

Phylogenetics and Sequence Analysis

Tree Building 1

Kurt Wollenberg, PhD

Phylogenetics Specialist

Bioinformatics and Computational Biosciences Branch

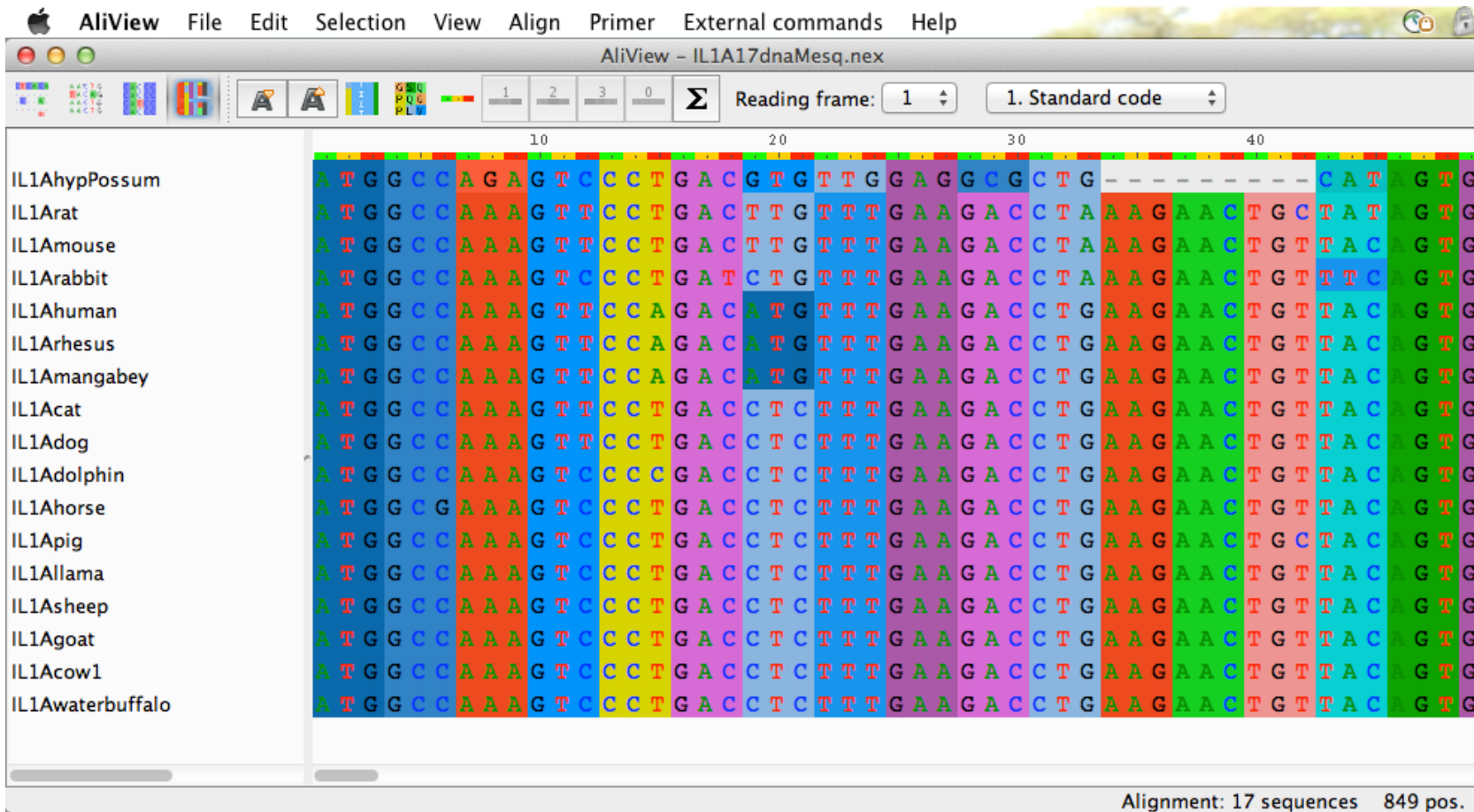
Office of Cyber Infrastructure and Computational Biology

March 2024



National Institute of
Allergy and
Infectious Diseases

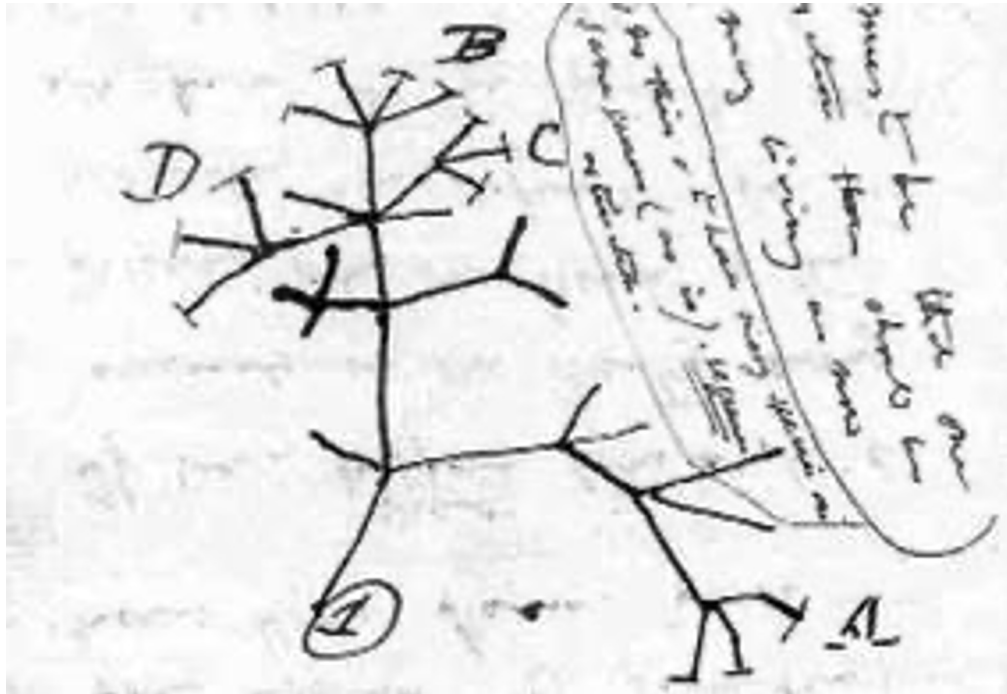
Multiple Sequence Alignment



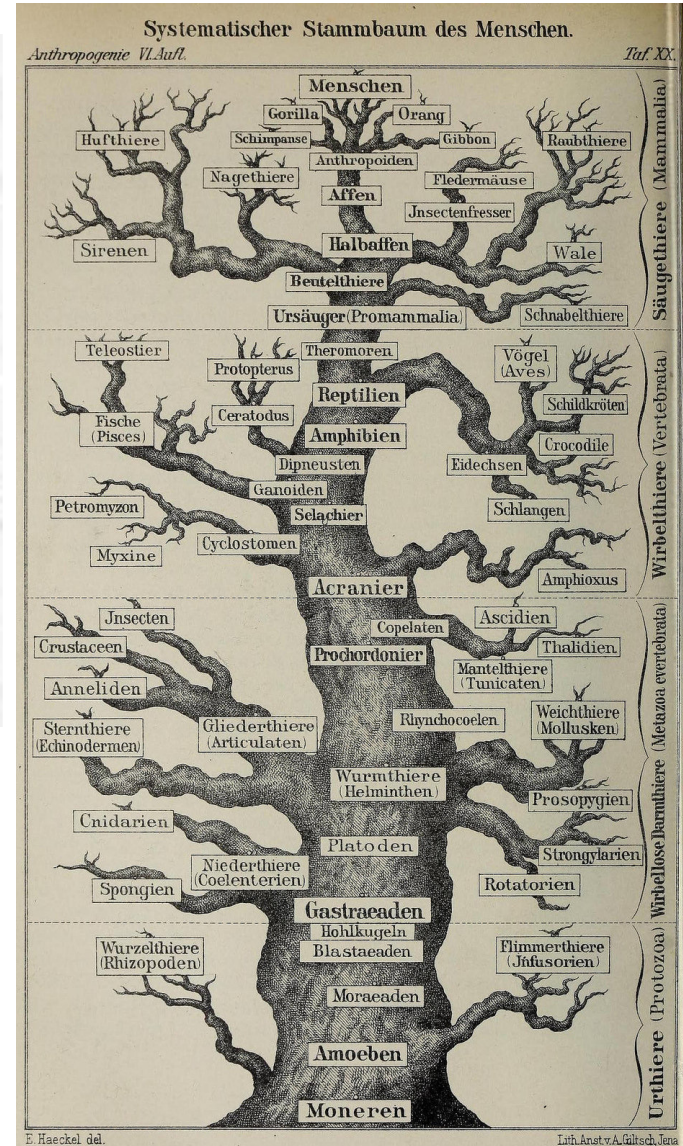
What's next?

