

# Práctica 4

Construcción de árboles  
con parsimonia y  
distancias

dos casos prácticos sencillos

hecho por:

José Manuel Sánchez Aquilué



MEGA



# Instalación

- <https://www.megasoftware.net/>



- `sudo dpkg -i path/to/the/deb/package`
- En windows, si se tienen problemas con la instalación, se recomienda deshabilitar el control de cuentas de usuarios
- [https://www.megasoftware.net/web\\_help\\_10/index.htm#t=Part\\_I\\_Getting\\_Started%2FInstalling\\_MEGA%2FInstalling\\_MEGA.htm](https://www.megasoftware.net/web_help_10/index.htm#t=Part_I_Getting_Started%2FInstalling_MEGA%2FInstalling_MEGA.htm)

# Molecular Evolutionary Genetics Analysis



File Analysis Help



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



RATES



CLOCKS



DIAGNOSE

NEXUS



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES



ANALYZE

PROTOTYPE



# Cargar secuencias

- File → Open A File/ Session
- Desde MEGA ofrece la opción de alinear o analizar secuencias
- Para analizar secuencias hace falta que estén alineadas, en caso contrario, dará error
- Antes de cargar los datos se nos preguntará por el tipo de secuencias (Proteínas, DNA) y el tipo de código genético(estándar, mitocondrial)

# Molecular Evolutionary Genetics Analysis



File Analysis Help



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



RATES



CLOCKS



DIAGNOSE



NEXUS



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES



ANALYZE

PROTOTYPE



# Cálculo de distancias

- MEGA permite seleccionar el tipo de distancia (número de diferencias, la p-distancia)
- Con o sin bootstrap
- La matriz de distancias resultante se puede exportar en excel, csv, mega o txt

**MX: Analysis Preferences**

Distance Estimation

Option	Setting
<b>ANALYSIS</b>	
Scope	→ Pairs of taxa
<b>ESTIMATE VARIANCE</b>	
Variance Estimation Method	→ None
No. of Bootstrap Replications	→ Not Applicable
<b>SUBSTITUTION MODEL</b>	
Substitutions Type	→ Nucleotide
Genetic Code Table	→ Not Applicable
Model/Method	→ No. of differences
Fixed Transition/Transversion Ratio	→ Not Applicable
Substitutions to Include	→ d: Transitions + Transversions
<b>RATES AND PATTERNS</b>	
Rates among Sites	→ Uniform Rates
Gamma Parameter	→ Not Applicable
Pattern among Lineages	→ Same (Homogeneous)
<b>DATA SUBSET TO USE</b>	
Gaps/Missing Data Treatment	→ Pairwise deletion
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"/> Non

? Help    X Cancel    ✓ OK

# Matriz de distancias por número de diferencias

MX: Pairwise Distances (apartado2\_0.fasta)

File Display Average Caption Help

	1	2	3	4	5	6	7	8	9	10
1. G65269.1 DIAPH2-3 utr-11 X chromosome Homo sapiens STS genomic										
2. G65268.1 MCM-5 X chromosome Homo sapiens STS genomic sequence	232.00									
3. G65267.1 DIAPH2-3 utr-3 X chromosome Homo sapiens STS genomic s	187.00	147.00								
4. G65266.1 MCM-14 X chromosome Homo sapiens STS genomic sequen	263.00	226.00	195.00							
5. G65265.1 MCM-12 X chromosome Homo sapiens STS genomic sequen	265.00	221.00	191.00	250.00						
6. G65264.1 MCM-10 X chromosome Homo sapiens STS genomic sequen	148.00	110.00	100.00	121.00	154.00					
7. G65263.1 DIAPH2-3 utr-7 X chromosome Homo sapiens STS genomic s	285.00	249.00	188.00	261.00	267.00	149.00				
8. G65262.1 DIAPH2-3 utr-12 X chromosome Homo sapiens STS genomic	266.00	221.00	192.00	229.00	238.00	139.00	252.00			
9. G65261.1 DIAPH2-3 utr-14 X chromosome Homo sapiens STS genomic	240.00	191.00	202.00	247.00	238.00	131.00	165.00	220.00		
10. G65260.1 MCM-8 X chromosome Homo sapiens STS genomic sequen	222.00	189.00	164.00	206.00	218.00	110.00	223.00	214.00	221.00	

# Matriz de p-distancias

MX: Pairwise Distances (apartado2\_0.fasta)

File Display Average Caption Help

	1	2	3	4	5	6	7	8	9	10
1. G65269.1 DIAPH2-3 utr-11 X chromosome Homo sapiens STS genomic										
2. G65268.1 MCM-5 X chromosome Homo sapiens STS genomic sequence	0.625									
3. G65267.1 DIAPH2-3 utr-3 X chromosome Homo sapiens STS genomic s	0.645	0.586								
4. G65266.1 MCM-14 X chromosome Homo sapiens STS genomic sequen	0.634	0.582	0.602							
5. G65265.1 MCM-12 X chromosome Homo sapiens STS genomic sequen	0.634	0.601	0.606	0.575						
6. G65264.1 MCM-10 X chromosome Homo sapiens STS genomic sequen	0.619	0.588	0.571	0.500	0.618					
7. G65263.1 DIAPH2-3 utr-7 X chromosome Homo sapiens STS genomic s	0.677	0.711	0.646	0.635	0.650	0.654				
8. G65262.1 DIAPH2-3 utr-12 X chromosome Homo sapiens STS genomic	0.665	0.674	0.662	0.604	0.626	0.618	0.655			
9. G65261.1 DIAPH2-3 utr-14 X chromosome Homo sapiens STS genomic	0.669	0.661	0.667	0.677	0.657	0.675	0.442	0.632		
10. G65260.1 MCM-8 X chromosome Homo sapiens STS genomic sequen	0.615	0.580	0.550	0.526	0.591	0.462	0.635	0.618	0.680	



# Molecular Evolutionary Genetics Analysis



File Analysis Help



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



RATES



CLOCKS



DIAGNOSE



NEXUS



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES



ANALYZE

PROTOTYPE



# Construcción de árboles por distancias

- Phylogeny → Construct/Test Neighbor-Joining Tree
- Phylogeny → Construct/Test UPGMA Tree
- Una vez generado, podemos guardar el árbol como newick, imagen, pdf o MEGA.

