

Save aligned sequences

The screenshot shows the MEGA11 software interface. The top menu bar includes options like Data, Edit, Search, Alignment, Web, Sequencer, Display, and Help. A context menu is open over a sequence alignment, specifically for the 17th site, with the following options:

- Create New
- Open
- Open a Recently Used File
- Close
- Phylogenetic Analysis
- Save Session
- Export Alignment
  - DNA Sequences
  - Protein Sequences
- Translate/Untranslate
- Genetic Code
- Reverse Complement
  - Reverse
  - Complement
- Quit

A sub-menu for "Export Alignment" is open, showing three format options: MEGA Format, FASTA Format (which is selected and highlighted in blue), and NEXUS/PAUP Format.

The main workspace displays a sequence alignment with 33 rows of DNA sequences. Each row is color-coded by species abbreviation from the left column. The sequences are aligned at the 17th position, where most species show a 'G' or 'A'. The alignment is presented in a grid format with each column representing a nucleotide position.

At the bottom, there are controls for "Site # 1701" (with dropdown arrows), a radio button for "with" (selected) and "w/o gaps", and a status message "Selected genetic code: Standard".

MEGA11 Data Edit Search Alignment Web Sequencer Display Help

M11: Alignment Explorer (australia\_0423.mas)

DNA Sequences Translated Protein Sequences

Species/Abbrv

1. 5a.2a A/India/Pun-NIV319324/2021	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T - - -
2. 5a.2a.1 A/Alagoas/28876/2022	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
3. A/Chungbuk/871/2023	G G G C A A T C G G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
4. A/Ulsan/867/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
5. A/Netherlands/10512/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
6. A/Catalonia/NSVH102069383/2023	G G G C A A T C A G G T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
7. A/Wisconsin/37/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
8. A/Manitoba/RV00381/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
9. A/Denmark/516/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
10. A/BosniaandHerzegovina/UKC-Tuzla-0	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
11. A/Saskatchewan/SKFLU297285/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T G A
12. A/Washington/23/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
13. A/Romania/547395/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
14. A/Lisboa/15/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
15. A/Arkhangelsk/CRIE/608/2023	G G G C A A T C A G G T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
16. A/Norway/04936/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
17. A/Amazonas/2023-004064-IEC/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
18. A/Moscow oblast/CRIE/616/2023	G G G C A A T C A G G T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
19. A/Arizona/24/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
20. A/Iowa/25/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
21. A/Melilla/1041/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
22. A/Ceuta/565/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
23. A/Catalonia/NSVH198267435/2023	G G G C A A T C A G G T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
24. A/South Dakota/21/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
25. A/Mato Grosso do Sul/IAL/C10477/2022	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
26. A/Montana/24/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
27. A/District Of Columbia/08/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
28. A/Minnesota/16/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
29. A/Victoria/380B/2023(2)	G A G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
30. A/Victoria/380C/2023(2)	G A G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
31. A/Victoria/453/2023	G G G C A G T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
32. A/Victoria/454/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A
33. A/Victoria/488/2023	G G G C A A T C A G C T T C T G G A T G T G C T C T A A T G G G T C T C T A C A G T G T A G A A T A T G T A T T T A A

Site # 1701      with      w/o gaps      Selected genetic code: Standard

Save file as

Save As: australia\_0423\_aligned.fasta

Tags:

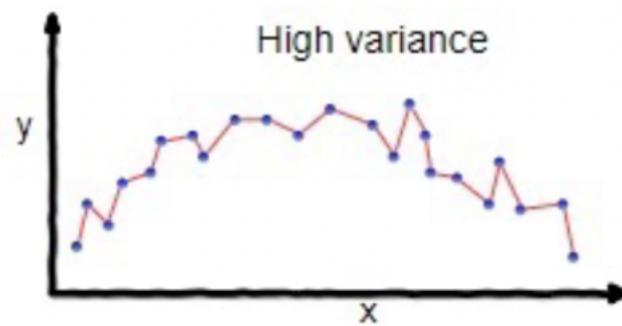
Where: 2023\_Timor-Leste

Format: FASTA Files

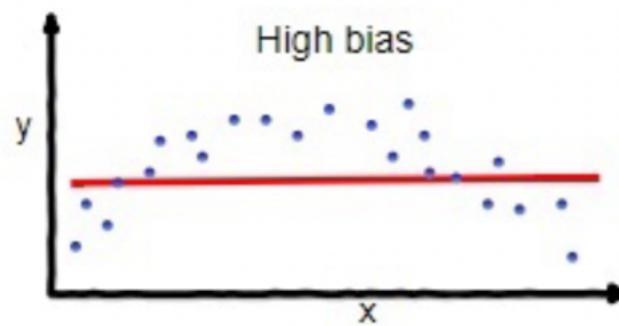
Cancel Save

Search for the best-fit nucleotide substitution model

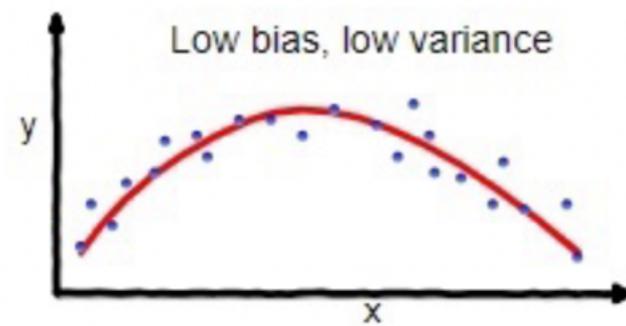
# Model Selection



overfitting



underfitting



Good balance

# Bias-variance trade-off

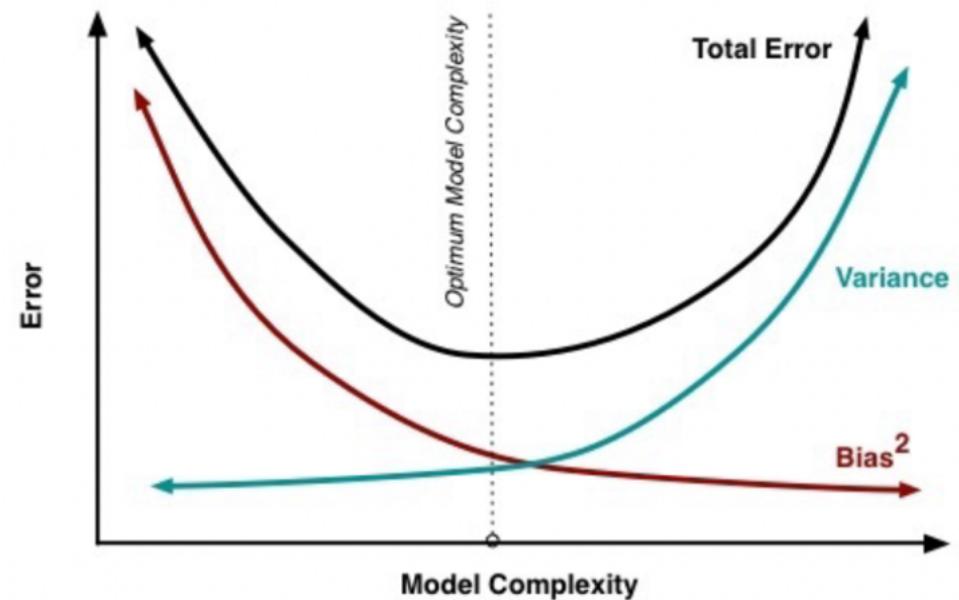
Adding more parameters always improves the fit of the model to the observed data

But it doesn't necessarily improve the model!

Goal is to find the best balance between bias and variance

Model selection methods:

- Likelihood-ratio test (LRT)
  - Compare nested models
- Akaike information criterion (AIC)
  - Compare non-nested models
- Bayesian information criterion (BIC)
  - Stronger penalty on number of parameters (compared to AIC)