Building trees with our MSA

What is a phylogenetic tree?



C. Darwin, 1837/8

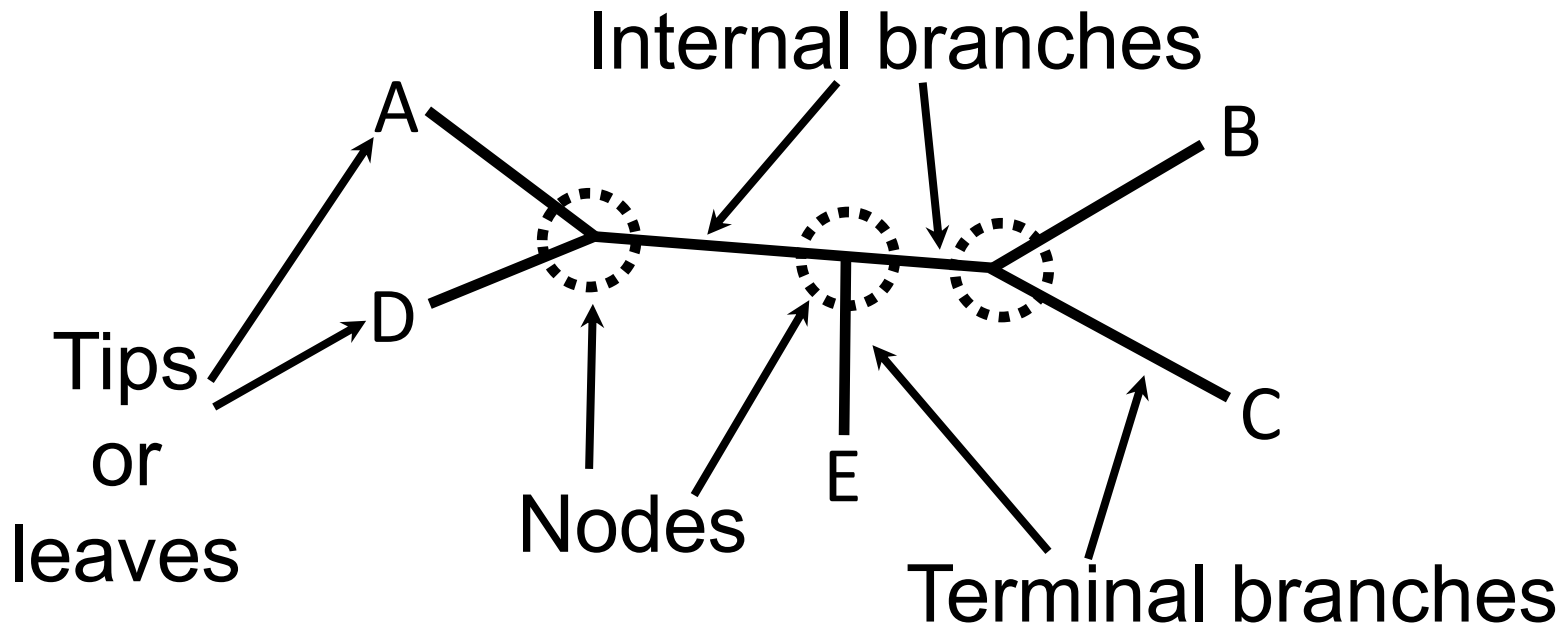


E. Haeckel, 1868

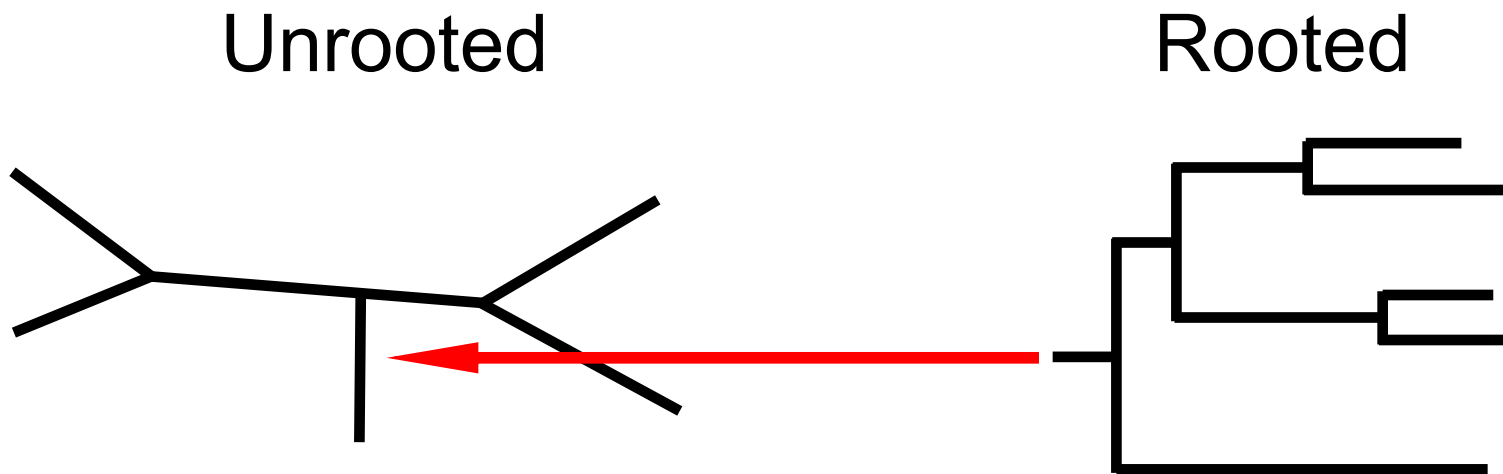
What is a phylogenetic tree?

- Reconstruction of biological history
- Based on similarities and differences among homologous attributes (characters) of the entities under scrutiny
- Molecular characters (sequences, usually) are most often found only in extant organisms

What is a phylogenetic tree?



What is a phylogenetic tree?



Two approaches to tree building

- Application of an algorithm to build the best tree from the data
- Evaluation of multiple possible best trees using an optimality criterion

The algorithm approach: Distance Methods

- Distance calculated based on a specific substitution model (J-C, Kimura, BLOSUM64, etc.)
- Distances from each sequence to all others are calculated and stored in a matrix
- Tree then calculated from the distance matrix using a specific tree-building algorithm

The algorithm approach: Distance Methods

Tree-Building Algorithms

- UPGMA
- Neighbor-Joining

The algorithm approach: Neighbor-joining Calculation

	A	B	C	D	E	R	R/(N-2)
A	-	0.1715	0.2147	0.3091	0.2326	0.9279	0.3093
B	-0.4766	-	0.2991	0.3399	0.2058	1.0163	0.3388
C	-0.4905	-0.4356	-	0.2795	0.3943	1.1876	0.3959
D	-0.4527	-0.4514	-0.5689	-	0.4289	1.3574	0.4525
E	-0.4972	-0.5535	-0.4221	-0.4441	-	1.2616	0.4205

$$\text{C to Node 1 distance} = 0.2795/2 + (0.3959 - 0.4525)/2 = 0.1114$$

$$\text{D to Node 1 distance} = 0.2795 - 0.1114 = 0.1681$$

$$\text{A to Node 1 distance} = (0.2147 + 0.3091 - 0.2795)/2 = 0.1222$$

$$\text{B to Node 1 distance} = (0.2991 + 0.3399 - 0.2795)/2 = 0.1798$$

$$\text{E to Node 1 distance} = (0.3943 + 0.4298 - 0.2795)/2 = 0.2719$$

The algorithm approach: Neighbor-joining Calculation

	A	B	E	Node 1	R	R/(N-2)
A	-	0.1715	0.2326	0.1222	0.5263	0.2631
B	-0.3701	-	0.2058	0.1798	0.5571	0.2785
E	-0.3856	-0.4278	-	0.2719	0.7103	0.3551
Node 1	-0.4278	-0.3856	-0.3701	-	0.5739	0.2869