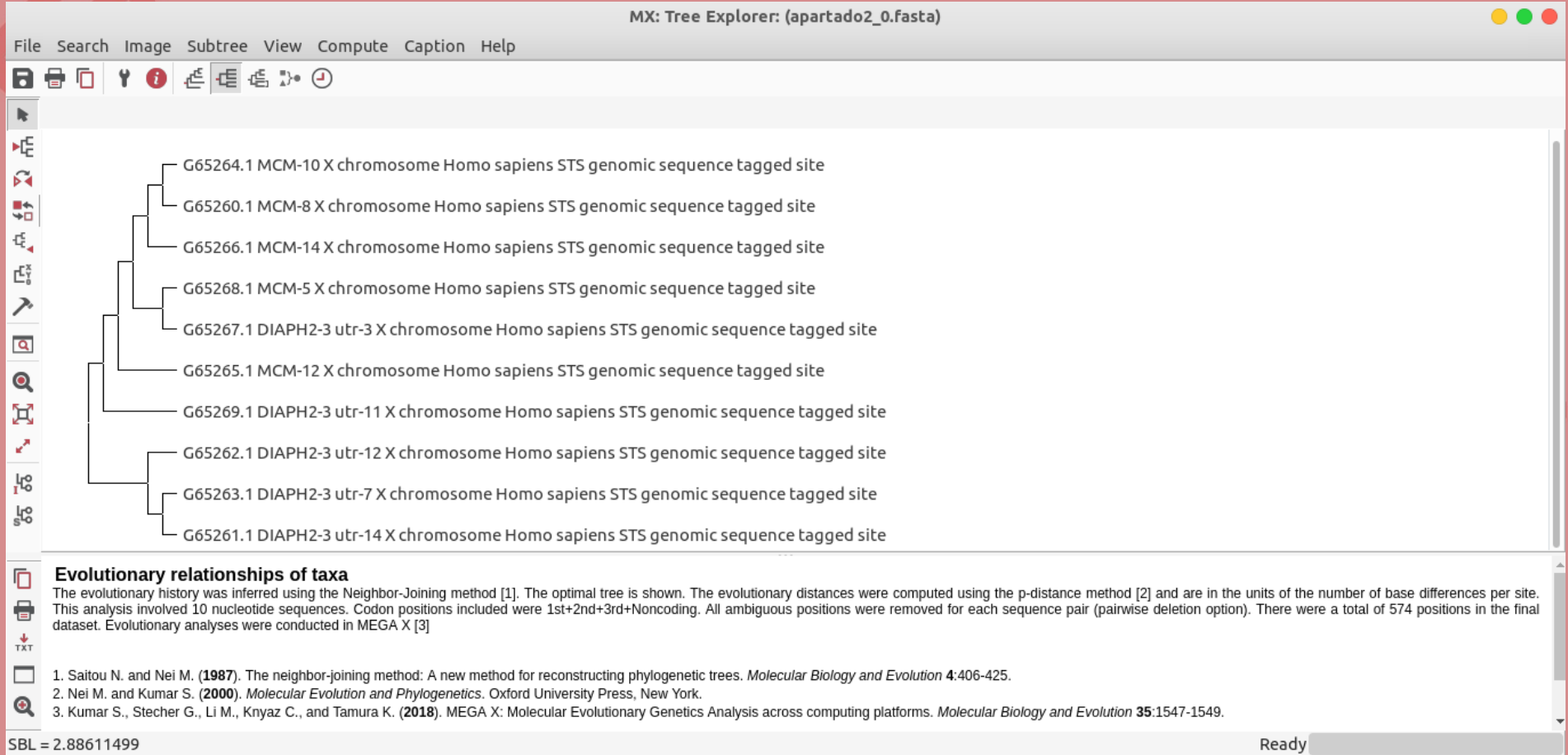
MX: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
<b>ANALYSIS</b>	
Statistical Method	→ Maximum Likelihood
<b>PHYLOGENY TEST</b>	
Test of Phylogeny	→ None
No. of Bootstrap Replications	→ Not Applicable
<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	→ Uniform Rates
No of Discrete Gamma Categories	→ Not Applicable
<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 Site
<b>TREE INFERENCE OPTIONS</b>	
ML Heuristic Method	→ Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML	→ Make initial tree automatically (Default - NJ/BioN
Initial Tree File	→ Not Applicable
Branch Swap Filter	→ None
<b>SYSTEM RESOURCE USAGE</b>	
Number of Threads	→ 3