MEGA11 File Analysis Help

Molecular Evolutionary Genetics Analysis

ALIGN DATA MODELS DISTANCE DIVERSITY PHYLOGENY USER TREE ANCESTORS SELECTION RATES CLOCKS DISEASE

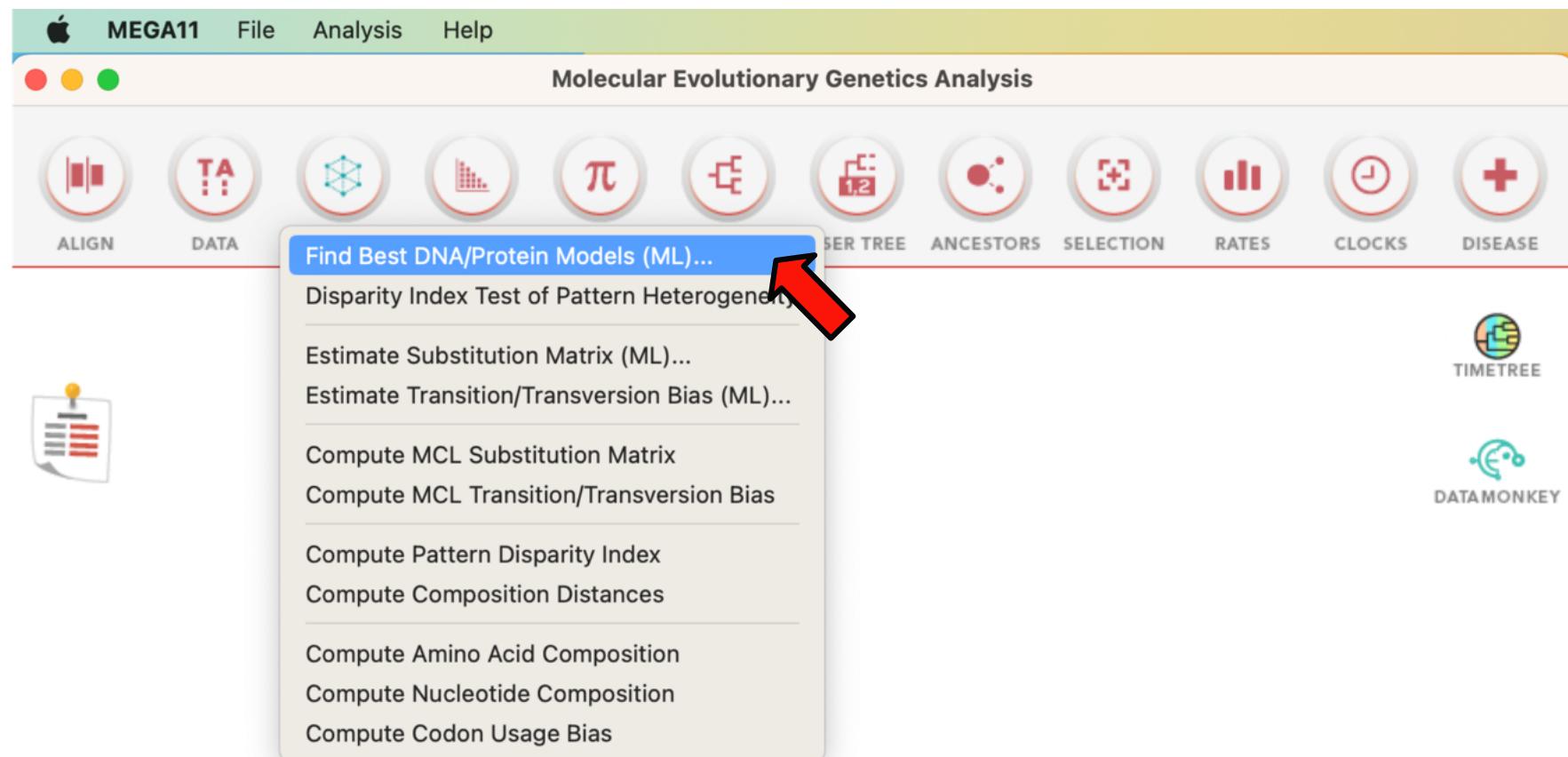
RECENT PUBLICATIONS

HELP DOCS EXAMPLES CITATION REPORT BUG UPDATES MEGA LINKS TOOLBAR PREFERENCES

ANALYZE PROTOTYPE

TIMETREE DATAMONKEY

The screenshot shows the main interface of MEGA11. At the top, there's a menu bar with 'MEGA11', 'File', 'Analysis', and 'Help'. Below the menu is a toolbar with various icons: ALIGN, DATA, MODELS (which is highlighted with a red box and a red arrow pointing to it), DISTANCE, DIVERSITY, PHYLOGENY, USER TREE, ANCESTORS, SELECTION, RATES, CLOCKS, and DISEASE. The title 'Molecular Evolutionary Genetics Analysis' is centered above the toolbar. Below the toolbar, there are two sections: 'RECENT PUBLICATIONS' and a row of links: HELP DOCS, EXAMPLES, CITATION, REPORT BUG, UPDATES, MEGA LINKS, TOOLBAR, and PREFERENCES. On the right side, there are logos for 'TIMETREE' and 'DATAMONKEY'. At the bottom right, there's a large red button with a white 'M' logo, labeled 'ANALYZE' and 'PROTOTYPE'.



#### RECENT PUBLICATIONS



MEGA11 File Analysis Help

Molecular Evolutionary Genetics Analysis

ALIGN DATA MODEL

TA Close Data

RECENT PUBLICATIONS

HELP DOCS EXAMPLES CITATION REPORT BUG UPDATES MEGA LINKS TOOLBAR PREFERENCES ANALYZE PROTOTYPE

CLOCKS DISEASE

TIMETREE DATAMONKEY

**M11: Analysis Preferences**

**Model Selection (ML)**

Option	Setting
Tree to Use →	Automatic (Neighbor-joining tree)
User Tree File →	Not Applicable
Statistical Method →	Maximum Likelihood
Substitutions Type →	Nucleotide
Genetic Code Table →	Not Applicable
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
Branch Swap Filter →	None
Number of Threads →	4

? Help    × Cancel    ✓ OK



MEGA11 File Analysis Help

Molecular Evolutionary Genetics Analysis

ALIGN DATA MODELS DISTANCE DIVERSITY PHYLOGENY USER TREE ANCESTORS SELECTION RATES CLOCKS DISEASE

TA Close Data

Close Data

PROGRESS

DETAILS STOP

STATUS/OPTIONS

M11: Progress

RUN STATUS	
Start time	13-6-23 14:57:57
Operation Run Time	00:00:14
Status	Making initial tree
Thread-1	GTR

TIMETREE

DATAMONKEY

The screenshot shows the MEGA11 software interface. At the top, there's a toolbar with various icons for analysis: Align, Data, Models, Distance, Diversity, Phylogeny, User Tree, Ancestors, Selection, Rates, Clocks, and Disease. Below the toolbar, there are two small windows: one for 'Close Data' and another for 'Recent Publications'. A central progress dialog box is open, titled 'M11: Progress', showing a progress bar at approximately 10%, a 'STOP' button, and a 'DETAILS' button. Below the progress bar is a 'STATUS/OPTIONS' section with a table of run status information:

RUN STATUS	
Start time	13-6-23 14:57:57
Operation Run Time	00:00:14
Status	Making initial tree
Thread-1	GTR

At the bottom, there's a 'RECENT PUBLICATIONS' section with links to Help Docs, Examples, Citation, Report Bug, Updates, MEGA Links, Toolbar, Preferences, Analyze, Prototype, and a large red 'Fix1' button.

RECENT PUBLICATIONS

HELP DOCS EXAMPLES

CITATION

REPORT BUG

UPDATES

MEGA LINKS

TOOLBAR

PREFERENCES



ANALYZE  
PROTOTYPE



MEGA11 File Edit View Help

Molecular Evolutionary Genetics Analysis

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

ALIGN

TA

TREE

RECENT

HELP DOCUMENTS

RESULTS

RESULTS

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	InL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)
HKY+G	148	10202.576	8763.176	-4233.410	n/a	0.40	4.49	0.352	0.236	0.190	0.221
HKY+I	148	10210.708	8771.307	-4237.476	0.44	n/a	4.48	0.352	0.236	0.190	0.221
HKY+G+I	149	10214.214	8765.090	-4233.365	0.13	0.52	4.49	0.352	0.236	0.190	0.221
TN93+G	149	10214.286	8765.162	-4233.401	n/a	0.39	4.49	0.352	0.236	0.190	0.221
TN93+I	149	10222.444	8773.320	-4237.480	0.44	n/a	4.48	0.352	0.236	0.190	0.221
TN93+G+I	150	10225.979	8767.132	-4233.383	0.12	0.51	4.49	0.352	0.236	0.190	0.221
T92+G	146	10227.438	8807.484	-4257.569	n/a	0.39	4.49	0.294	0.294	0.206	0.206
HKY	147	10228.007	8798.329	-4251.989	n/a	n/a	4.47	0.352	0.236	0.190	0.221
T92+I	146	10235.660	8815.706	-4261.680	0.44	n/a	4.48	0.294	0.294	0.206	0.206
GTR+G	152	10236.716	8758.422	-4227.023	n/a	0.39	4.47	0.352	0.236	0.190	0.221
T92+G+I	147	10239.091	8809.414	-4257.531	0.14	0.54	4.49	0.294	0.294	0.206	0.206
TN93	148	10239.686	8800.285	-4251.964	n/a	n/a	4.47	0.352	0.236	0.190	0.221
GTR+I	152	10244.968	8766.675	-4231.150	0.44	n/a	4.48	0.352	0.236	0.190	0.221
GTR+G+I	153	10248.342	8760.325	-4226.972	0.14	0.54	4.47	0.352	0.236	0.190	0.221

MEGA11 File Edit View Help

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

Results

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	f(T)	f(C)	f(G)
HKY+G	148	0.236	0.190	0.221
HKY+I	148	0.236	0.190	0.221
HKY+G+I	149	0.236	0.190	0.221
TN93+G	149	0.236	0.190	0.221
TN93+I	149	0.236	0.190	0.221
TN93+G+I	150	0.236	0.190	0.221
T92+G	146	0.294	0.206	0.206
HKY	147	0.236	0.190	0.221
T92+I	146	0.294	0.206	0.206
GTR+G	152	0.236	0.190	0.221
T92+G+I	147	0.294	0.206	0.206
TN93	148	0.236	0.190	0.221
GTR+I	152	0.236	0.190	0.221
GTR+G+I	153	0.236	0.190	0.221

M11: Select Output Format

SELECT OUTPUT FORMAT

XLSX: MS Excel workbook (2007+)  
 XLS: MS Excel Workbook (all versions)  
 ODS: Open/Libre Office Workbook  
 CSV: Comma-Separated-Values

Save to Disk  Display Results

OUTPUT FILENAME

australia\_aligned\_0423\_model\_selection.txt

DIRECTORY

/Users/clydedapat/Desktop/2023\_Timor-Leste/

...

RECENT

HELP DOCS EXAMPLES CITATION REPORT BUG UPDATES MEGA LINKS TOOLBAR PREFERENCES PROTOTYPE







MEGA-result0

Office Update To keep up-to-date with security updates, fixes, and improvements, choose Check for Updates.

A1 Table. Maximum Likelihood fits of 24 different nucleotide substitution models

1 Table. Maximum Likelihood fits of 24 different nucleotide substitution models

NOTE.— Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented [1]. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (R) are shown for each model, as well. They are followed by nucleotide frequencies (f) and rates of base substitutions (r) for each nucleotide pair. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. This analysis involved 73 nucleotide sequences. Codon positions included were

2 1st+2nd+3rd+Noncoding. There were a total of 1701 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2][3]

3

4 Abbreviations: TR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor./div>

5 1. Nei M. and Kumar S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.

6 2. Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution <https://doi.org/10.1093/molbev/msab120>.

7 3. Stecher G., Tamura K., and Kumar S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Molecular Biology and Evolution 37:1237-1239.

Disclaimer: Although utmost care has been taken to ensure the correctness of the caption, the caption text is provided "as is" without any warranty of any kind. Authors advise the user to carefully check the caption prior to its use for any purpose and report any errors or problems to the authors immediately ([www.megasoftware.net](http://www.megasoftware.net)). In no event shall the authors and their employers be liable for any damages, including but not limited to special, consequential, or other damages. Authors specifically disclaim all other warranties expressed or implied, including but not limited to the determination of suitability of this caption text for a specific purpose, use, or application.

Info Caption HKY+G HKY+I HKY+G+I TN93+G TN93+I TN93+G+I T92+G HKY T92+I GTR+G T92+G+I TN93 +

Ready - + 100%

MEGA-result0

Search Sheet

Share

Home Insert Page Layout Formulas Data Review View

Paste Cut Copy Undo Redo

Arial 10 A A Wrap Text General

B I U Merge & Center Conditional Formatting

C Cell Styles

D Delete

E Format as Table

F Sort & Filter

G Check for Updates

X Office Update To keep up-to-date with security updates, fixes, and improvements, choose Check for Updates.

