

Phylogenetic Analysis



WHO Collaborating Centre
for Reference and
Research on Influenza
VIDRL



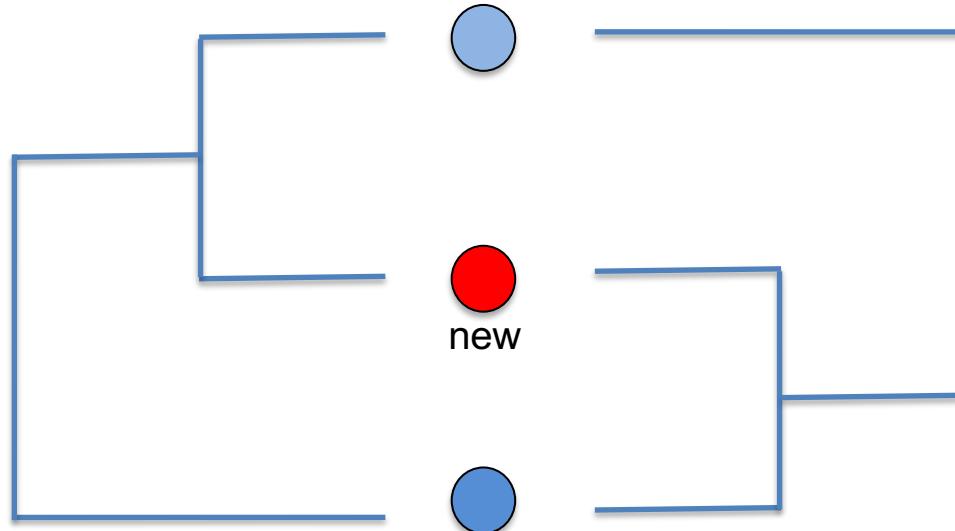
A joint venture between The University of Melbourne and The Royal Melbourne Hospital



Objectives

- To gain a basic understanding of phylogenetics
- To know the different methods of creating a phylogenetic tree

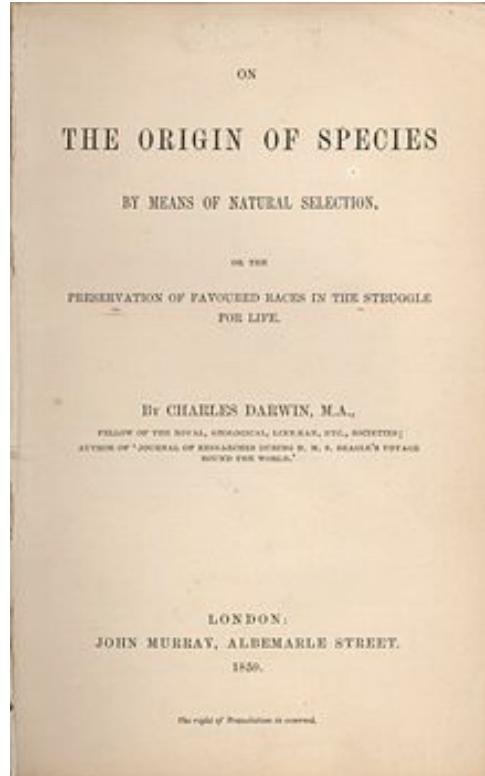
Why do phylogenetics?



Theory of Evolution: Darwin



Charles Darwin
1809-1882
https://en.wikipedia.org/wiki/Charles_Darwin



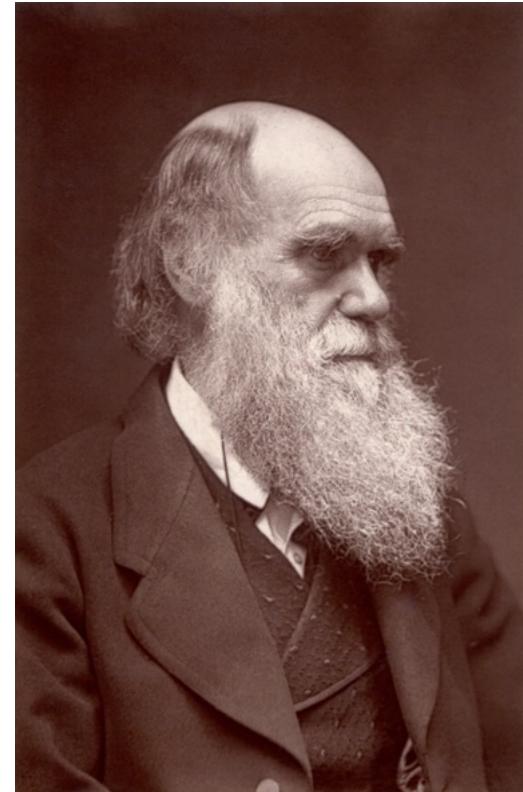
Origin of Species. First edition (1859)

The single most important idea
in human history

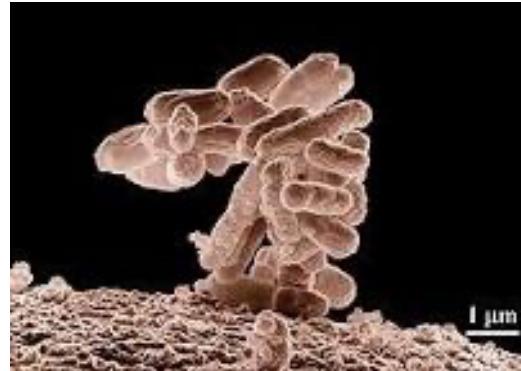
Natural Selection: Darwin's postulates

1. Individuals within species vary in their characteristics
2. Some of these variations are heritable
3. In each generation, more young are reproduced than can survive to reproduce
4. The survival and reproduction of individuals is not random: Individuals who survive and go on to reproduce are those with the most favorable variation. (They are naturally selected.)

Consequence: Populations change over time – they adapt to their environment



Darwin didn't know about



Microorganisms



DNA

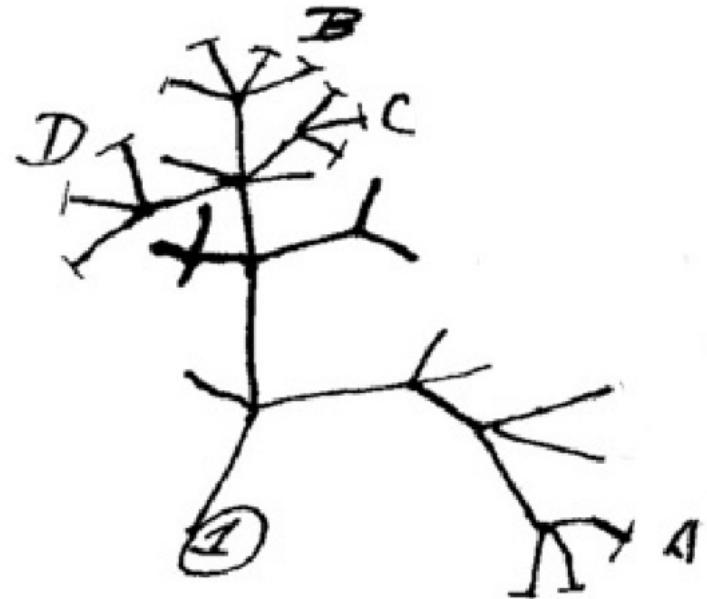
Phylogenetics

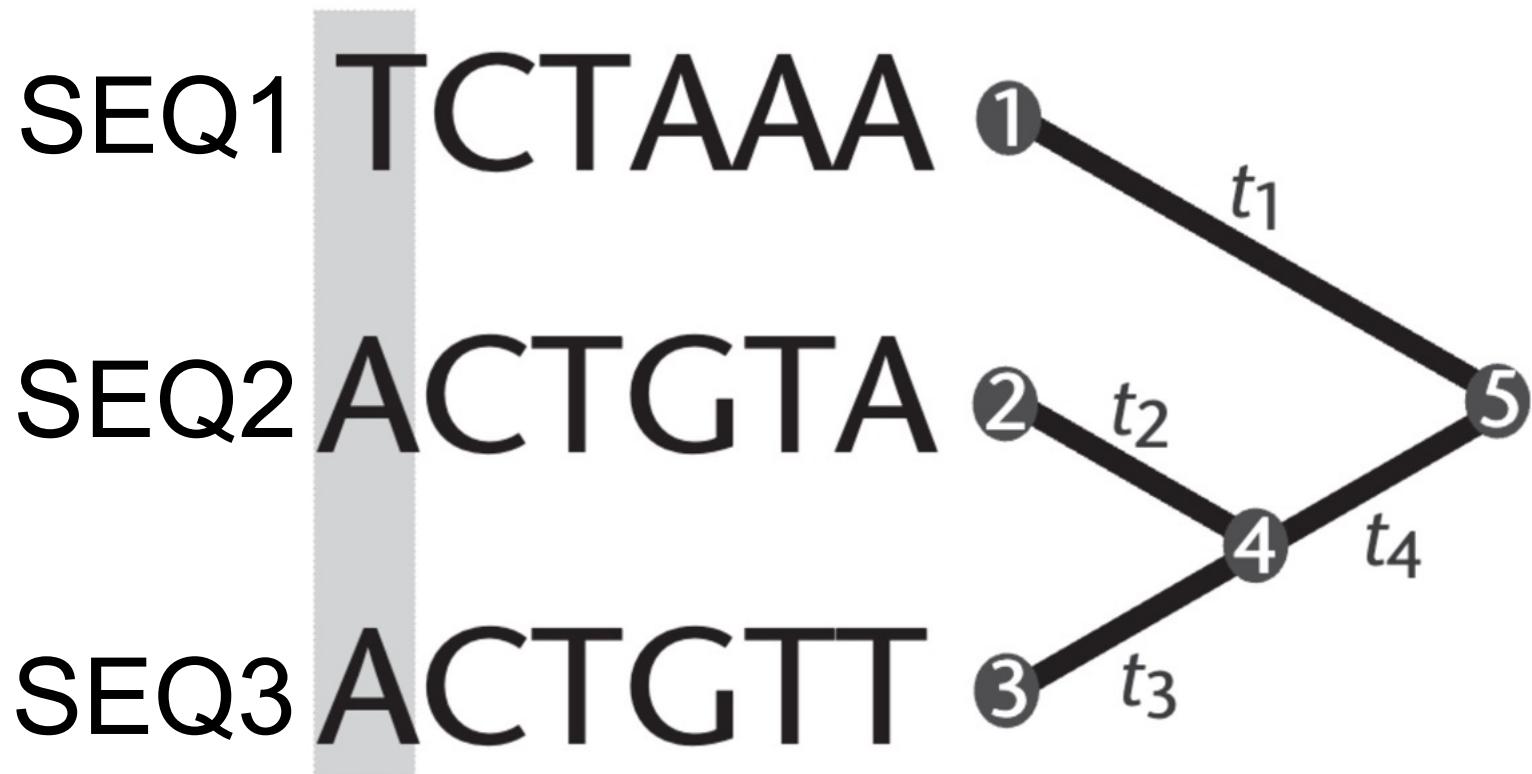
Phylogenetics is the study of evolutionary relationships among biological entities – often species, individuals or genes (which may be referred to as taxa)

Use phylogenetic trees as models of evolution

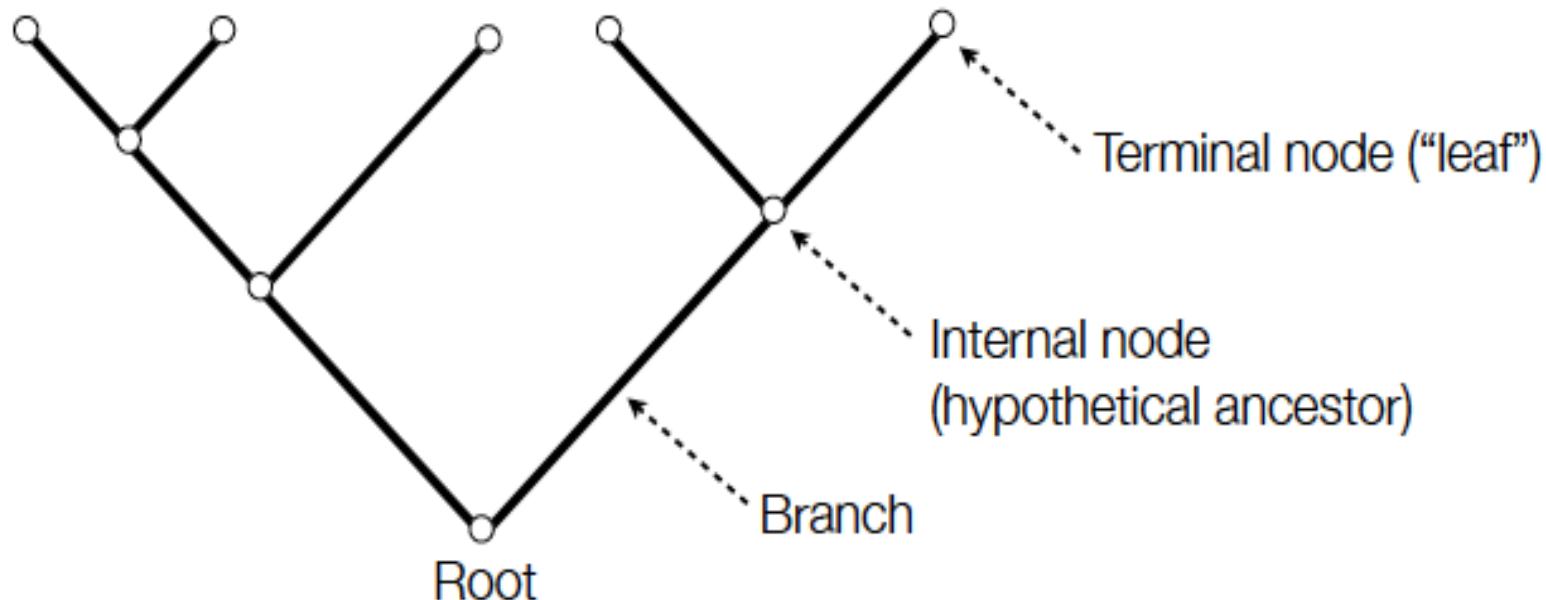


I think

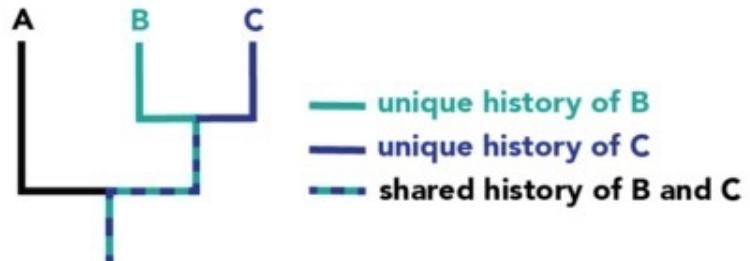
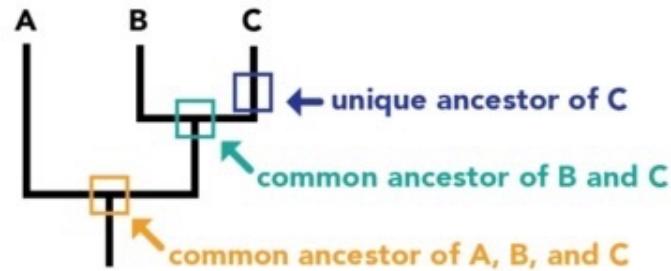
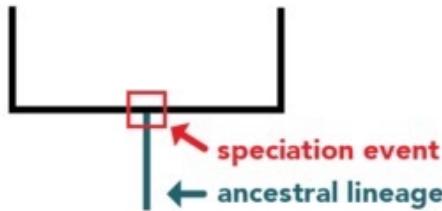
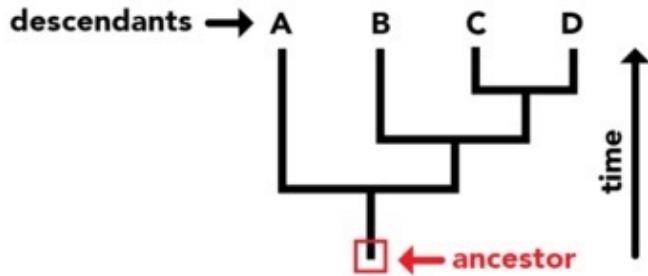


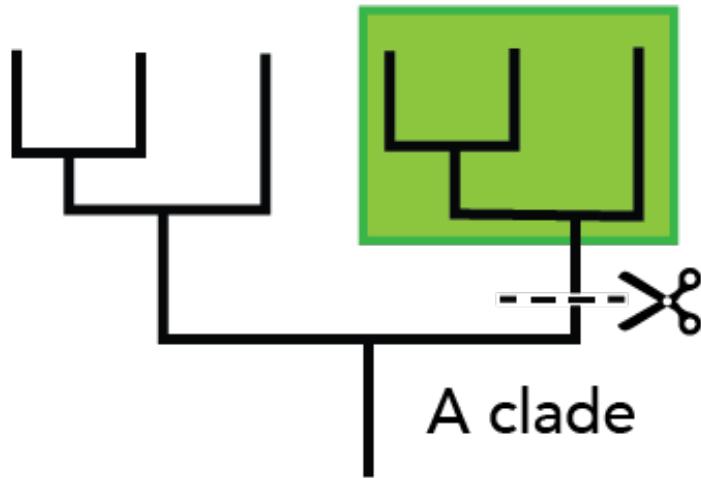


Trees: terminology

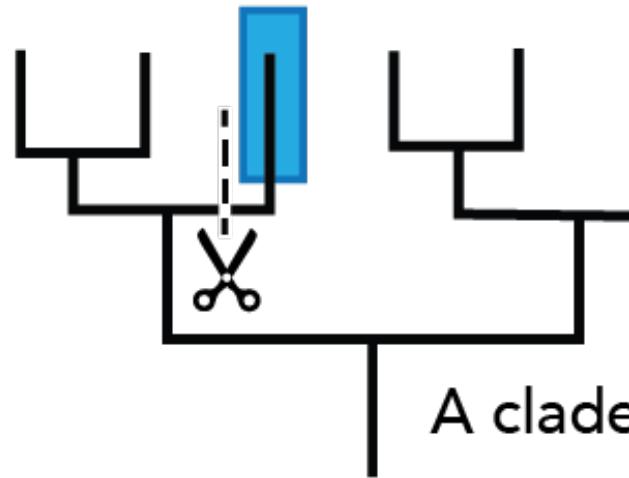


Reading trees

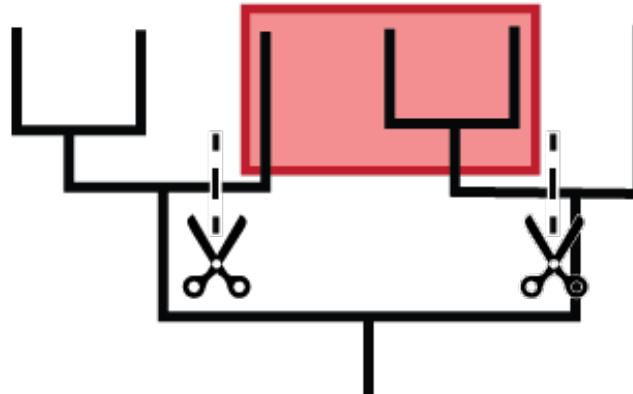




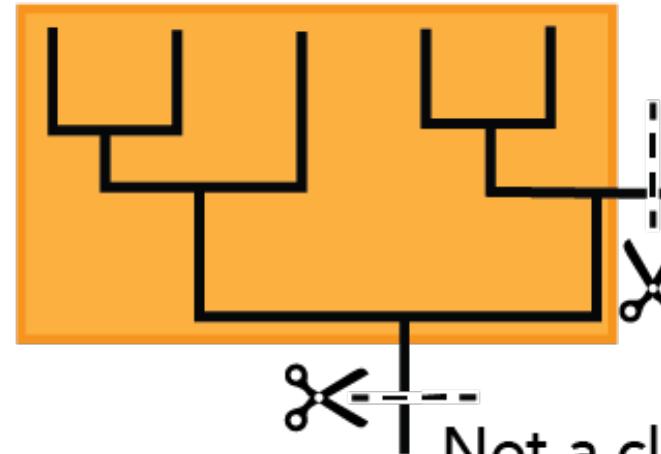
A clade



A clade

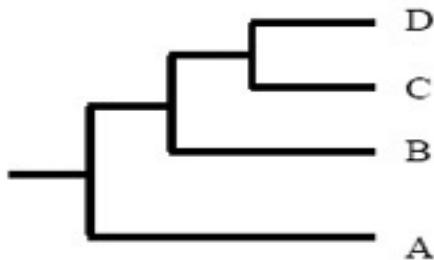


Not a clade



Not a clade

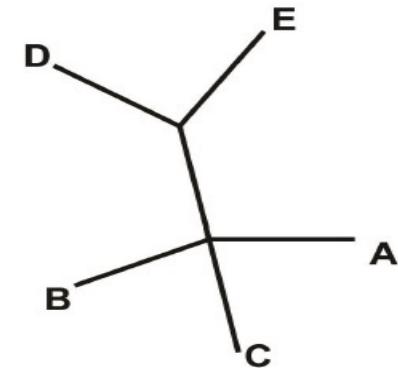
Trees: types



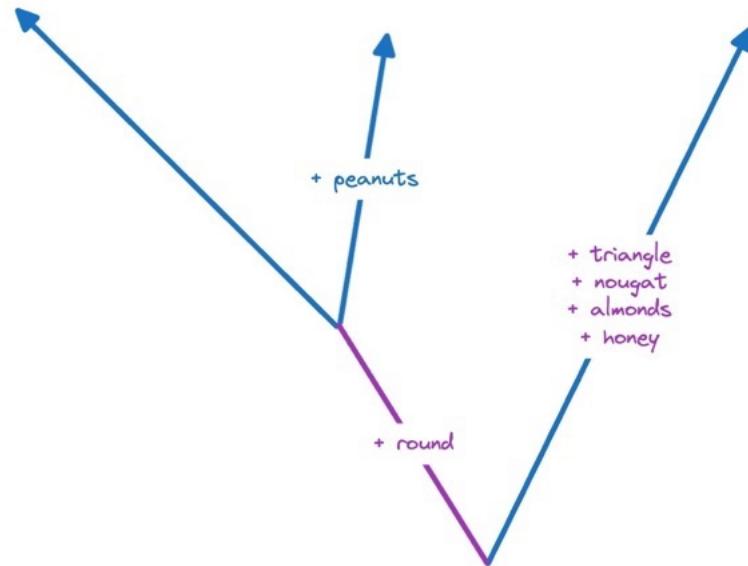
Traditional Tree



Circular/Radial Tree



Unrooted Tree

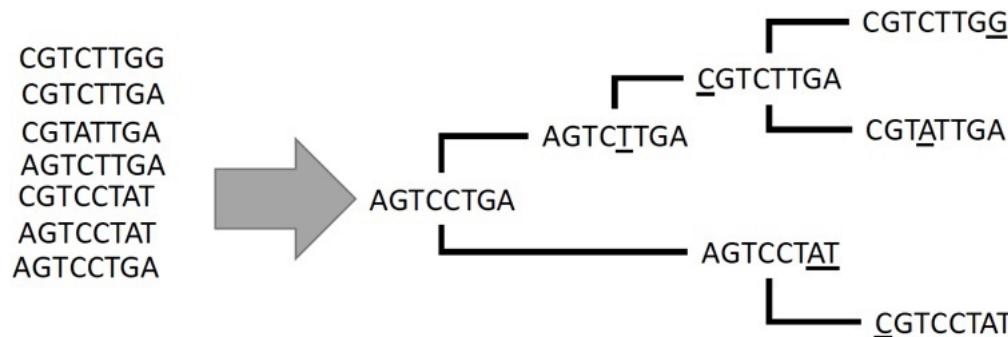


<https://acesse.dev/gDTeA>

Assumptions in phylogenetics

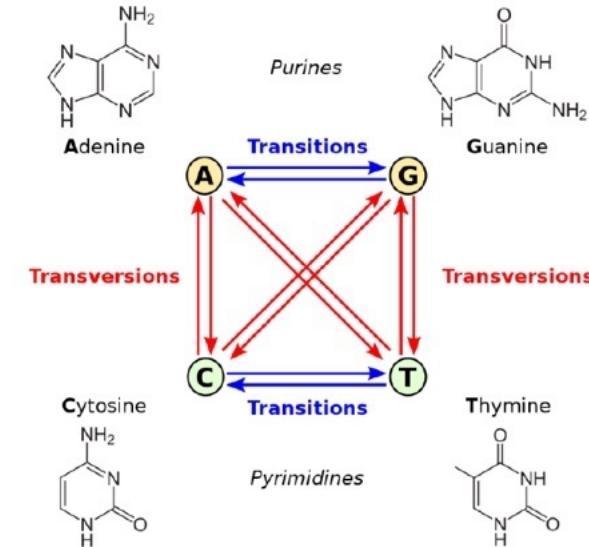
Several fundamental assumptions are made in phylogenetic methods

