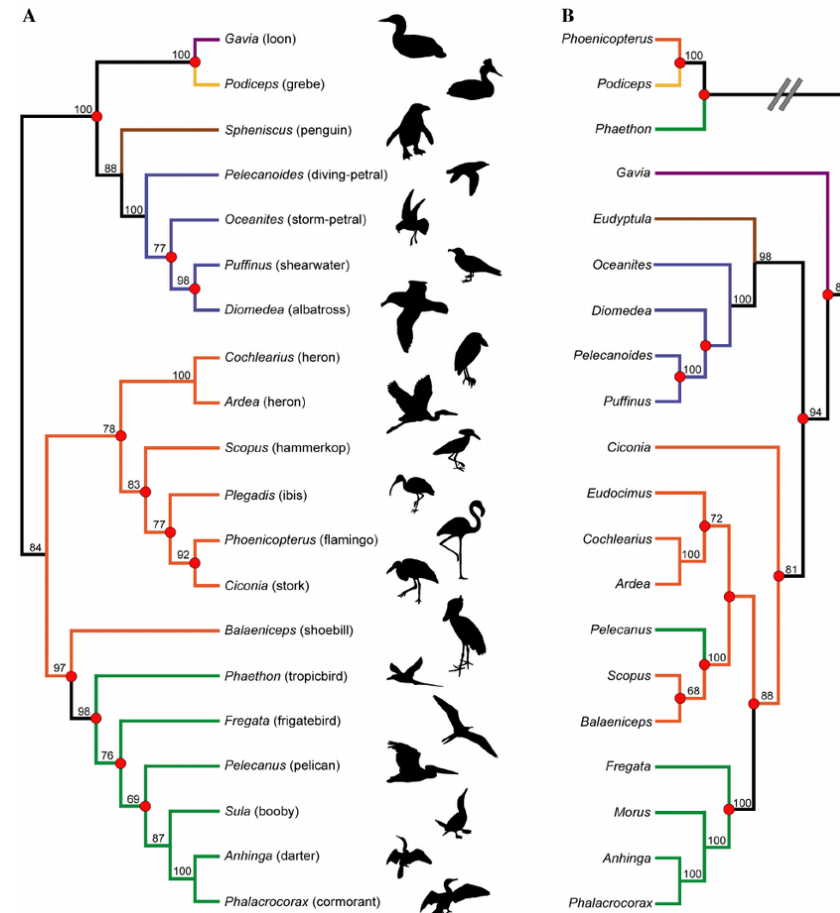


# Métodos basados en caracteres

## Parsimonia:

- El menor número de cambios para explicar mejor los datos
- No toma en cuenta modelos evolutivos
  - Devolución de caracteres:  $C \rightarrow T \rightarrow C$
  - Homoplasia

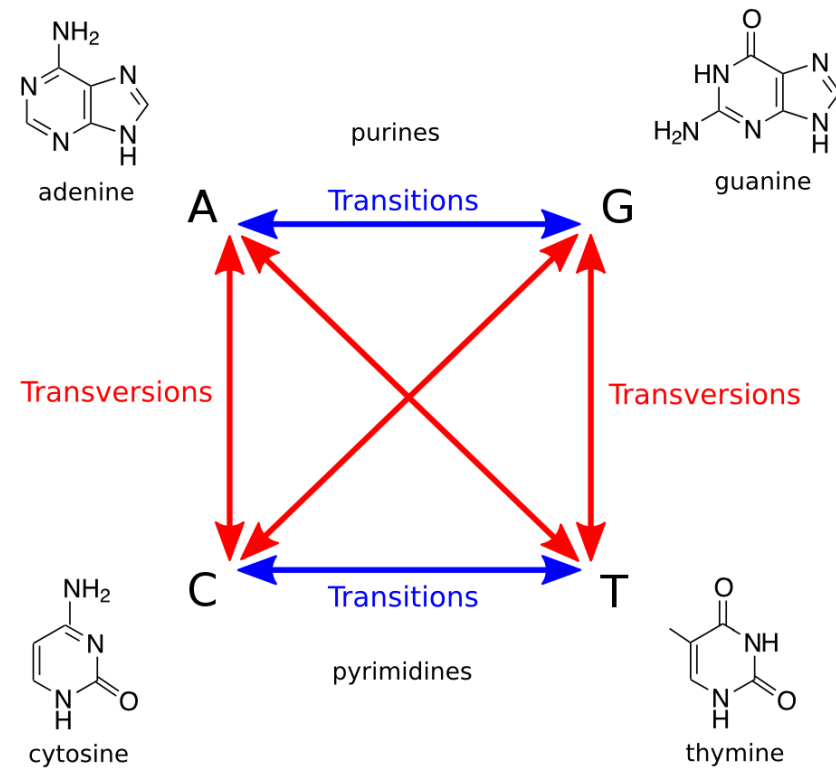




DOI:10.1371/journal.pone.0013354



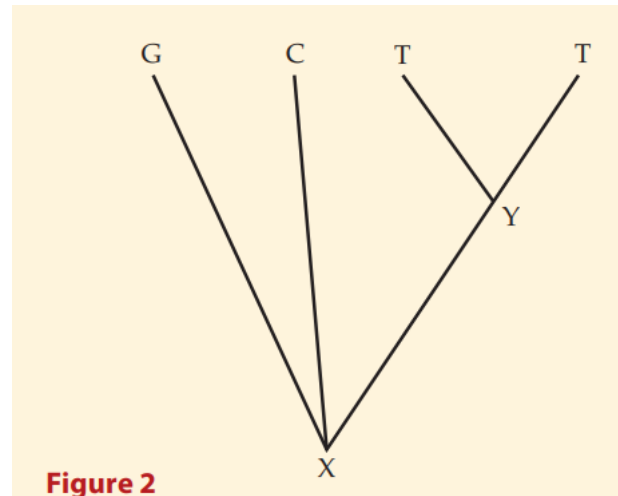
# Maximum likelihood

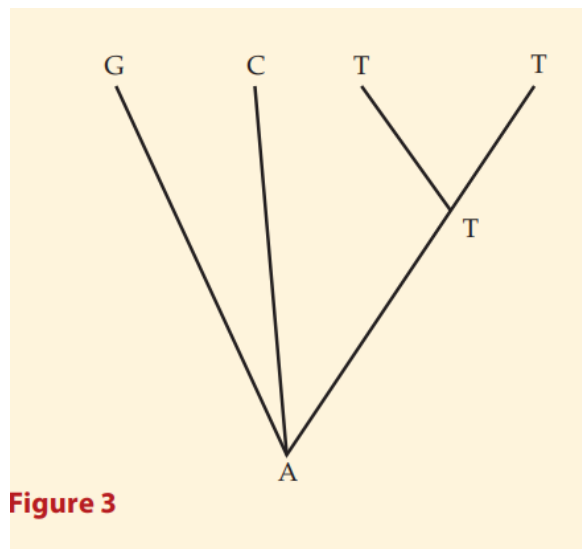




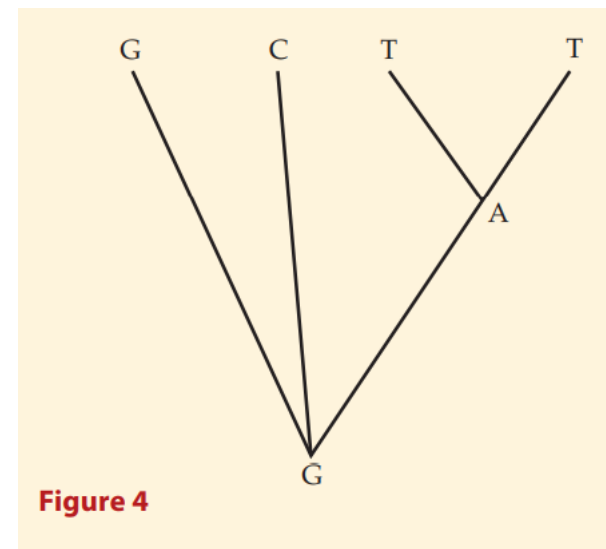


1 TCAAAAATGGCTTTATTTCG<sup>\*</sup>CTTAATGCCGTTAACCTTGCGGGGGCCATG  
2 TCCGTGATGGATTTATTTCTGCAATGCCTGTCATCTTATTCTCAAGTATC  
3 TTCGTGATGGATTTATTGCTGGTATGCCAGTCATCCTTTTCTCATCTATC  
4 TTCGTGACGGGTTTATCTCGCAATGCCGGTCATCCTATTTTCGAGTATT





$$P_{\text{Fig3}} = P_A \times P_{AG} \times P_{AC} \times P_{AT} \times P_{TT} \times P_{TT}$$



P = ?