A1

From|To

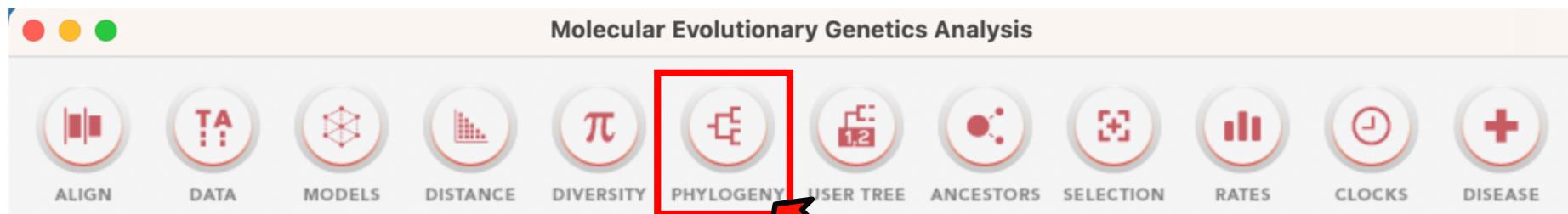
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	From To	A	T	C	G												
2	A	-	0.021569852	0.017391739	0.18086829												
3	T	0.032180537	-	0.155579815	0.020218651												
4	C	0.032180537	0.192955614	-	0.020218651												
5	G	0.287874723	0.021569852	0.017391739	-												
6																	
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	