- Each aligned site represents a set of orthologous characters i.e. a gene in different species that evolved from a common ancestor by speciation
- Sites in alignment evolve independently
- Lineages evolve independently
- Relationships amongst sequences can be represented by a bifurcating (division into two branches) tree



Assumptions in phylogenetics

- Transitions are seen more often than transversions
- G/C → A/T is more commonly seen than A/T → G/C
- Each base → base mutation rate can be modeled separately
- However, **statistical assumptions** like **independence of sites** are made that are not always true



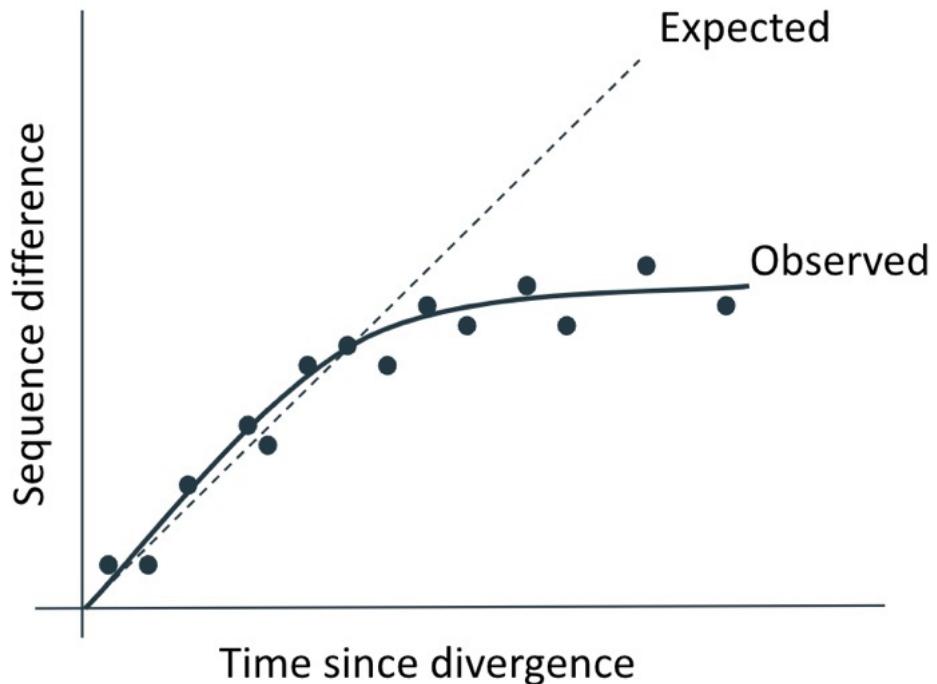
Nucleotide substitution models

- How nucleotides change to other nucleotides.
- Used for distance correction.
- Continuous-time Markov process

#	A	B	C	D
[[0.25,0.25,0.25,0.25], # A				
[0.25,0.25,0.25,0.25], # B				
[0.25,0.25,0.25,0.25], # C				
[0.25,0.25,0.25,0.25]] # D				

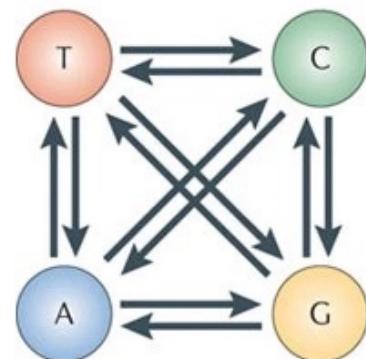
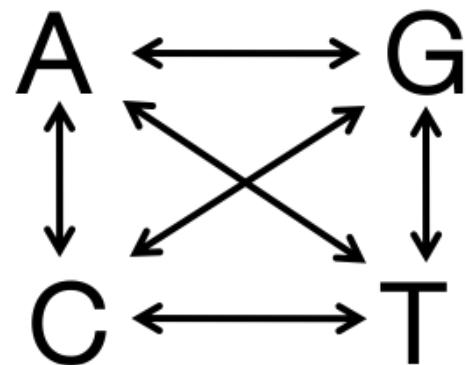


Multiple substitutions



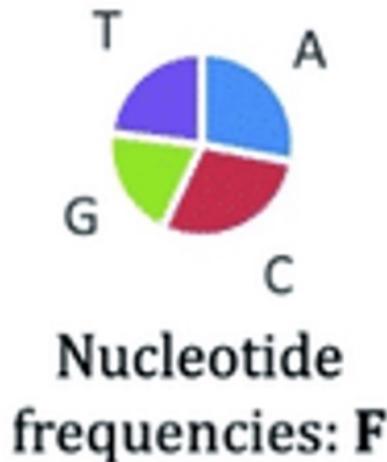
ACCGGTC
↓ ↓ ↓
TCCCCGTC
↓ ↓
GCCGGTA A

Rate Matrix



Base Frequencies

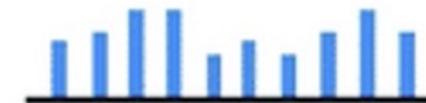
$$\pi_A + \pi_C + \pi_G + \pi_T = 1$$



Site Rates

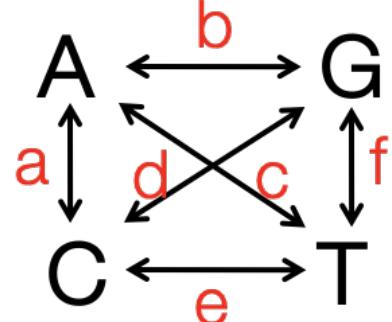
+ I + G

- C **GAAACATG**
A **AGA** G - - A **TA**
T **AGA** GC - T **TG**
A **AGA** AAAA **ATG**



- C G A A A C A T G
A A G A G - - A T A
T A G A G C - T T G
A A G A A A A A T G

Rate Matrix



Base Frequencies

$$\pi_A + \pi_C + \pi_G + \pi_T = 1$$

Site Rates

$$+ I + G$$

JC

$$a=b=c=d=e=f$$

$$\pi_A = \pi_C = \pi_G = \pi_T$$

No I or G

0 free

parameters

HKY

$$a=c=d=f, b=e$$

$$\pi_A, \pi_C, \pi_G, \pi_T$$

No I or G

4 free

parameters

GTR

$$a, b, c, d, e, f$$

$$\pi_A, \pi_C, \pi_G, \pi_T$$

No I or G

8 free

parameters

GTR+I+G

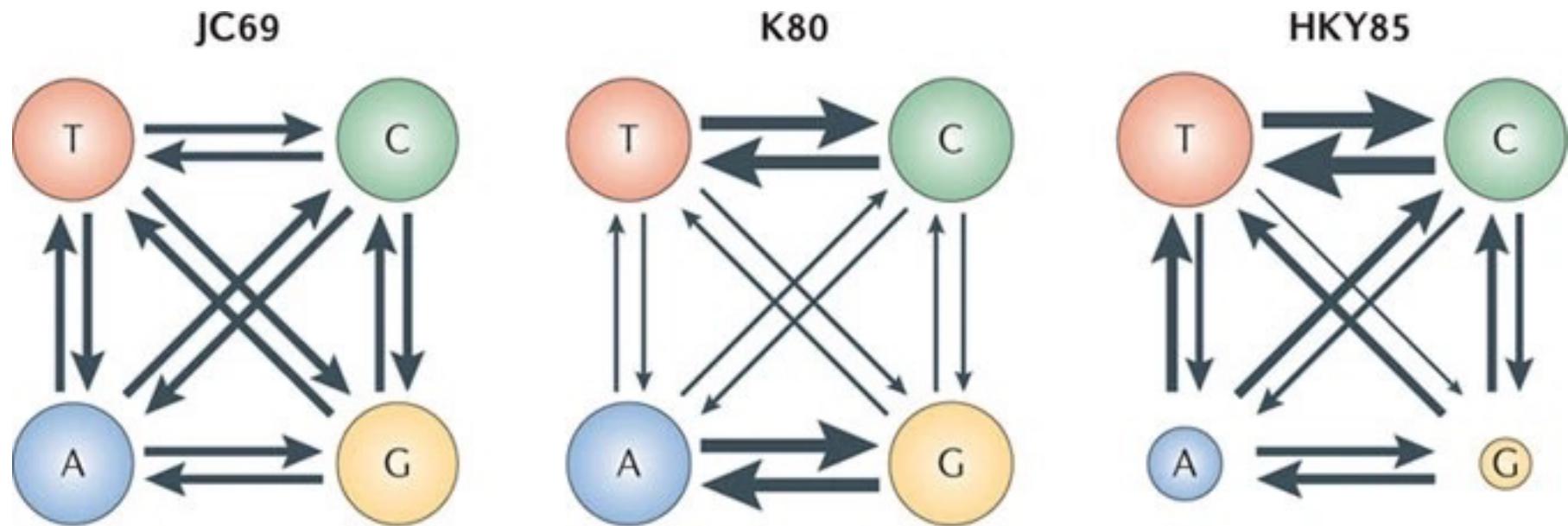
$$a, b, c, d, e, f$$

$$\pi_A, \pi_C, \pi_G, \pi_T$$

I, G

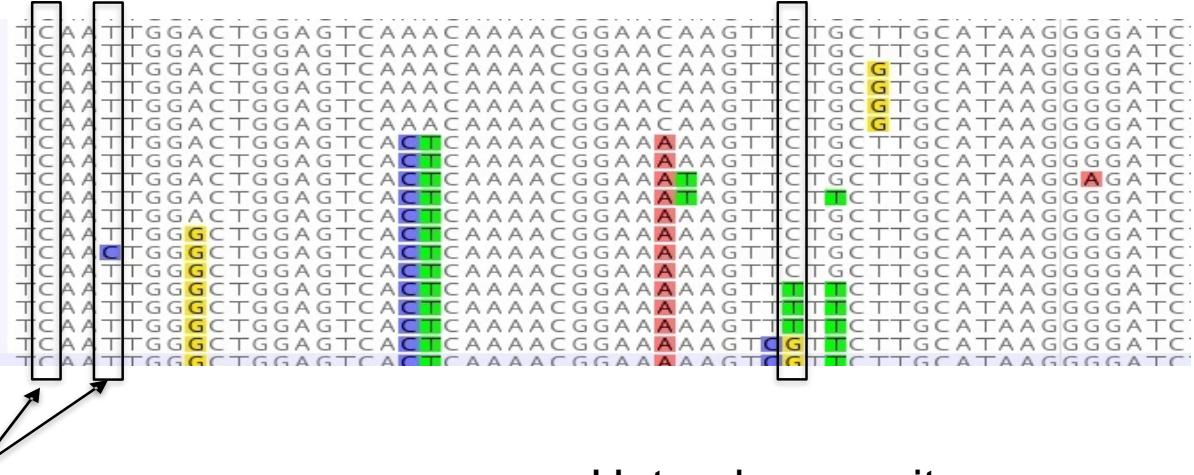
10 free

parameters



Homologous and heterologous sites

- ➡ 126. A/New_Caledonia/99/2019_jul
- ➡ 127. A/South_Australia/39/2019e_feb
- ➡ 128. 131K_A/Victoria/23/2018_oct
- ➡ 129. A/Sri_Lanka/25/2019_jun
- ➡ 130. A/Sri_Lanka/27/2019_jul
- ➡ 131. A/Victoria/2/2019_jan
- ➡ 132. 135K_A/FJI/71/2017_sep
- ➡ 133. 135K_A/VICTORIA/653/2017e
- ➡ 134. 135N_A/PHILIPPINES/13/2017_sep
- ➡ 135. 135N_A/SYDNEY/22/2018e_mar
- ➡ 136. A/Malaysia/RP0049/2019
- ➡ 137. A/South_Africa/R03989/2019_apr
- ➡ 138. A/Hong_Kong/681/2018e
- ➡ 139. A/Timor_Leste/5/2019_mar
- ➡ 140. A/Macau/603415/2019_may
- ➡ 141. A/Victoria/213/2019_aug
- ➡ 142. cdcA/Hong_Kong/45/2019
- ➡ 143. A/South_Australia/2/2019e_jan
- ➡ 144. A/South_Australia/4/2019e_jan



Homologous sites:

- Inherited from the common ancestor of all the species in the alignment
- Not informative for inferring information on the amount of evolution i.e. branch lengths, times, evolutionary rates

- Both type of sites are useful in phylogenetic studies as are insertions & deletions

Heterologous sites:

- Informative for inferring information on the amount of evolution

Rate variation across sites

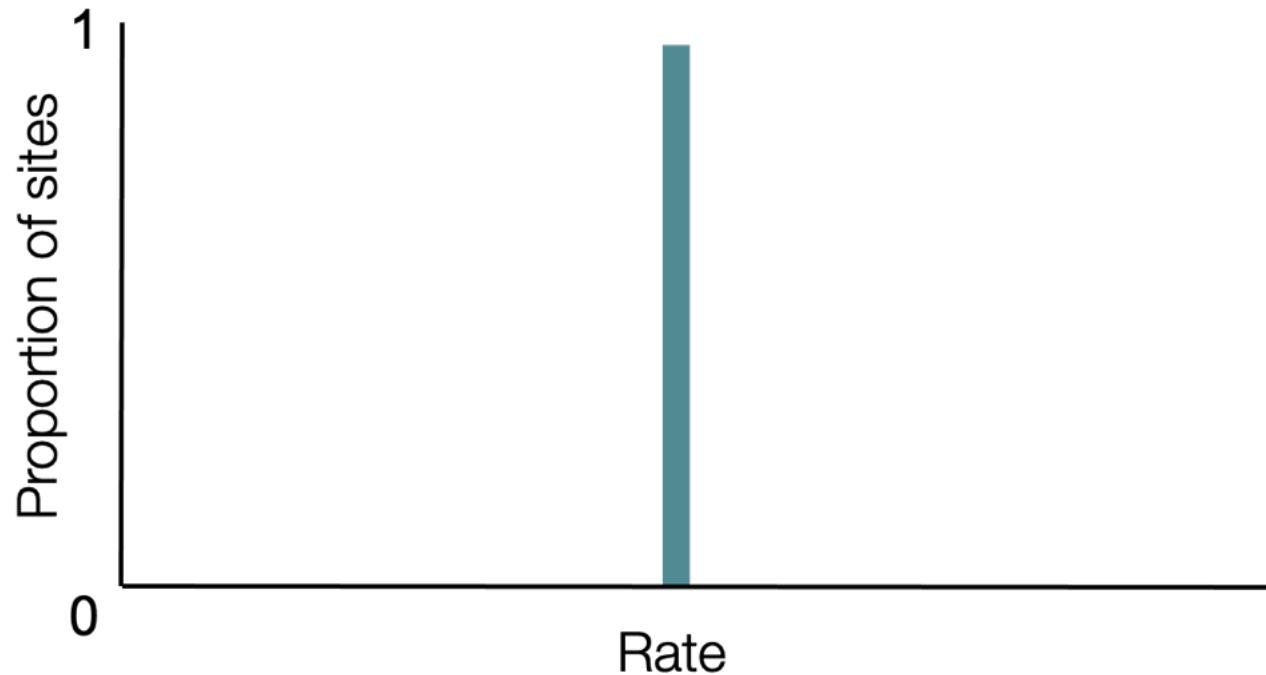
The figure displays five DNA sequence alignments, each consisting of two rows. The top row shows a reference sequence, and the bottom row shows a variant sequence. The sequences are color-coded to represent evolutionary rates: green for Medium, orange for Slow, and blue for Fast. Vertical lines below the sequences categorize positions into Medium, Slow, and Fast rates.

Position	Reference Sequence	Variant Sequence
1	CTAT	CTAT
2	--	--
3	GGC	GGC
4	ACC	AA
5	CC	CC
6	A	C
7	GCCC	AG
8	ATGCAT	CCCATACAT
9	-GGT	-GCT
10		
11	CTATGTGG	CTATGTGG
12	CAACC	AAACC
13	AG	AG
14	GCCC	CCCAG
15	ATGCAT	ATACAT
16	-GCT	-GCT
17		
18	ATATGTGG	ATATGTGG
19	CGCAG	CGCAG
20	CCAG	CCAG
21	-----	-----
22	GCATAGGT	GCATAGGT
23		
24	ATATGTGG	ATATGTGG
25	CGCAG	CGCAG
26	CCAG	CCAG
27	GCAT	CCCCAT
28	ATAGGT	GCATAGGT

Legend: Medium (green), Slow (orange), Fast (blue).