Extraídos del libro: Phylogenetics tree made easy

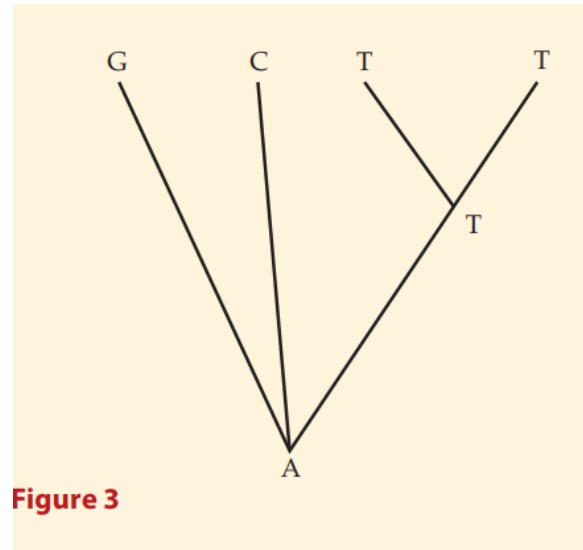


Figure 3

$$P_{\text{Fig3}} = P_A \times P_{AG} \times P_{AC} \times P_{AT} \times P_{TT} \times P_{TT}$$

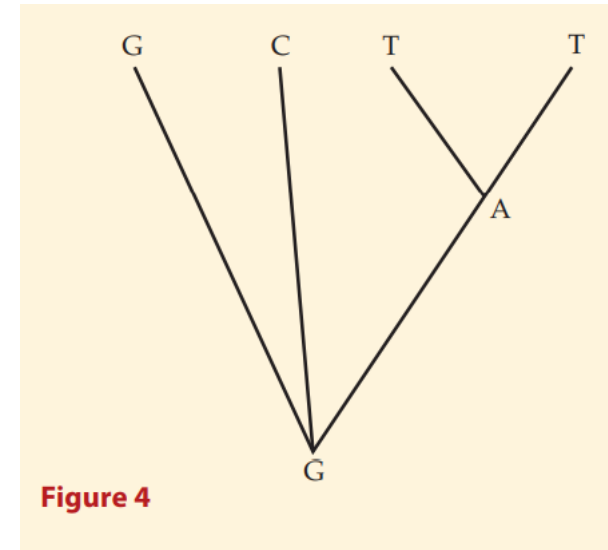


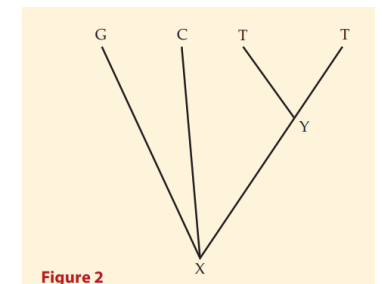
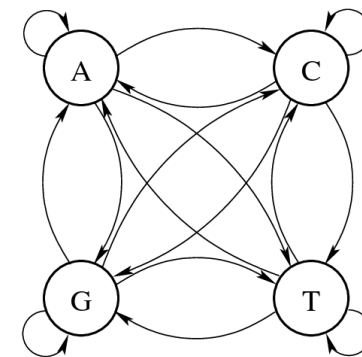
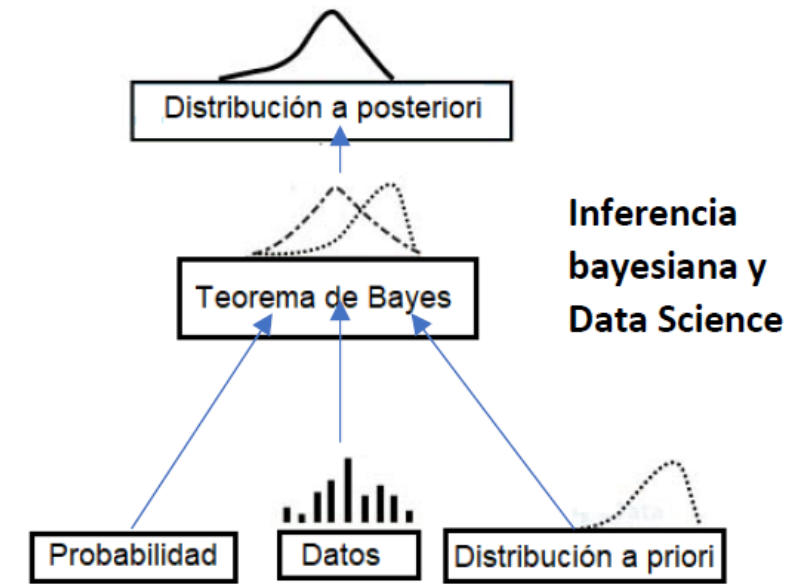
Figure 4

$$P = P^G \times P^{GG} \times P^{GC} \times P^{GA} \times P^{AT} \times P^{AT}$$