HKY+G HKY+I HKY+G+I TN93+G TN93+I TN93+G+I T92+G HKY T92+I GTR+G T92+G+I TN93 +

Ready

100%

Construct a maximum likelihood tree



#### RECENT PUBLICATIONS



HELP DOCS EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS

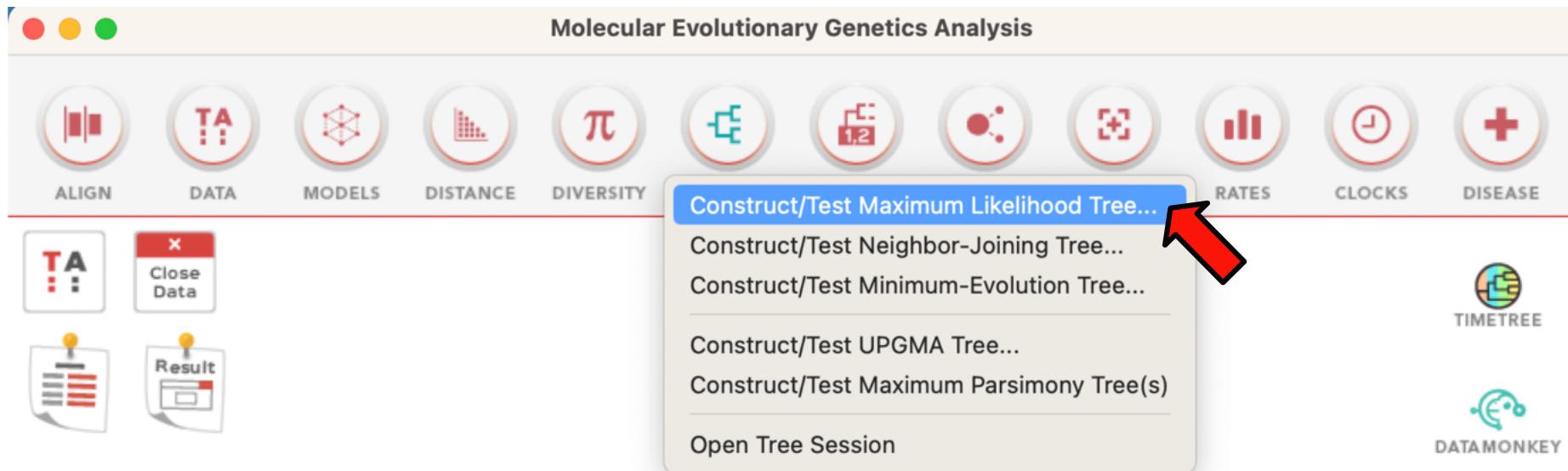


TOOLBAR



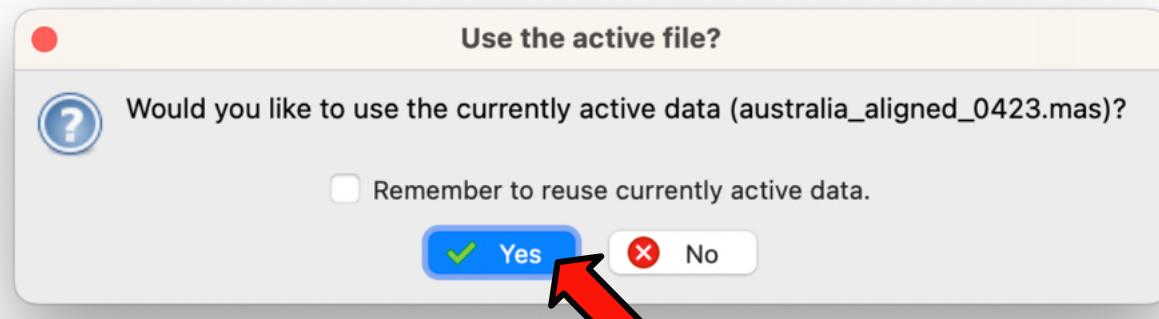
PREFERENCES





#### RECENT PUBLICATIONS





#### RECENT PUBLICATIONS



ANALYZE  
PROTOTYPE



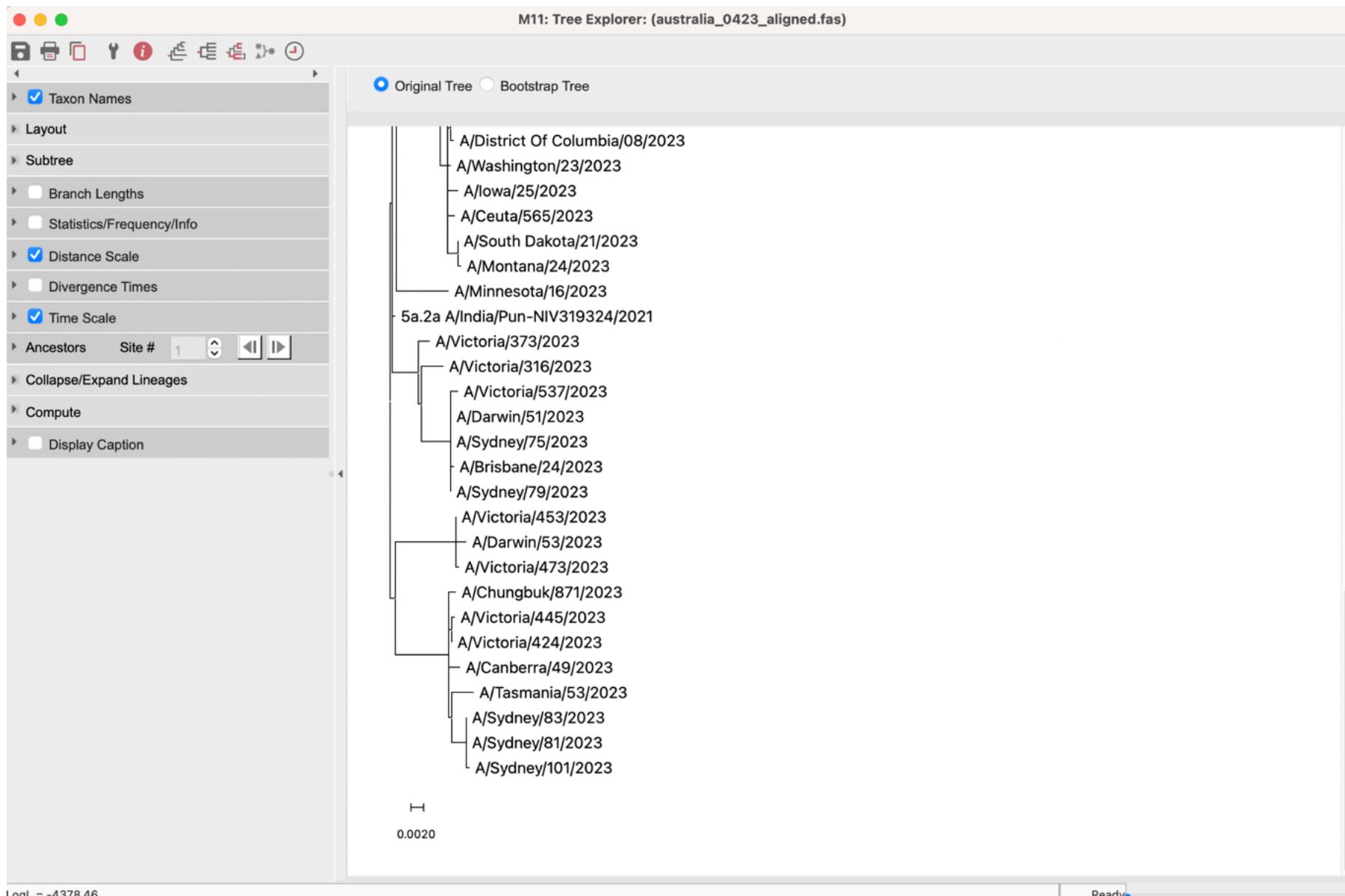
M11: Analysis Preferences

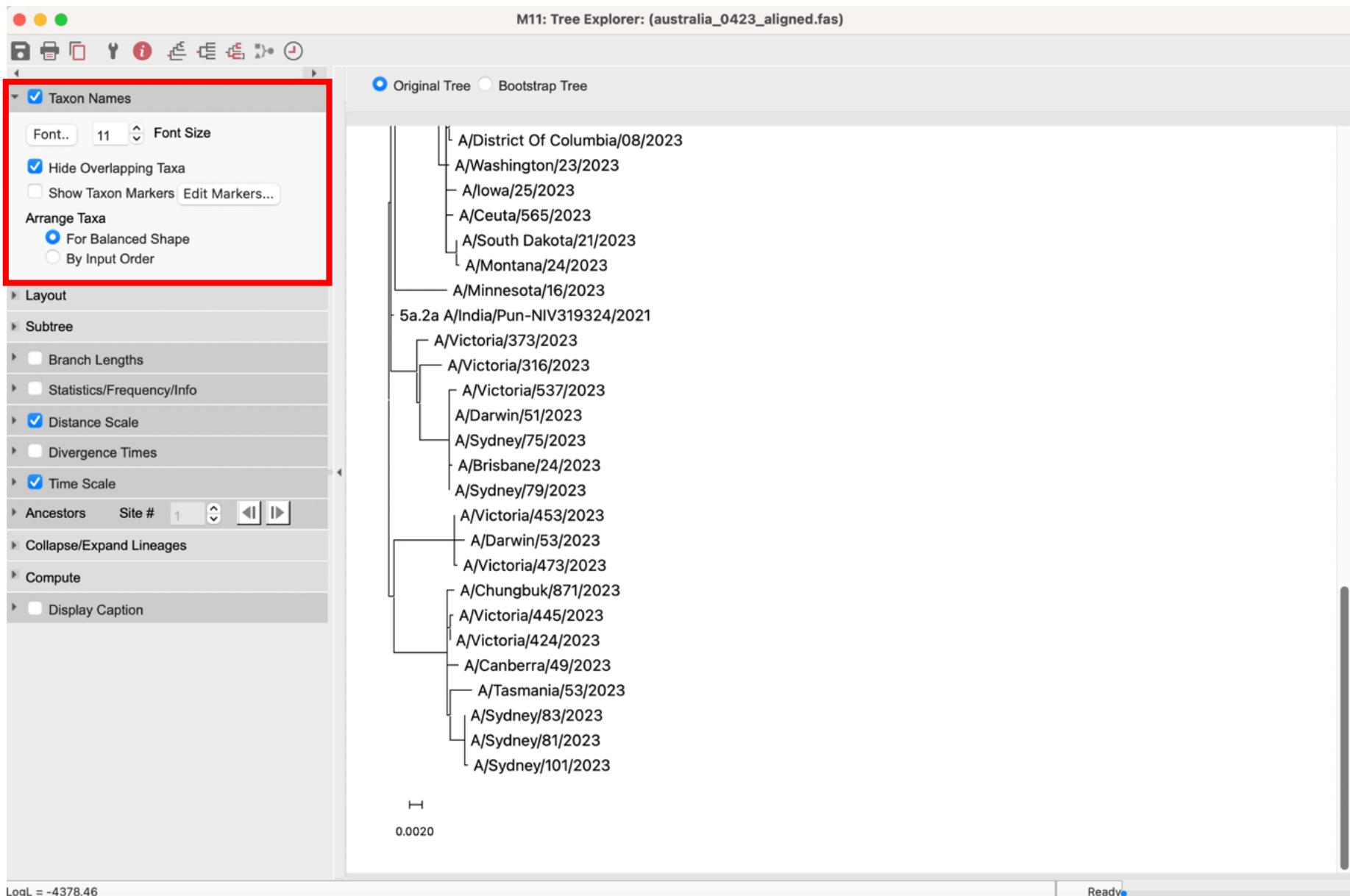
Phylogeny Reconstruction

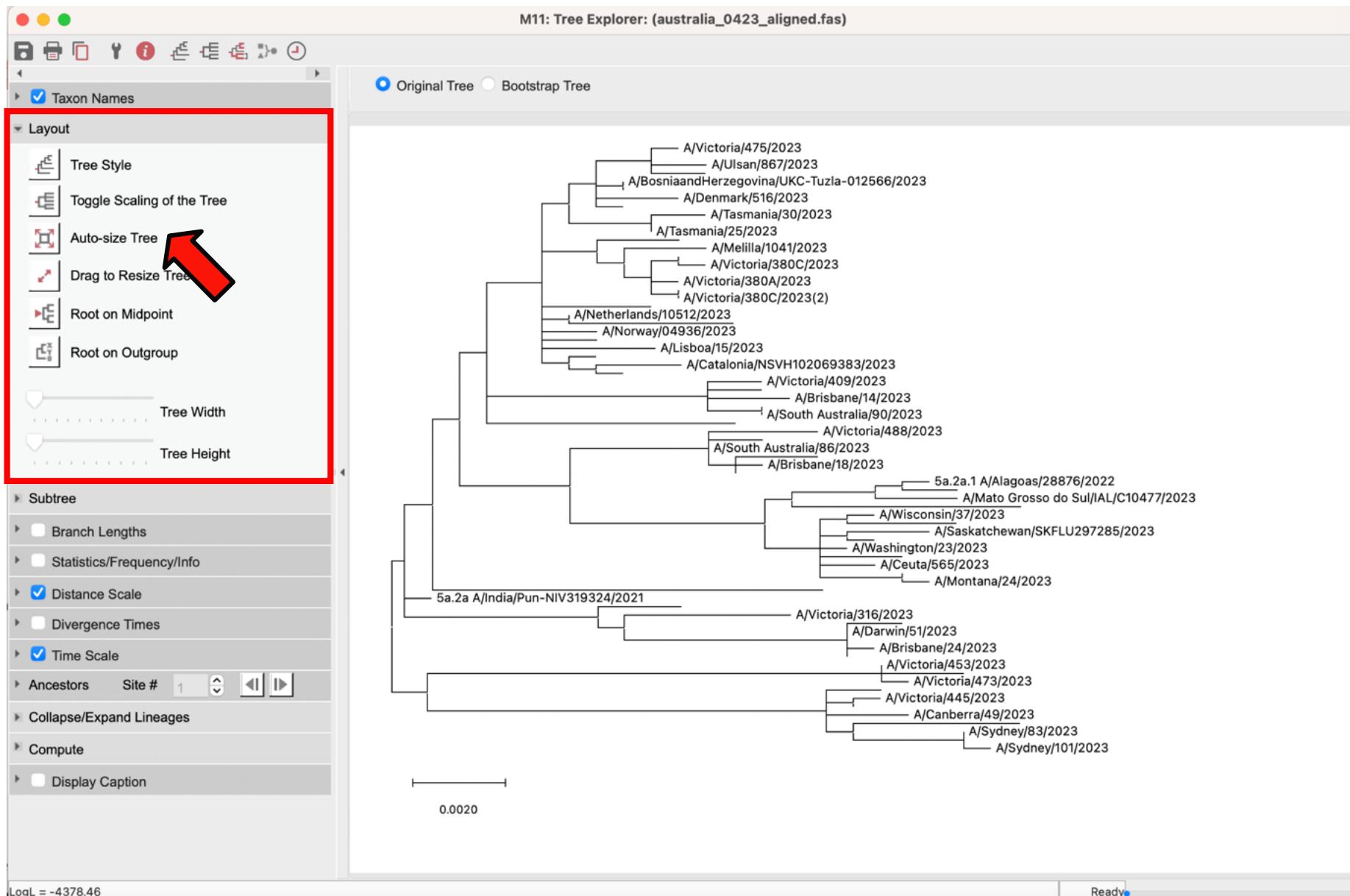
Option	Setting
Statistical Method	→ Maximum Likelihood
Test of Phylogeny	→ Bootstrap method
No. of Bootstrap Replications	→ 100
Substitutions Type	→ Nucleotide
Genetic Code Table	→ Not Applicable
Model/Method	→ Hasegawa-Kishino-Yano model
Rates among Sites	→ Gamma Distributed (G) 
No of Discrete Gamma Categories	→ 5
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
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
Number of Threads	→ 4