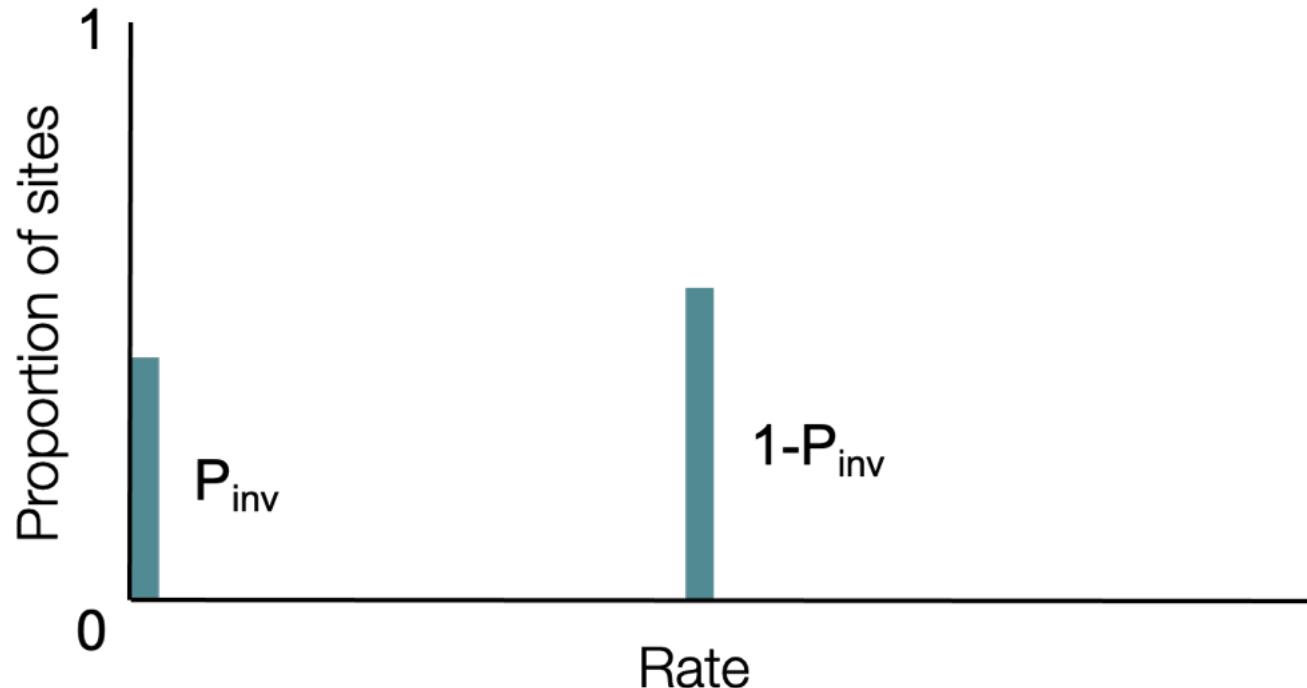
Rate variation across sites

Equal rates among sites (e.g., JC, GTR, HKY models)



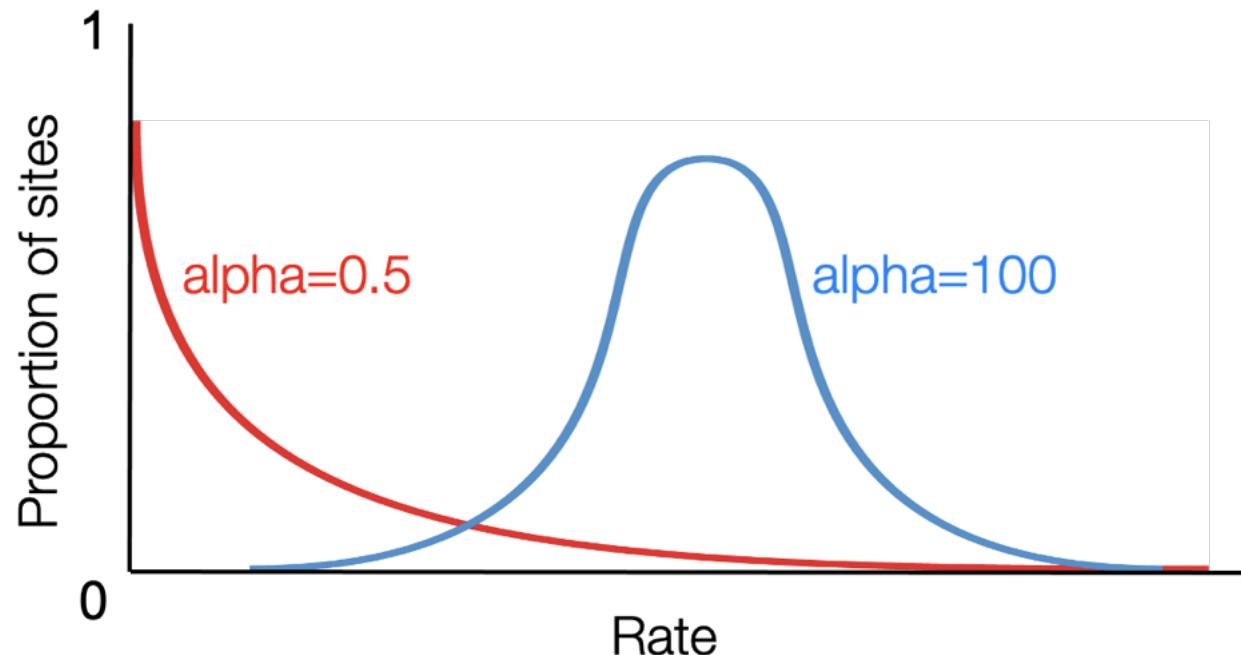
Rate variation across sites

- Proportion of invariable sites (e.g., JC+I, GTR+I, HKY+I models)



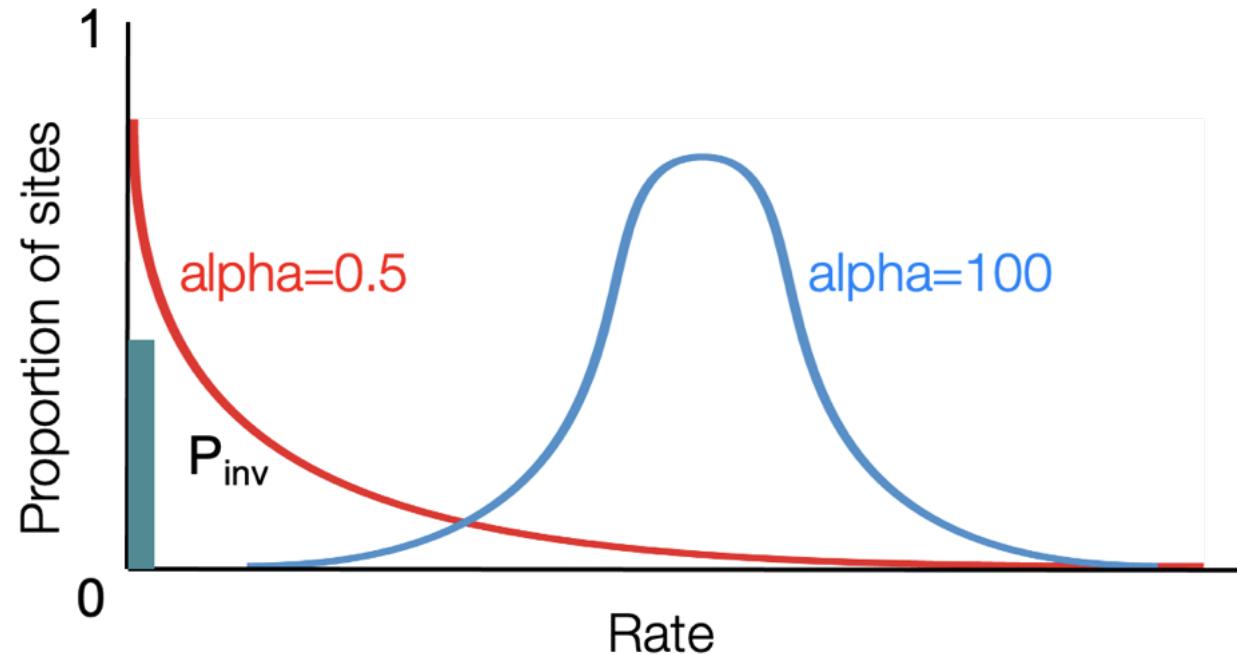
Rate variation across sites

- Gamma-distributed rate variation among sites
(e.g., JC+G, GTR+G, HKY+G models)



Rate variation across sites

- Gamma-distributed rate variation among sites and a proportion of invariable sites (e.g., JC+G+I, GTR+G+I, HKY+G+I models)



Popular phylogenetic methods for inferring trees

- Maximum parsimony
- Distance-based methods
- Maximum likelihood
- Bayesian inference

Model based

Maximum parsimony

Identifies the tree topology that can explain the sequence data, using the smallest number of inferred substitution events

Not model based

Now rarely used for analysing genetic data (revived!)

Cannot estimate evolutionary rates or timescales

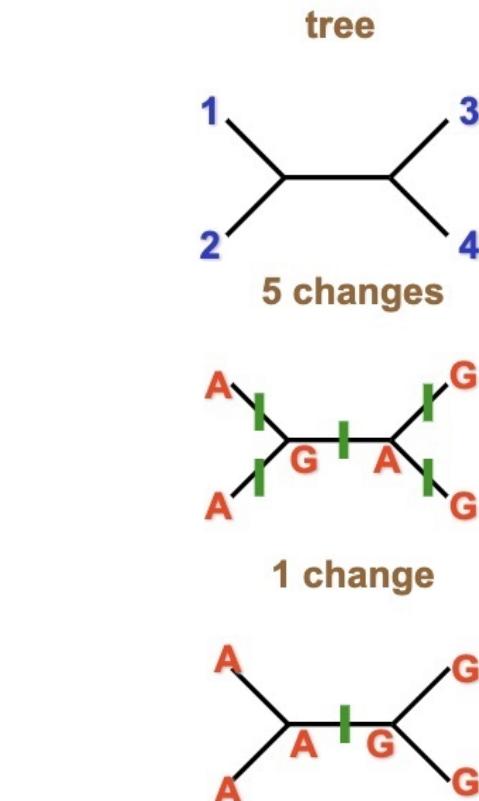
Does not correct for multiple substitutions at the same site

- This leads to a problem known as ‘long-branch attraction’

Long branches in the tree tend to group together



Taxon-1	ATATT
Taxon-2	ATCGT
Taxon-3	GCAGT
Taxon-4	GCCGT



Neighbour Joining method

Neighbor joining takes as input a distance matrix specifying the distance between each pair of taxa.

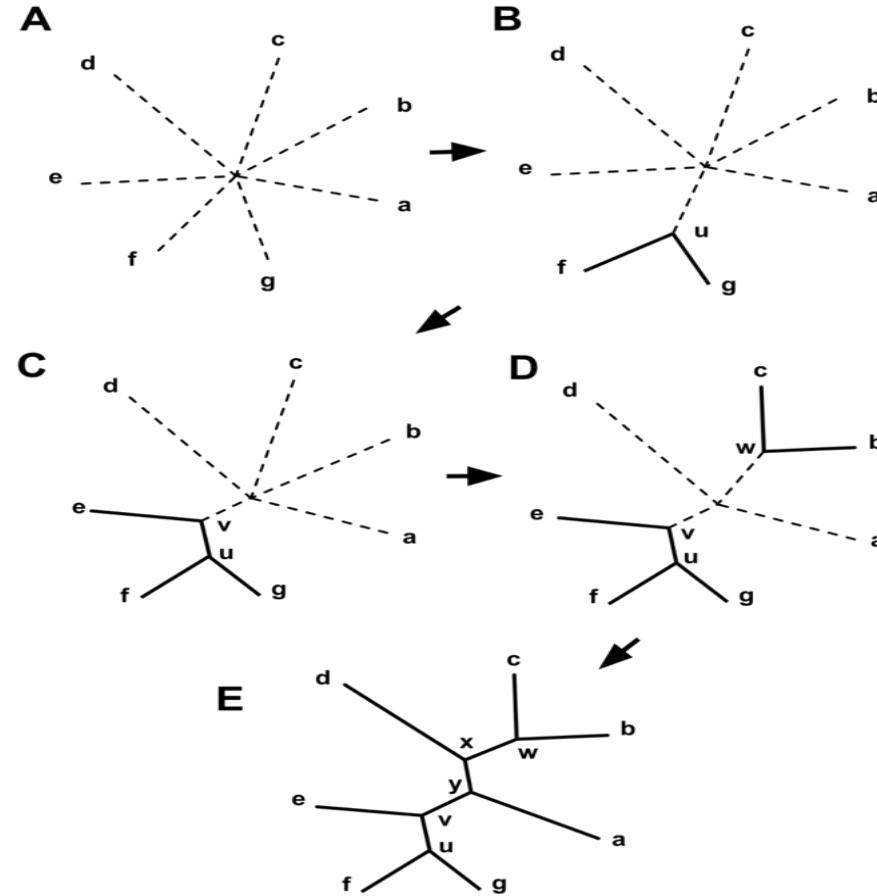
Seq A	ATC ATCG ATTT GCA
Seq B	TTC CCTCG ATAT GCA
Seq C	ATG ATGG ATAT GCA
Seq D	TTCAAC GAATT GCA
Seq E	ATG ATGG ATTT GCA
Seq F	TTCAAC GAAAT GCA



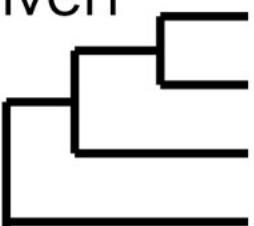
	Matrix: # Differences <input type="button" value="▼"/> Decimal Places: <input type="button" value="0"/> Style: Heatmap and Numbers <input type="button" value="▼"/>											
1_V1A_B/BRISBANE/60/20...	18	39	39	38	37	38	35	35	37	32	28	
B/VICTORIA/502/2015e		41	41	40	39	40	37	37	39	34	30	
B/Brisbane/22/2019_aug	39		0	1	10	21	24	24	26	21	29	
B/New_Caledonia/10/2019...	41			1	10	21	24	24	26	21	29	
B/Washington/02/2019_jan	39	0										
V1A.2_B/Florida/39/2018_j...	41	1	1		9	20	23	23	25	20	28	
B/Perth/1036/2019_jun	38	40										
B/Perth/1036/2019_jun		10	10	9		17	22	22	24	19	27	
B/Canberra/38/2019_jun	37	39										
B/Perth/32/2019_jun	38	40	21	21	20	17	23	23	25	20	28	
B/Darwin/69/2019_jun	35	37	24	24	23	22	23		0	2	5	
B/Malaysia/RP0467/2019_f...	35	37	24	24	23	22	23	0		2	5	
B/Perth/18/2019_apr	37	39	26	26	25	24	25	2	2		7	
B/Perth/18/2019_apr	32	34	21	21	20	19	20	5	5	7		
B/Perth/18/2019_apr	28	30	29	29	28	27	28	25	25	26	22	

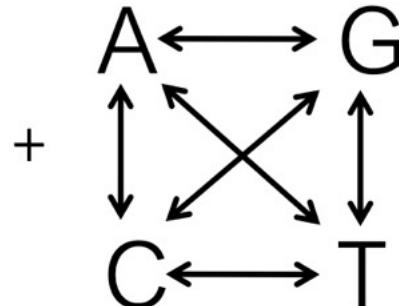
Neighbour Joining method

Connects 2 taxa
with smallest
distance



Maximum likelihood

Given
 Sample1
Sample2
Sample3
Sample4



Probability of?

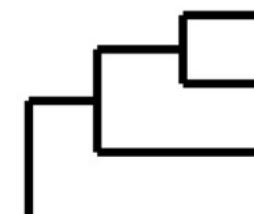
Sample1	CG T TAG G TAC A CT
Sample2	CG A TAG T TCA C CT
Sample3	CG T TAG T TT A CC
Sample4	C ATTGG G TT A CT

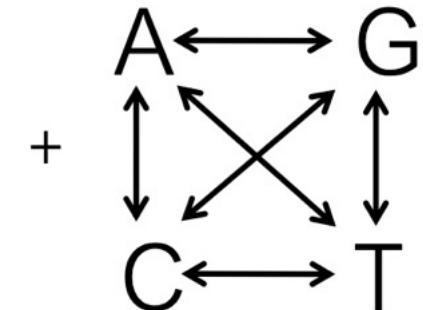
Bayesian inference

Given

Sample1 CG**T**TAG**G**TAC**A**CT
Sample2 CG**A**TAG**T**TCA**C**CT
Sample3 CG**T**TAG**T**TT**A**CC
Sample4 **C**ATTGG**G**TT**A**CT

Probability of?

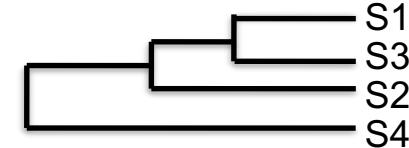
 Sample1
Sample2
Sample3
Sample4



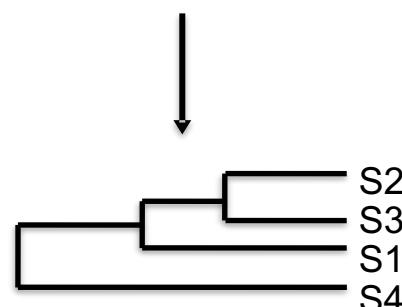
Maximum Likelihood

- Maximum Likelihood is a statistical method of estimating the probability distributions to assign probabilities to a particular possible phylogenetic tree
- It allows for varying rates of evolution across both lineages & nucleotide sites

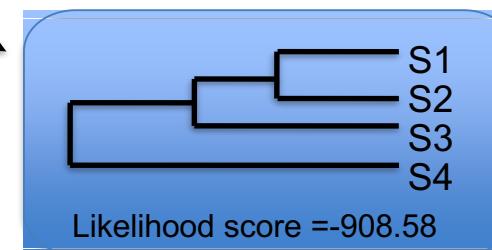
S1 CGTTAGTACACT
S2 CGATAGTTCACT
S3 CGTTAGTTTACC
S4 CATTGGTTTACT



Likelihood score = -1203.83



Likelihood score = -1241.47



Likelihood score = -908.58

Bootstrap

Felsenstein's application of the bootstrap method to evolutionary trees is one of the most cited scientific papers of all time.

The bootstrap method, which is based on resampling and replications, is used extensively to assess the robustness of phylogenetic inferences.

Felsenstein, Joseph. "Confidence limits on phylogenies: an approach using the bootstrap." *evolution* 39.4 (1985): 783-791.

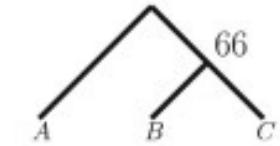
MSA

A	A	C	T	T
B	G	G	A	T
C	G	G	C	C

Original Data

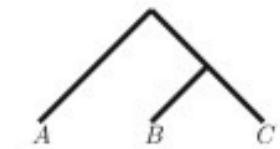


Inferred Tree



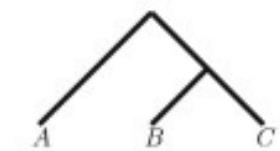
Bootstrap Replicate #1

A	A	C	T	C
B	G	G	A	G
C	G	G	C	G



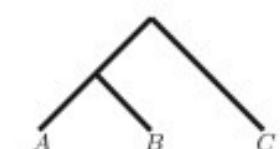
Bootstrap Replicate #2

A	C	A	T	A
B	G	G	A	G
C	G	G	C	G



Bootstrap Replicate #3

A	T	T	T	T
B	A	T	T	A
C	C	C	C	C



Which method?

	Pros	Cons
Clustering • Neighbour Joining	<ul style="list-style-type: none">FastExcellent start tree	<ul style="list-style-type: none">Only a single treeDoesn't consider sequence evolutionCompress sequence information
Tree Searching • Maximum Likelihood	<ul style="list-style-type: none">Evaluate all possible treesConsider sequence evolutionBetter estimates of branch lengths	<ul style="list-style-type: none">SlowComputationally intenseIf you use the method incorrectly, you will get the incorrect answer (with high statistical support)

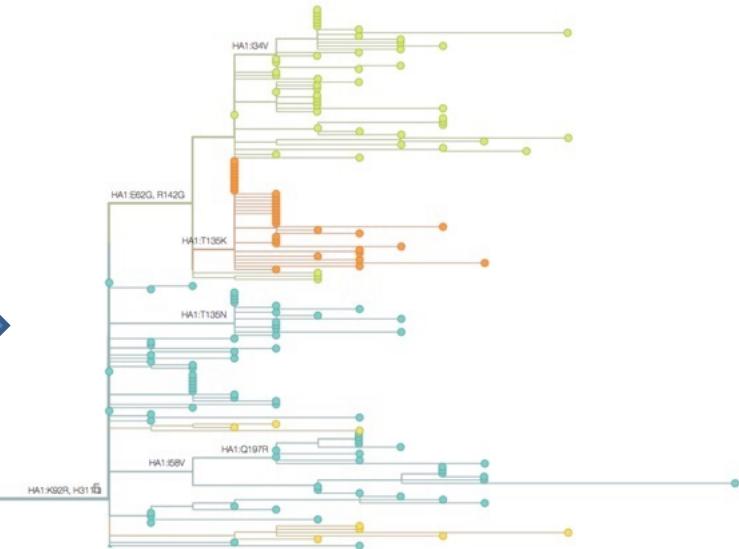
Practical: Phylogenetic Analysis

“Data” is generally a sequence alignment

1. A/HongKong/VB17037915/2017
 2. A/Hebei/27407/2017
 3. A/Jiangsu/60447/2016
 4. A/Shanghai/51907/2013
 5. A/Chicken/Jiangxi/1848
 6. A/Zhejiang/5/2017
 7. A/Guangdong/17SF033/2017
 8. A/Anhui/1883/2014
 9. A/Guangxi/1810/2017
 10. A/Beijing/9149/2014
 11. A/Zhejiang/22/2013
 12. A/Environment/Nanjing/2913
 13. A/Anhui/33225/2015
 14. A/Anhui/1876/2014
 15. A/Environment/InnerMongolia/2865
 16. A/Chicken/Dongguan/1143
 17. A/Environment/Henan/SC232/2013
 18. A/Fujian/12/2017
 19. A/Guangxi/5/2017
 20. A/Beijing/40610/2015
 21. A/Zhejiang/7807/2014
 22. A/HongKong/5942/2013
 23. A/Henan/14905/2017
 24. A/Fujian/56600/2016
 25. A/Anhui/13423/2017
 26. A/Guangdong/60923/2016
 27. A/Zhejiang/15/2016
 28. A/Yunnan/13500/2017
 29. A/Environment/Shandong-Linyi/EV01/2015
 30. A/Henan/11159/2017
 31. A/Zhejiang/33/2014
 32. A/Jiangsu/6306/2014
 33. A/XinjiangBintuan/99117/2014
 34. A/Yunnan/13501/2017
 35. A/Changsha/1/2013
 36. A/HongKong/125/2017
 37. A/Jiangxi/27569/2014
 38. A/Environment/Shandong/SD039/2013
 39. A/Anhui/33228/2015
 40. A/Chicken/Shenzhen/749/2013
 41. A/Beijing/1-A/2013
 42. A/Fujian/16/2014
 43. A/Taiwan/1/2017
 44. A/Hubei/34007/2015
 45. A/Environment/Hunan/3971
 46. A/Fujian/8/2017
 47. A/Environment/Guangdong/4740