Extraídos del libro: Phylogenetics tree made easy

# Inferencia bayesiana

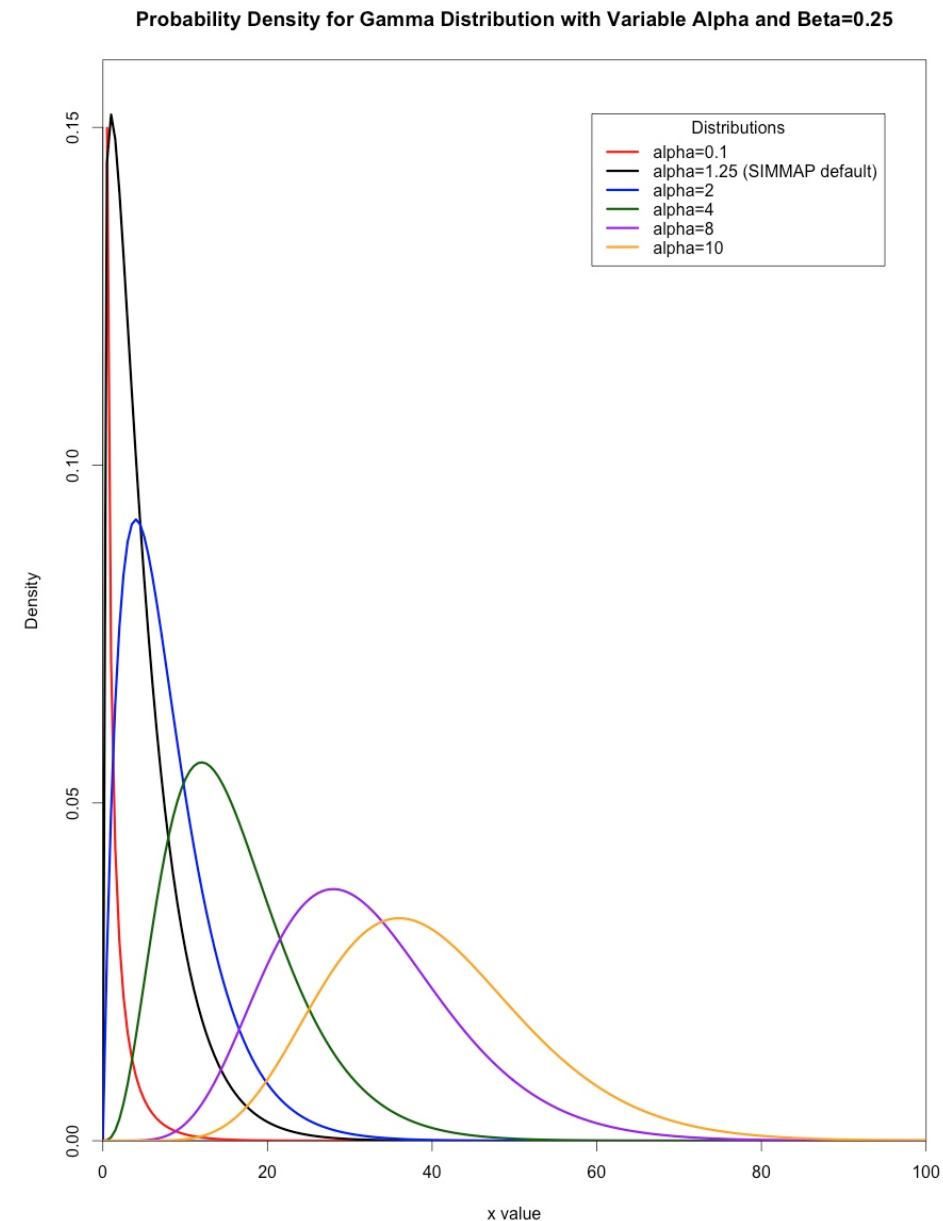
- Probabilidad a posteriori
  - Calcula la probabilidad de que un árbol sea correcto basado en una serie de datos y modelos previos
  - Utiliza un algoritmo llamado Cadenas de Markov
  - Toma un árbol inicial y empieza a crear generaciones de nuevos árboles con modificaciones al primero.
  - Basado en toda la información previa define cuál es el árbol con la mayor probabilidad a posterior de cumplir con la hipótesis





# Distribución gamma

		Second Letter					
		U	C	A	G		
1st letter	U	UUU   Phe UUC   UUA   Leu UUG	UCU   UCC   Ser UCA   UCG	UAU   Tyr UAC   UAA   Stop UAG   Stop	UGU   Cys UGC   UGA   Stop UGG   Trp	U C A G	3rd letter
	C	CUU   CUC   Leu CUA   CUG	CCU   CCC   Pro CCA   CCG	CAU   His CAC   CAA   Gln CAG	CGU   CGC   Arg CGA   CGG	U C A G	
	A	AUU   AUC   Ile AUA   AUG   Met	ACU   ACC   Thr ACA   ACG	AAU   Asn AAC   AAA   Lys AAG	AGU   Ser AGC   AGA   Arg AGG	U C A G	
	G	GUU   GUC   Val GUA   GUG	GCU   GCC   Ala GCA   GCG	GAU   Asp GAC   GAA   Glu GAG	GGU   GGC   Gly GGA   GGG	U C A G	





## Cómo construir un árbol?

- Gen (o genes) “neutrales” que sean compartidos entre todas las especies a utilizar
- 16S, 18S, COI, etc.
- Los genes a utilizar pueden variar dependiendo del grupo de interés
- Outgroup taxonómico: Especie que sirve como base para plantar la raíz del árbol filogenético



# GenBank

Nucleotide    
[Create alert](#) [Advanced](#)

Items: 1 to 20 of 92534

<< First < Prev Page 1

☐ [Citrobacter freundii strain B38 chromosome, complete sequence](#)

1. 5,134,500 bp circular DNA

Accession: CP016762.1 GI: 1050862732

[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Citrobacter freundii partial 16S rRNA gene, strain HLR20](#)

2. 1,359 bp linear DNA

Accession: HG974539.1 GI: 646226450

[PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Citrobacter freundii IRQBAS-155 gene for 16S rRNA, partial sequence](#)

3. 1,419 bp linear DNA



- Pegar la secuencia en un bloc de notas

### Citrobacter freundii partial 16S rRNA gene, strain HLR20

GenBank: HG974539.1

[GenBank](#) [Graphics](#)

```
>HG974539.1 Citrobacter freundii partial 16S rRNA gene, strain HLR20
GGAGCTTGCTCCTTGGGTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAACTGCCCGATGGAGGGGG
ATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCT
TGCCATCGGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTA
GCTGGTCTGAGAGGATGACCAGCCACACTGGAACCTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGT
GGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTG
TAAAGTACTTTTCAGCGAGGAGGAAGGCGTTGTGGTTAATAACCGCAGCGATTGACGTTACTCGCAGAAGA
AGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGG
CGTAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCC
GAAACTGGCAGGCTAGAGTCTTGAGAGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATC
TGGAGGAATACCGGTGGCGAAGGCGGCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAG
CAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGC
GTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAACTCAAATGA
ATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTAC
TCTTGACATCCAGAGAACTTAGCAGAGATGCTTTGGTGCCTTCGGGAACTCTGAGACAGGTGCTGCATGG
CTGTCGTAGCTCGTGTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGC
CAGCGGTTAGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAG
TCATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCATATACAAAGAGAAGCGACCTCGCG
AGAGCAAGCGGACCTCATAAAGTATGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGG
AATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCTTGACACACCGCCCGTCA
CACCATGGGAGTGGGTTGCAAAAGAAGTA
```