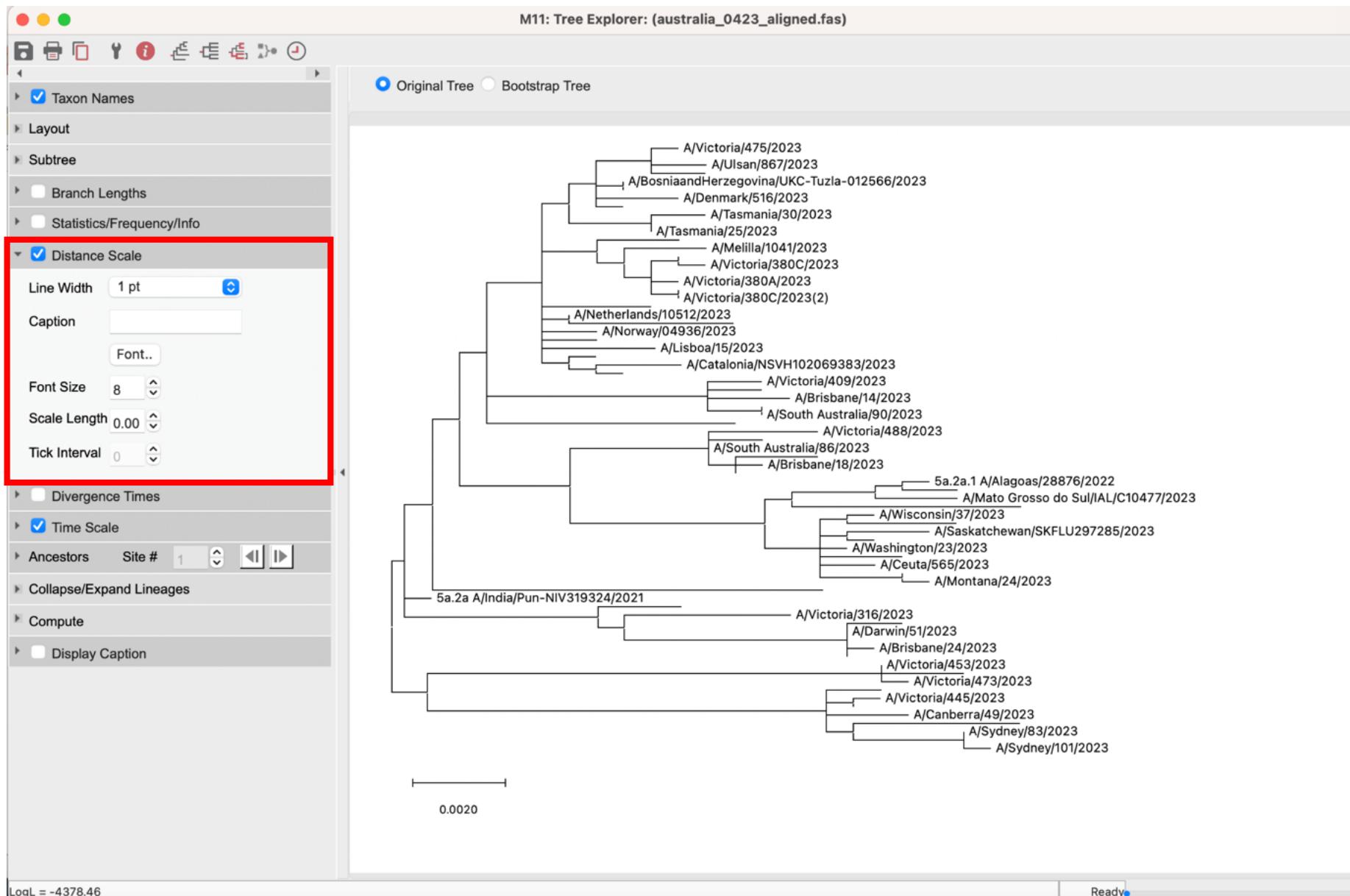
 Help    Cancel    OK

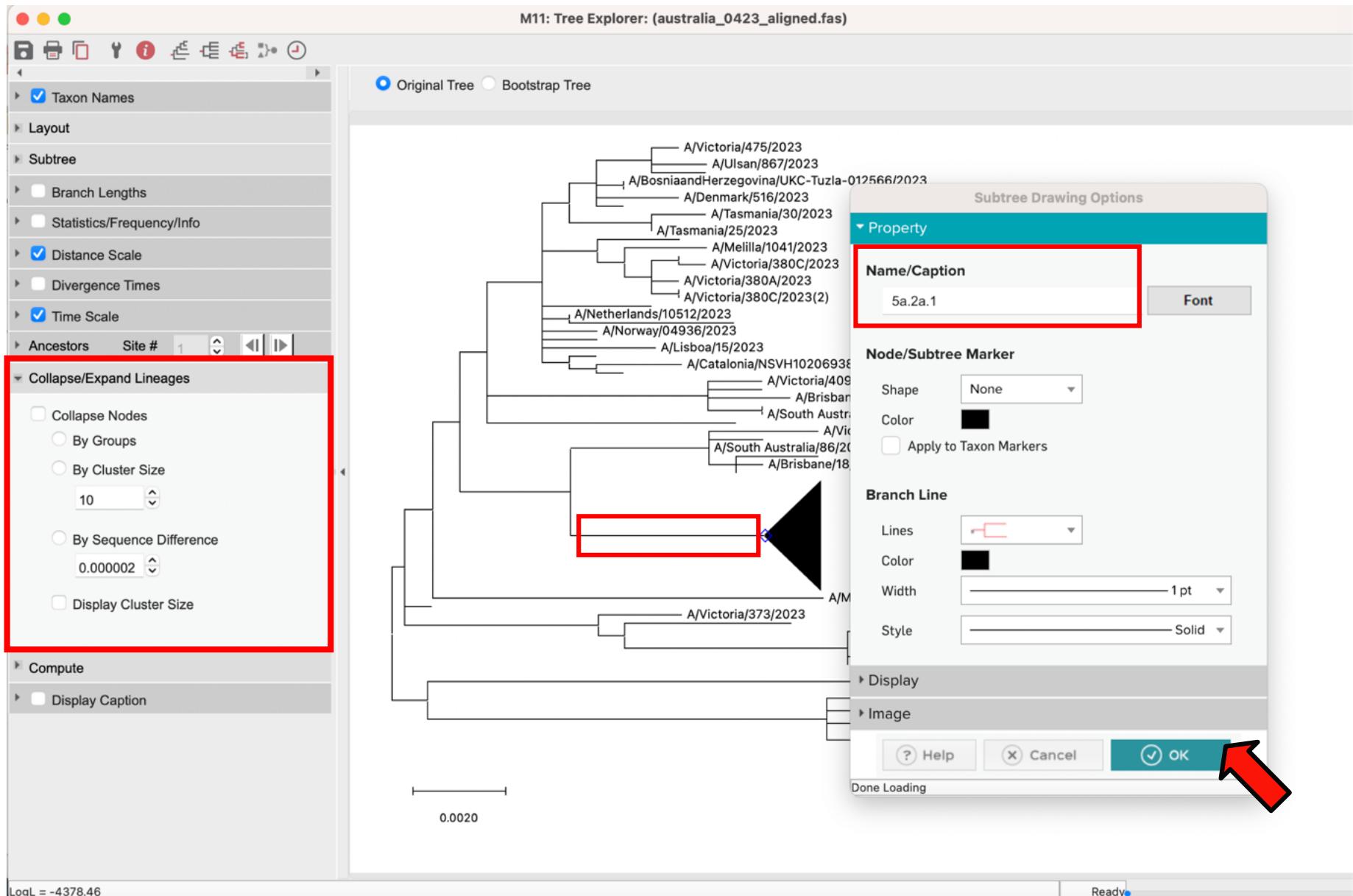


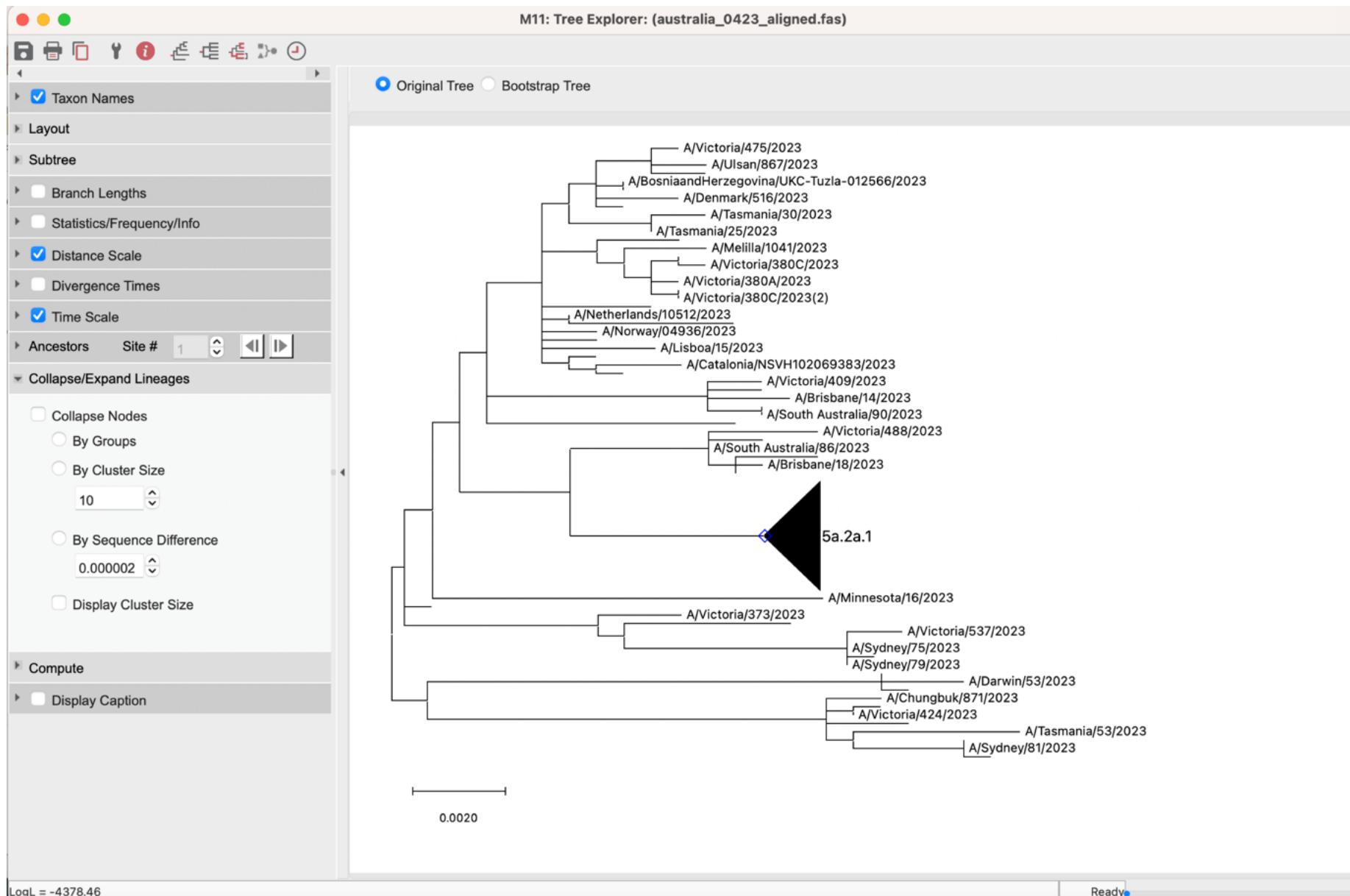


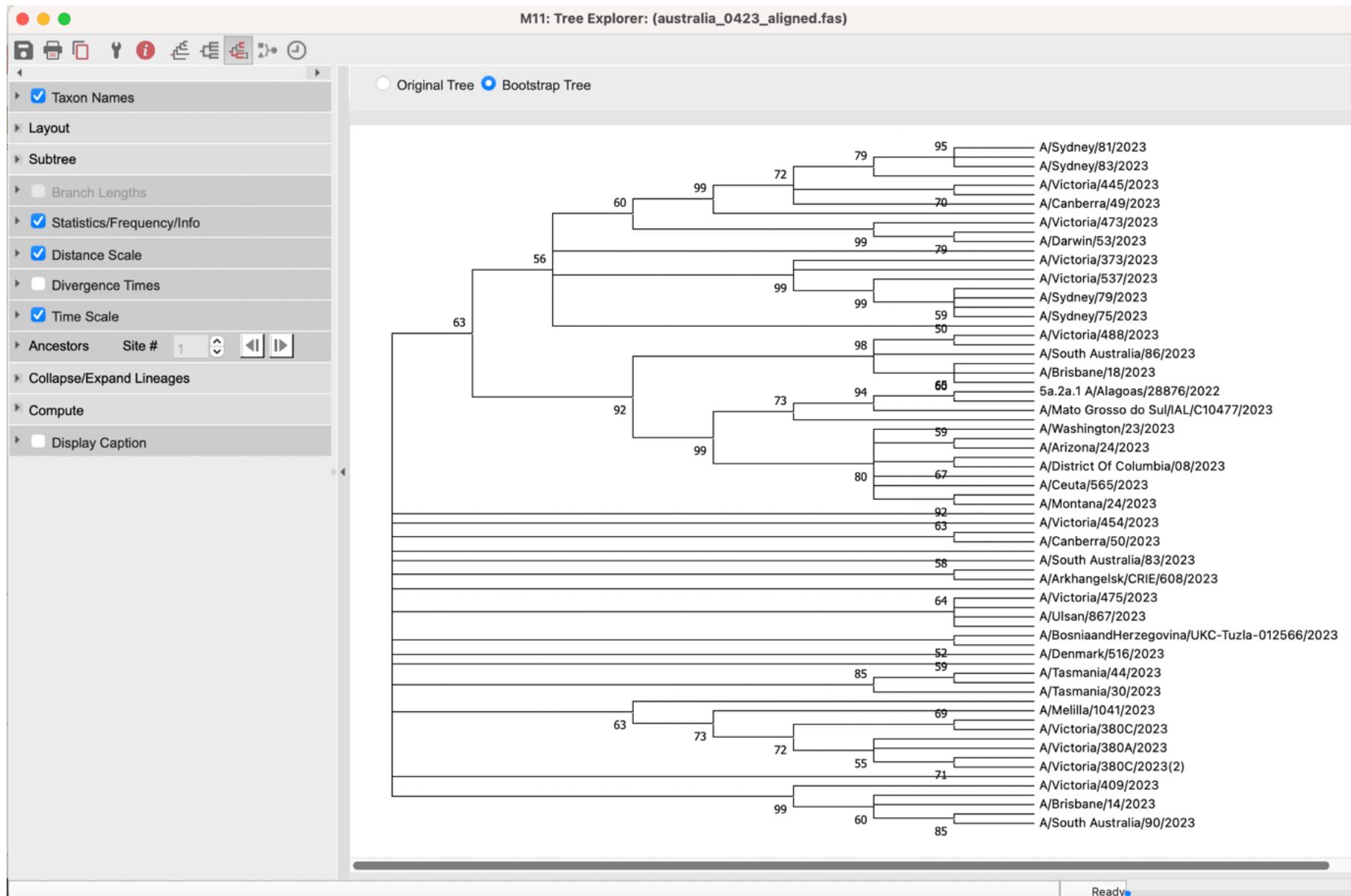


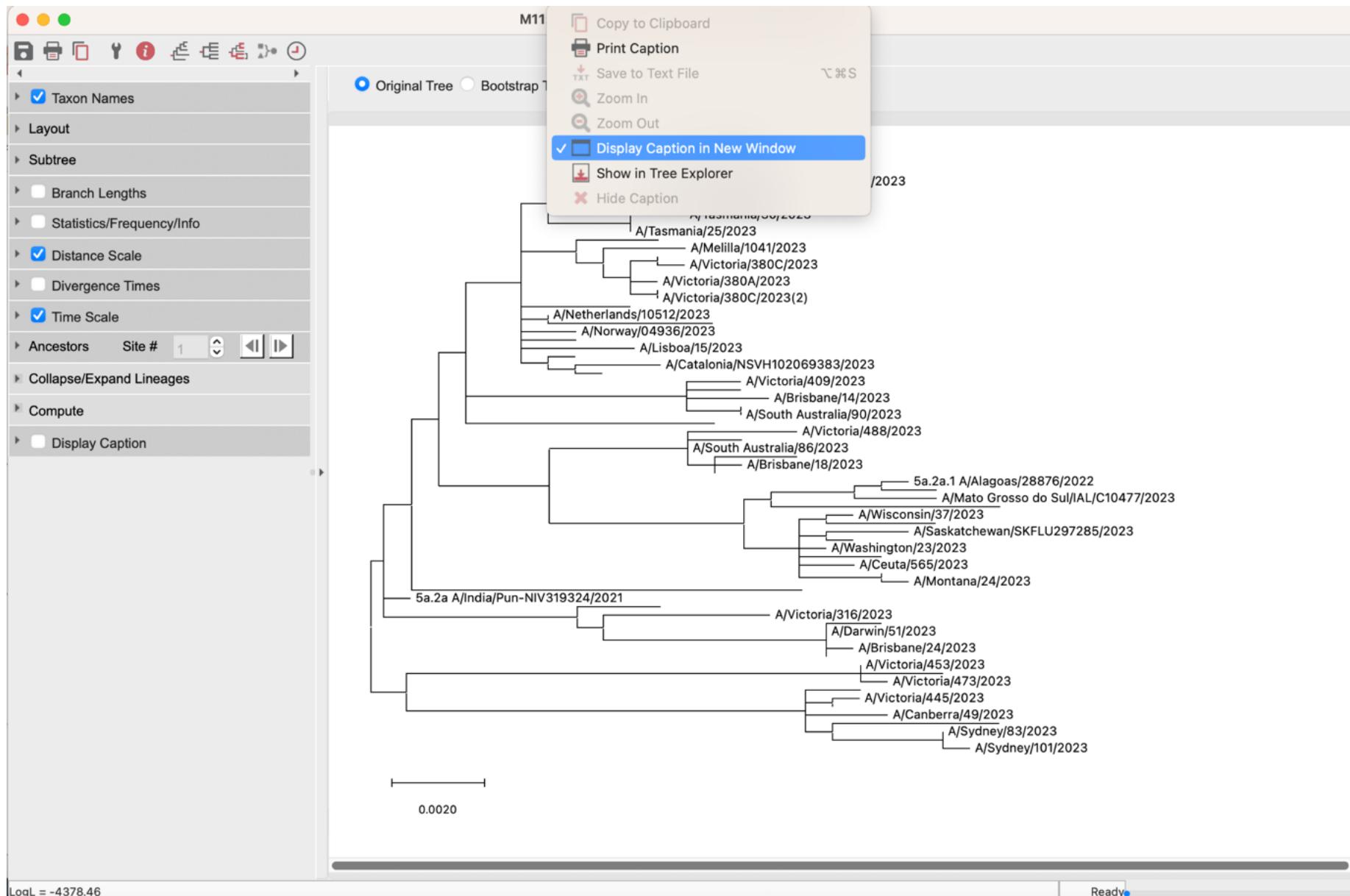


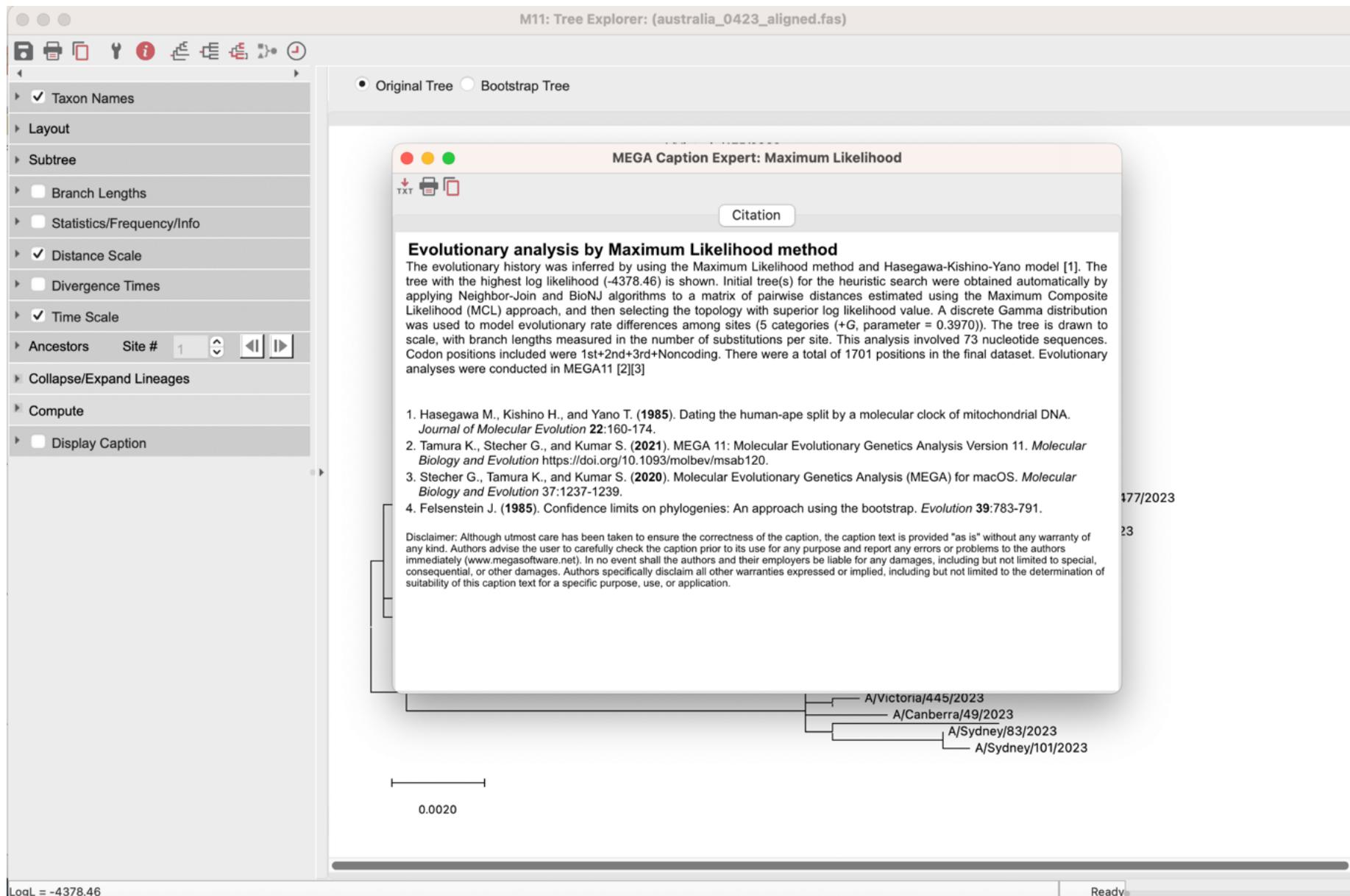