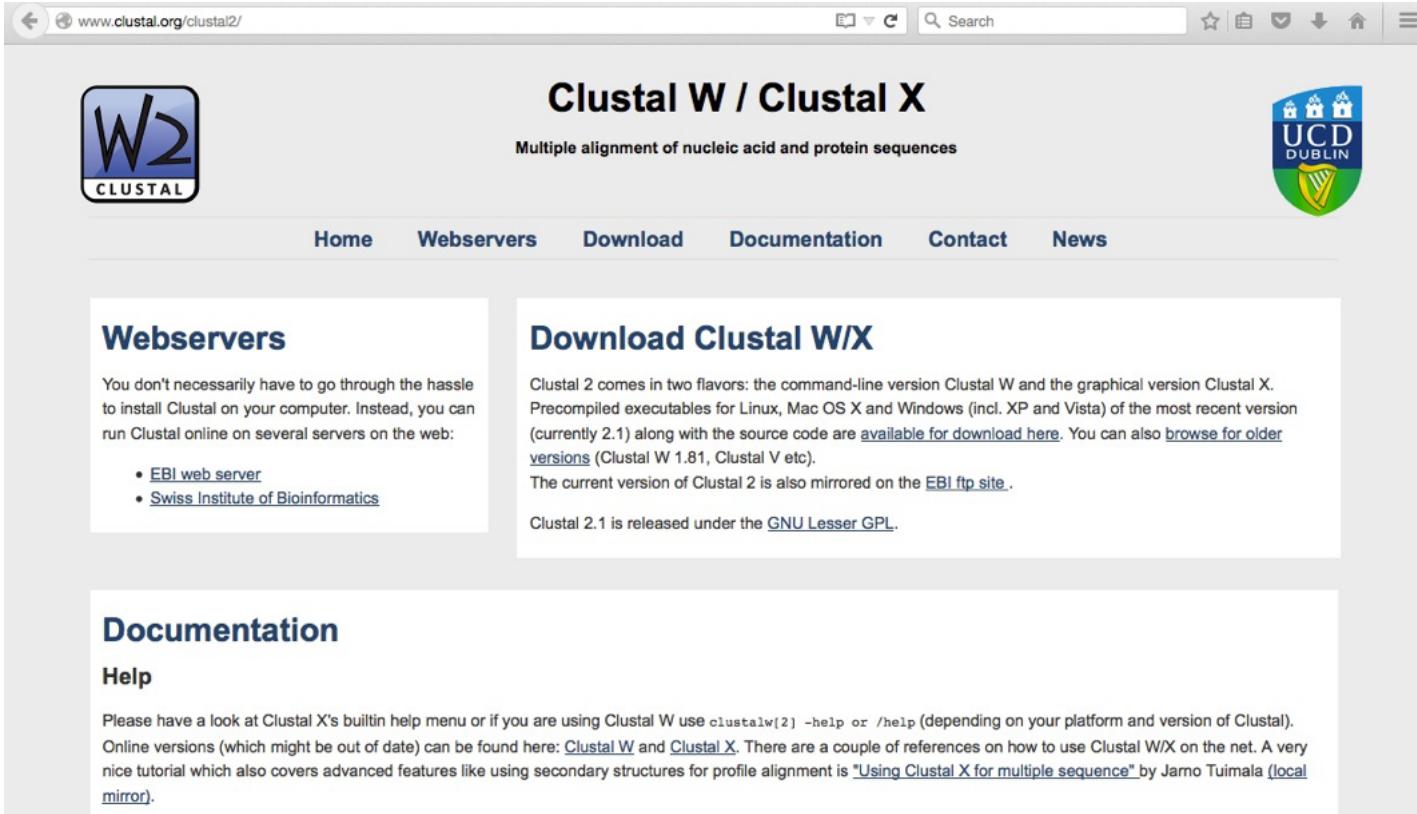
Phylogenetic analysis

inference



alignment

Multiple sequence alignment



The screenshot shows the homepage of the Clustal W / Clustal X website. At the top left is the W2 CLUSTAL logo. In the center is the title "Clustal W / Clustal X" and the subtitle "Multiple alignment of nucleic acid and protein sequences". At the top right is the UCD DUBLIN logo. Below the title is a navigation bar with links: Home, Webservers, Download, Documentation, Contact, and News. The main content area is divided into three sections: "Webservers" (with links to EBI web server and Swiss Institute of Bioinformatics), "Download Clustal W/X" (describing the two versions and providing download links), and "Documentation" (with a "Help" section and a note about the help menu). The URL in the browser bar is www.clustal.org/clustal2/.

www.clustal.org/clustal2/

Clustal W / Clustal X

Multiple alignment of nucleic acid and protein sequences

UCD DUBLIN

Home Webservers Download Documentation Contact News

Webservers

You don't necessarily have to go through the hassle to install Clustal on your computer. Instead, you can run Clustal online on several servers on the web:

- [EBI web server](#)
- [Swiss Institute of Bioinformatics](#)

Download Clustal W/X

Clustal 2 comes in two flavors: the command-line version Clustal W and the graphical version Clustal X. Precompiled executables for Linux, Mac OS X and Windows (incl. XP and Vista) of the most recent version (currently 2.1) along with the source code are [available for download here](#). You can also [browse for older versions](#) (Clustal W 1.81, Clustal V etc.).

The current version of Clustal 2 is also mirrored on the [EBI ftp site](#).

Clustal 2.1 is released under the [GNU Lesser GPL](#).

Documentation

Help

Please have a look at Clustal X's builtin help menu or if you are using Clustal W use `clustalw[2] -help` or `/help` (depending on your platform and version of Clustal). Online versions (which might be out of date) can be found here: [Clustal W](#) and [Clustal X](#). There are a couple of references on how to use Clustal W/X on the net. A very nice tutorial which also covers advanced features like using secondary structures for profile alignment is "[Using Clustal X for multiple sequence](#)" by Jarno Tuimala ([local mirror](#)).

Multiple sequence alignment

The screenshot shows the MUSCLE web interface on the EMBL-EBI website. The page has a dark header with the EMBL-EBI logo, a search bar, and links for Services, Research, Training, and About us. Below the header is a teal banner with the word "MUSCLE". The main content area has a dark header with "Input form", "Web services", "Help & Documentation", "Share", and "Feedback". A breadcrumb navigation shows "Tools > Multiple Sequence Alignment > MUSCLE". The main content is titled "Multiple Sequence Alignment". It explains that MUSCLE stands for MUltiple Sequence Comparison by Log- Expectation and compares its performance with ClustalW2 or T-Coffee. There are three main steps: "STEP 1 - Enter your input sequences", "STEP 2 - Set your Parameters", and "STEP 3 - Submit your job". Step 1 contains a text area for pasting sequences and a file upload section. Step 2 allows setting the output format (set to ClustalW) and provides a link to view or change default settings. Step 3 includes a checkbox for being notified by email.

www.ebi.ac.uk/Tools/msa/muscle/

EMBL-EBI

MUSCLE

Input form Web services Help & Documentation Share Feedback

Tools > Multiple Sequence Alignment > MUSCLE

Multiple Sequence Alignment

MUSCLE stands for MUltiple Sequence Comparison by Log- Expectation. MUSCLE is claimed to achieve both better average accuracy and better speed than ClustalW2 or T-Coffee, depending on the chosen options.

STEP 1 - Enter your input sequences

Enter or paste a set of sequences in any supported format:

Or upload a file: No file selected.

STEP 2 - Set your Parameters

OUTPUT FORMAT:

The default settings will fulfill the needs of most users and, for that reason, are not visible.

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

Be notified by email. (Tick this box if you want to be notified by email when the results are available)

MEGA

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

The screenshot shows the official website for MEGA (Molecular Evolutionary Genetics Analysis). The header features the MEGA logo with the letters M, E, G, A in red, green, blue, and yellow respectively, followed by the text "Molecular Evolutionary Genetics Analysis". Below the header is a navigation bar with links for "home", "features", "publications", "manual", and "feedback". A decorative banner at the top displays various biological icons representing different species and processes. A prominent dark banner in the center contains a red bell icon and text about testing the new macOS version. Below this is a descriptive paragraph: "Sophisticated and user-friendly software suite for analyzing DNA and protein sequence data from species and populations." At the bottom, there are dropdown menus for "macOS", "Graphical (GUI)", and "MEGA 7 (32-bit)", along with a large green "DOWNLOAD" button. The main content area is divided into three columns: "Sequence Analyses" (Phylogeny Inference, Model Selection, Dating and Clocks, Ancestral States, Selection and Tests, Sequence Alignment), "Statistical Methods" (Maximum Likelihood, Distance Methods, Ordinary Least Squares, Maximum Parsimony, Composite Likelihood, Bayesian), and "Powerful Visual Tools" (Alignment/Trace Editor, Tree Explorer, Data Explorers, Legend Generator, Gene Duplication Wizard, Timetree Wizard).

Molecular Evolutionary
Genetics Analysis

home features publications manual feedback

Volunteers are needed for testing the new macOS version of MEGA X. If you are a macOS user and would like to help, please [sign up here](#). Note: this version is only for 64-bit macOS systems and is only intended for testing.

Sophisticated and user-friendly software suite for analyzing DNA and protein sequence data from species and populations.

macOS Graphical (GUI) MEGA 7 (32-bit) DOWNLOAD

Sequence Analyses

- Phylogeny Inference
- Model Selection
- Dating and Clocks
- Ancestral States
- Selection and Tests
- Sequence Alignment

Statistical Methods

- Maximum Likelihood
- Distance Methods
- Ordinary Least Squares
- Maximum Parsimony
- Composite Likelihood
- Bayesian

Powerful Visual Tools

- Alignment/Trace Editor
- Tree Explorer
- Data Explorers
- Legend Generator
- Gene Duplication Wizard
- Timetree Wizard



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



RATES



CLOCKS



DISEASE



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE

Load FASTA file

Molecular Evolutionary Genetics Analysis



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



RATES



CLOCKS



DISEASE



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE



Molecular Evolutionary Genetics Analysis



ALIGN



Open a File/Session...



Concatenate Sequence Alignments

Explore Active Data

F4

Export Data...

Save Data Session to File...

Select Genetic Code Table...

Select Genes and Domains...

Select Taxa and Groups...



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE

Molecular Evolutionary Genetics Analysis



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



RATES



CLOCKS



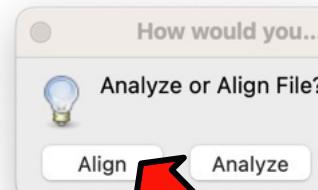
DISEASE



TIMETREE



DATAMONKEY



RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE

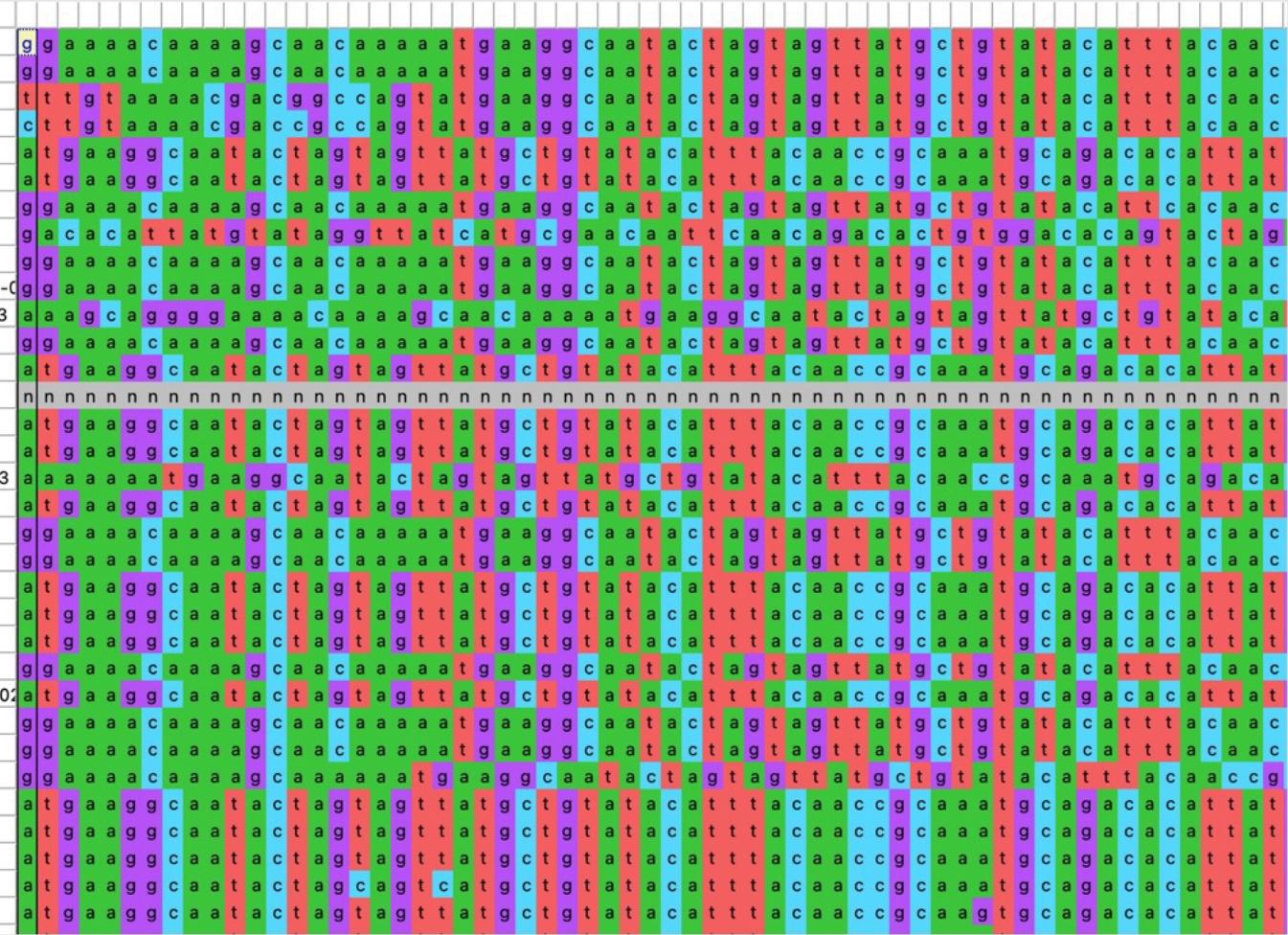
M11: Alignment Explorer (australia_0423.fasta)



DNA Sequences Translated Protein Sequences

Species/Abbrv

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



Site # 1



with



w/o gaps

Selected genetic code: Standard

MEGA11 Data Edit Search Alignment Web Sequencer Display Help

W Align by ClustalW
W Align by ClustalW (Codons)
Align by MUSCLE
Align by MUSCLE (Codons)
Mark/Unmark Site
Align Marked Sites
Unmark All Sites
Delete Gap-Only Sites
✓ Auto-Fill Gaps

australia_0423.fasta)

ed Protein Sequences

Species/Abbrv

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

Site # 1 with w/o gaps Selected genetic code: Standard



DNA Sequences Translated Protein Sequences

Species/Abbrv

1. 5a.2a A/India/Pun-N	g a a a a c a a a a g c a a c a a a a t g a a g g c a a t a c a t g t a t a c a t t a c a a a c c g c a a a t g
2. 5a.2a.1 A/Alagoas/2	g a a a a c a a a a a g c a a c a a a a t g a a g g c a a t a c a t t a c a a a c c g c a a a t g
3. A/Chungbuk/871/20	t t g t a a a a
4. A/Ulsan/867/2023	c t t g t a a a a
5. A/Netherlands/1051	a t g a a g g c a
6. A/Catalonia/NSVH10	a t g a a g g c a
7. A/Wisconsin/37/202	g g a a a a c a a a
8. A/Manitoba/RV0038	g a c a c a t t a
9. A/Denmark/516/202	g g a a a a c a a a
10. A/BosniaandHerzeg	g g a a a a c a a a
11. A/Saskatchewan/SK	a a g c a g g g
12. A/Washington/23/2	g g a a a a c a a a
13. A/Romania/547395	a t g a a g g c a
14. A/Lisboa/15/2023	n n n n n n n n n
15. A/Arkhangelsk/CRI	a t g a a g g c a
16. A/Norway/04936/2	a t g a a g g c a
17. A/Amazonas/2023-	a a a a a a a t g
18. A/Moscow oblast/C	a t g a a g g c a
19. A/Arizona/24/2023	g g a a a a c a a a
20. A/Iowa/25/2023	g g a a a a c a a a
21. A/Melilla/1041/2023	a t g a a g g c a
22. A/Ceuta/565/2023	a t g a a g g c a
23. A/Catalonia/NSVH1	a t g a a g g c a a t a c a t g t a t a c a t t a c a a a c c g o
24. A/South Dakota/21	g g a a a a c a a a a g c a a c a a a a t g a a g g c a a t a c a t g t a t g c
25. A/Mato Grosso do	a t g a a g g c a a t a c a t g t a t g c t g t a t a c a t t a c a a a c c g c a a t g c
26. A/Montana/24/2022	g g a a a a c a a a a a g c a a c a a a a a t g a a g g c a a t a c a t g t a t g c t g t a t a c a a a c c g c a a t g
27. A/District Of Colum	g g a a a a c a a a a a g c a a c a a a a a t g a a g g c a a t a c a t g t a t g c t g t a t a c a a a c c g c a a t g
28. A/Minnesota/16/2020	g g a a a a c a a a a a g c a a a a a a t g a a g g c a a t a c a t g t a t g c t g t a t a c a a a c c g c a g a t g c a
29. A/Victoria/380B/2020	a t g a a g g c a a t a c a t g t a t g c t g t a t a c a t t a c a a a c c g c a a t g c a g a c a c a t t a g t a t a g g t
30. A/Victoria/380C/2020	a t g a a g g c a a t a c a t g t a t g c t g t a t a c a t t a c a a a c c g c a a t g c a g a c a c a t t a g t a t a g g t
31. A/Victoria/453/2020	a t g a a g g c a a t a c a t g t a t g c t g t a t a c a t t a c a a a c c g c a a t g c a g a c a c a t t a g t a t a g g t
32. A/Victoria/454/2020	a t g a a g g c a a t a c a t g c a g t c a t g c t g t a t a c a t t a c a a a c c g c a a a t g c a g a c a c a t t a g t a t a g g t
33. A/Victoria/488/2020	a t g a a g g c a a t a c a t g t a t g c t g t a t a c a t t a c a a a c c g c a a g t g c a g a c a c a t t a g t a t a g g t