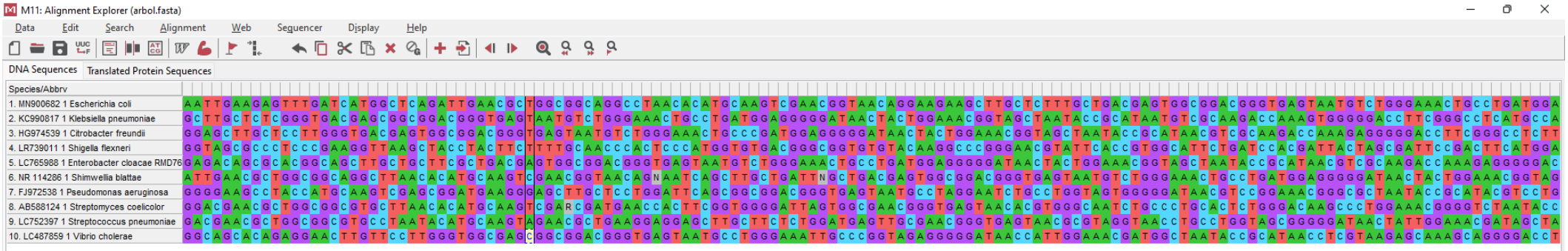
- Combinar todas las secuencias en un solo archivo

```
>MN900682_1 Escherichia coli
AATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCTAACACATGCAAGTCGAACGGTA
ACAGGAAGAAGCTTGCTCTTTGCTGACGAGTGGCGGACGGGTAGTAATGTCTGGGAAACTGCCTGATGG
AGGGGGATAACTACTGGAACCGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGG
GCCTCTTGCCATCGGATGTGCCAGATGGGATTAGCTAGTAGGTGGGTAAACGGCTCACCTAGGCGACGA
TCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAGTGAAGACAGGTCCAGACTCCTACGGAGGC
AGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTC
GGGTTGTAAGTACTTTTCAGCGGGGAGGAAGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGC
AGAAGAAGCACCAGCTAACTCCGTGCCAGCAGCCGCGTAATACGGAGGGTGCAAGCGTTAATCGGAATT
ACTGGGCGTAAAGCGCAGCAGGCGGTTTGTAAAGTCAGATGTGAAATCCCGGGCTCAACCTGGGAAC
GCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGTAGAATTCAGGTGTAGCGGTGAAATGCGTA
GAGATCTGGAGGAATACCGGTGGCGAAGCGGCCCTTGGACGAAGACTGACGCTCAGGTGCGAAAGCGT
GGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCGTAACGATGTCGACTTGGAGGTTGTGCCCT
TGAGGCGTGGCTTCGGAGCTAACGCGTTAAGTCGACCGCTGGGGAGTACGGCCGAAGGTTAAACCTC
AAATGAATTGACGGGGGCCGACAAAGCGGTGGAGCATGTGGTTTAAATTCGATGCAACGCGAAGAACCCT
ACCTGGTCTTGACATCCACGGAAGTTTTTCAGAGATGAGAATGTGCCTTCGGGAACCGTGAGACAGGTGCT
GCATGGCTGTCGTGAGCTCGTGTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCTTATCCTT
TGTGCGCAGCGGTCGGCCGGGAACCAAGGAGACTGCCAGTGATAAATGGAGGAAGGTGGGGATGAC
GTCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAGAGAAGCGAC
CTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGA
AGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCTTGTACACACCGC
CCGTACACCAATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACTTT
GTGATTCATGACTGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCT
>KC990817_1 Klebsiella pneumoniae
GCTTGCTCTCGGGTGACGAGCGCGGACGGGTAGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAA
CTACTGGAACCGTAGCTAATACCGCATAATGTGCGCAAGACCAAAGTGGGGGACCTTCGGGCCATGCC
ATCAGATGTGCCAGATGGGATTAGCTAGTAGGTGGGTAAACGGCTCACCTAGGCGACGATCCCTAGCTG
GTCTGAGAGGATGACCAGCCACACTGGAAGTGAAGACAGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG
AATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGAAA
GCACCTTCAGCGGGGAGGAAGGCATAAGGTTAATAACCTTGTGATTGACGTTACCCGCAAGAAGCA
CCGGCTAACTCCGTGCCAGCAGCCGCGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA
AAGCGCACGAGCGGTCTGTCAAGTCGGATGTGAAATCCCGGGCTCAACCTGGGAAGTGCATTGAAA
CTGGCAGGCTAGAGTCTTGTAGAGGGGGTAGAATTCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGA
GGAATACGGTGGCGAAGCGGCCCTTGGACAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAA
CAGGATTAGATACCTGCTGATGTCACGCGGTAAACGATGTCGATTGAGGTTGTGCTTGAAGCGTGG
```

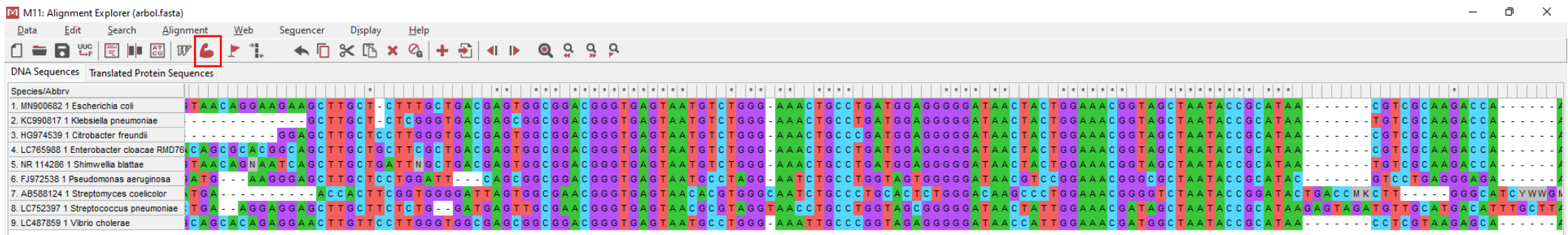
Ln 45, Col 25



# Alineamiento



Muscle

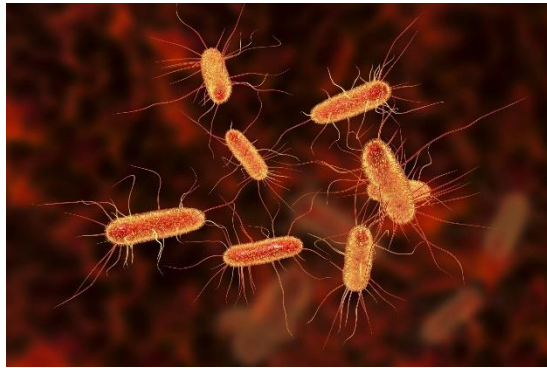


Clustalw

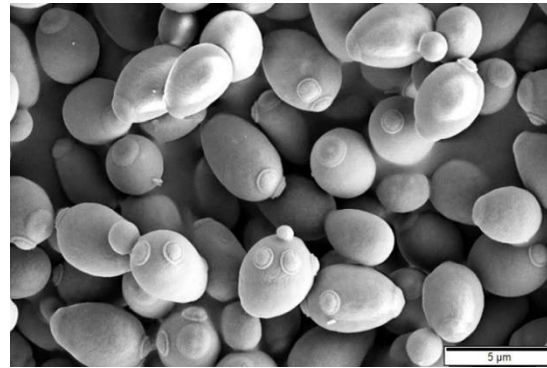


# Mejor modelo evolutivo

- Frecuencia de tetranucleóticos (ATGC)
- Contenido GC



*E. coli* GC: 50%



*S. cerevisiae* GC: 40%



*D. melanogaster* GC: 53%



## DNA models

### Base substitution rates

IQ-TREE includes all common DNA models (ordered by complexity):

Model	df	Explanation	Code
JC or JC69	0	Equal substitution rates and equal base frequencies (Jukes and Cantor, 1969).	000000
F81	3	Equal rates but unequal base freq. (Felsenstein, 1981).	000000
K80 or K2P	1	Unequal transition/transversion rates and equal base freq. (Kimura, 1980).	010010
HKY or HKY85	4	Unequal transition/transversion rates and unequal base freq. (Hasegawa, Kishino and Yano, 1985).	010010
TN or TN93	5	Like <b>HKY</b> but unequal purine/pyrimidine rates (Tamura and Nei, 1993).	010020
TNe	2	Like <b>TN</b> but equal base freq.	010020
K81 or K3P	2	Three substitution types model and equal base freq. (Kimura, 1981).	012210
K81u	5	Like <b>K81</b> but unequal base freq.	012210
TPM2	2	AC=AT, AG=CT, CG=GT and equal base freq.	010212
TPM2u	5	Like <b>TPM2</b> but unequal base freq.	010212
TPM3	2	AC=CG, AG=CT, AT=GT and equal base freq.	012012
TPM3u	5	Like <b>TPM3</b> but unequal base freq.	012012
TIM	6	Transition model, AC=GT, AT=CG and unequal base freq.	012230
TIMe	3	Like <b>TIM</b> but equal base freq.	012230
TIM2	6	AC=AT, CG=GT and unequal base freq.	010232
TIM2e	3	Like <b>TIM2</b> but equal base freq.	010232
TIM3	6	AC=CG, AT=GT and unequal base freq.	012032
TIM3e	3	Like <b>TIM3</b> but equal base freq.	012032
TVM	7	Transversion model, AG=CT and unequal base freq.	012314
TVMe	4	Like <b>TVM</b> but equal base freq.	012314
SYM	5	Symmetric model with unequal rates but equal base freq. (Zharkikh, 1994).	012345
GTR	8	General time reversible model with unequal rates and unequal base freq. (Tavare, 1986).	012345

The last column **Code** is a 6-digit code defining the equality constraints for 6 *relative* substitution rates: A-C, A-G, A-T, C-G, C-T and G-T.