# Export tree file

# File formats

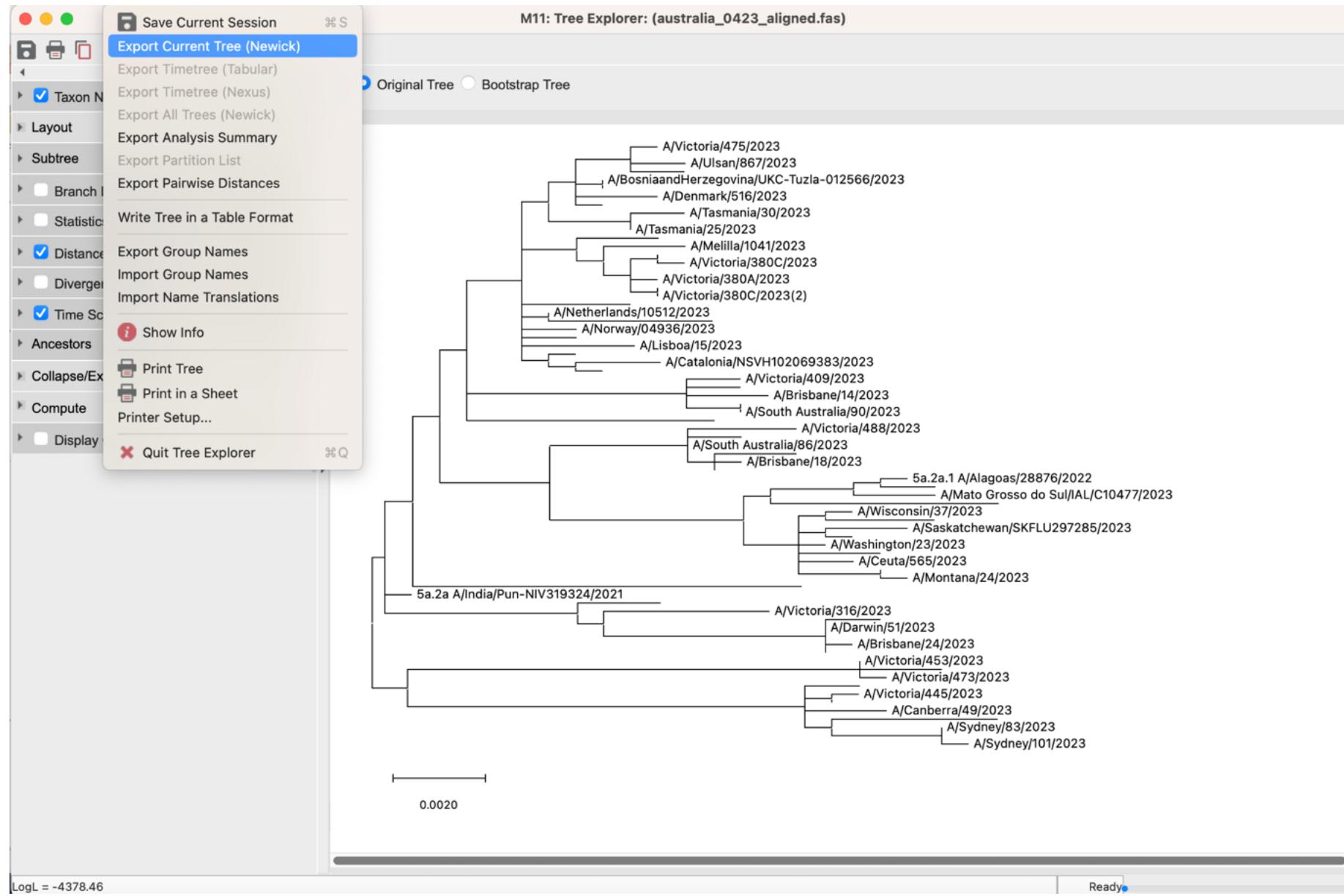
```
(,,(,));
(A,B,(C,D));
(A,B,(C,D)E)F;
(:0.1,:0.2,(:0.3,:0.4):0.5);
(:0.1,:0.2,(:0.3,:0.4):0.5):0.0;
(A:0.1,B:0.2,(C:0.3,D:0.4):0.5);
(A:0.1,B:0.2,(C:0.3,D:0.4)E:0.5)F;
((B:0.2,(C:0.3,D:0.4)E:0.5)F:0.1)A;
```