? Help X Cancel ✓ OK

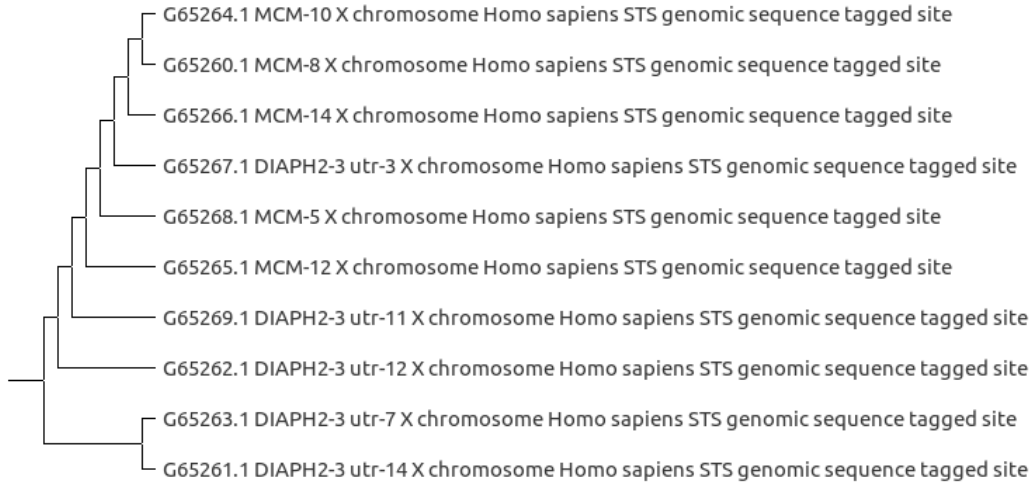
# NJ



# UPGMA

MX: Tree Explorer: (apartado2\_0.fasta)

File Search Image Subtree View Compute Caption Help



## Evolutionary relationships of taxa

The evolutionary history was inferred using the UPGMA method [1]. The optimal tree is shown. The evolutionary distances were computed using the p-distance method [2] and are in the units of the number of base differences per site. This analysis involved 10 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 574 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [3]

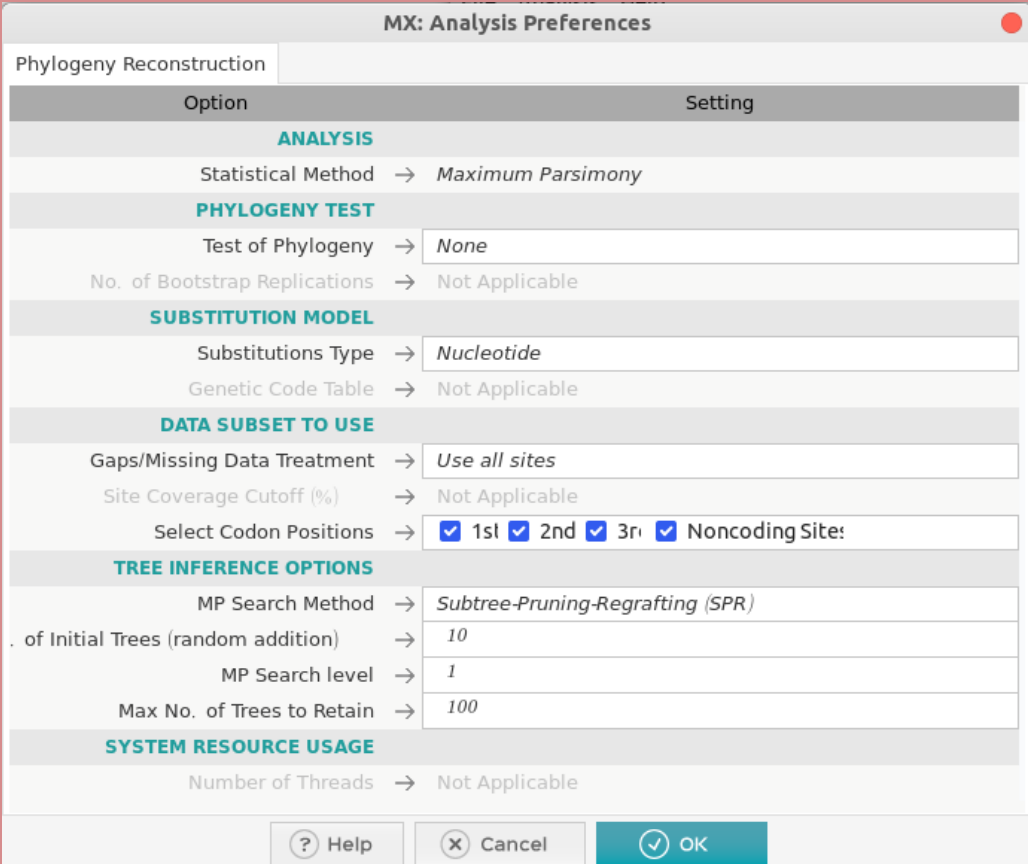
1. Sneath P.H.A. and Sokal R.R. (1973). *Numerical Taxonomy*. Freeman, San Francisco.
2. Nei M. and Kumar S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
3. Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549.

SBL = 2.88174103

Ready

# Construcción de árboles por máxima parsimonia

- Phylogeny → Construct/Test Maximum Parsimony Tree
- Conviene utilizar bootstrap
- Probar valores entre 10 y 100 iteraciones
- Para comparar la fiabilidad estadística de los distintos tipos de árboles habrá que ejecutarlos con bootstrap