$$\text{A to Node 2 distance} = 0.1222/2 + (0.2631 - 0.2869)/2 = 0.0492$$

$$\text{Node 1 to Node 2 distance} = 0.1222 - 0.0492 = 0.0730$$

$$\text{B to Node 2 distance} = (0.1715 + 0.1798 - 0.1222)/2 = 0.1146$$

$$\text{E to Node 2 distance} = (0.2326 + 0.2719 - 0.1222)/2 = 0.1912$$

The algorithm approach: Neighbor-joining Calculation

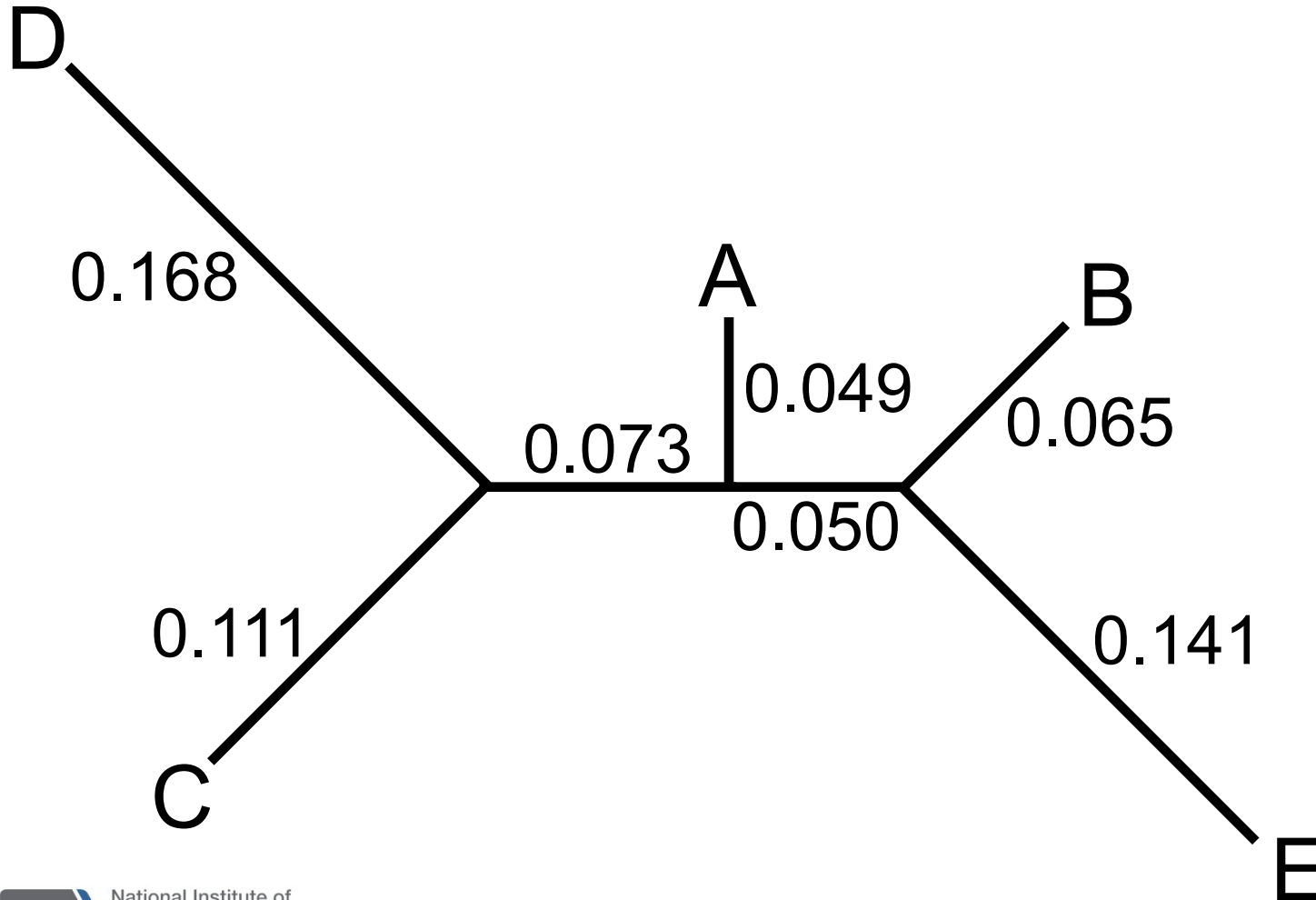
	B	E	Node 2	R	R/(N-2)
B	-	0.2058	0.1146	0.3204	0.3204
E	-0.5116	-	0.1912	0.3970	0.3970
Node 2	-0.5116	-0.5116	-	0.3058	0.3058

B to Node 3 distance = $0.1146/2 + (0.3204 - 0.3058)/2 = 0.0646$

Node 2 to Node 3 distance = $0.1146 - 0.0646 = 0.0500$

E to Node 3 distance = $(0.2058 + 0.1912 - 0.1146)/2 = 0.1412$

The algorithm approach: Neighbor-joining Calculation



The optimality criterion approach

- Build a tree or trees
- Evaluate the tree(s) using a specific numerical optimality criterion
- Most common optimality criteria
 - Maximum parsimony
 - Maximum likelihood
- Explore tree space to find the optimal tree

Optimality Criterion: Parsimony

Occam's Razor: The simplest explanation is the preferred explanation.

The tree requiring the minimal number of changes is the optimal tree.

A step is any change in the data from one state to another

The optimality criterion approach

- Build the initial tree
 - Construct a neighbor-joining tree
 - Stepwise addition
- Calculate the tree score
 - Count steps (parsimony)
 - Calculate likelihood of the data given the tree
- Explore tree space
 - Branch swapping
 - Tree bisection and reconnection (TBR)
- Is this the best tree? (Stopping criteria)