**010010** means that A-G rate is equal to C-T rate (corresponding to **1** in the code) and the remaining four substitution rates are equal



# Model selection may not be a mandatory step for phylogeny reconstruction

[Shiran Abadi](#), [Dana Azouri](#), [Tal Pupko](#) & [Itay Mayrose](#)

*Nature Communications* **10**, Article number: 934 (2019) | [Cite this article](#)

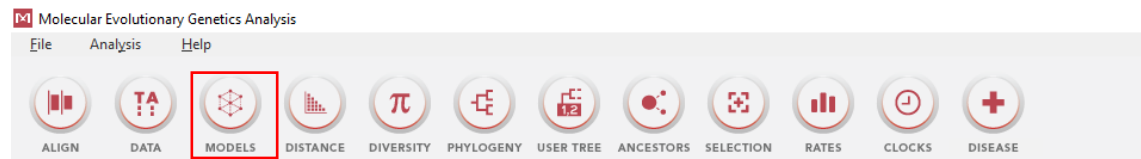
## Table 1 Model selection criteria procedures

From: [Model selection may not be a mandatory step for phylogeny reconstruction](#)

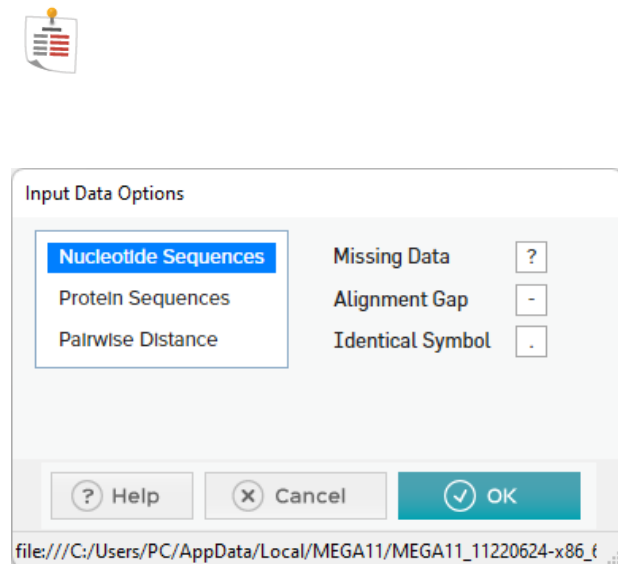
Criterion	Procedure
AIC	ML is computed for every candidate model and the model with minimal $\{-2\ell + 2K\}$ is selected
AICc	Based on AIC but penalizes also for the data size. Namely, the model with minimal $\{AIC + \frac{2K(K+1)}{n-K-1}\}$ is selected; advised to be used instead of AIC when $\frac{n}{K} < 40$
BIC	ML is computed for every candidate model and the model with minimal $\{-2\ell + K \ln n\}$ is selected
DT	Based on BIC but incorporates relative branch-length error as a performance measure
hLRT/dLRT	Sequential likelihood ratio tests between pairs of nested models until one cannot be rejected. Topologies are fixed to allow nesting. While in hLRT the order in which parameters are added is defined a priori, in dLRT all models that differ in one parameter are compared in parallel and the hierarchy proceeds with the model that maximizes the log-likelihood difference. Thus, dLRT enables a different order of hypotheses testing for different datasets
BF	The ratio between the marginal likelihood of two models. A ratio above 10 implies strong support for the model at the numerator



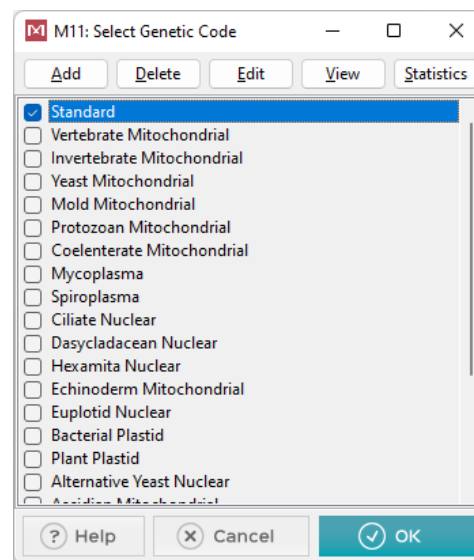
1)



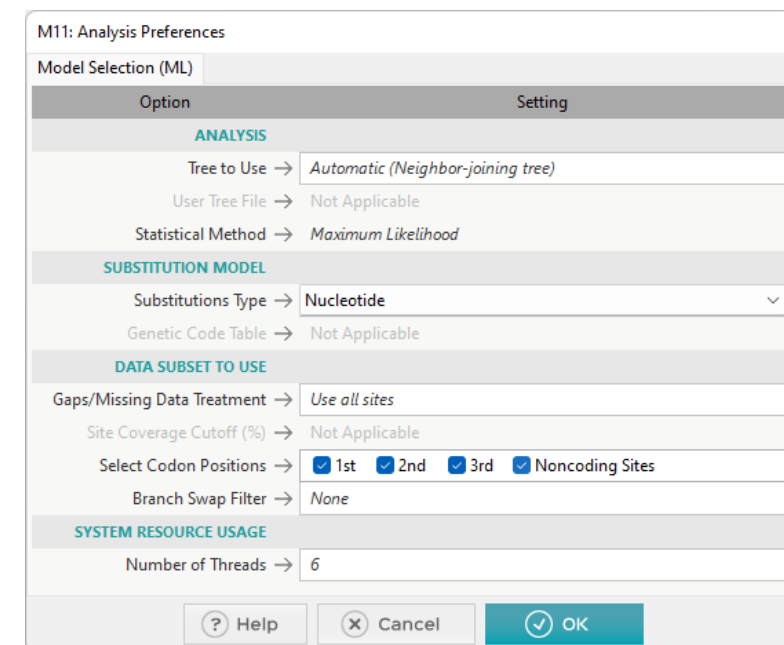
2)



3)



4)







MEGA Caption Expert: Find Best-Fit Substitution Model (ML)