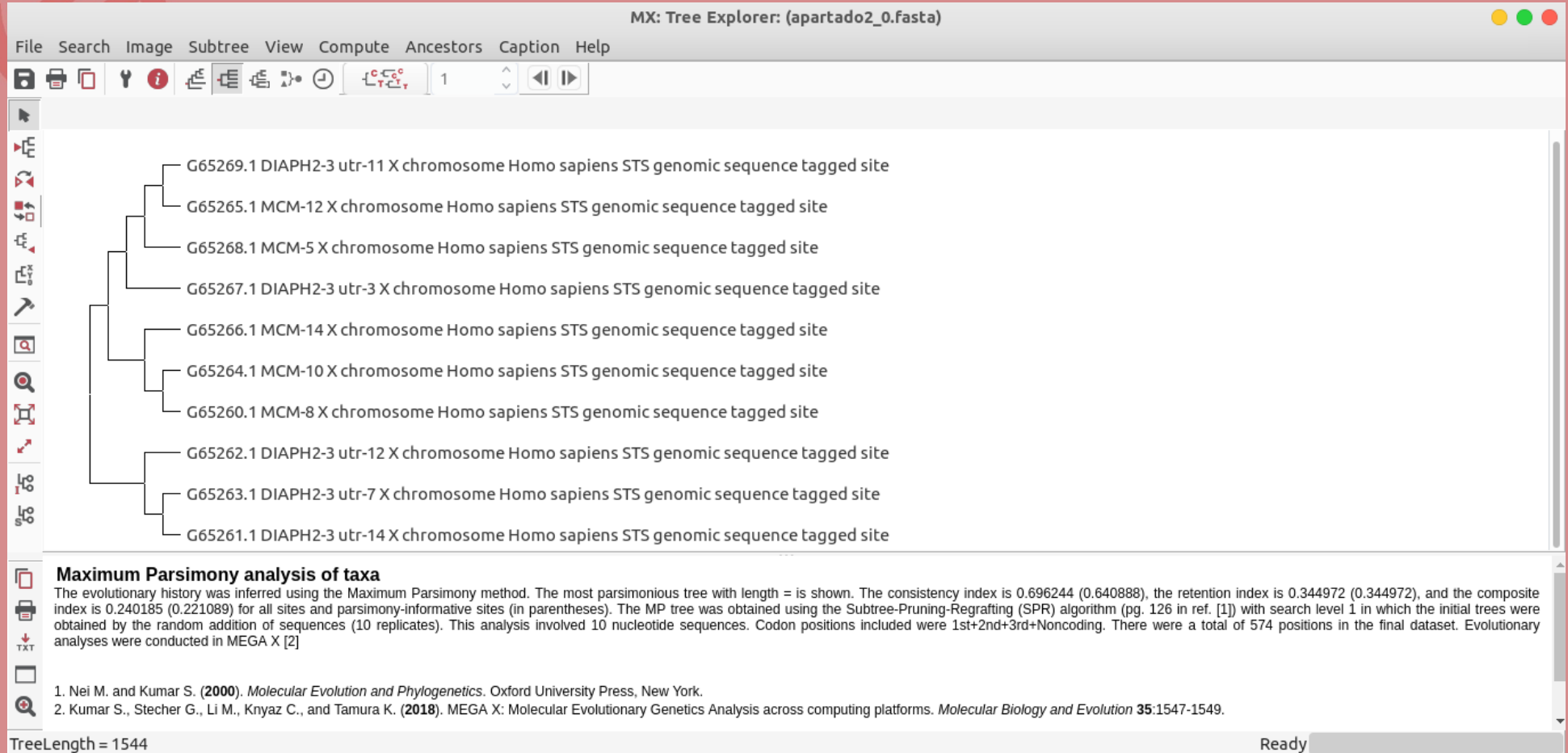
**MX: Analysis Preferences**

Phylogeny Reconstruction

Option	Setting
<b>ANALYSIS</b>	
Statistical Method	→ <i>Maximum Parsimony</i>
<b>PHYLOGENY TEST</b>	
Test of Phylogeny	→ <i>None</i>
No. of Bootstrap Replications	→ <i>Not Applicable</i>
<b>SUBSTITUTION MODEL</b>	
Substitutions Type	→ <i>Nucleotide</i>
Genetic Code Table	→ <i>Not Applicable</i>
<b>DATA SUBSET TO USE</b>	
Gaps/Missing Data Treatment	→ <i>Use all sites</i>
Site Coverage Cutoff (%)	→ <i>Not Applicable</i>
Select Codon Positions	→ <input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Site:
<b>TREE INFERENCE OPTIONS</b>	
MP Search Method	→ <i>Subtree-Pruning-Regrafting (SPR)</i>
No. of Initial Trees (random addition)	→ <i>10</i>
MP Search level	→ <i>1</i>
Max No. of Trees to Retain	→ <i>100</i>
<b>SYSTEM RESOURCE USAGE</b>	
Number of Threads	→ <i>Not Applicable</i>

? Help    X Cancel    OK

# Ejemplo



ETE  
treeview



# Introducción

- Es una herramienta online, no necesita instalación
- Se le pasa un árbol en formato newick y un alineamiento en formato FASTA y muestra una visualización de ambos
- Hecha con ETE toolkit, una librería de python
- <http://etetoolkit.org/treeview/>



# Tipos de visualizaciones

- Solo el árbol
- El árbol junto a las categorías taxonómicas a las que pertenece cada especie
- El árbol junto al alineamiento (por bloques o en formato condensado)
- A continuación se mostrarán ejemplos