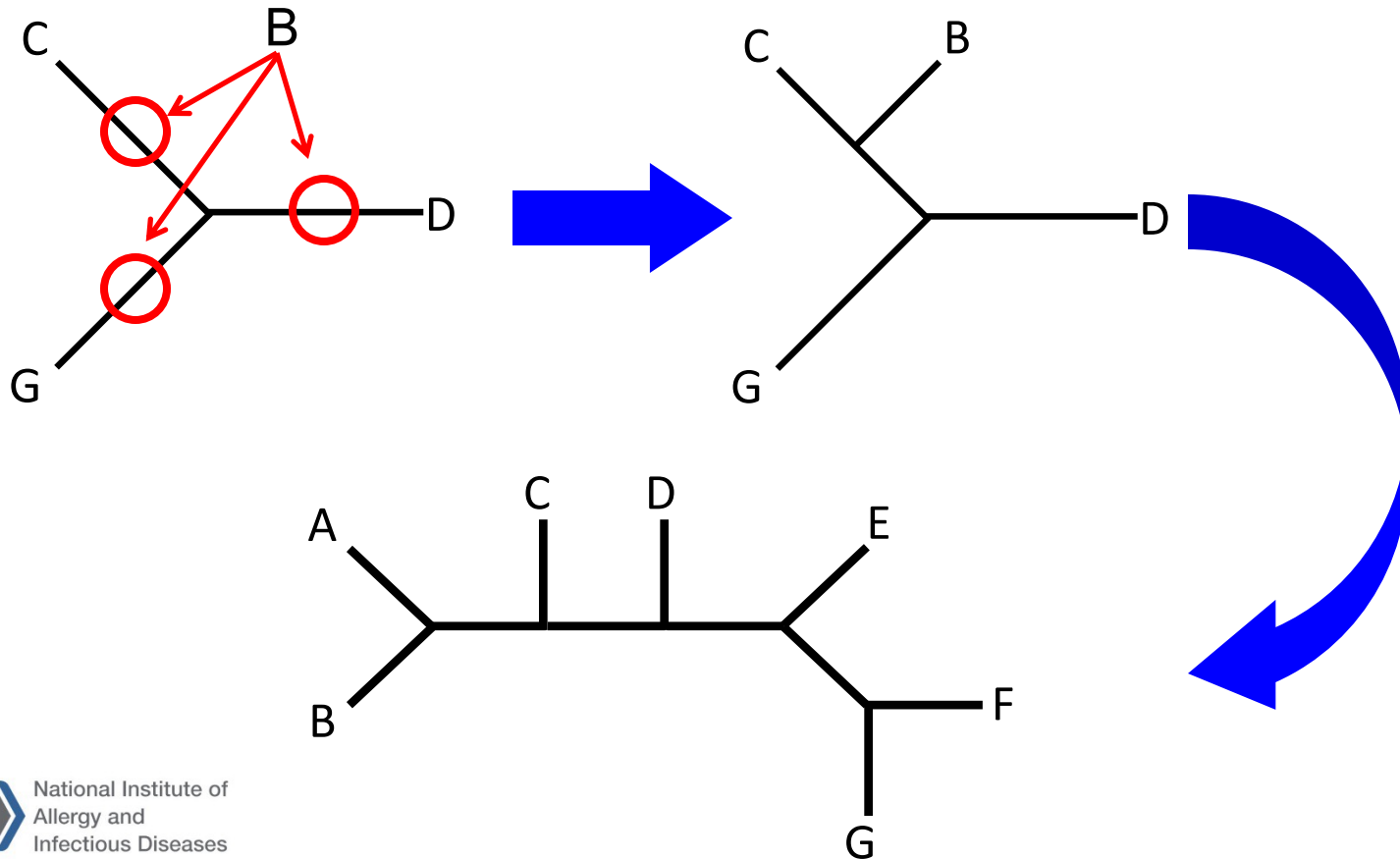
The optimality criterion approach

Building the initial tree

- Stepwise addition
 - Choose three taxa and join
 - Random, or closest
- Select a new taxon to add
- Calculate the optimal 4-taxa tree
- Repeat until all taxa are joined

The optimality criterion approach

Building the initial tree



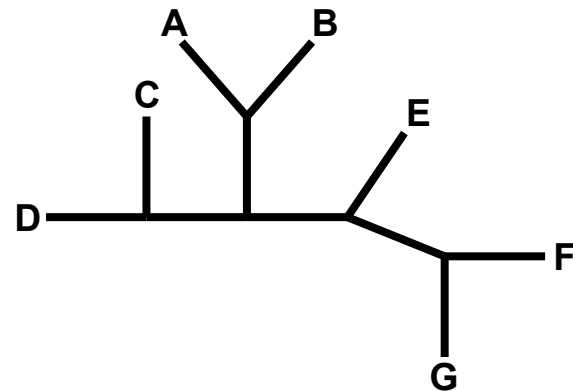
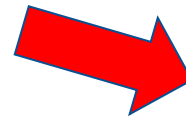
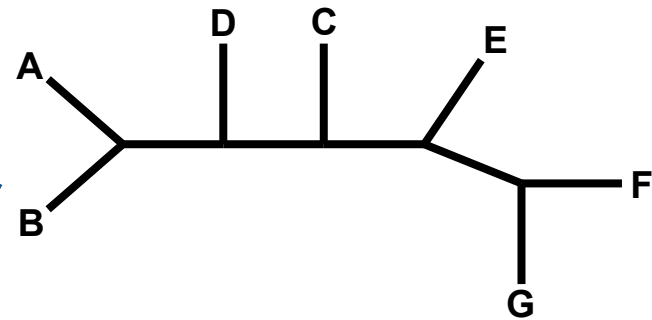
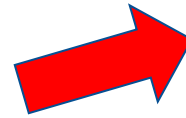
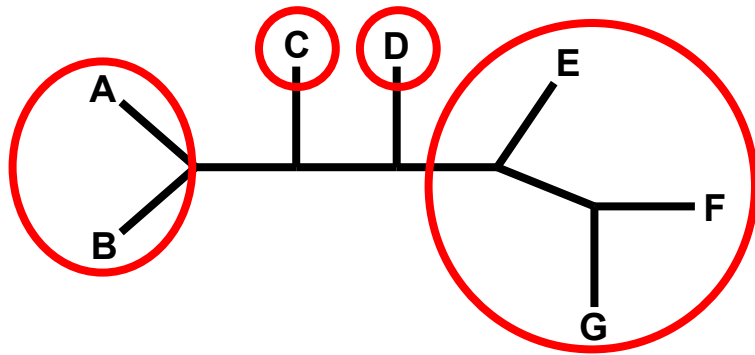
The optimality criterion approach

Exploring tree space: Branch swapping

- Nearest neighbor interchange
- Subtree pruning and regrafting
- Tree bisection and reconnection