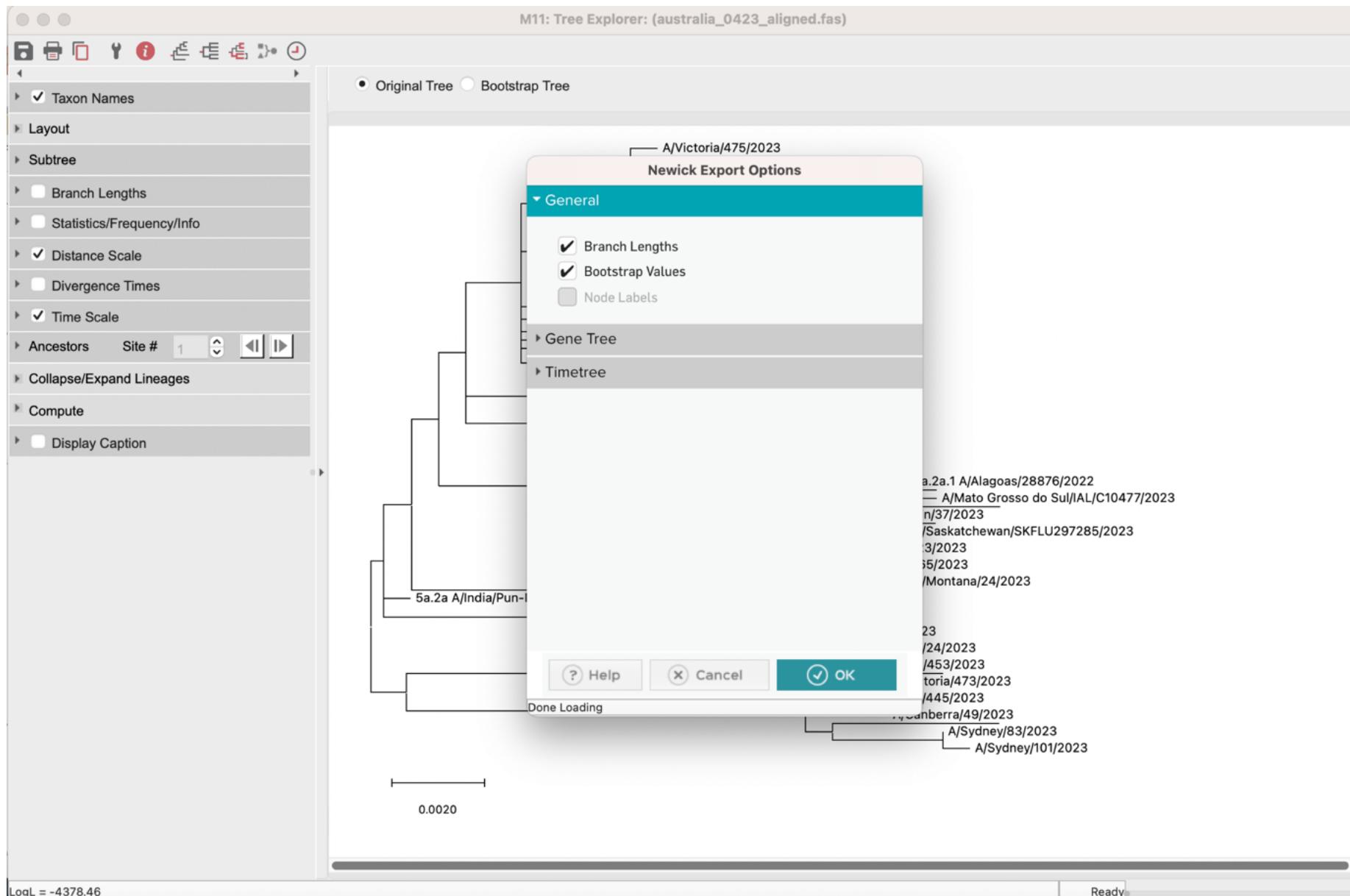
Newick (Newick's restaurant)