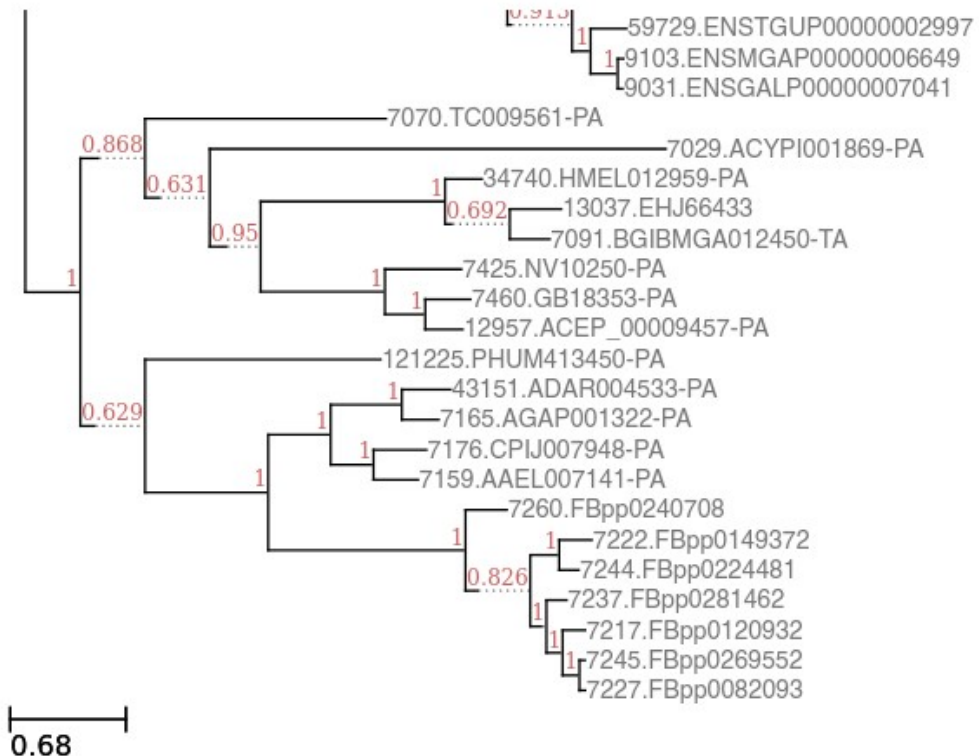
View tree!

Alignment image type:

Do not display alignment ▼

- ☐ Resolve taxonomic ids

Share (beta): <http://etetoolkit.org/treeview/?treeid=2d28bfcbcd37bd91980b303f599f97d3&algid=ce443ed1e53858bf4e11d1e069c7a927>



View tree!

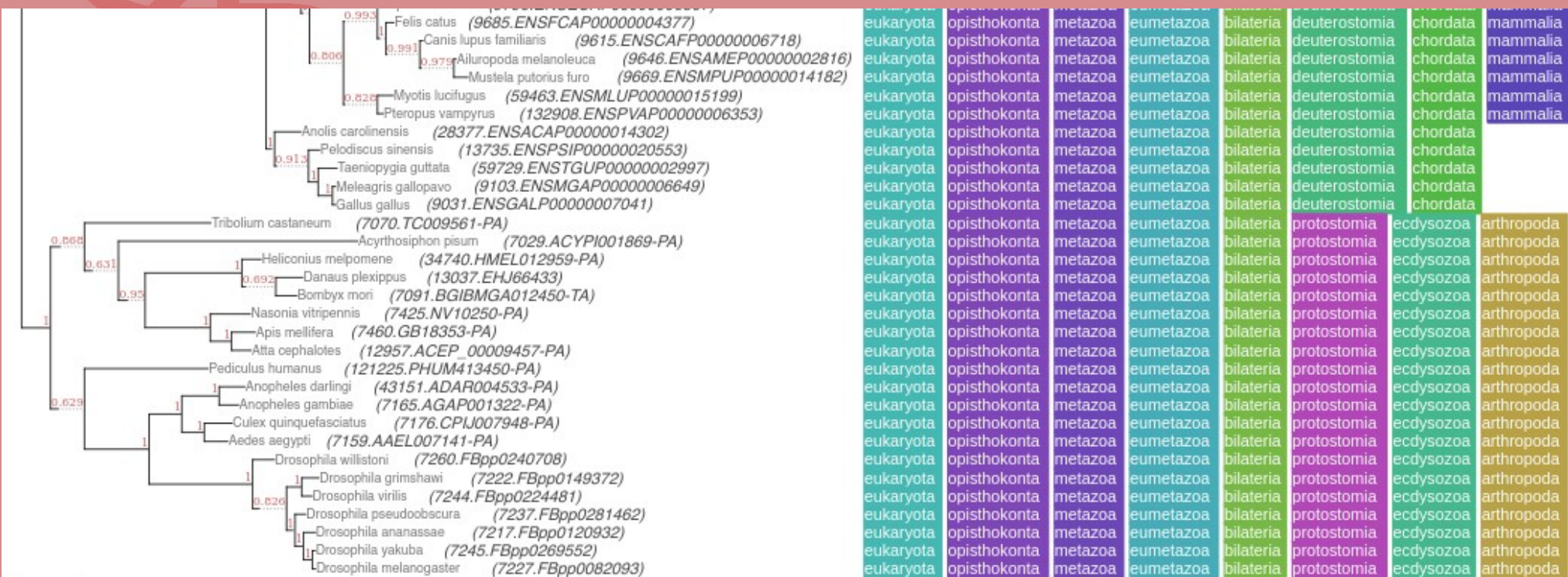
Alignment image type:

Do not display alignment ▾



Resolve taxonomic ids

Share (beta): <http://et toolkit.org/treeview/?treeid=2d28bfbcb37bd91980b303f599f97d3&algid=ce443ed1e53858bf4e11d1e069c7a927>