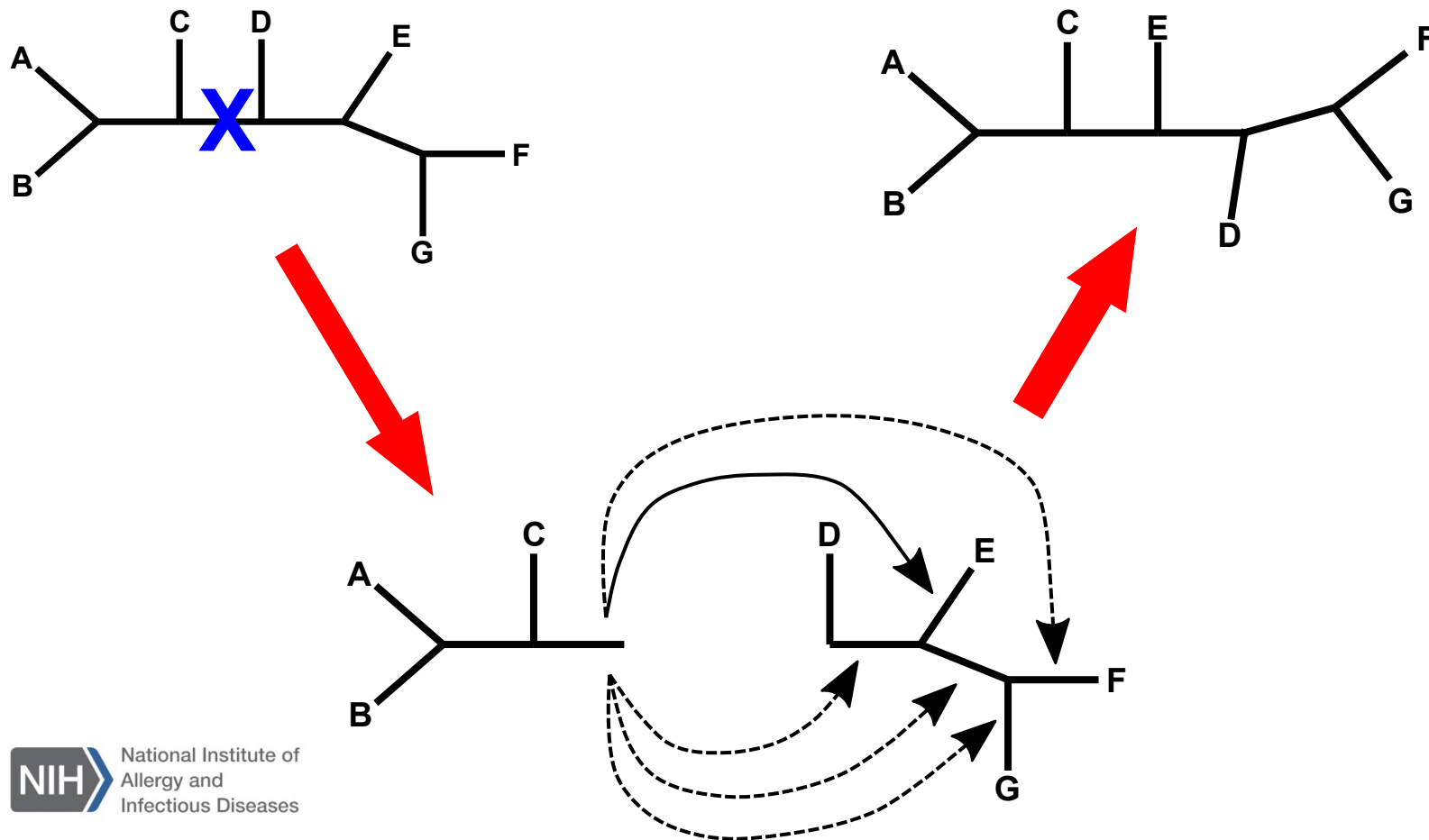
The optimality criterion approach

Branch swapping: Nearest Neighbor Interchange



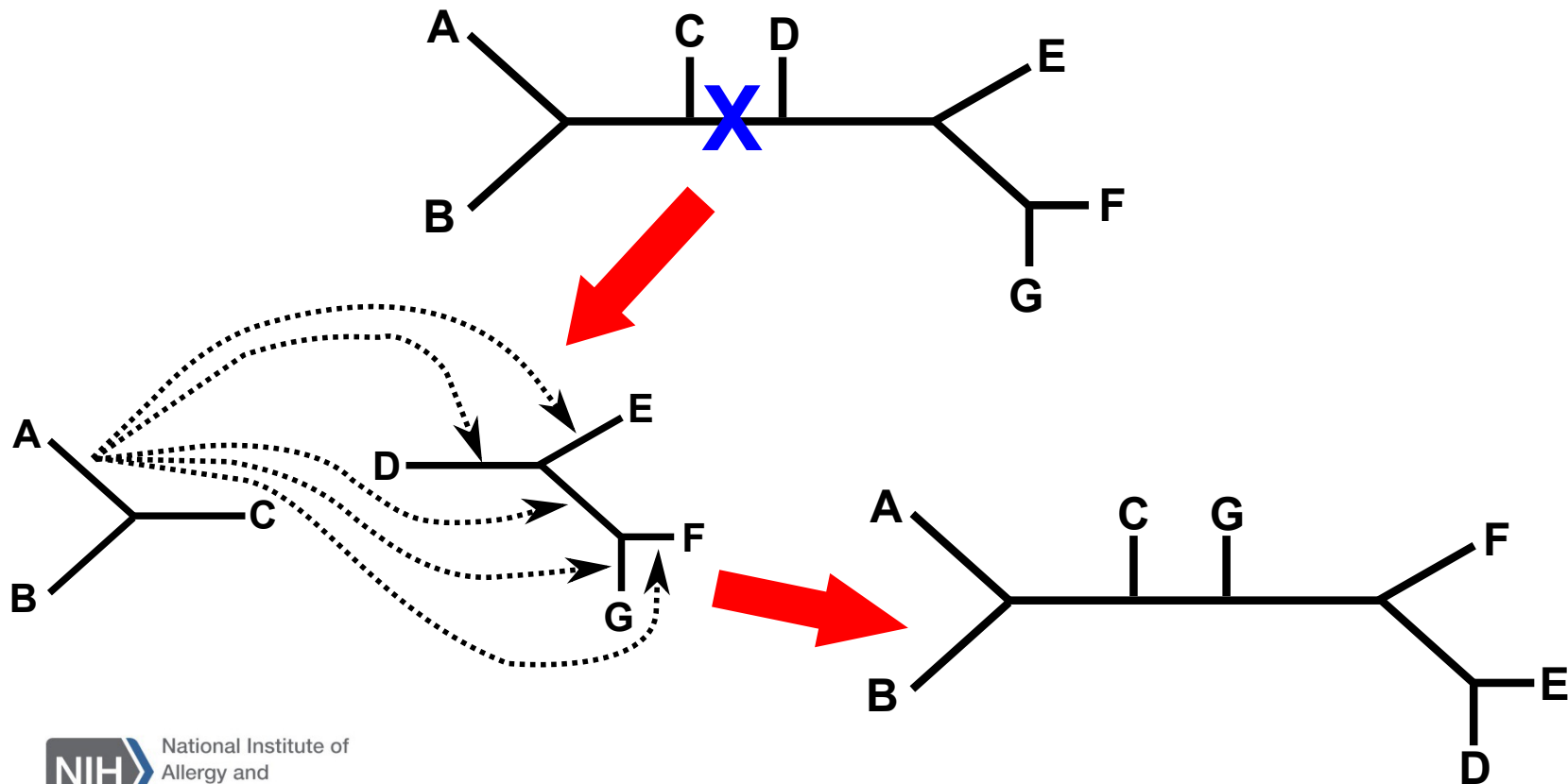
The optimality criterion approach

Branch swapping: Subtree pruning and regrafting



The optimality criterion approach

Branch swapping: Tree bisection and reconnection



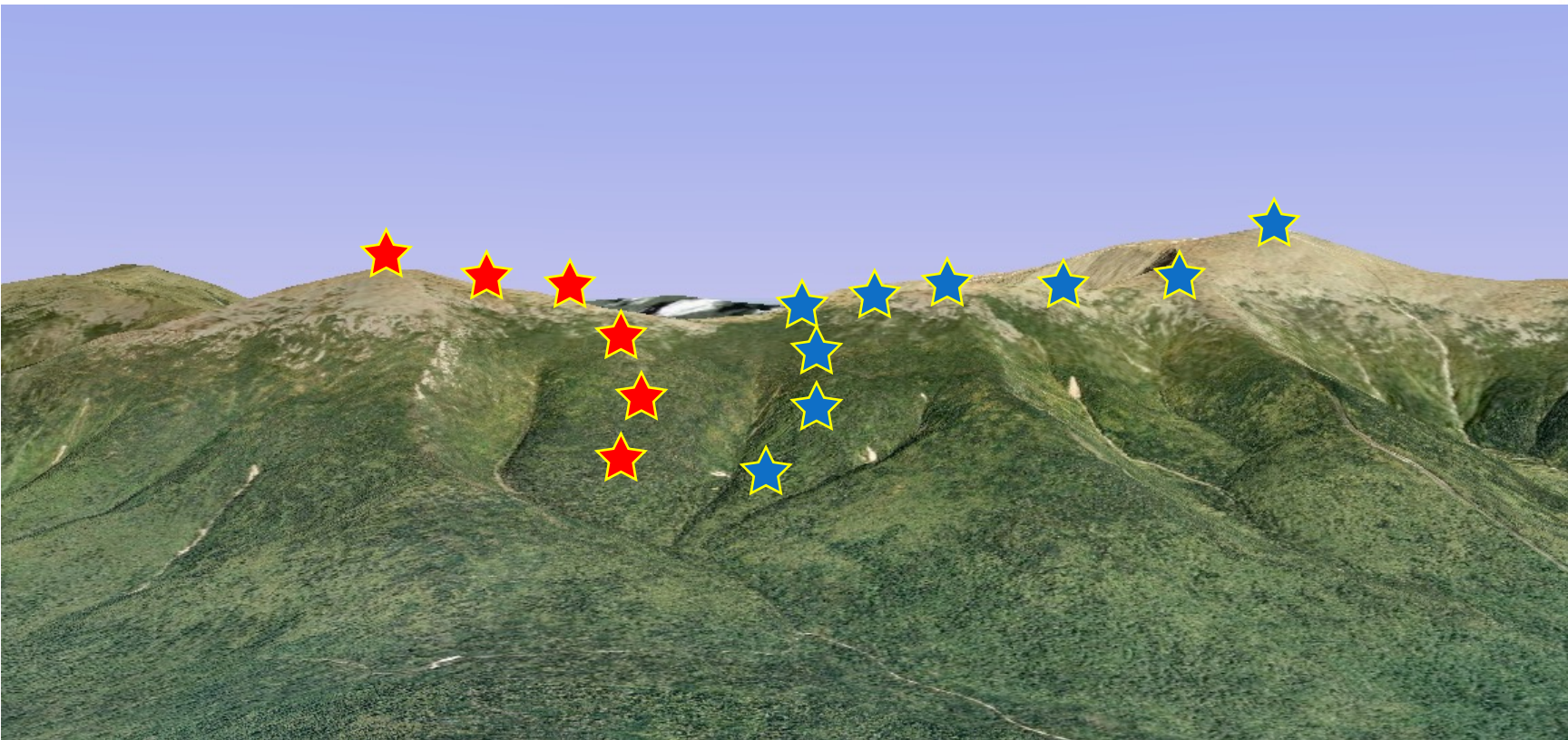
The optimality criterion approach

Exploring tree space

Beware! Hill climbing can often lead to local maxima rather than a global solution.