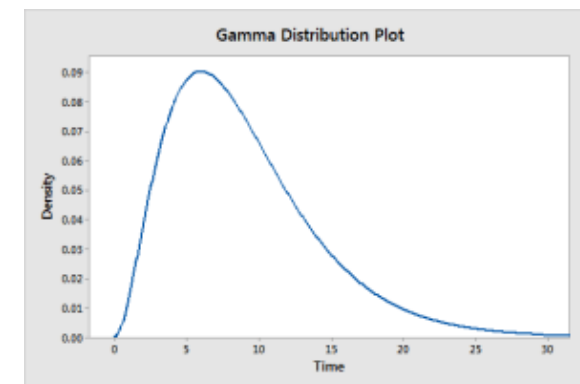
File Edit View Help

TXT

Results

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	$\hat{\pi}(A)$	$\hat{\pi}(T)$	$\hat{\pi}(C)$	$\hat{\pi}(G)$	$\hat{\pi}(AT)$	$\hat{\pi}(AC)$	$\hat{\pi}(AG)$	$\hat{\pi}(TA)$	$\hat{\pi}(TC)$	$\hat{\pi}(TG)$	$\hat{\pi}(CA)$	$\hat{\pi}(CT)$	$\hat{\pi}(CG)$	$\hat{\pi}(GA)$	$\hat{\pi}(GT)$	$\hat{\pi}(GC)$
TN93+G	19	10564.156	10425.668	-5193.799	n/a	0.33	1.61	0.249	0.201	0.231	0.318	0.038	0.043	0.123	0.047	0.217	0.060	0.047	0.188	0.060	0.096	0.038	0.043
TN93+G+I	20	10573.449	10427.675	-5193.799	0.00	0.32	1.61	0.249	0.201	0.231	0.318	0.038	0.043	0.123	0.047	0.217	0.060	0.047	0.189	0.060	0.096	0.038	0.043
T92+G	16	10574.367	10457.736	-5212.843	n/a	0.32	1.58	0.225	0.225	0.275	0.275	0.043	0.053	0.169	0.043	0.169	0.053	0.043	0.138	0.053	0.138	0.043	0.053
T92+G+I	17	10583.660	10459.743	-5212.843	0.00	0.32	1.58	0.225	0.225	0.275	0.275	0.043	0.053	0.169	0.043	0.169	0.053	0.043	0.138	0.053	0.138	0.043	0.053
K2+G	15	10583.896	10474.551	-5222.254	n/a	0.32	1.56	0.250	0.250	0.250	0.250	0.049	0.049	0.152	0.049	0.152	0.049	0.049	0.152	0.049	0.152	0.049	0.049
HKY+G	18	10588.406	10457.203	-5210.570	n/a	0.32	1.63	0.249	0.201	0.231	0.318	0.039	0.045	0.195	0.048	0.142	0.062	0.048	0.123	0.062	0.153	0.039	0.045
GTR+G	22	10588.645	10428.302	-5192.104	n/a	0.33	1.62	0.249	0.201	0.231	0.318	0.049	0.033	0.123	0.061	0.217	0.059	0.036	0.189	0.058	0.096	0.037	0.042
K2+G+I	16	10593.189	10476.558	-5222.254	0.00	0.32	1.56	0.250	0.250	0.250	0.250	0.049	0.049	0.152	0.049	0.152	0.049	0.049	0.152	0.049	0.152	0.049	0.049
HKY+G+I	19	10597.700	10459.211	-5210.570	0.00	0.32	1.63	0.249	0.201	0.231	0.318	0.039	0.045	0.195	0.048	0.142	0.062	0.048	0.123	0.062	0.153	0.039	0.045
GTR+G+I	23	10597.937	10430.310	-5192.104	0.00	0.32	1.62	0.249	0.201	0.231	0.318	0.049	0.033	0.123	0.061	0.217	0.059	0.036	0.189	0.058	0.096	0.037	0.042
JC+G	14	10718.233	10616.175	-5294.068	n/a	0.36	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
TN93+I	19	10721.891	10583.402	-5272.666	0.33	n/a	1.31	0.249	0.201	0.231	0.318	0.043	0.049	0.121	0.053	0.192	0.068	0.053	0.167	0.068	0.094	0.043	0.049
JC+G+I	15	10727.526	10618.182	-5294.069	0.00	0.36	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
T92+I	16	10737.576	10620.945	-5294.447	0.33	n/a	1.30	0.225	0.225	0.275	0.275	0.049	0.059	0.156	0.049	0.156	0.059	0.049	0.128	0.059	0.128	0.049	0.059
K2+I	15	10745.762	10636.418	-5303.187	0.33	n/a	1.29	0.250	0.250	0.250	0.250	0.055	0.055	0.141	0.055	0.141	0.055	0.055	0.141	0.055	0.141	0.055	0.055
GTR+I	22	10746.242	10585.899	-5270.903	0.33	n/a	1.31	0.249	0.201	0.231	0.318	0.052	0.041	0.121	0.065	0.191	0.072	0.045	0.166	0.062	0.094	0.045	0.045
HKY+I	18	10753.046	10621.843	-5292.890	0.33	n/a	1.31	0.249	0.201	0.231	0.318	0.044	0.051	0.179	0.055	0.130	0.070	0.055	0.113	0.070	0.140	0.044	0.051
TN93	18	10847.591	10716.388	-5340.162	n/a	n/a	1.21	0.249	0.201	0.231	0.318	0.045	0.052	0.122	0.056	0.180	0.071	0.056	0.157	0.071	0.095	0.045	0.052
JC+I	14	10865.799	10763.742	-5367.852	0.33	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
T92	15	10865.986	10756.641	-5363.299	n/a	n/a	1.21	0.225	0.225	0.275	0.275	0.051	0.062	0.151	0.051	0.151	0.062	0.051	0.124	0.062	0.124	0.051	0.062
GTR	21	10871.715	10718.656	-5338.286	n/a	n/a	1.21	0.249	0.201	0.231	0.318	0.053	0.044	0.122	0.065	0.179	0.077	0.047	0.156	0.066	0.095	0.049	0.048
K2	14	10873.060	10771.003	-5371.482	n/a	n/a	1.21	0.250	0.250	0.250	0.250	0.057	0.057	0.137	0.057	0.137	0.057	0.057	0.137	0.057	0.137	0.057	0.057
HKY	17	10882.988	10759.070	-5362.507	n/a	n/a	1.21	0.249	0.201	0.231	0.318	0.046	0.053	0.172	0.057	0.125	0.073	0.057	0.109	0.073	0.135	0.046	0.053
JC	13	10986.432	10891.662	-5432.814	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083





1)



2)

M11: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
<b>ANALYSIS</b>	
Statistical Method →	Maximum Likelihood
<b>PHYLOGENY TEST</b>	
Test of Phylogeny →	Bootstrap method
No. of Bootstrap Replications →	100
<b>SUBSTITUTION MODEL</b>	
Substitutions Type →	Nucleotide
Genetic Code Table →	Not Applicable
Model/Method →	Tamura-Nei model
<b>RATES AND PATTERNS</b>	
Rates among Sites →	Gamma Distributed (G)
No of Discrete Gamma Categories →	5
<b>DATA SUBSET TO USE</b>	
Gaps/Missing Data Treatment →	Use all sites
Site Coverage Cutoff (%) →	Not Applicable
Select Codon Positions →	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
<b>TREE INFERENCE OPTIONS</b>	
ML Heuristic Method →	Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML →	Make initial tree automatically (Default - NJ/BioNJ)
Initial Tree File →	Not Applicable
Branch Swap Filter →	None
<b>SYSTEM RESOURCE USAGE</b>	
Number of Threads →	6

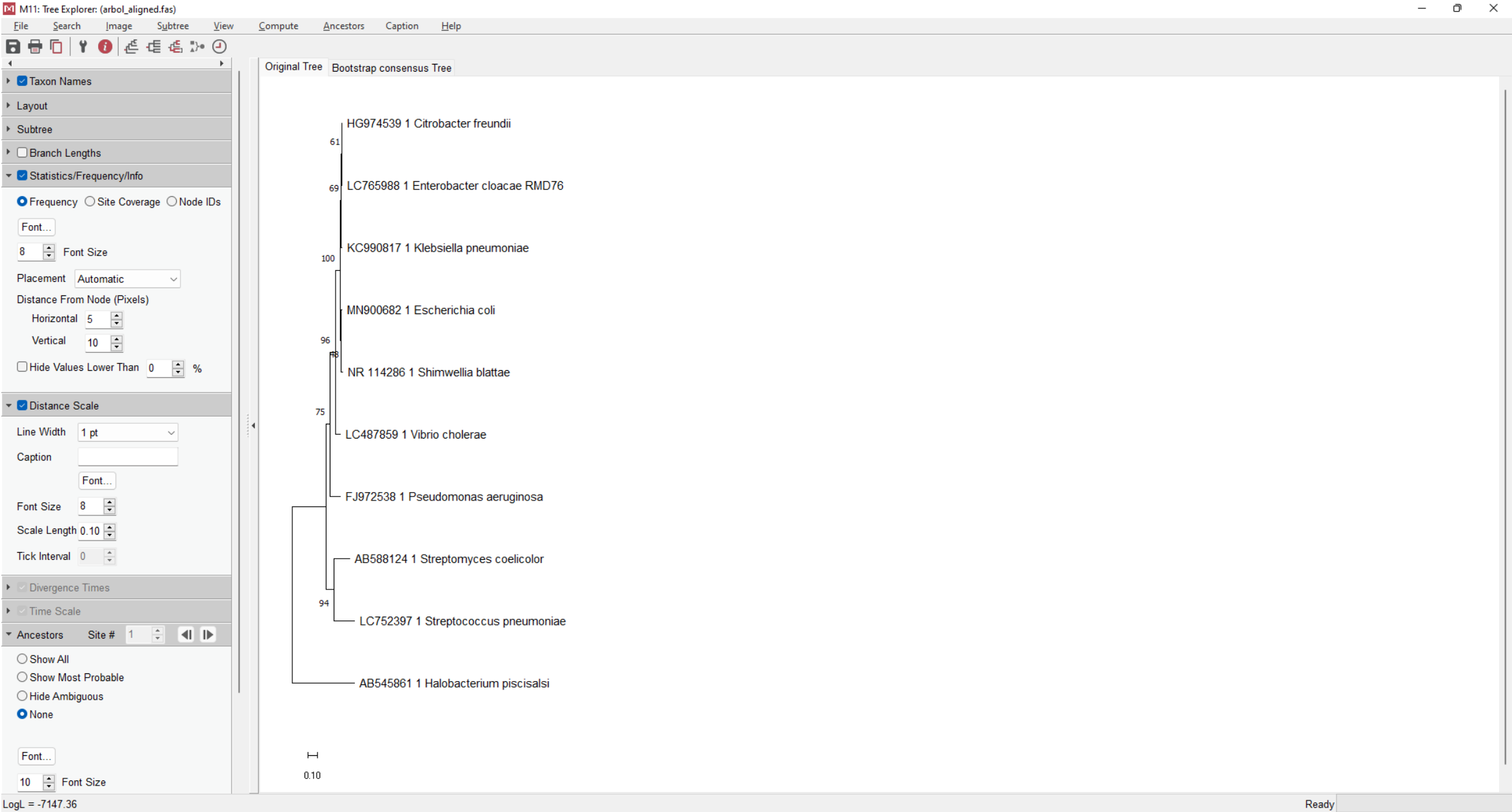
3)