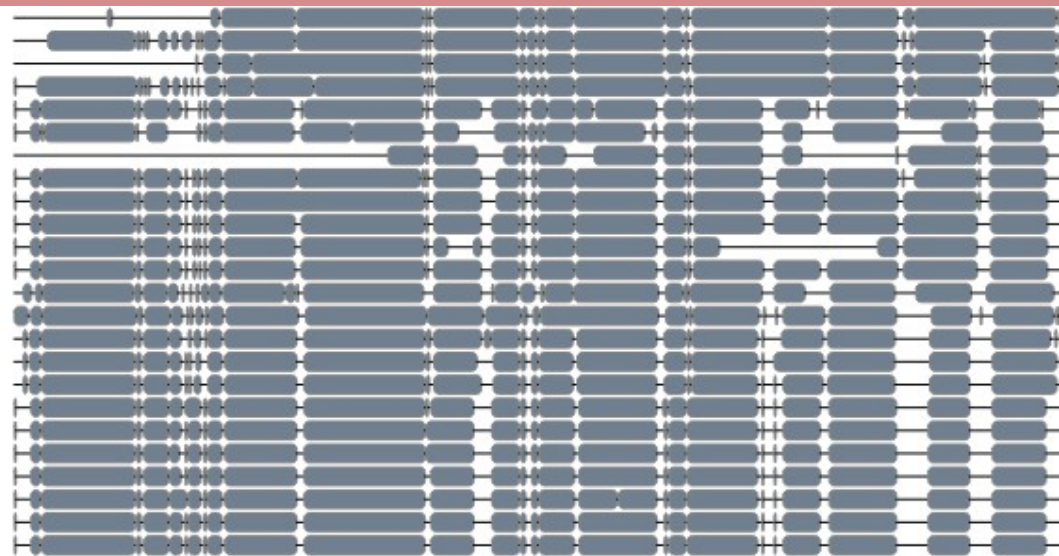
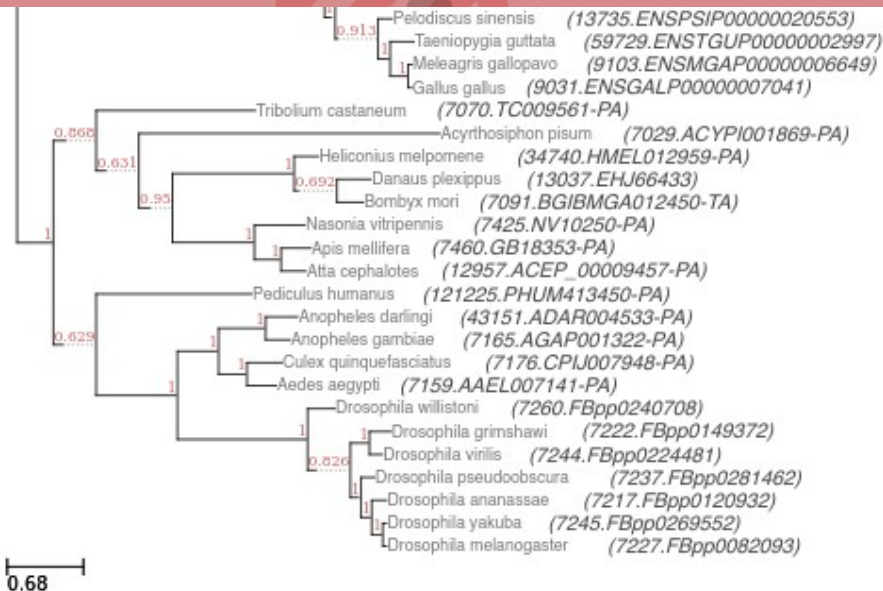
View tree!

Alignment image type: Condensed format



☒ Resolve taxonomic ids

Share (beta): <http://etetoolkit.org/treeview/?treeid=2d28bfcbcd37bd91980b303f599f97d3&algid=ce443ed1e53858bf4e11d1e069c7a927>





View tree!

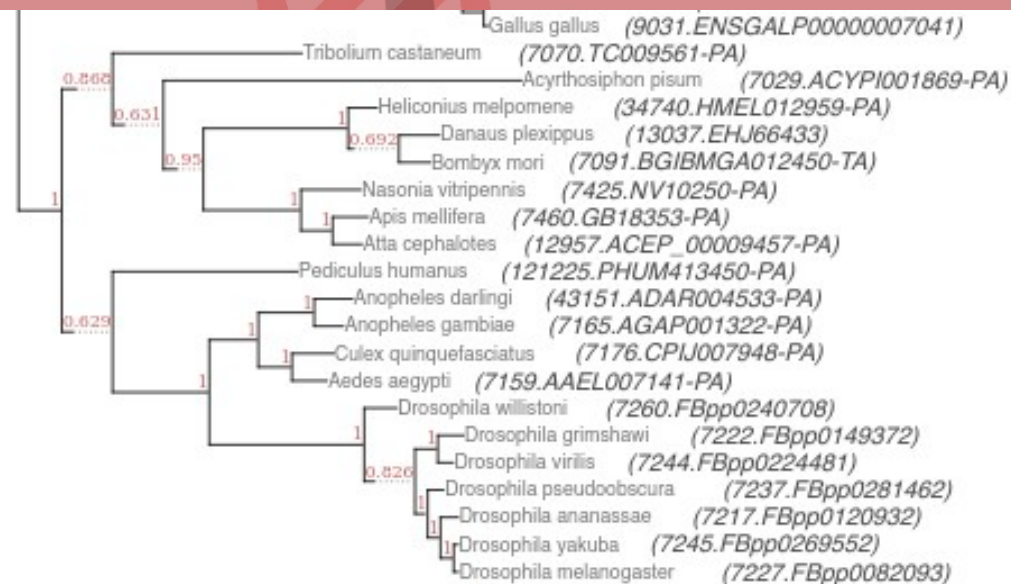
Alignment image type:

Condensed format



☒ Resolve taxonomic ids

Share (beta): <http://etetoolkit.org/treeview/?treeid=2d28bfcbcd37bd91980b303f599f97d3&algid=ce443ed1e53858bf4e11d1e069c7a927>