The optimality criterion approach

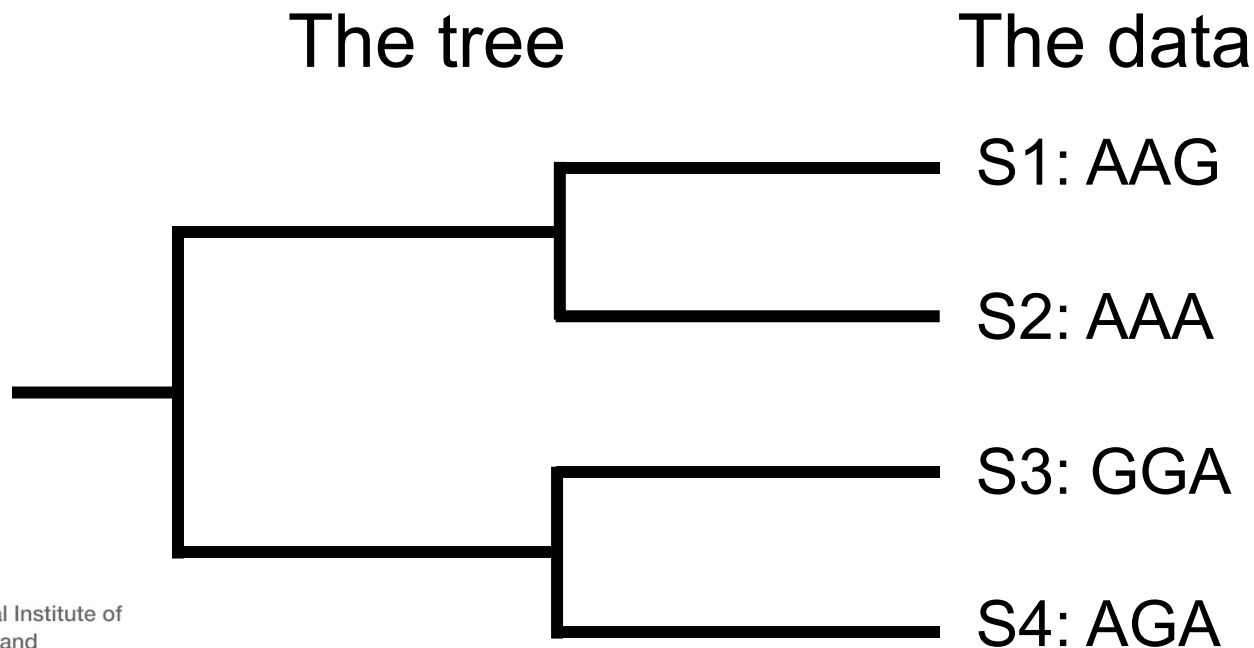
Exploring tree space



The optimality criterion approach

Is this tree optimal?

Counting changes (Fitch parsimony)



The optimality criterion approach

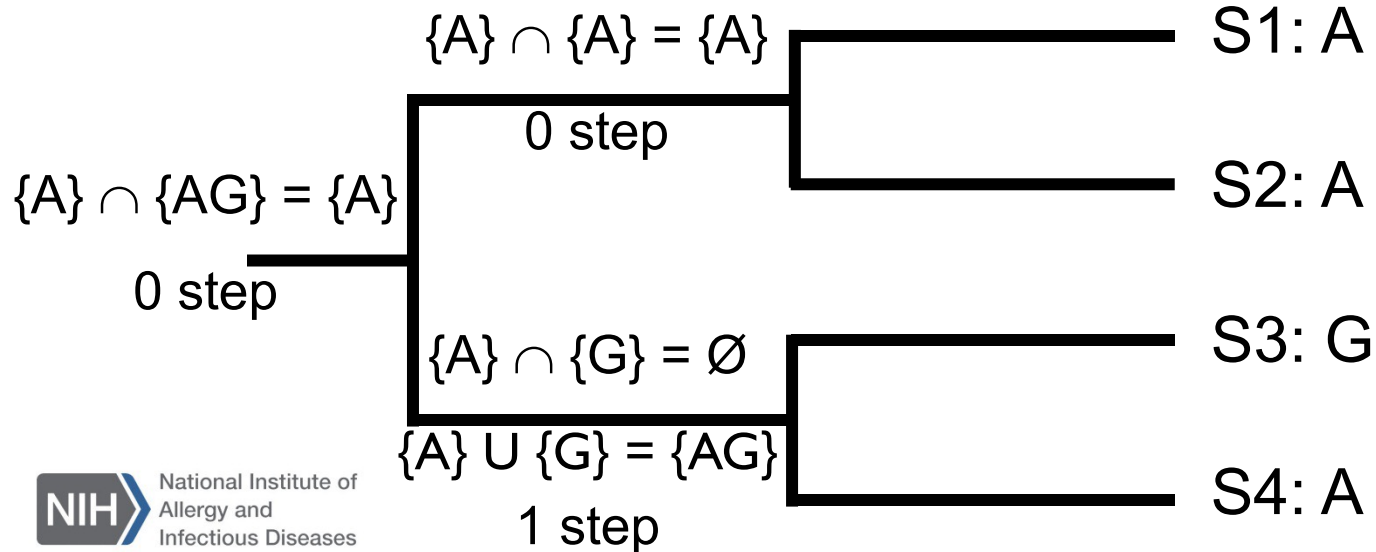
Is this tree optimal?

Counting changes (Fitch parsimony)

Position 1

The tree

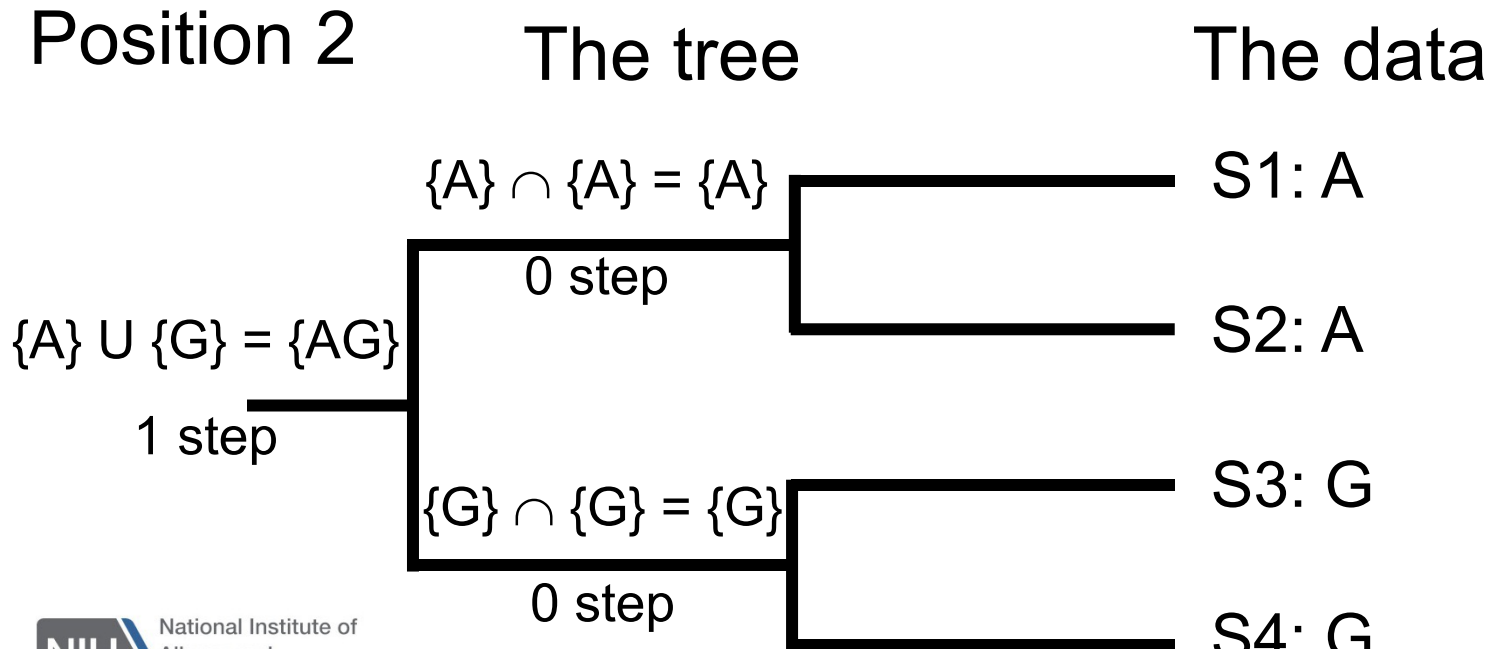
The data



The optimality criterion approach

Is this tree optimal?