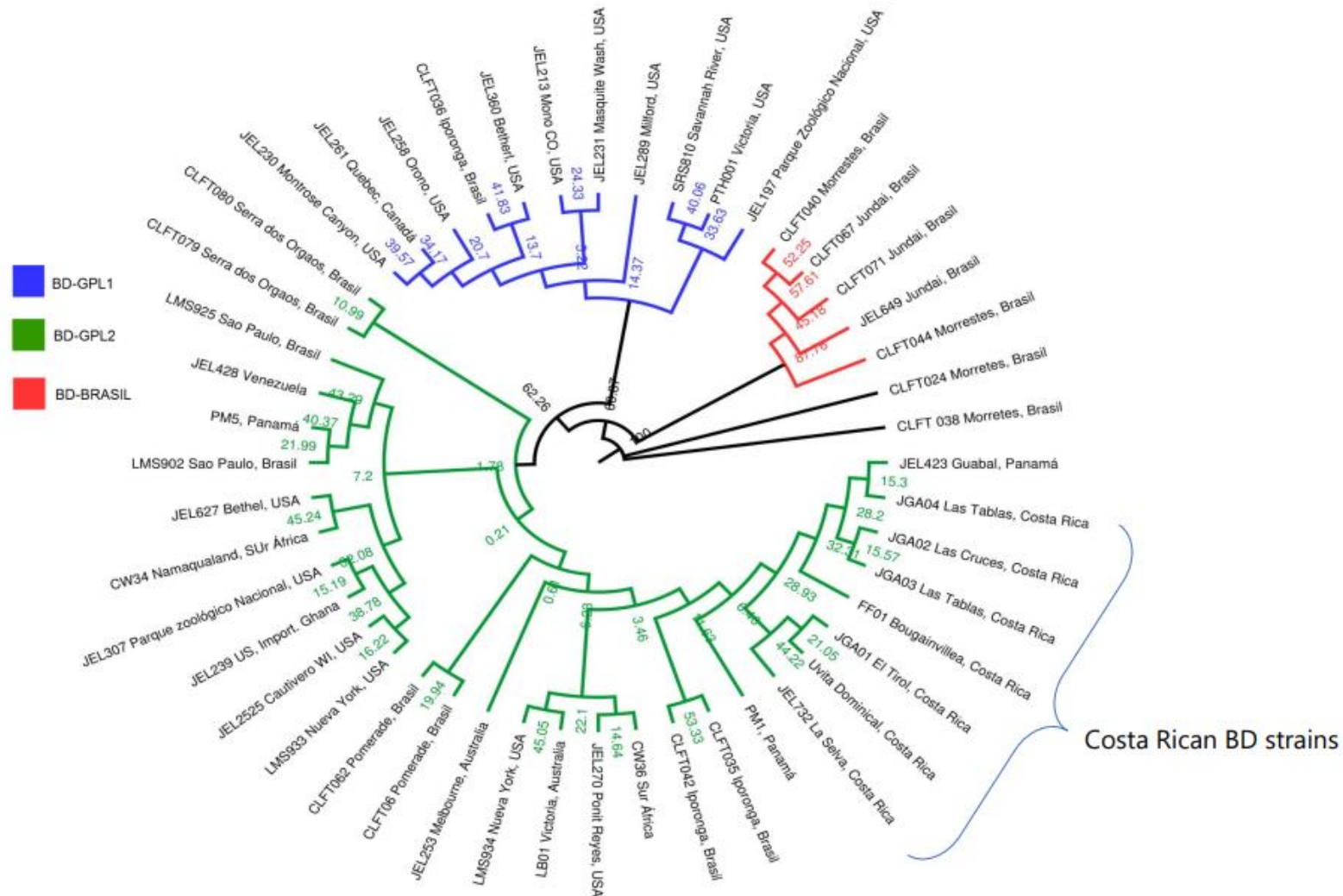
M11: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
<b>ANALYSIS</b>	
Statistical Method →	Maximum Likelihood
<b>PHYLOGENY TEST</b>	
Test of Phylogeny →	Bootstrap method
No. of Bootstrap Replications →	100
<b>SUBSTITUTION MODEL</b>	
Substitutions Type →	Nucleotide
Genetic Code Table →	Not Applicable
Model/Method →	Tamura-Nei model
<b>RATES AND PATTERNS</b>	
Rates among Sites →	Gamma Distributed (G)
No of Discrete Gamma Categories →	Uniform Rates
<b>DATA SUBSET TO USE</b>	
Gaps/Missing Data Treatment →	Use all sites
Site Coverage Cutoff (%) →	Not Applicable
Select Codon Positions →	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
<b>TREE INFERENCE OPTIONS</b>	
ML Heuristic Method →	Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML →	Make initial tree automatically (Default - NJ/BioNJ)
Initial Tree File →	Not Applicable
Branch Swap Filter →	None
<b>SYSTEM RESOURCE USAGE</b>	
Number of Threads →	6







**Figure 3.** Dendrogram showing the genetic relationships between BD strains based on MLST. The dendrogram was created using seven molecular markers employing the "Neighbor-Joining" method and a bootstrap of 10000 subsamples.