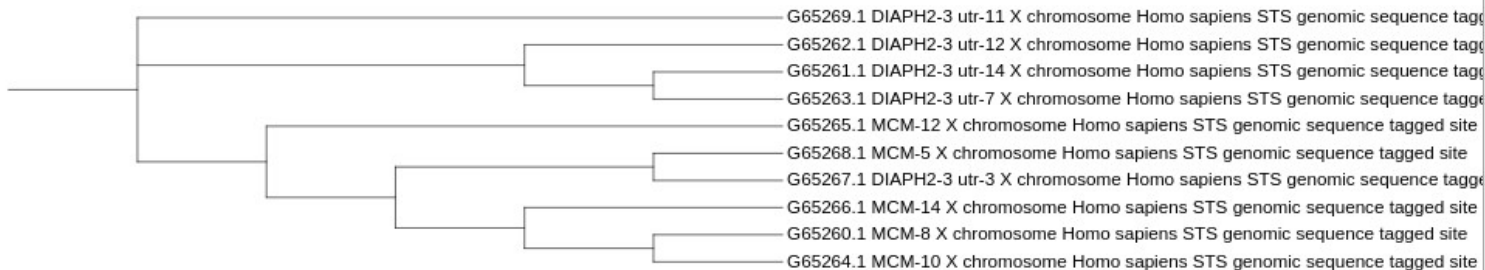
0.68

itol



# Introducción

- Herramienta online de edición y visualización de arboles
- Soporta distintos formatos: Newick, Nexus o PhyloXML
- <https://itol.embl.de/upload.cgi>



Controls

Basic

Advanced

Datasets

Export

Display mode

Circular

Normal

Unrooted

Parameters

0

° rotation

Invert

Yes

No

Slanted

Yes

No

Labels

Aligned

At tips

Off

Label rotation

On

Off

Label alignment

Left

Right

Label shift

0

Label font

Arial

Add

Font style

20

px

Font style

B

I

Branch lines

1

px

Curved

Normal

Branch gradients

On

Off

Dashed lines

0.3

px

Dashed lines

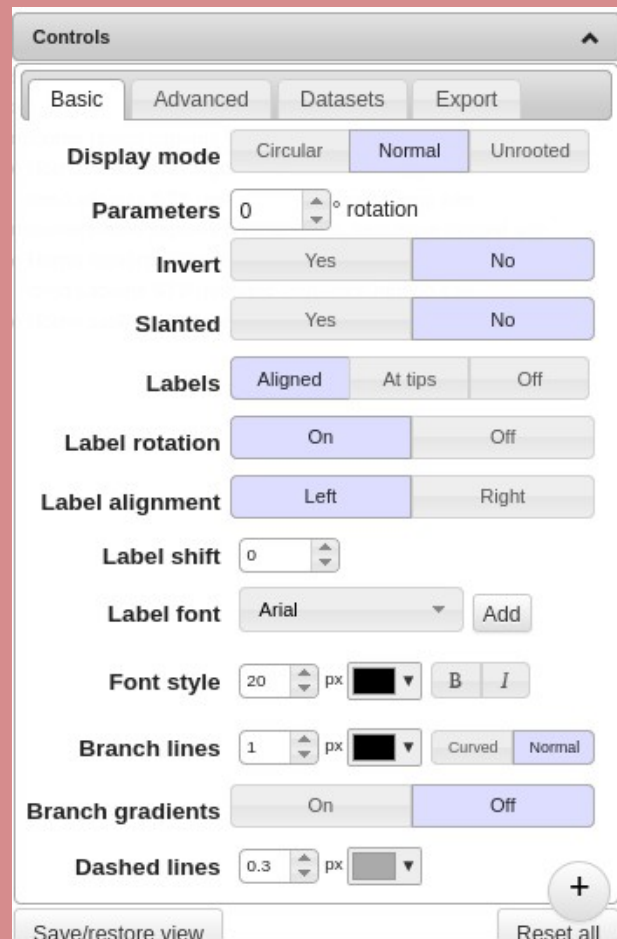
Save/restore view

Reset all



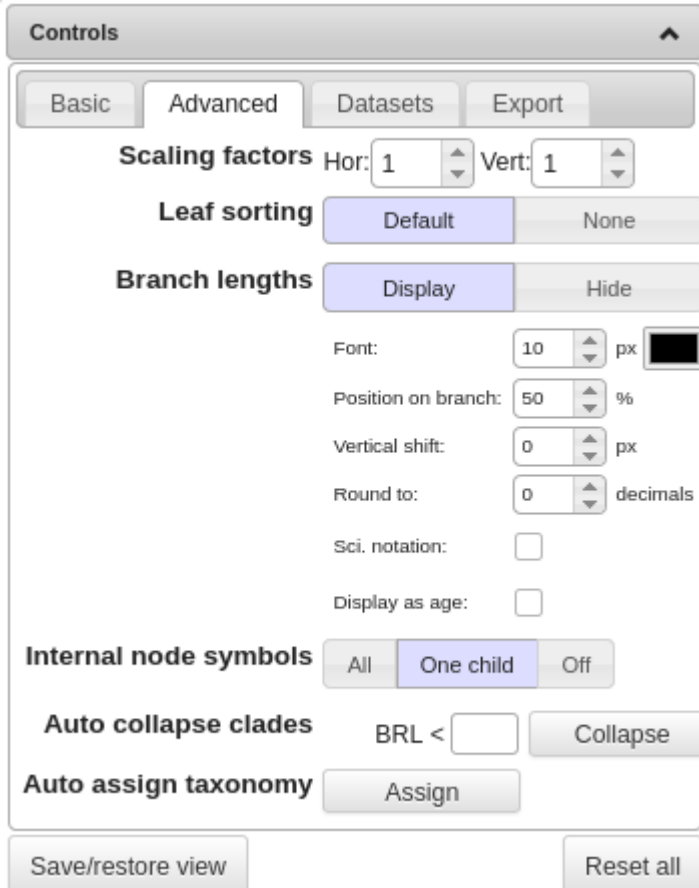
# Ajustes básicos

- Tipo de árbol: normal, circular
- Rotación, inversión, inclinación...
- Ajustes de etiquetas: fuente, color, tamaño, rotación, alineación...
- Ajustes de ramas: tamaño y color de líneas...