MUSCLE Alignment Options

Option Setting

GAP PENALTIES

Gap Open -400.00Gap Extend 0.00

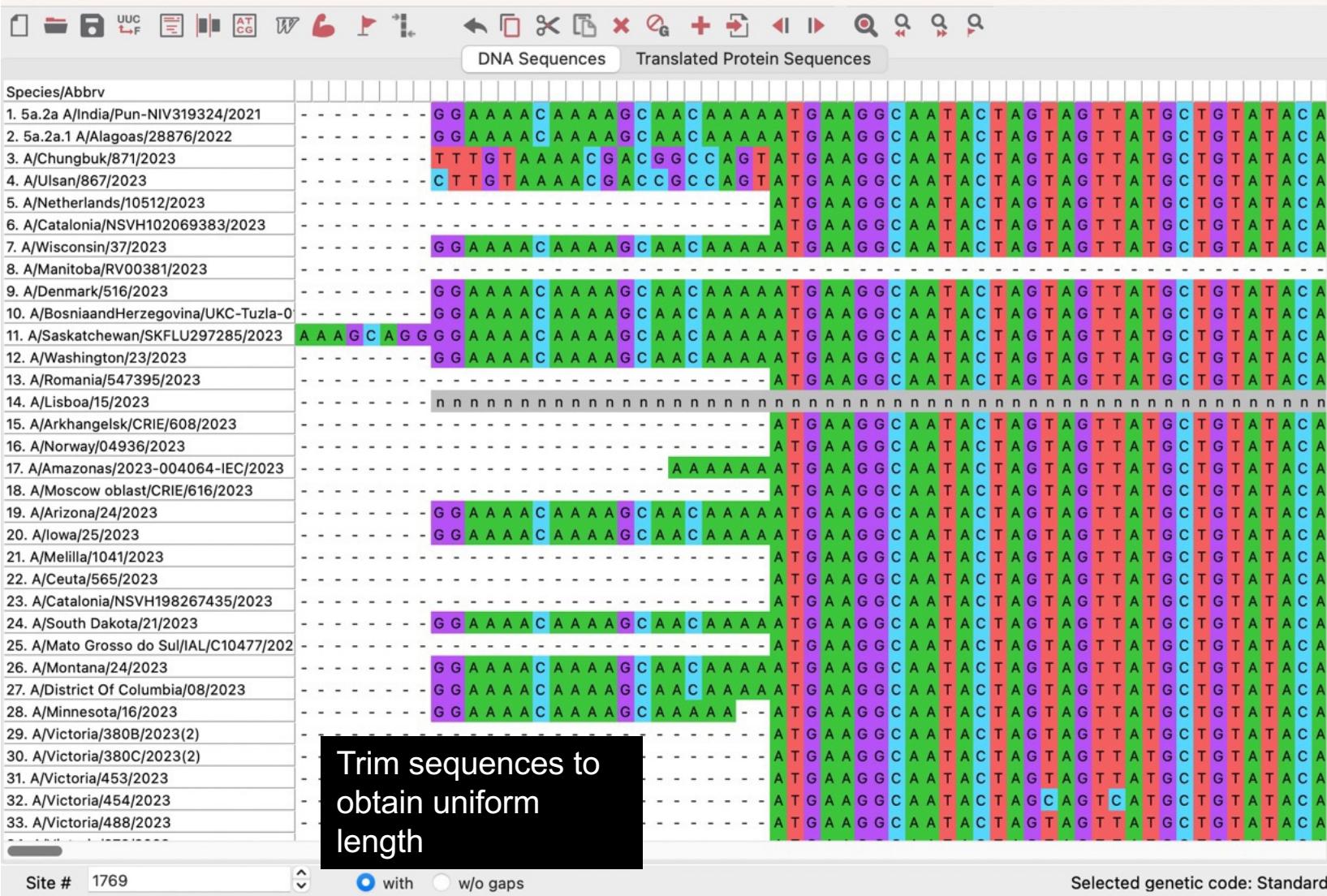
MEMORY/ITERATIONS

Max Memory in MB 2048Max Iterations 16

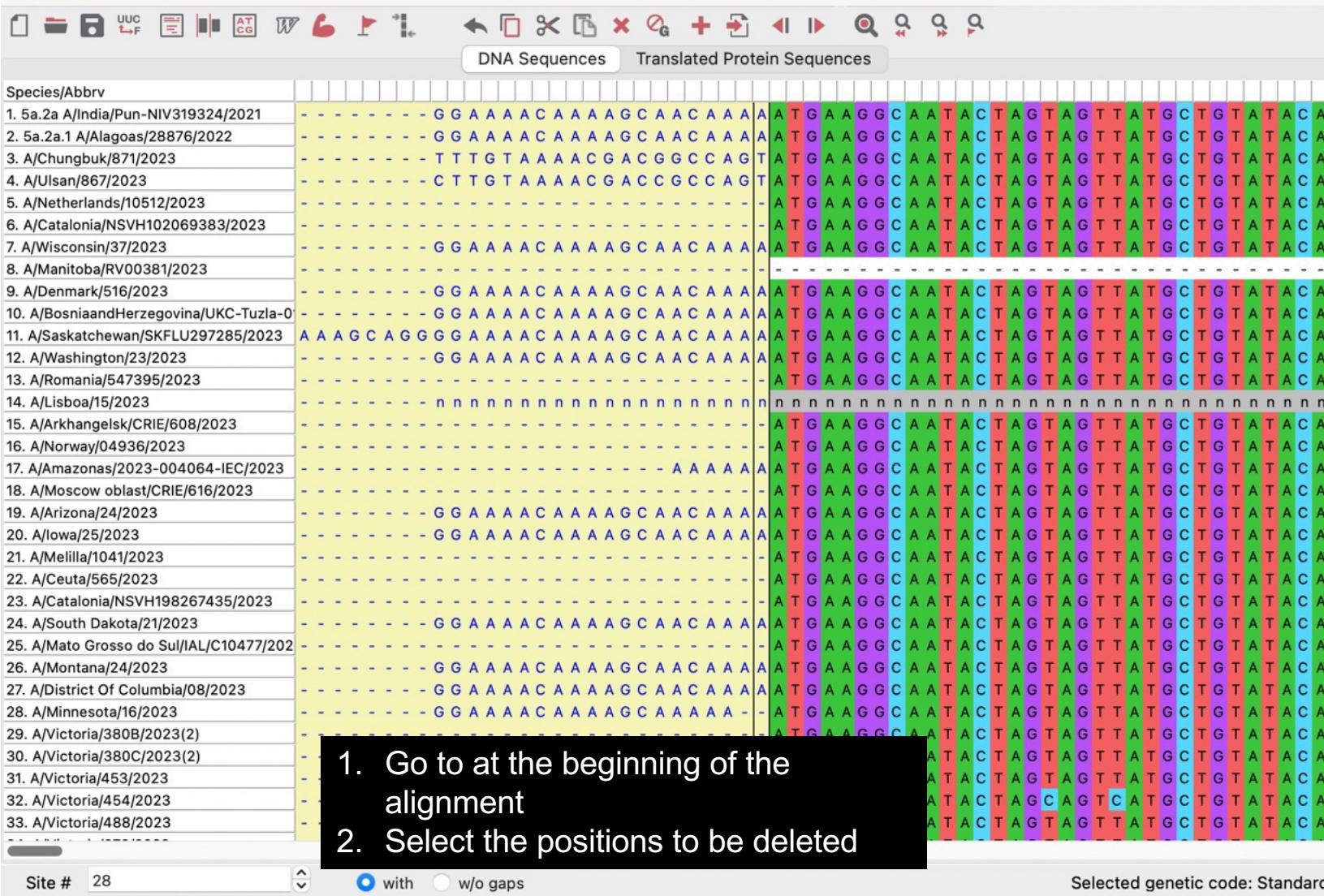
ADVANCED OPTIONS

Cluster Method (Iterations 1,2) UPGMACluster Method (Other Iterations) UPGMAMin Diag Length (Lambda) 24 Help

M11: Alignment Explorer (australia_0423.fasta)



M11: Alignment Explorer (australia_0423.fasta)



undo ⌘ Z
Copy ⌘ C
Cut ⌘ X
Paste ⌘ V
Delete ⌘ D Red arrow pointing here
Delete Gaps ⌘ G
Insert Blank Sequence ⌘ N
Insert Sequence From File ⌘ I
Select Site(s)
Select Sequence(s)
Select All ⌘ A
Allow Base Editing
Modify All Bases To Upper Case

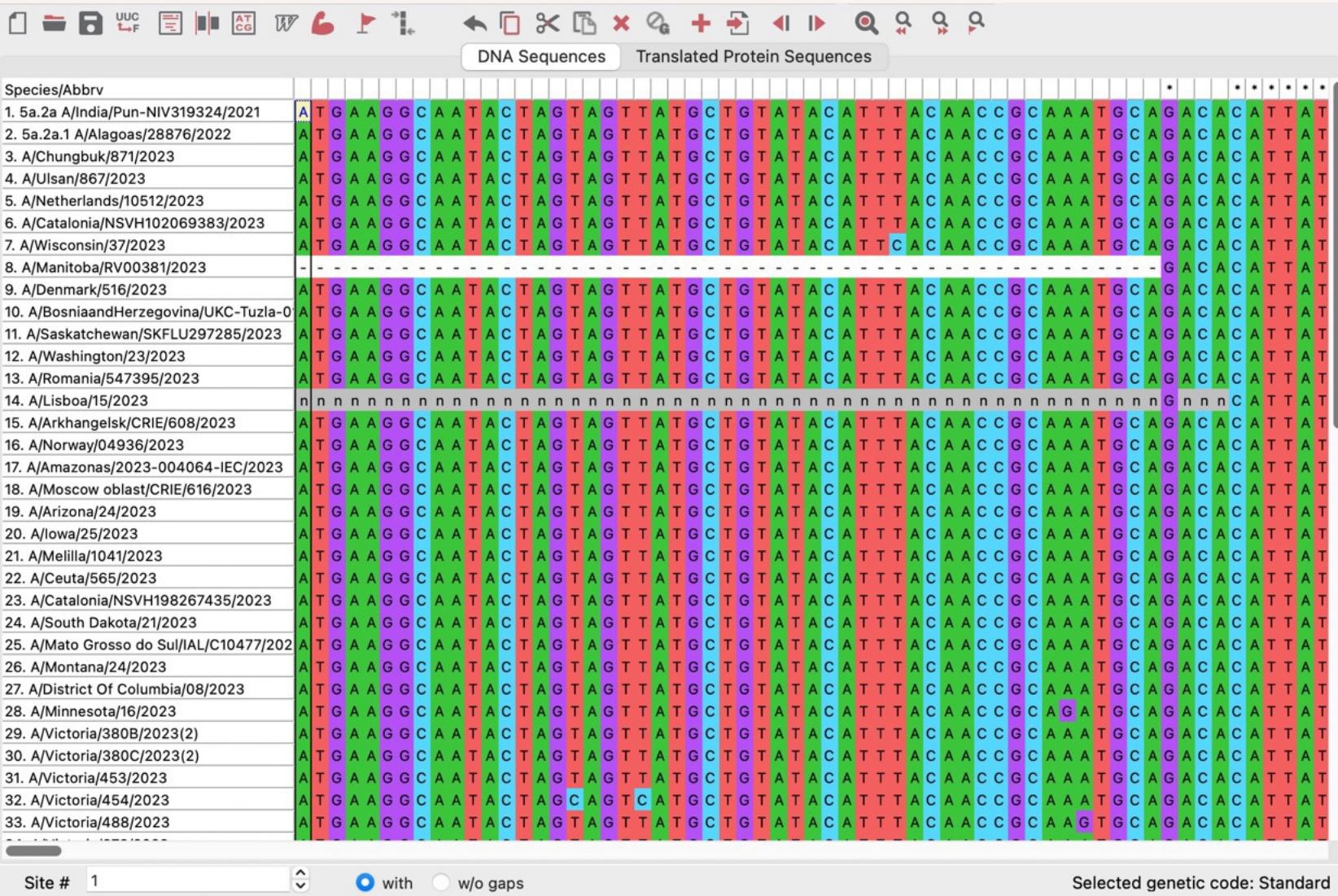
Species/Abbrv
1. 5a.2a A/India/Pun-NIV319
2. 5a.2a.1 A/Alagoas/28876/
3. A/Chungbuk/871/2023
4. A/Ulsan/867/2023
5. A/Netherlands/10512/2022
6. A/Catalonia/NSVH102069
7. A/Wisconsin/37/2023
8. A/Manitoba/RV00381/2022
9. A/Denmark/516/2023
10. A/BosniaandHerzegovina/
11. A/Saskatchewan/SKFLU2
12. A/Washington/23/2023
13. A/Romania/547395/2023
14. A/Lisboa/15/2023
15. A/Arkhangelsk/CRIE/608
16. A/Norway/04936/2023
17. A/Amazonas/2023-004064-IEC/2023
18. A/Moscow oblast/CRIE/616/2023
19. A/Arizona/24/2023
20. A/Iowa/25/2023
21. A/Melilla/1041/2023
22. A/Ceuta/565/2023
23. A/Catalonia/NSVH198267435/2023
24. A/South Dakota/21/2023
25. A/Mato Grosso do Sul/IAL/C10477/2022
26. A/Montana/24/2023
27. A/District Of Columbia/08/2023
28. A/Minnesota/16/2023
29. A/Victoria/380B/2023(2)
30. A/Victoria/380C/2023(2)
31. A/Victoria/453/2023
32. A/Victoria/454/2023
33. A/Victoria/488/2023

Site # 28 with w/o gaps Selected genetic code: Standard

Sequence Explorer (australia_0423.fasta)

Translated Protein Sequences

M11: Alignment Explorer (australia_0423.fasta)





DNA Sequences Translated Protein Sequences

Species/Abbrv

41. A/Tasmania/30/2023	G T G T A G A A T A T G T A T T T A A -
42. A/Brisbane/13/2023	G T G T A G A A T A T G T A T T T A A -
43. A/South Australia/85/2023	G T G T A G A A T A T G T A T T T A A -
44. A/Canberra/43/2023	G T G T A G A A T A T G T A T T T A A -
45. A/South Australia/86/2023	G T G T A G A A T A T G T A T T T A A -
46. A/South Australia/90/2023	G T G T A G A A T A T G T A T T T A A -
47. A/Canberra/50/2023	G T G T A G A A T A T G T A T T T A A -
48. A/Canberra/49/2023	G T G T A G A A T A T G T A T T T A A -
49. A/South Australia/83/2023	G T G T A G A A T A T G T A T T T A A -
50. A/Tasmania/25/2023	G T G T A G A A T A T G T A T T T A A -
51. A/Victoria/380A/2023	G T G T A G A A T A T G T A T T T A A -
52. A/Victoria/380B/2023	G T G T A G A A T A T G T A T T T A A -
53. A/Canberra/47/2023	G T G T A G A A T A T G T A T T T A A -
54. A/Victoria/382/2023	G T G T A G A A T A T G T A T T T A A -
55. A/Victoria/473/2023	G T G T A G A A T A T G T A T T T A A -
56. A/Brisbane/24/2023	G T G T A G A A T A T G T A T T T A A -
57. A/Victoria/490/2023	G T G T A G A A T A T G T A T T T A A -
58. A/Darwin/53/2023	G T G T A G A A T A T G T A T T T A A -
59. A/Victoria/445/2023	G T G T A G A A T A T G T A T T T A A -
60. A/Sydney/83/2023	G T G T A G A A T A T G T A T T T A A -
61. A/Darwin/51/2023	G T G T A G A A T A T G T A T T T A A -
62. A/Victoria/409/2023	G T G T A G A A T A T G T A T T T A A -
63. A/Sydney/75/2023	G T G T A G A A T A T G T A T T T A A -
64. A/Sydney/81/2023	G T G T A G A A T A T G T A T T T A A -
65. A/Sydney/79/2023	G T G T A G A A T A T G T A T T T A A -
66. A/Victoria/424/2023	G T G T A G A A T A T G T A T T T A A -
67. A/Tasmania/53/2023	G T G T A G A A T A T G T A T T T A A -
68. A/Tasmania/54/2023	G T G T A G A A T A T G T A T T T A A -
69. A/Victoria/380C/2023	G T G T A G A A T A T G T A T T T A A -
70. A/Victoria/540/2023	G T G T A G A A T A T G T A T T T A A -
71. A/Victoria/537/2023	G T G T A G A A T A T G T A T T T A A -
72. A/Victoria/380/2023	G T G T A G A A T A T G T A T T T A A -
73. A/Sydney/101/2023	G T G T A G A A T A T G T A T T T A A -

1. Go to at the end of the alignment
2. Select the positions to be deleted

Species/Abbrv

- 41. A/Tasmania/30/2023
- 42. A/Brisbane/13/2023
- 43. A/South Australia/85/20
- 44. A/Canberra/43/2023
- 45. A/South Australia/86/20
- 46. A/South Australia/90/20
- 47. A/Canberra/50/2023
- 48. A/Canberra/49/2023
- 49. A/South Australia/83/20
- 50. A/Tasmania/25/2023
- 51. A/Victoria/380A/2023
- 52. A/Victoria/380B/2023
- 53. A/Canberra/47/2023
- 54. A/Victoria/382/2023
- 55. A/Victoria/473/2023
- 56. A/Brisbane/24/2023
- 57. A/Victoria/490/2023
- 58. A/Darwin/53/2023
- 59. A/Victoria/445/2023
- 60. A/Sydney/83/2023
- 61. A/Darwin/51/2023
- 62. A/Victoria/409/2023
- 63. A/Sydney/75/2023
- 64. A/Sydney/81/2023
- 65. A/Sydney/79/2023
- 66. A/Victoria/424/2023
- 67. A/Tasmania/53/2023
- 68. A/Tasmania/54/2023
- 69. A/Victoria/380C/2023
- 70. A/Victoria/540/2023
- 71. A/Victoria/537/2023
- 72. A/Victoria/380/2023
- 73. A/Sydney/101/2023

Undo ⌘ Z
Copy ⌘ C
Cut ⌘ X
Paste ⌘ V
Delete ⌘ D
Delete Gaps ⌘ ⌫ G
Insert Blank Sequence ⌘ N
Insert Sequence From File ⌘ I
Select Site(s)
Select Sequence(s)
Select All ⌘ A
Allow Base Editing
Modify All Bases To Upper Case

	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
41.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
42.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
43.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
44.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
45.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
46.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
47.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
48.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
49.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
50.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
51.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
52.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
53.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
54.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
55.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
56.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
57.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
58.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
59.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
60.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
61.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
62.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
63.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
64.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
65.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
66.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
67.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
68.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
69.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
70.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
71.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
72.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A
73.	G	T	G	T	A	G	A	T	A	T	G	T	A	T	T	T	T	A	A

M11: Alignment Explorer (australia_0423.mas)



DNA Sequences Translated Protein Sequences

Species/Abbrv

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

Site # 1701

 with w/o gaps

Selected genetic code: Standard

Save aligned sequences

MEGA11 Data Edit Search Alignment Web Sequencer Display Help

Create New ⌘ N

Open ⌘ O >

Open a Recently Used File >

Close

Phylogenetic Analysis

Save Session ⌘ S

Export Alignment >

DNA Sequences

Protein Sequences

Translate/Untranslate

Genetic Code

Reverse Complement

Reverse

Complement

Quit ⌘ F4

M11: Alignment Explorer (australia_0423.mas)

DNA Sequences Translated Protein Sequences

MEGA Format

FASTA Format

NEXUS/PAUP Format

Site # 1701

with w/o gaps

Selected genetic code: Standard

M11: Alignment Explorer (australia_0423.mas)



DNA Sequences Translated Protein Sequences

Species/Abbrv

1. 5a.2a A/India/Pun-NIV319324/2021

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

7. A/Wisconsin/37/2023

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

16. A/Norway/04936/2023

GGG 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

GGG 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

GGG 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

19. A/Arizona/24/2023

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

24. A/South Dakota/21/2023

GGG 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

GGG 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

GGG 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

GGG 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

GGG 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

Save file as

Save As: australia_0423_aligned.fasta

Tags:

Where: 2023_Timor-Leste

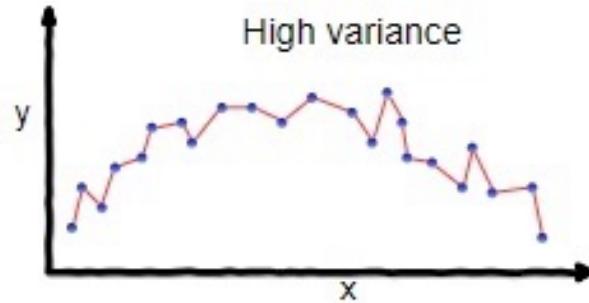
Format: FASTA Files

Cancel

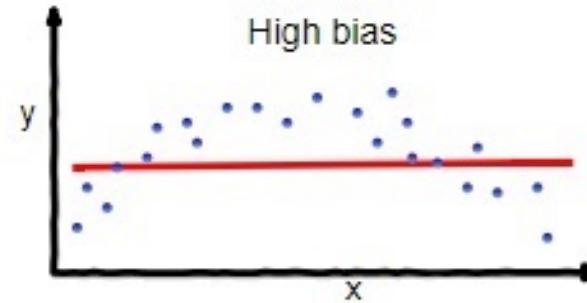
Save

Search for the best-fit nucleotide substitution model

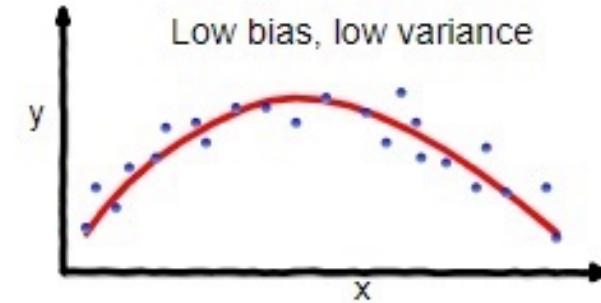
Model Selection



overfitting



underfitting



Good balance

Bias-variance trade-off

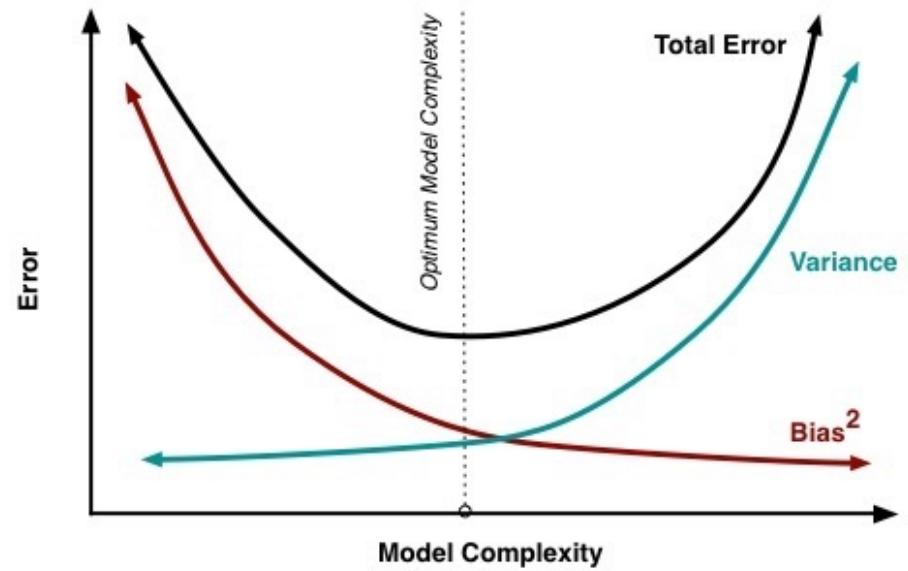
Adding more parameters always improves the fit of the model to the observed data

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)



Molecular Evolutionary Genetics Analysis



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



RATES



CLOCKS



DISEASE



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE



Molecular Evolutionary Genetics Analysis



ALIGN



DATA



TRE

ANCESTORS

SELECTION

RATES

CLOCKS

DISEASE

Find Best DNA/Protein Models (ML)...

Disparity Index Test of Pattern Heterogeneity

Estimate Substitution Matrix (ML)...

Estimate Transition/Transversion Bias (ML)...