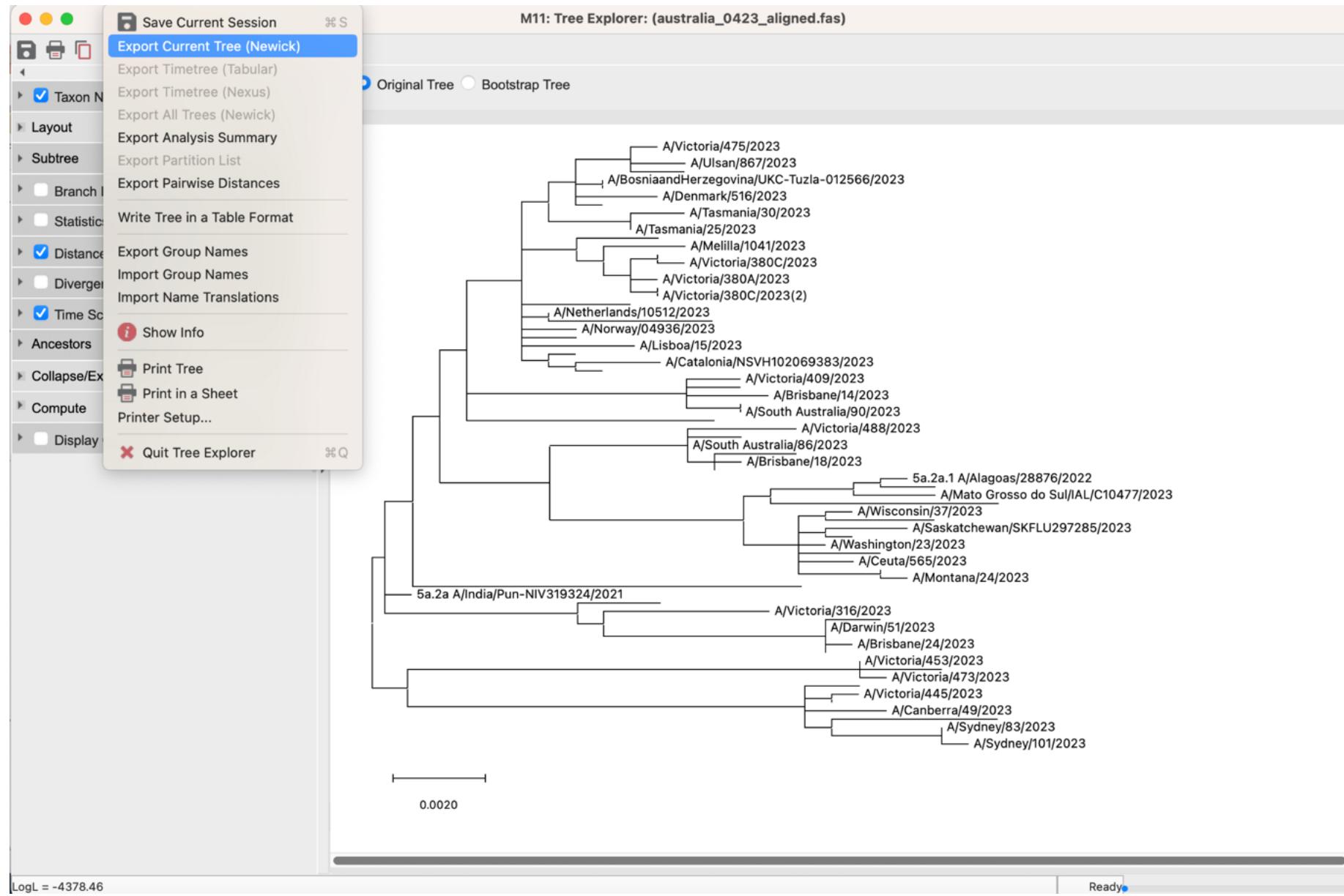
```
#NEXUS
Begin TAXA;
Dimensions ntax=4;
TaxLabels SpaceDog SpaceCat SpaceOrc SpaceElf;
End;

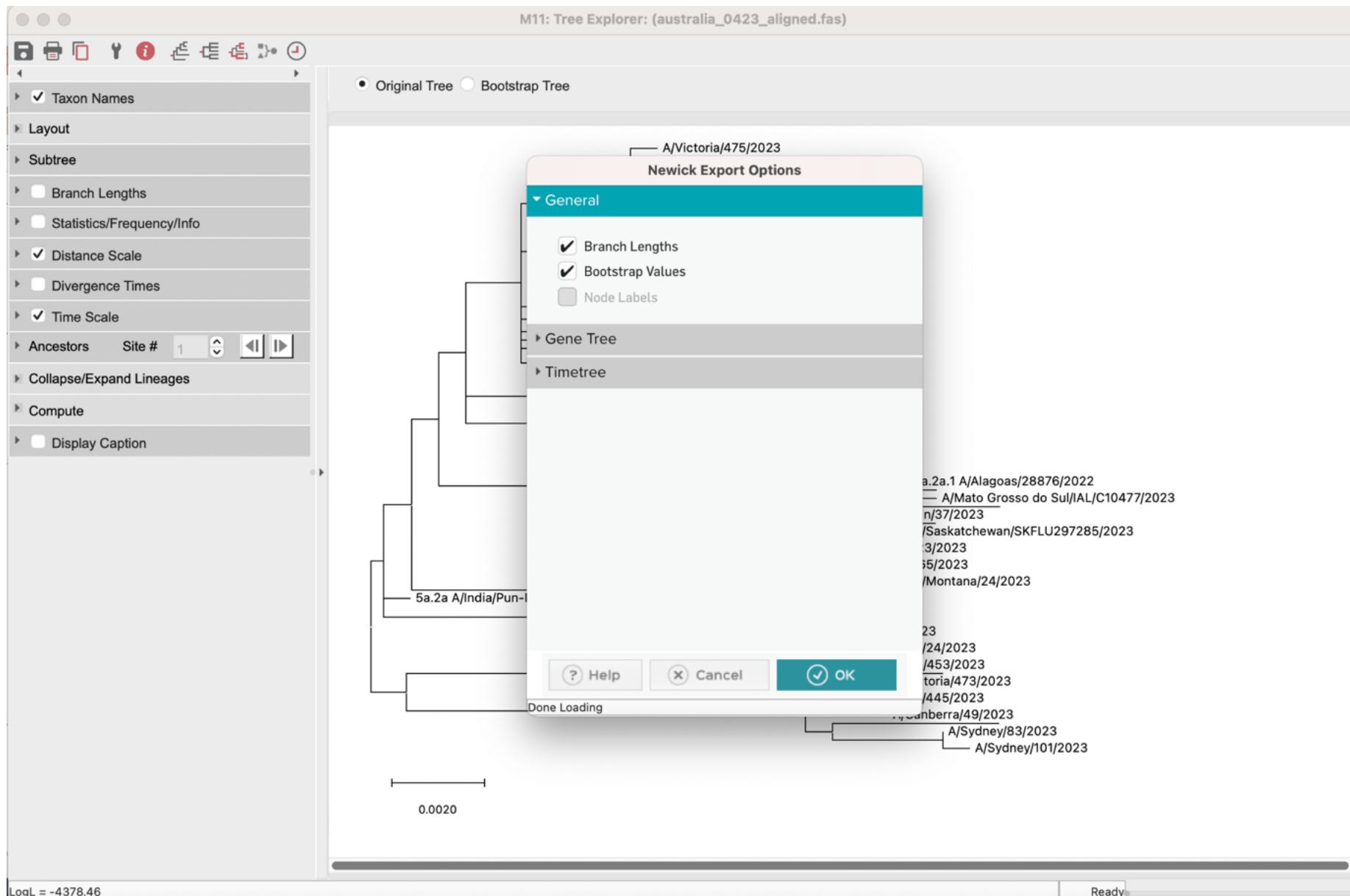
Begin data;
Dimensions nchar=15;
Format datatype=dna missing=? gap=- matchchar=.;
Matrix
[ When a position is a "matchchar", it means that it is the same as
the first entry at the same position. ]
SpaceDog  atgctagctagctcg
SpaceCat   .....??.-a.
SpaceOrc   ...t.....-g. [ same as atgttagctag-tgg ]
SpaceElf   ...t.....-a.
;
End;

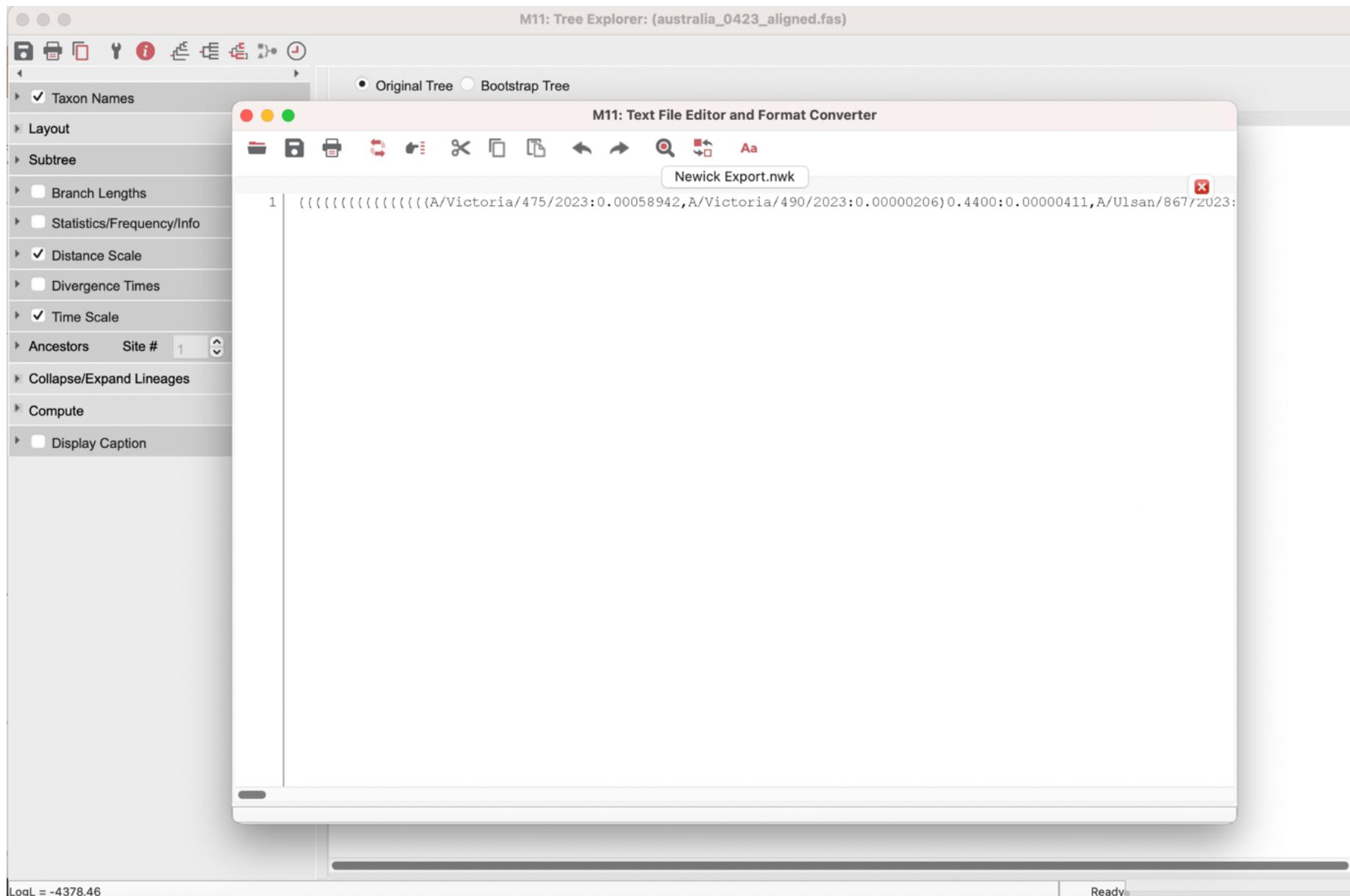
BEGIN TREES;
Tree tree1 = (((SpaceDog,SpaceCat),SpaceOrc,SpaceElf));
END;
```



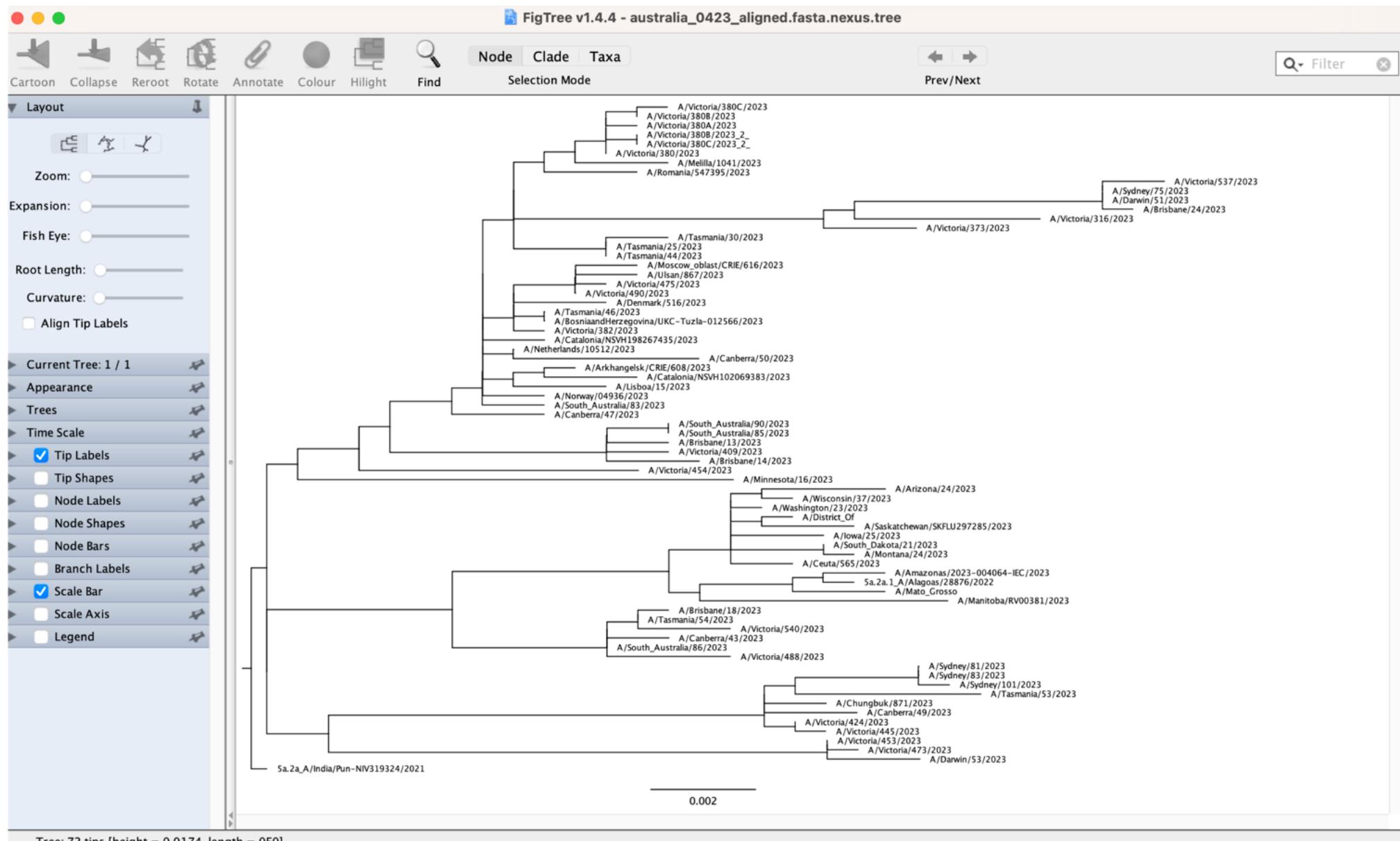








# Import tree file in FigTree



# Thank you !

[clyde.dapat@influenzacentre.org](mailto:clyde.dapat@influenzacentre.org)



WHO Collaborating Centre  
for Reference and  
Research on Influenza  
**VIDRL**



A joint venture between The University of Melbourne and The Royal Melbourne Hospital