# Ajustes avanzados

- Escalado horizontal y vertical
- Mostrar longitud y ajustar el formato de esta
- Compresión de clados (agrupación común) automática
- Asignación taxonómica automática

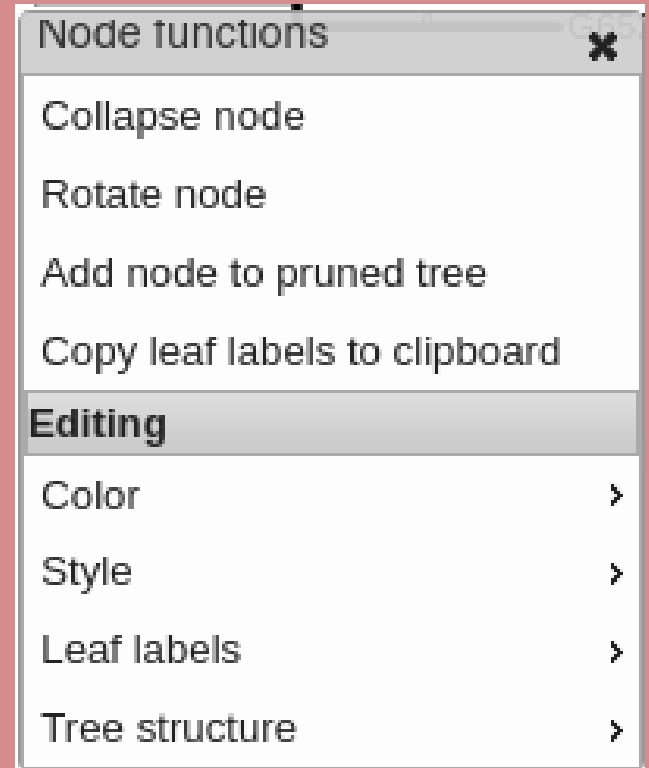


The screenshot shows the 'Controls' window with the 'Advanced' tab selected. The settings are as follows:

- Scaling factors:** Hor: 1, Vert: 1
- Leaf sorting:** Default
- Branch lengths:** Display
- Font:** 10 px
- Position on branch:** 50 %
- Vertical shift:** 0 px
- Round to:** 0 decimals
- Sci. notation:** ☐
- Display as age:** ☐
- Internal node symbols:** One child
- Auto collapse clades:** BRL < [ ] Collapse
- Auto assign taxonomy:** Assign
- Buttons:** Save/restore view, Reset all

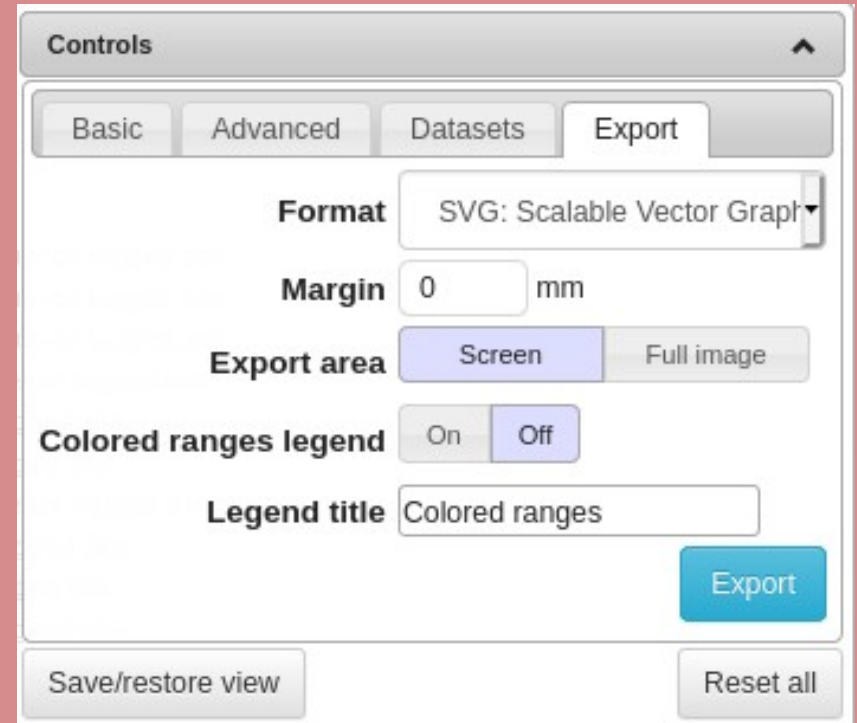
# Ajustes de nodo

- Compresión de nodo
- Rotación de nodo
- Poda
- Copiar hojas
- Edición: cambiar color, estilo, ajustes de etiqueta...
- Cambiar la raíz del nodo
- Comprimir, mover o eliminar clado



# Exportación

- Formato SVG, EPS, PDF, PNG
- Formatos textuales Newick, Nexus, PhyloXML



**Controls**

Basic Advanced Datasets **Export**

**Format** SVG: Scalable Vector Graph

**Margin** 0 mm

**Export area** Screen Full image

**Colored ranges legend** On Off

**Legend title** Colored ranges

Export

Save/restore view Reset all

MUCHAS  
GRACIAS