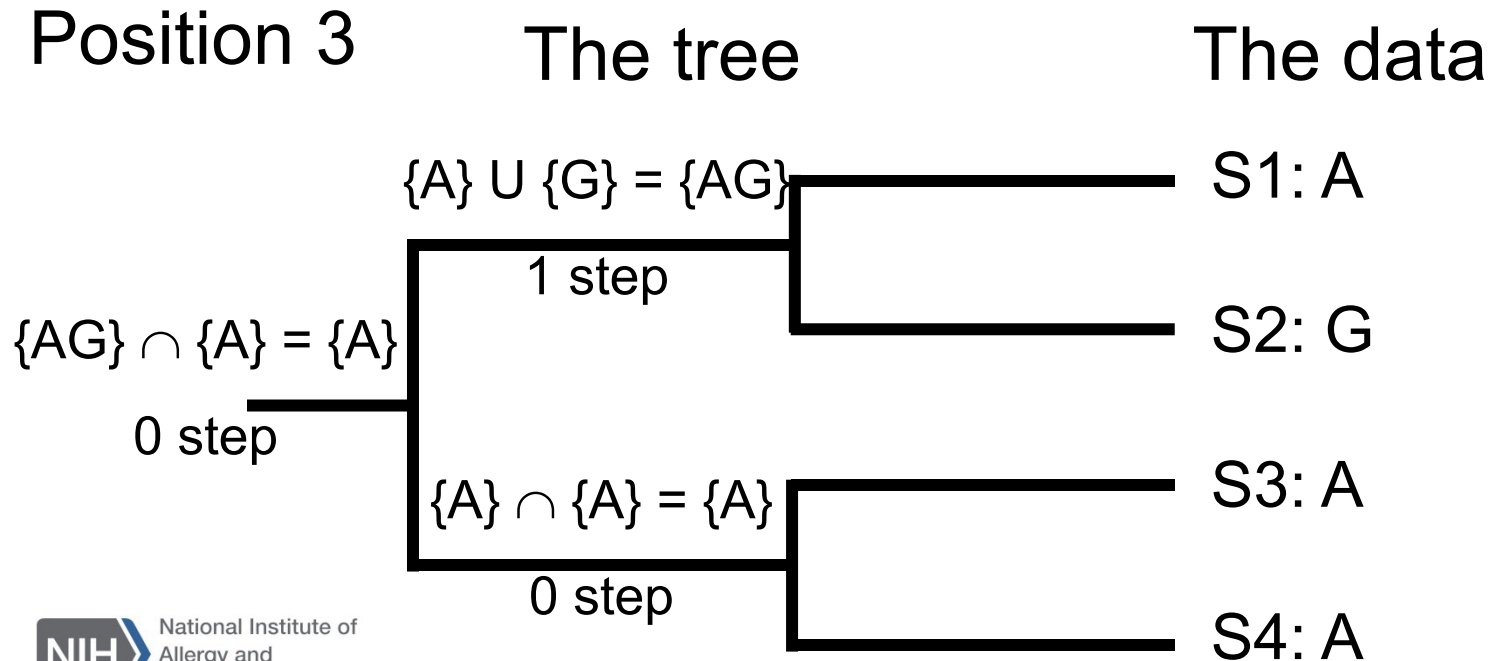
Counting changes (Fitch parsimony)



The optimality criterion approach

Is this tree optimal?

Counting changes (Fitch parsimony)



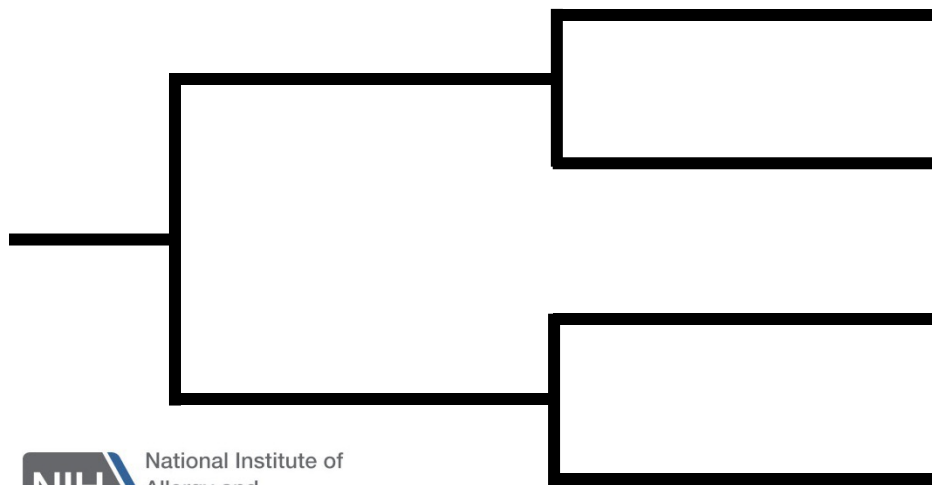
The optimality criterion approach

Is this tree optimal?

Counting changes (Fitch parsimony)

The tree

The data



S1: AAG

S2: AAA

S3: GGA

S4: AGA

Total
steps:
3

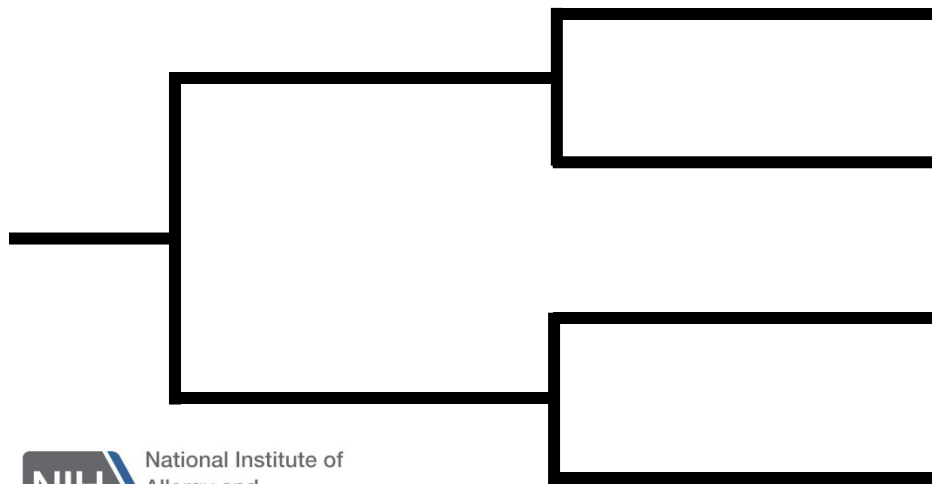
The optimality criterion approach

Is this tree optimal?

Counting changes (Fitch parsimony)

The tree

The data



S1: AAG

S3: GGA

S2: AAA

S4: AGA

Total
steps:
4

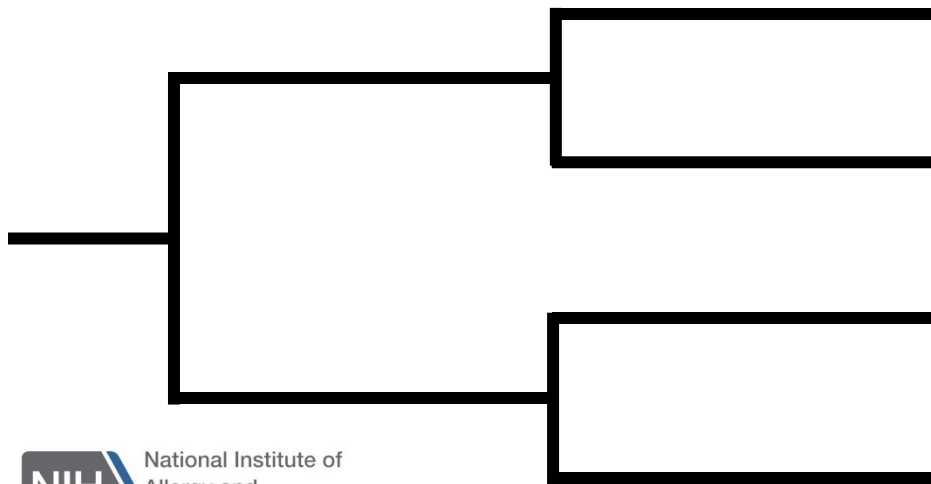
The optimality criterion approach

Is this tree optimal?