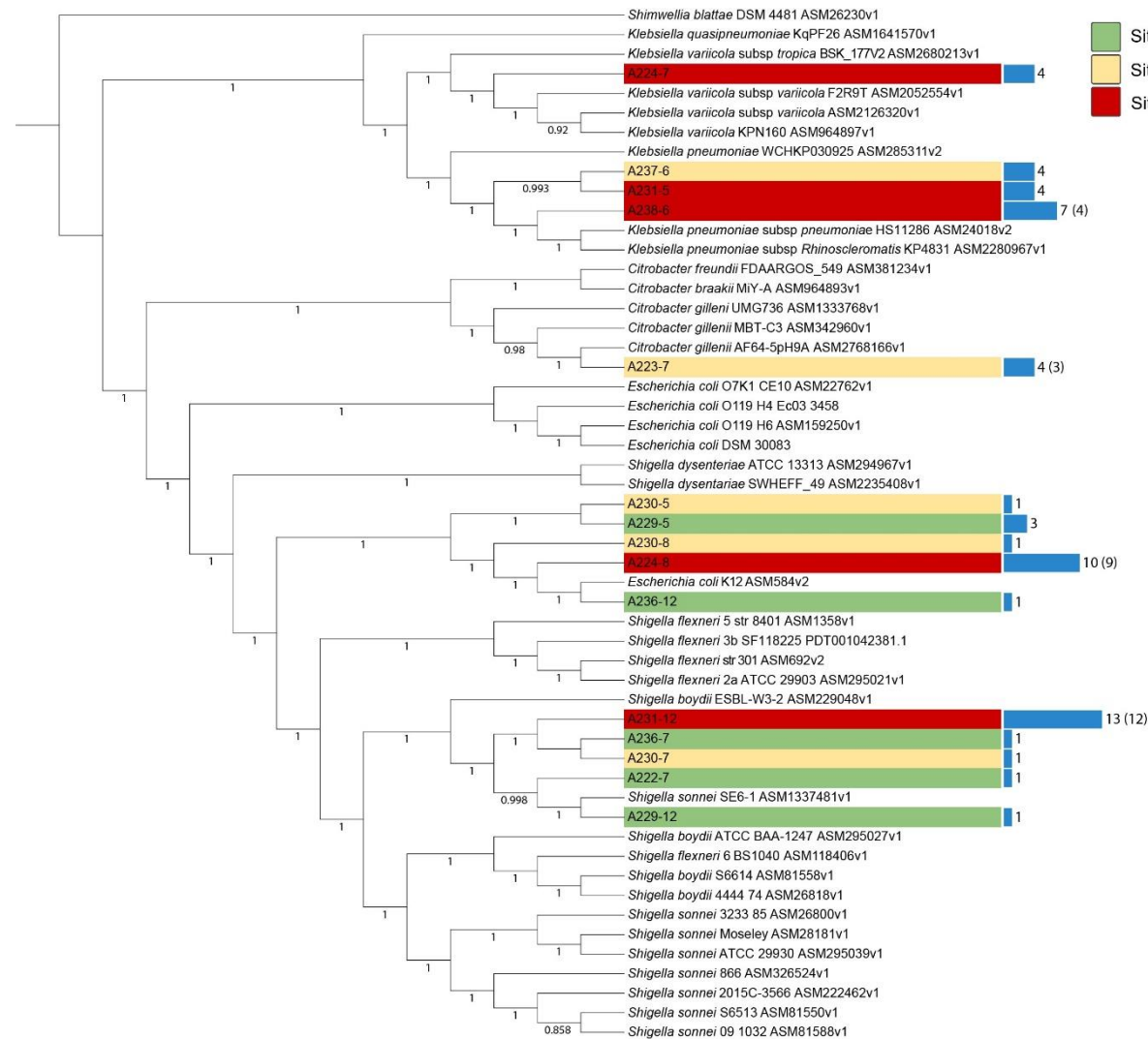
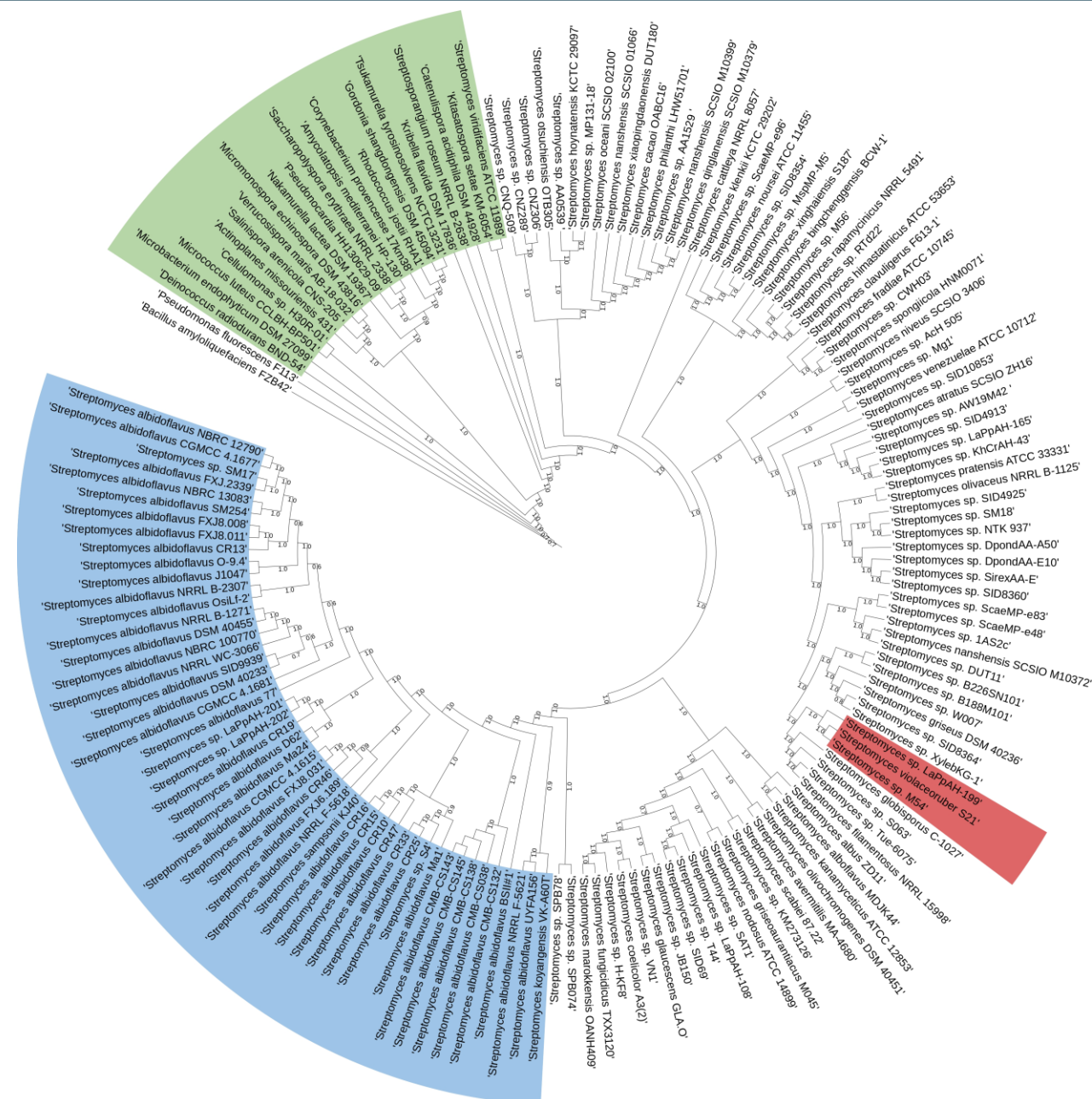


Fig4. Árbol filogenómico representando las relaciones evolutivas de las bacterias aisladas. En los nodos se muestran los valores de bootstrap, los colores representan los sitios de dónde se aislaron las bacterias y las barras en azul representan el número de ARGs presentes (los números entre paréntesis representan los ARGs encontrados en plásmidos). Se utilizaron 1489 genes de copia única y el método de máxima verosimilitud con el modelo evolutivo GTR.



- Árbol filogenómico mostrando las relaciones evolutivas del género Streptomyces. En Verde se muestran los genomas más basales utilizados como outgroup, en azul el grupo de Streptomyces albidoflavus y en rojo el clado asociado al aislamiento de interés. Se utilizaron 71 genes de copia única y se realizó el método de inferencia bayesiana con el modelo evolutivo GTR con distribución gamma y proporción de sitios invariables.





Gracias 😊

**If you had a time  
machine what would  
you do?  
Me:**