Compute MCL Substitution Matrix

Compute MCL Transition/Transversion Bias

Compute Pattern Disparity Index

Compute Composition Distances

Compute Amino Acid Composition

Compute Nucleotide Composition

Compute Codon Usage Bias



TIMETREE



DATAMONKEY



RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE



Molecular Evolutionary Genetics Analysis



ALIGN



DATA



MODEL

Close
Data

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 → 1st 2nd 3rd Noncoding Sites

Branch Swap Filter → None

Number of Threads → 4



Help



Cancel



OK

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE

CLOCKS



DISEASE



TIMETREE



DATA MONKEY





Molecular Evolutionary Genetics Analysis



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



RATES



CLOCKS



DISEASE



M11: Progress

PROGRESS •

DETAILS STOP

STATUS/OPTIONS

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



RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE



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

ALIGN

TA

TABLE

RECENT

HELP DOCUMENT

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



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

Results

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameter		M11: Select Output Format	f(T)	f(C)	f(G)
HKY+G	148		<input checked="" type="radio"/> XLSX: MS Excel workbook (2007+) <input type="radio"/> XLS: MS Excel Workbook (all versions) <input type="radio"/> ODS: Open/Libre Office Workbook <input type="radio"/> CSV: Comma-Separated-Values	0.236	0.190	0.221
HKY+I	148		<input checked="" type="radio"/> Save to Disk <input checked="" type="radio"/> Display Results	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		OUTPUT FILENAME australia_aligned_0423_model_selection.txt	0.236	0.190	0.221
T92+I	146		DIRECTORY /Users/clydedapat/Desktop/2023_Timor-Leste/	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	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



RECENT



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

Check for Updates

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

[Check for Updates](#)

MEGA-result0

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

D26

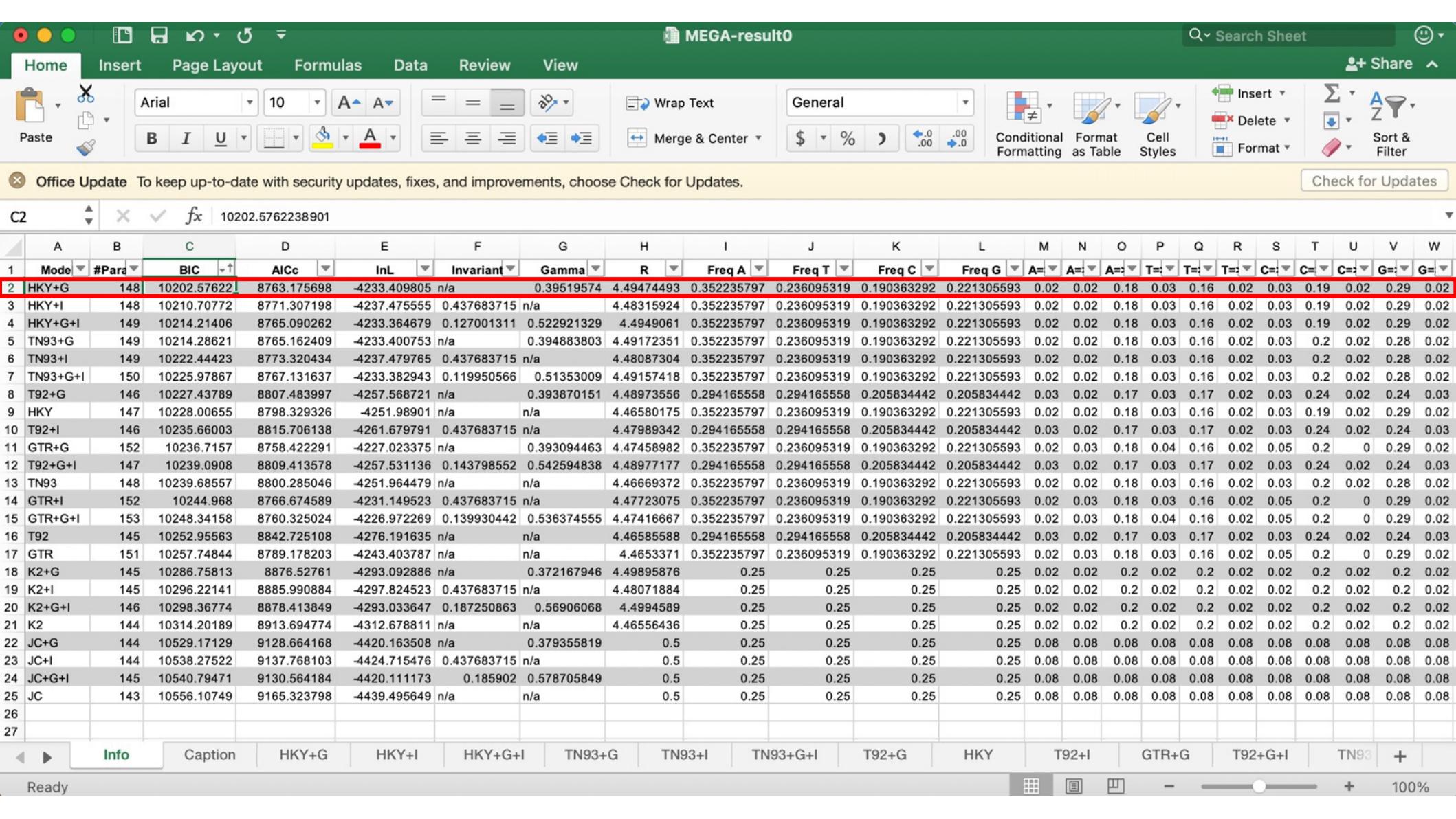
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Mode	#Paras	BIC	AICc	InL	Invariant	Gamma	R	Freq A	Freq T	Freq C	Freq G	A=	A=	T=	T=	T=	T=	C=	C=	G=	G=	
2	HKY+G	148	10202.57622	8763.175698	-4233.409805	n/a	0.39519574	4.49474493	0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.19	0.02	0.29	0.02
3	HKY+I	148	10210.70772	8771.307198	-4237.475555	0.437683715	n/a	4.48315924	0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.19	0.02	0.29	0.02
4	HKY+G+I	149	10214.21406	8765.090262	-4233.364679	0.127001311	0.522921329	4.4949061	0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.19	0.02	0.29	0.02
5	TN93+G	149	10214.28621	8765.162409					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.2	0.02	0.28	0.02
6	TN93+I	149	10222.44423	8773.320434					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.2	0.02	0.28	0.02
7	TN93+G+I	150	10225.97867	8767.131637					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.2	0.02	0.28	0.02
8	T92+G	146	10227.43789	8807.483997					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.2	0.02	0.28	0.02
9	HKY	147	10228.00655	8798.329326					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.19	0.02	0.29	0.02
10	T92+I	146	10235.66003	8815.706138					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.24	0.02	0.24	0.03
11	GTR+G	152	10236.7157	8758.422291					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
12	T92+G+I	147	10239.0908	8809.413578					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.24	0.02	0.24	0.03
13	TN93	148	10239.68557	8800.285046					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.2	0.02	0.28	0.02
14	GTR+I	152	10244.968	8766.674589					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.24	0.02	0.24	0.03
15	GTR+G+I	153	10248.34158	8760.325024					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
16	T92	145	10252.95563	8842.725108					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.24	0.02	0.24	0.03
17	GTR	151	10257.74844	8789.178203					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.02	0.18	0.03	0.16	0.02	0.03	0.2	0.02	0.28	0.02
18	K2+G	145	10286.75813	8876.52761					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
19	K2+I	145	10296.22141	8885.990884					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
20	K2+G+I	146	10298.36774	8878.413849					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.24	0.02	0.24	0.03
21	K2	144	10314.20189	8913.694774					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
22	JC+G	144	10529.17129	9128.664168					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
23	JC+I	144	10538.27522	9137.768103					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
24	JC+G+I	145	10540.79471	9130.564184					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
25	JC	143	10556.10749	9165.323798					0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.24	0.02	0.24	0.03
26									0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02
27									0.352235797	0.236095319	0.190363292	0.221305593	0.02	0.03	0.18	0.04	0.16	0.02	0.05	0.2	0	0.29	0.02

Info Caption HKY+G

Clear Filter

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

Ready



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 ($\ln L$), 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 1st+2nd+3rd+Noncoding. There were a total of 1701 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2][3]

2

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.

8 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). 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.

9

10

11

12

13

14

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

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

Check for Updates

	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																	

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%

Construct a maximum likelihood tree



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



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



Construct/Test Maximum Likelihood Tree...



Construct/Test Neighbor-Joining Tree...



Construct/Test Minimum-Evolution Tree...



RATES

CLOCKS



DISEASE



Construct/Test Maximum Likelihood Tree...

Construct/Test Neighbor-Joining Tree...

Construct/Test Minimum-Evolution Tree...

Construct/Test UPGMA Tree...

Construct/Test Maximum Parsimony Tree(s)

Open Tree Session



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

ANALYZE
PROTOTYPE



ALIGN



DATA



MODELS



DISTANCE



DIVERSITY



PHYLOGENY



USER TREE



ANCESTORS



SELECTION



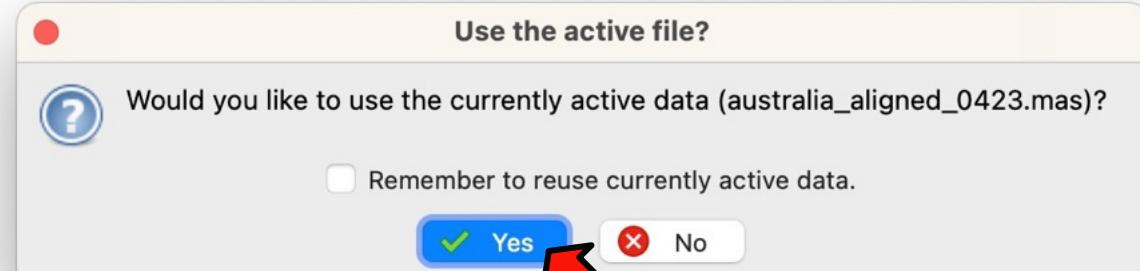
RATES



CLOCKS



DISEASE



TIMETREE



DATAMONKEY

RECENT PUBLICATIONS



HELP DOCS



EXAMPLES



CITATION



REPORT BUG



UPDATES



MEGA LINKS



TOOLBAR



PREFERENCES

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

 OK



File Edit View Insert Tools Window Help

Taxon Names

Layout

Subtree

Branch Lengths

Statistics/Frequency/Info

Distance Scale

Divergence Times

Time Scale

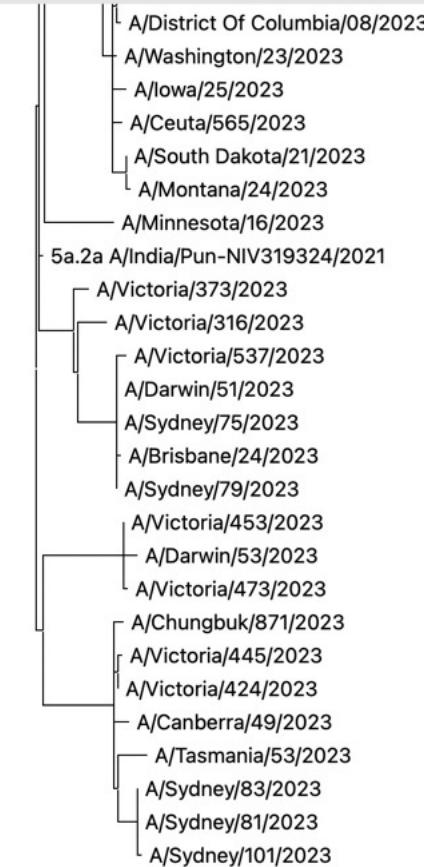
Ancestors Site # 1

Collapse/Expand Lineages

Compute

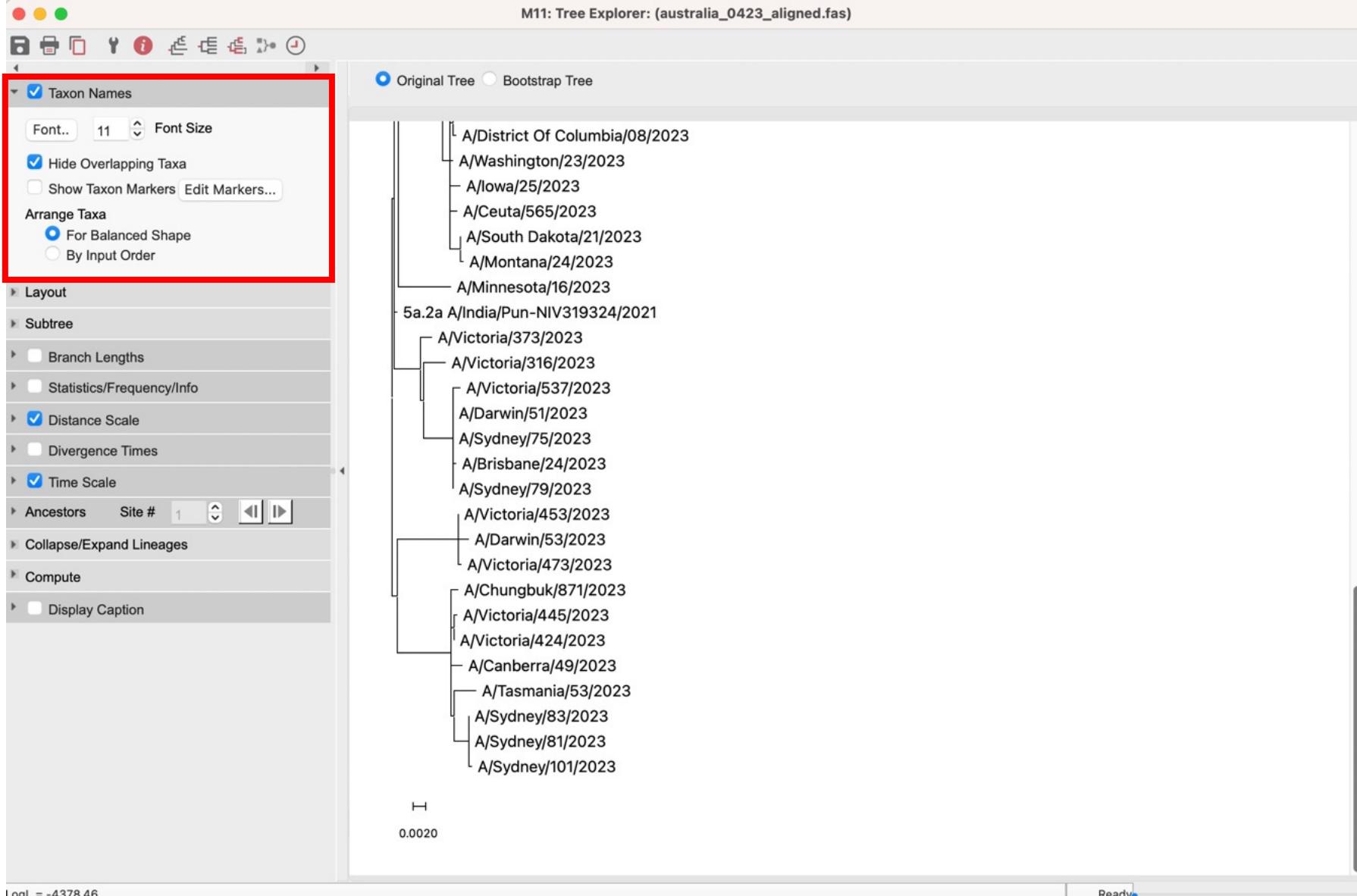
Display Caption

Original Tree Bootstrap Tree



H

0.0020





File Edit View Insert Tools Window Help

Taxon Names

Original Tree Bootstrap Tree

Layout

- Tree Style
- Toggle Scaling of the Tree
- Auto-size Tree 
- Drag to Resize Tree
- Root on Midpoint
- Root on Outgroup

Tree Width

Tree Height

Subtree

Branch Lengths

Statistics/Frequency/Info

Distance Scale

Divergence Times

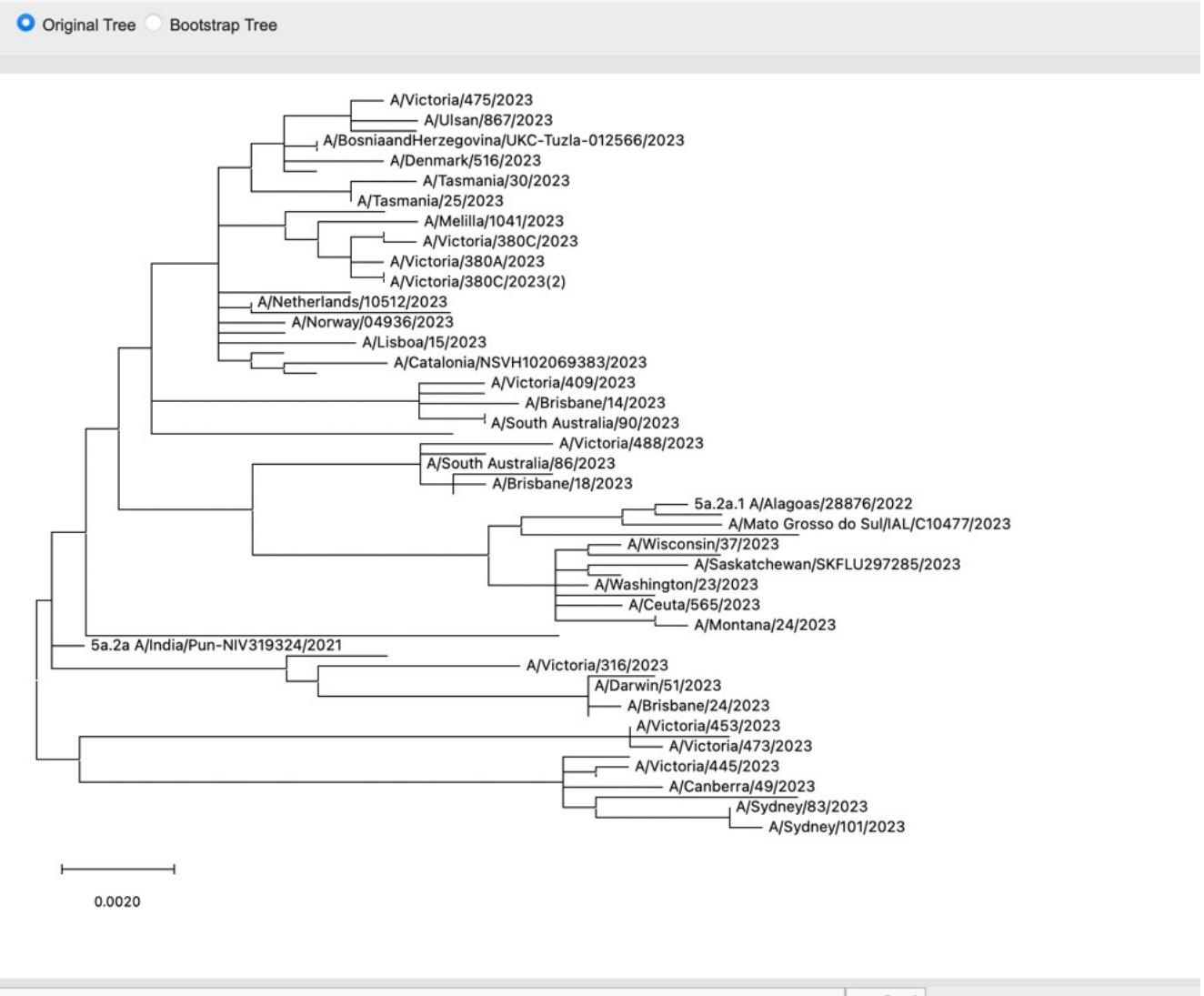
Time Scale

Ancestors Site # 1  

Collapse/Expand Lineages

Compute

Display Caption





File Edit View Insert Tools Window Help

Taxon Names

Layout

Subtree

Branch Lengths

Statistics/Frequency/Info

Distance Scale

Line Width 1 pt

Caption

Font...

Font Size 8

Scale Length 0.00

Tick Interval 0

Divergence Times

Time Scale

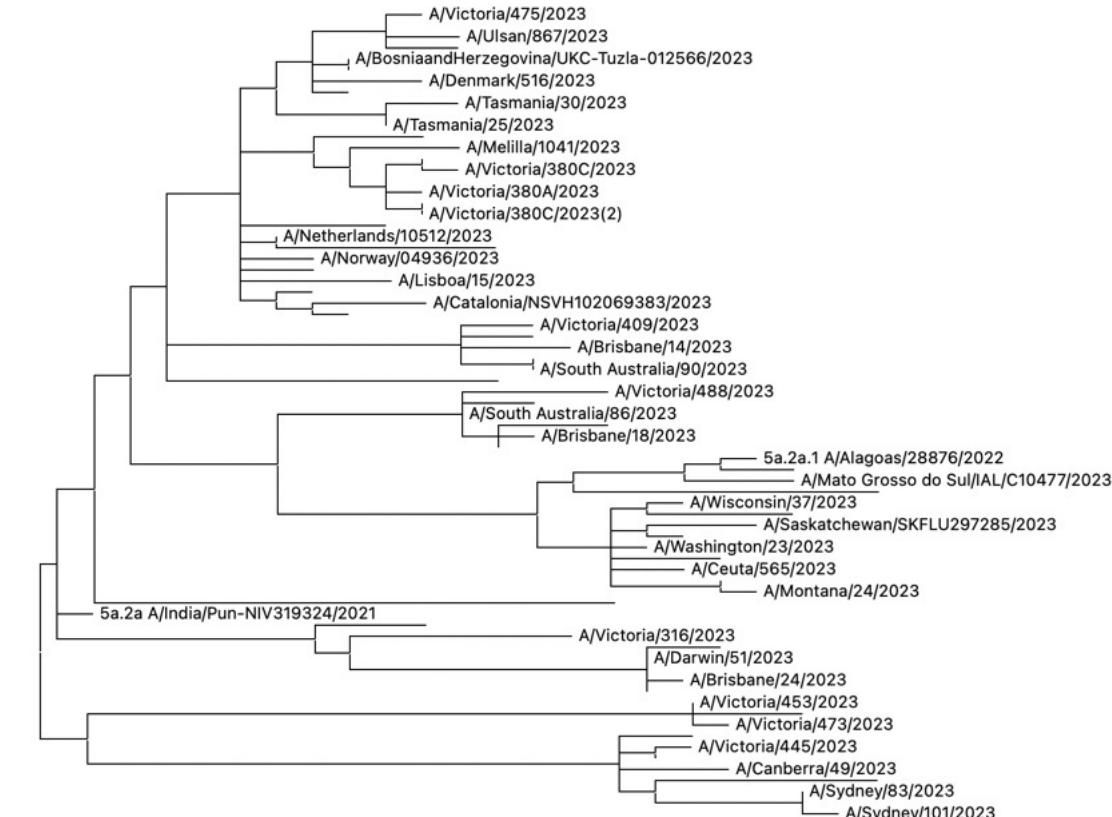
Ancestors Site # 1

Collapse/Expand Lineages

Compute

Display Caption

Original Tree Bootstrap Tree



0.0020

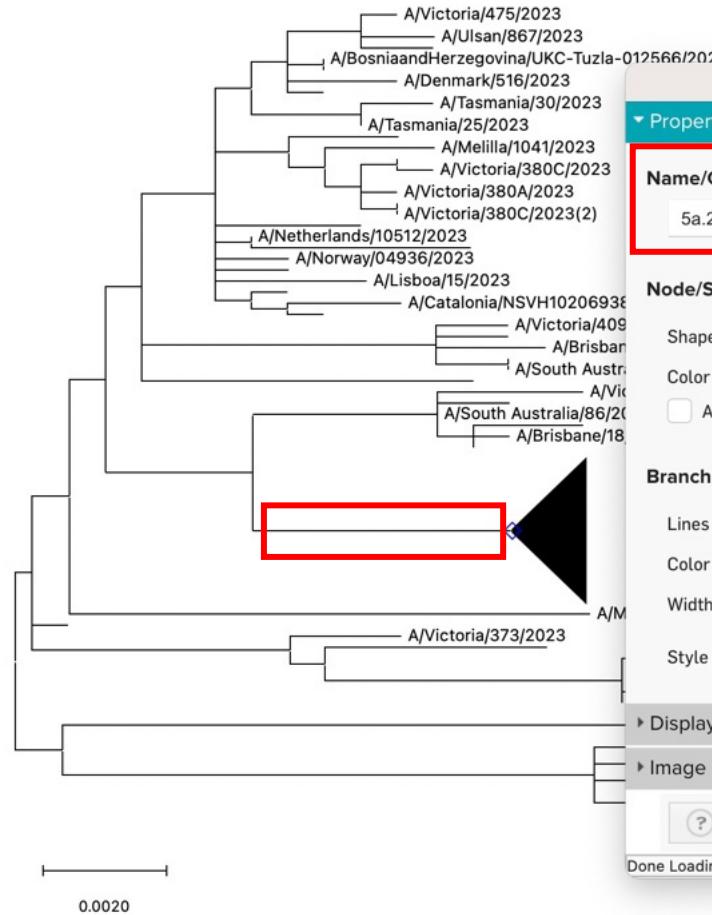


File Edit View Insert Taxon Names Layout Subtree Branch Lengths Statistics/Frequency/Info Distance Scale Divergence Times Time Scale Ancestors Site # 1 ▲ ▶ ▶ Collapse/Expand Lineages

- Collapse Nodes
- By Groups
- By Cluster Size 10
- By Sequence Difference 0.000002
- Display Cluster Size

Compute Display Caption

Original Tree Bootstrap Tree



Subtree Drawing Options

Property

Name/Caption

5a.2a.1

Font

Node/Subtree Marker

Shape

None

Color

█

Apply to Taxon Markers

Branch Line

Lines

█

Color

█

Width

1 pt

Style

Solid

Display

Image

? Help

X Cancel

OK

Done Loading



File Edit View Insert Tools Window Help

Taxon Names

Layout

Subtree

Branch Lengths

Statistics/Frequency/Info

Distance Scale

Divergence Times

Time Scale

Ancestors Site # 1 ▲ ▶ ▶▶

Collapse/Expand Lineages

Collapse Nodes

By Groups

By Cluster Size
10 ▲ ▼

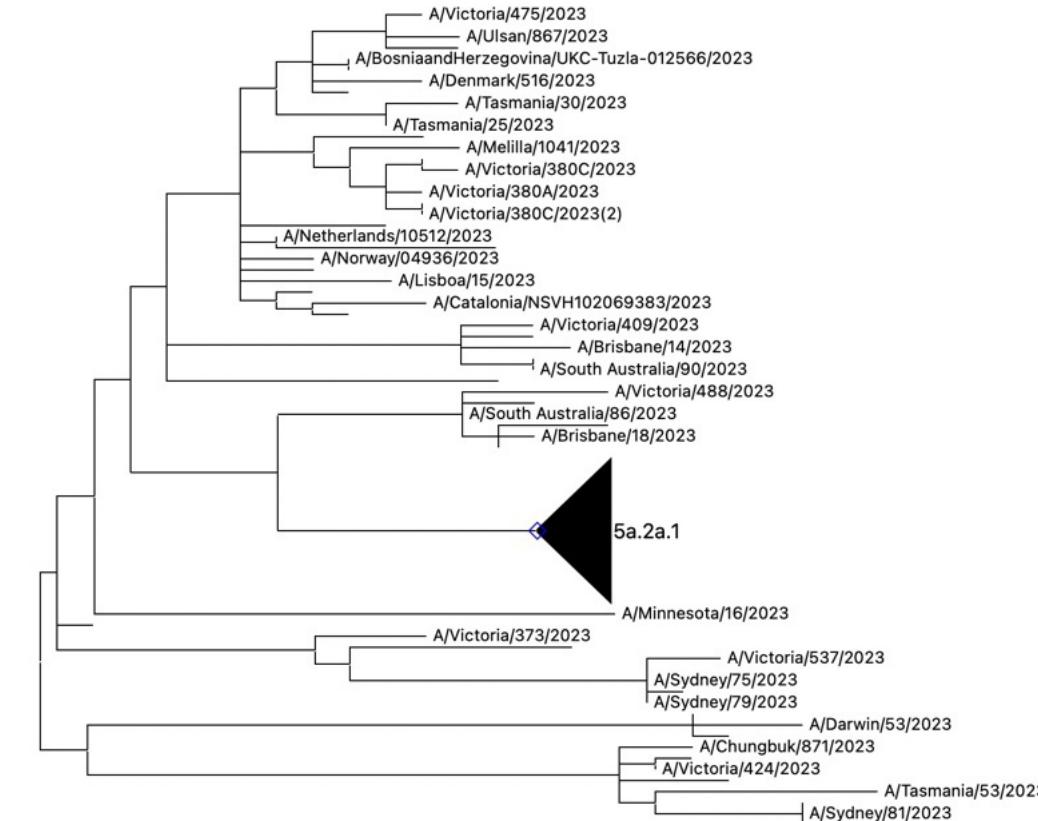
By Sequence Difference
0.000002 ▲ ▼

Display Cluster Size

Compute

Display Caption

Original Tree Bootstrap Tree



0.0020



↶ ↷ ⌂ ⌃ ⌄ ⌅ ⌆ ⌇ ⌈ ⌉ ⌊ ⌋

Taxon Names

Layout

Subtree

Branch Lengths

Statistics/Frequency/Info

Distance Scale

Divergence Times

Time Scale

Ancestors Site # 1 ⌂ ⌃ ⌄ ⌅ ⌆ ⌇

Collapse/Expand Lineages

Compute

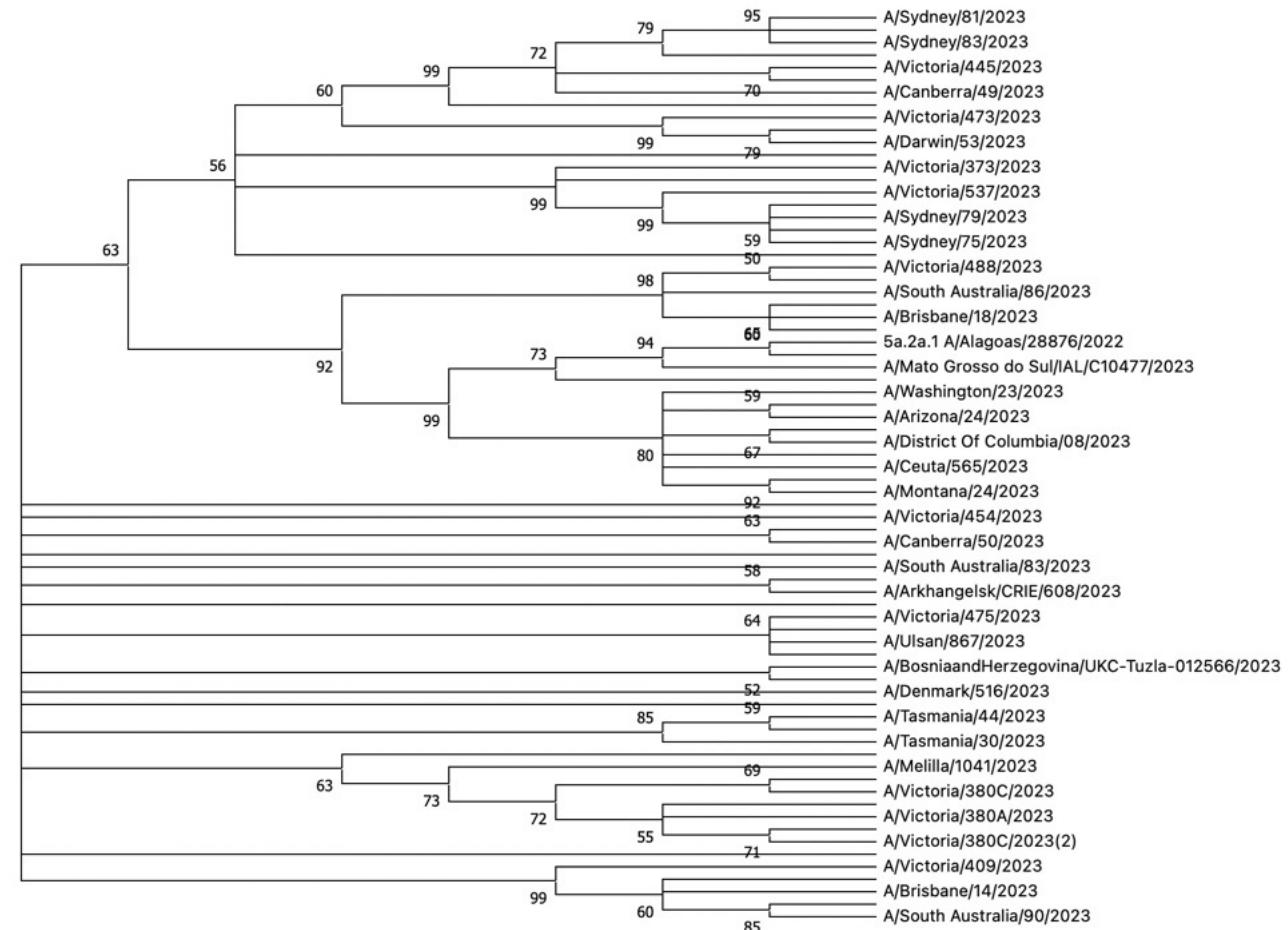
Display Caption

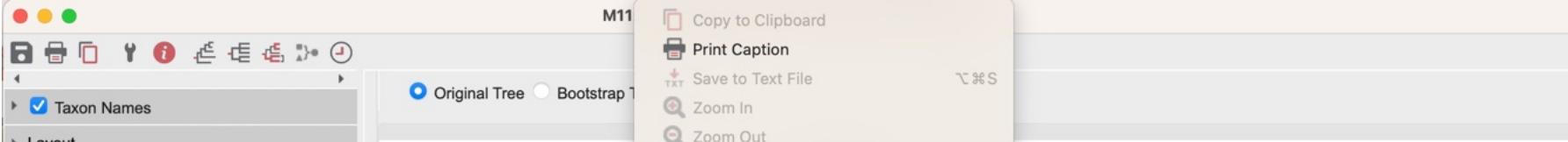


Original Tree

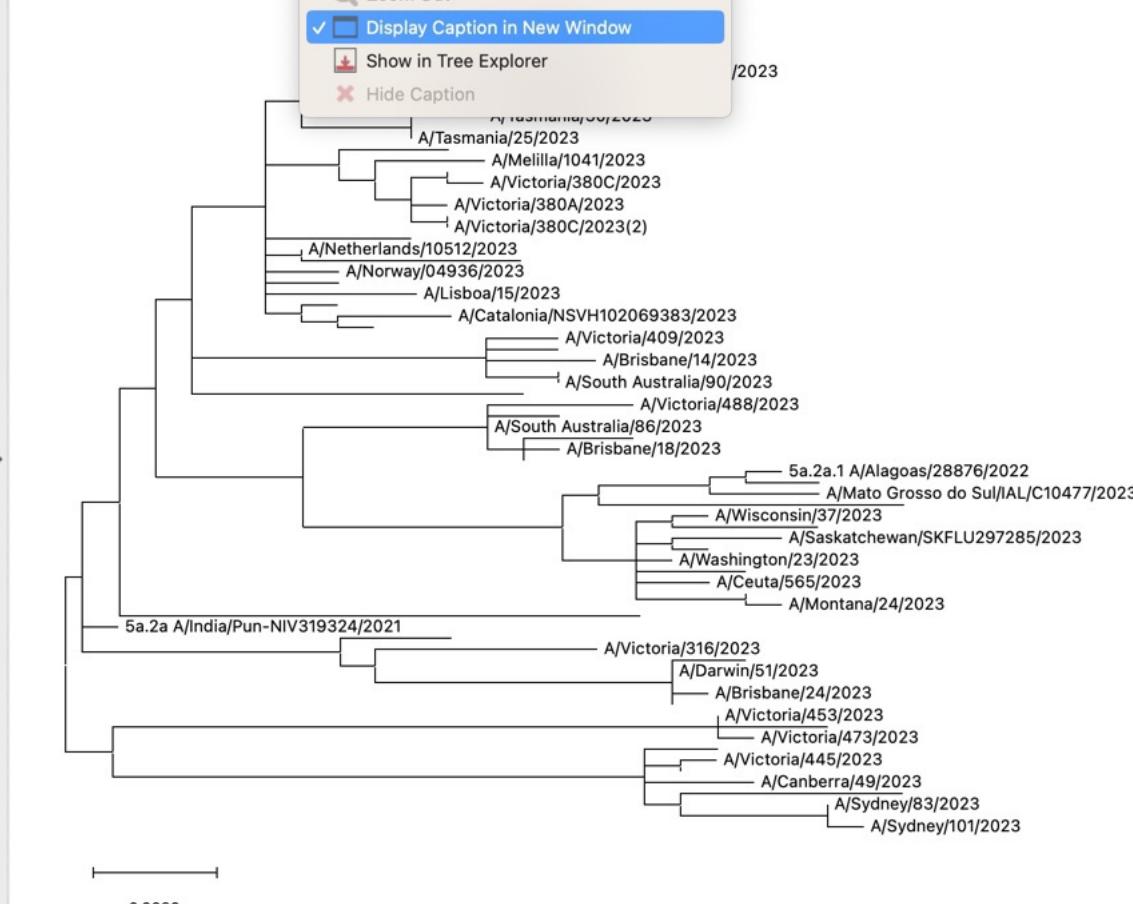


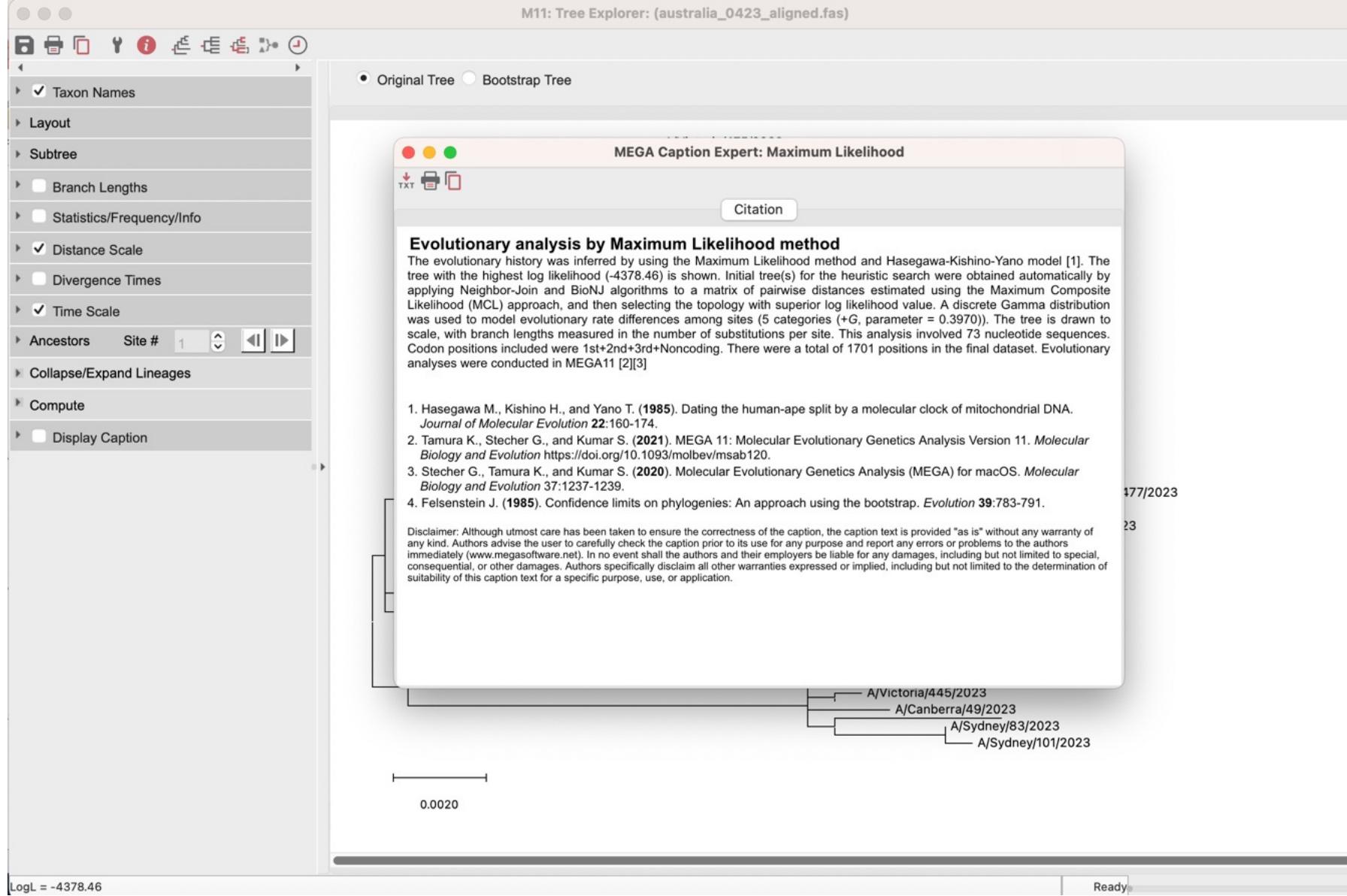
Bootstrap Tree





- > Taxon Names
- > Layout
- > Subtree
- > Branch Lengths
- > Statistics/Frequency/Info
- > Distance Scale
- > Divergence Times
- > Time Scale
- Ancestors Site # 1
- > Collapse/Expand Lineages
- > Compute
- > Display Caption





Export tree file

File formats

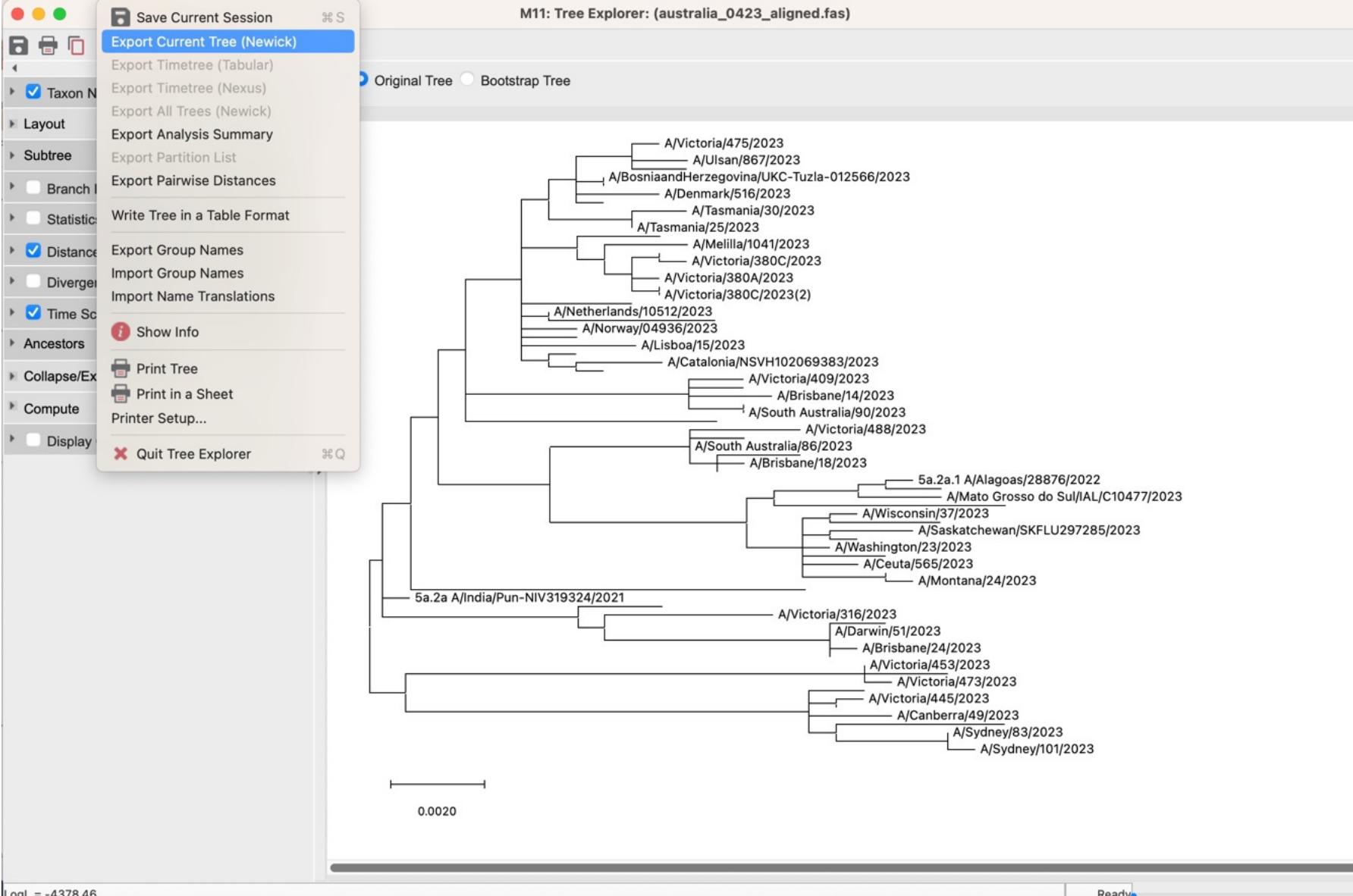
```
(,,(,));
(A,B,(C,D));
(A,B,(C,D)E)F;
(:0.1,:0.2,(:0.3,:0.4):0.5);
(:0.1,:0.2,(:0.3,:0.4):0.5):0.0;
(A:0.1,B:0.2,(C:0.3,D:0.4):0.5);
(A:0.1,B:0.2,(C:0.3,D:0.4)E:0.5)F;
((B:0.2,(C:0.3,D:0.4)E:0.5)F:0.1)A;
```

Newick (Newick's restaurant)

```
#NEXUS
Begin TAXA;
    Dimensions ntax=4;
    TaxLabels SpaceDog SpaceCat Space0rc 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.
        Space0rc   ...t.....-g. [ same as atgttagctag-tgg ]
        SpaceElf    ...t.....-a.
    ;
End;

BEGIN TREES;
    Tree tree1 = (((SpaceDog,SpaceCat),Space0rc,SpaceElf));
END;
```



File Edit View Insert Tools Window Help

Original Tree Bootstrap Tree

Taxon Names

Layout

Subtree

Branch Lengths

Statistics/Frequency/Info

Distance Scale

Divergence Times

Time Scale

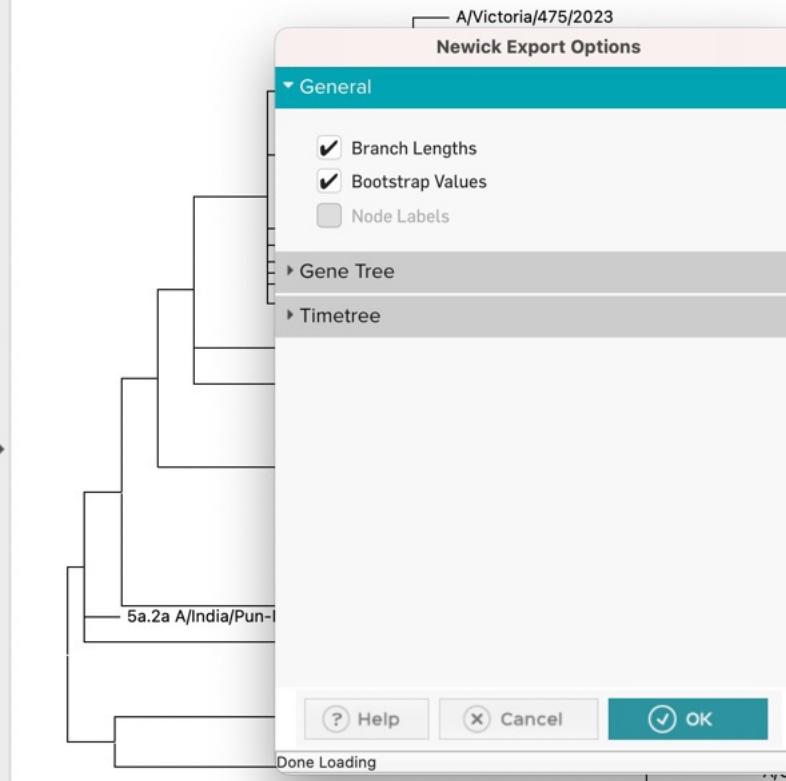
Ancestors Site # 1 ▲ ▶ ▶

Collapse/Expand Lineages

Compute

Display Caption

0.0020

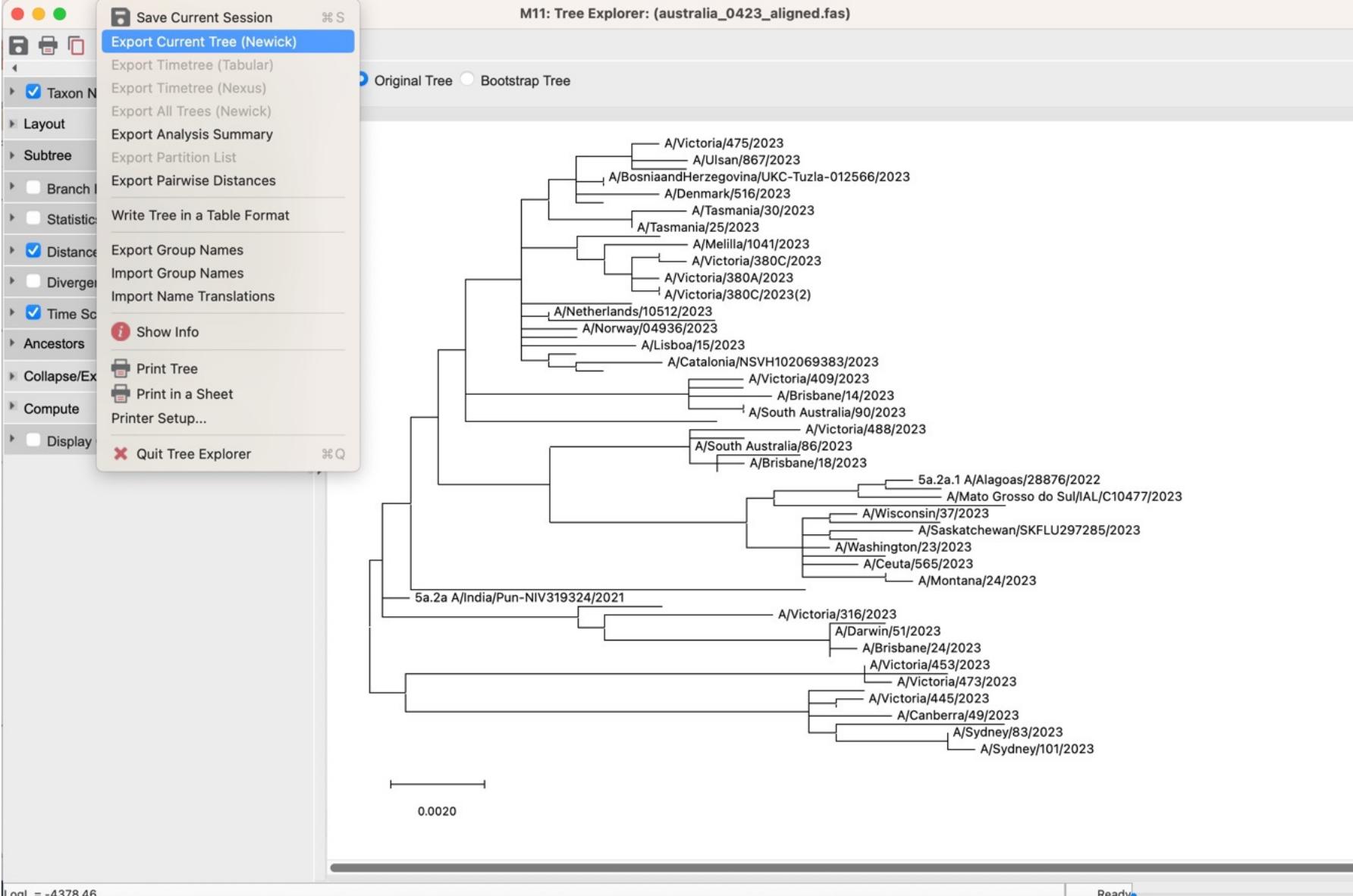


```

a.2a.1 A/Alagoas/28876/2022
|— A/Mato Grosso do Sul/IAL/C10477/2023
n/37/2023
/Saskatchewan/SKFLU297285/2023
3/2023
5/2023
/Montana/24/2023

23
/24/2023
/453/2023
toria/473/2023
/445/2023
.../Canberra/49/2023
A/Sydney/83/2023
|— A/Sydney/101/2023

```



The screenshot shows the M11 Tree Explorer application window. On the left, a sidebar contains various settings: Taxon Names (checked), Layout, Subtree, Branch Lengths (unchecked), Statistics/Frequency/Info (unchecked), Distance Scale (checked), Divergence Times (unchecked), Time Scale (checked), Ancestors (Site # 1), Collapse/Expand Lineages, Compute, and Display Caption (unchecked). Below these are buttons for Help, Cancel, and OK.

The main area displays a phylogenetic tree with several lineages. A specific node is highlighted with a gray background. The tree is rooted at the top. One branch leads to a lineage labeled "5a.2a A/India/Pun-". Another branch leads to a lineage labeled "A/Victoria/475/2023".

A context menu is open over the "A/Victoria/475/2023" node, titled "Newick Export Options". The "General" tab is selected, showing checked options for "Branch Lengths" and "Bootstrap Values", and an unchecked option for "Node Labels". Other tabs include "Gene Tree" and "Timetree".

On the right side of the tree, there is a detailed view of the "A/Victoria/475/2023" lineage, listing its history:

- a.2a.1 A/Alagoas/28876/2022
- A/Mato Grosso do Sul/IAL/C10477/2023
- n/37/2023
- /Saskatchewan/SKFLU297285/2023
- 3/2023
- 5/2023
- /Montana/24/2023

Below this, the lineage continues with:

- 23
- /24/2023
- /453/2023
- toria/473/2023
- /445/2023
- .,/Canberra/49/2023
- A/Sydney/83/2023
- A/Sydney/101/2023

At the bottom of the tree, a scale bar indicates a distance of 0.0020.



Original Tree Bootstrap Tree

Taxon Names

Layout

Subtree

Branch Lengths

Statistics/Frequency/Info

Distance Scale

Divergence Times

Time Scale

Ancestors Site # 1

Collapse/Expand Lineages

Compute

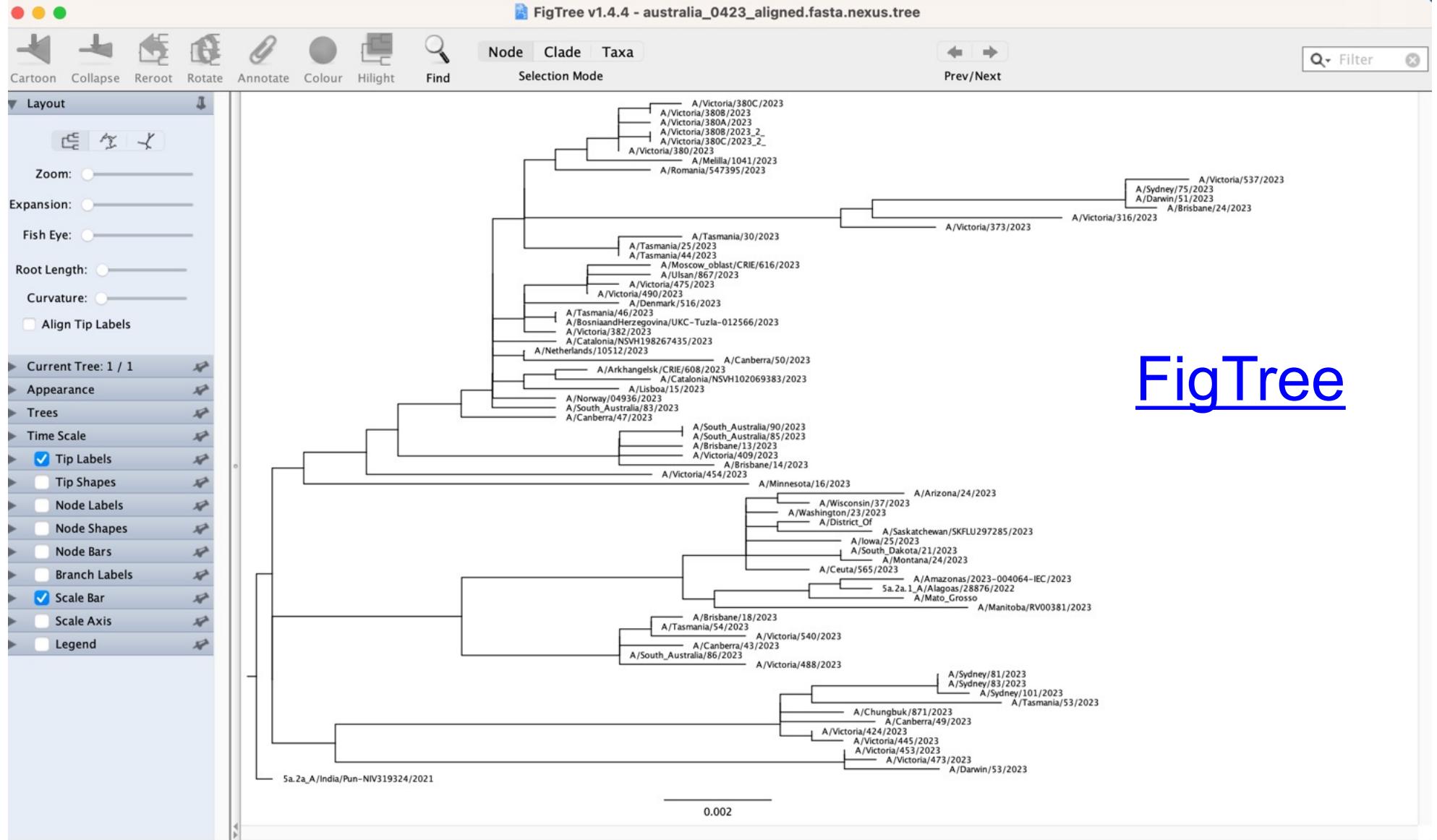
Display Caption

M11: Text File Editor and Format Converter

Newick Export.nwk

```
1 | ((((((((((((A/Victoria/475/2023:0.00058942,A/Victoria/490/2023:0.00000206)0.4400:0.00000411,A/Ulsan/867/2023:
```

Visualizing Trees



FigTree

UShER: Ultrafast Sample placement on Existing tRee

UShER: Ultrafast Sample placement on Existing tRee

[view in Genome Browser](#)

[view downsampled global tree in Nextstrain](#)

View subtree [subtree 1](#) in

Nextstrain
MicrobeTrace

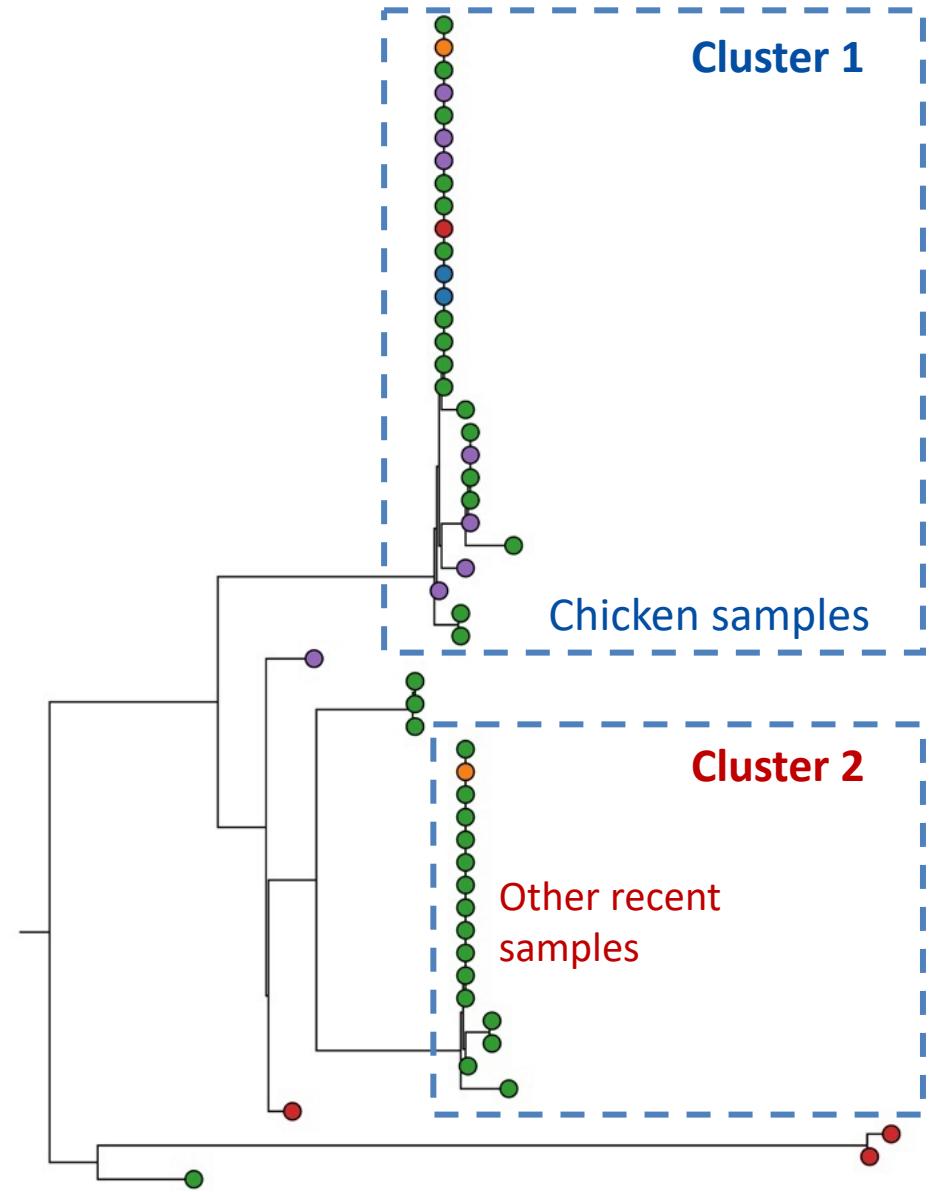
If you have metadata you wish to display, click a 'view subtree in Nextstrain' button, and then you can drag on a CSV file to [add it to the tree view](#).

Note: The Nextstrain subtree views, and Download files below, are temporary files and will expire within two days. Please download the Nextstrain subtree J drag-dropped onto <https://auspice.us>.

Downloads: | Global phylogenetic tree with your sequences | [TSV summary of sequences and placements](#) | [TSV summary of Spike mutations](#) | ZIP file of s

Fasta Sequence	Size (?)	#Ns (?)	#Mixed (?)	Bases aligned (?)	Inserted bases (?)	Deleted bases (?)	#SNVs used for placement (?)	#Masked SNVs (?)	Nextstrain clade (?)	Pango lineage (?)
Philippines@JN.1.1@EPI_ISL_18986949@2024.027	29903	4	0	29714 (?)	0	0	119 (?)	1 (?)	23I (BA.2.86)	JN.1

Subtree 1:

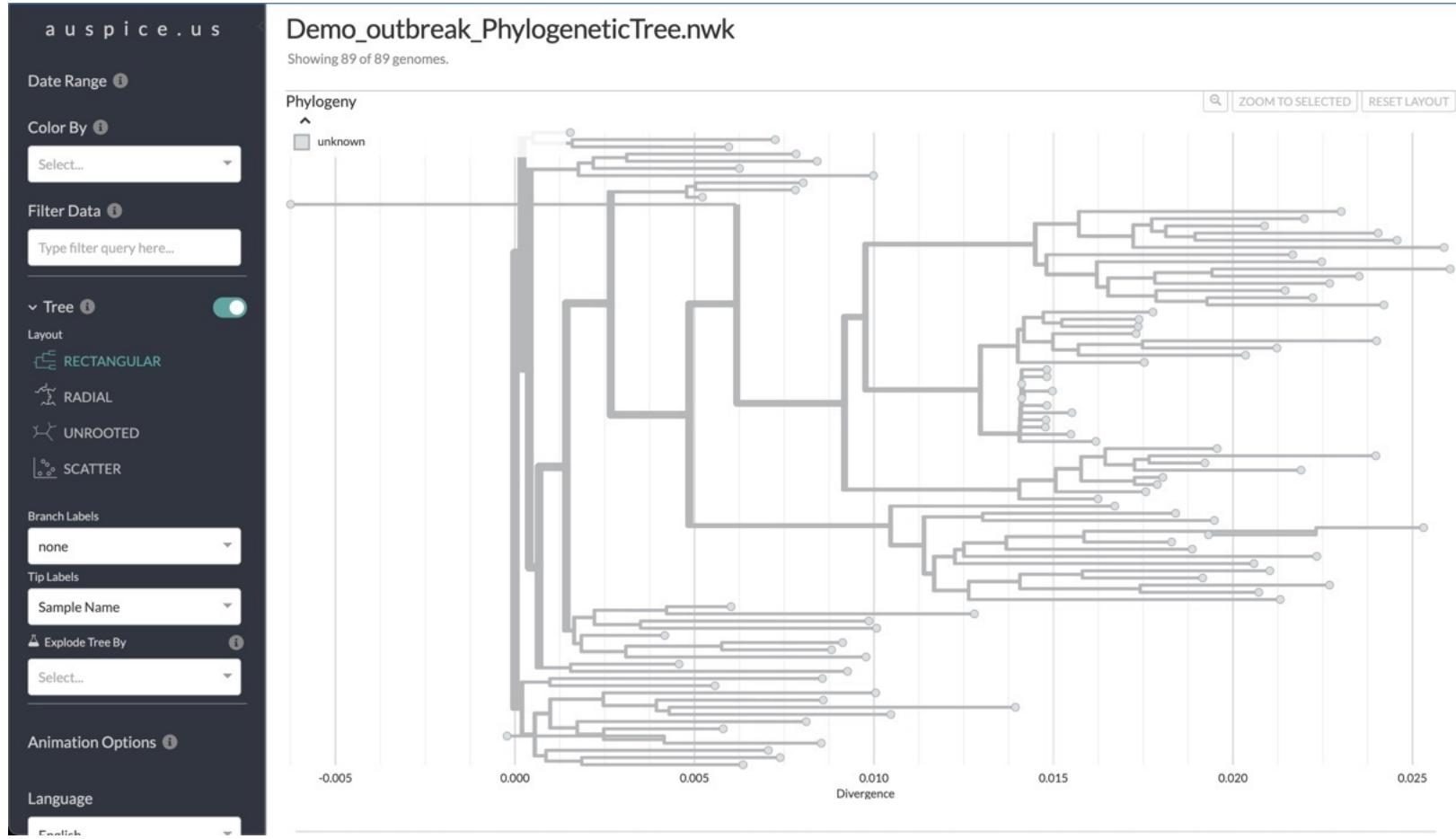


Phylogenetic clustering

<https://microbetrace.cdc.gov/MicrobeTrace>



<https://auspice.us>



<https://taxonum.org>