Counting changes (Fitch parsimony)

The tree

The data



S1: AAG

S4: AGA

S2: AAA

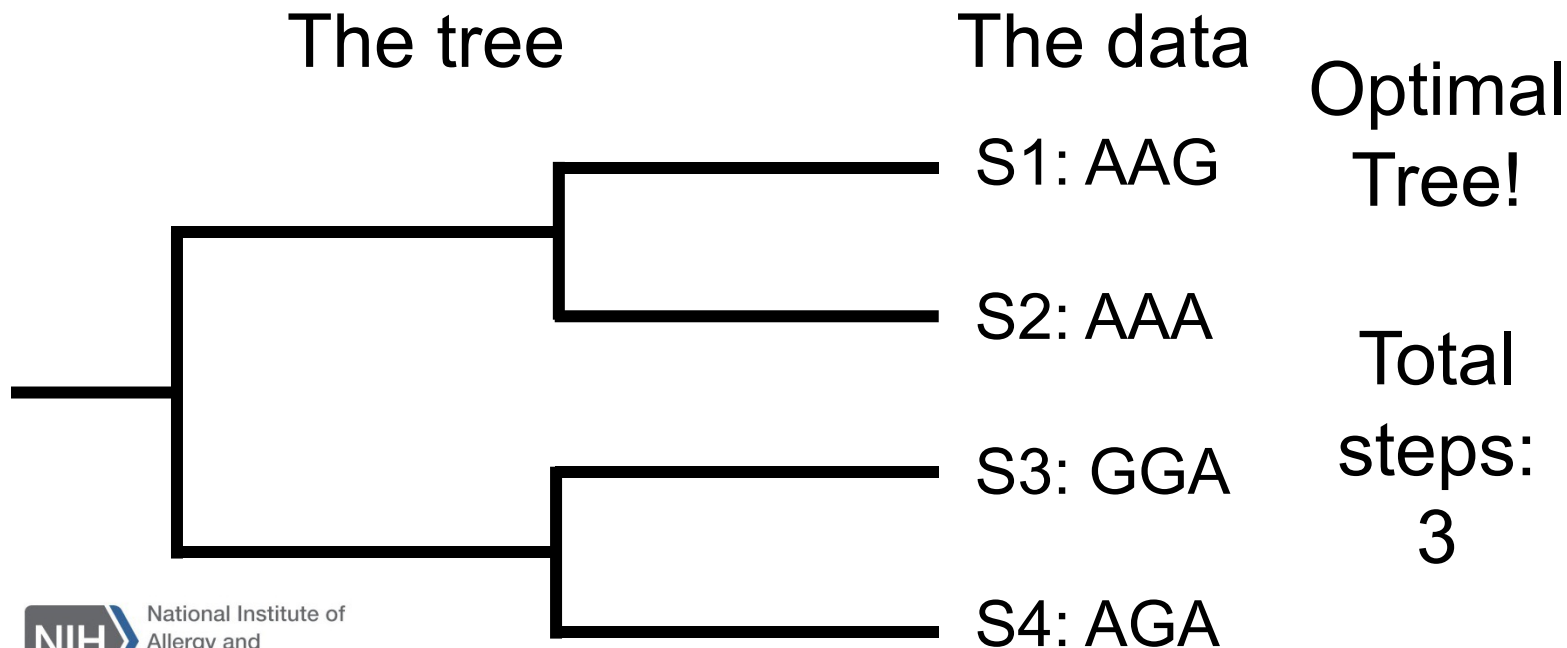
S3: GGA

Total
steps:
4

The optimality criterion approach

Is this tree optimal?

Counting changes (Fitch parsimony)



Multiple Sequence Alignment

File formats for MSA export/Phylogenetics input

FASTA

```
>HCVT050 Hepatitis C virus gene for polyprotein, complete cds,  
GGTCTTGGTCTACTGTGAGCGAGGAGGCCGGTGAGGACGTCGTCTGCTGCTCGATGTCCT  
ACACATGGACAGGCGCCTTAATCACGCCATGCGGCGCGGAGGAGACTAAGCTGCCCATCA  
ATGCGTTGAGCAACTCTTTGCTGCGTCACCACAACATGGTCTATGCCACAACATCCCGCA  
GCGCAAGCCAGCGGCAGAGAAAGGTCACCTTTGACAGACTGCAAGTCCTGGACGACCACT  
ACCGGGACGTGCTCAAGGAGATGAAGGCGAAGGCGTCCACAGTTAAGGCTAAACTTCTAT  
CCGTAGAAGAAGCCTGCAAGCTGACGCCCCCACATTCGGCCAGATCCAAGTTTGGCTATG  
GGGCAAAGGACGTCCGGAACCTATCCAGCAAGGCCGTTAACCACATCCGCTCCGTGTGGA  
AGGACTTGCTGGAAGACACTGAAACACCAATAAACACCACCATCATGGCAAAAAGTGAGG  
TCTTCTGTGTTCAACCAGAGAAAGGAGGCCGCAAGCCAGCTCGCCTTATCGTATTCCCAG
```


Multiple Sequence Alignment

File formats for MSA export/Phylogenetics input

PHYLIP

1st line: Number of sequences(space)Number of sites

2nd line: Sequence ID (10 characters max) Sequence

```
9 1823
HCVT050 GGTCTTGGTCTACTGTGAGCGAGGAGGCCGGTGAGGACGTCGTCTGCTGC
HCVT142 GGTCTTGGTCTACCGTGAGTGAGGAGGCCACTGAGGACGTCGTCTGCTGC
HCVT169 GGTCTTGGTCTACCGTGAGCGAGGAGGCTAGTGAGGACGTCGTCTGCTGC
SE03071689GGTCGTGGTCCACCGTGAACGAGGAGGCTGGTGAGGACGTCGTCTGCTGC
HCVT221 GGTCTTGGTCTACCGTGAGCGAGGAGGCCAGTGAAGACGTTGTCTGCTGC
MD2_2 GGTCTTGGTCTACTGTAAGCGAGGAGGCTAGTGAGGACGTCGTCTGCTGC
HCV1b GGTCTTGGTCTACCGTGAGCGAAGAGGCTGGTGAGGATGTCGTCTGCTGC
Contig0001GGTCTTGGTCTACCGTGAGCGAGGAGGCTAGTGAGGACGTCGTCTGCTGC
HCVT140 GGTCTTGGTCTACTGTGAGCGAGGAGGCTAGTGAGGATGTCGTCTGTTGC
```

Multiple Sequence Alignment

File formats for MSA export/Phylogenetics input

NEXUS

```
#NEXUS
[Name: HCVT050           Len: 1823 Check: 5A341084]
[Name: HCVT142          Len: 1823 Check: AB5C0B76]
[Name: HCVT169          Len: 1823 Check: 7EAF66DA]
[Name: SE03071689       Len: 1823 Check: 1EFF8405]
[Name: HCVT221          Len: 1823 Check: 3D0C96F0]
[Name: MD2_2            Len: 1823 Check: 1E2A0948]
[Name: HCV1b            Len: 1823 Check: BC29D7FB]
[Name: Contig0001       Len: 1823 Check: CD240524]
[Name: HCVT140          Len: 1823 Check: 2A5C0D4E]

begin data;
  dimensions ntax=9 nchar=1823;
  format datatype=dna interleave missing=-;
matrix
HCVT050 GGTCTTGGTCTACTGTGAGC GAGGAGGCCGGTGAGGACGT
HCVT142 GGTCTTGGTCTACCGTGAGT GAGGAGGCCACTGAGGACGT
HCVT169 GGTCTTGGTCTACCGTGAGC GAGGAGGCTAGTGAGGACGT
SE0307168 GGTTCGTGGTCCACCGTGAAC GAGGAGGCTGGTGAGGACGT
HCVT221 GGTCTTGGTCTACCGTGAGC GAGGAGGCCAGTGAAGACGT
MD2_2   GGTCTTGGTCTACTGTAAGC GAGGAGGCTAGTGAGGACGT
HCV1b   GGTCTTGGTCTACCGTGAGC GAAGAGGCTGGTGAGGATGT
Contig000 GGTCTTGGTCTACCGTGAGC GAGGAGGCTAGTGAGGACGT
HCVT140 GGTCTTGGTCTACTGTGAGC GAGGAGGCTAGTGAGGATGT
;
end;
```

Multiple Sequence Alignment

Phylogenetics program input data guidelines

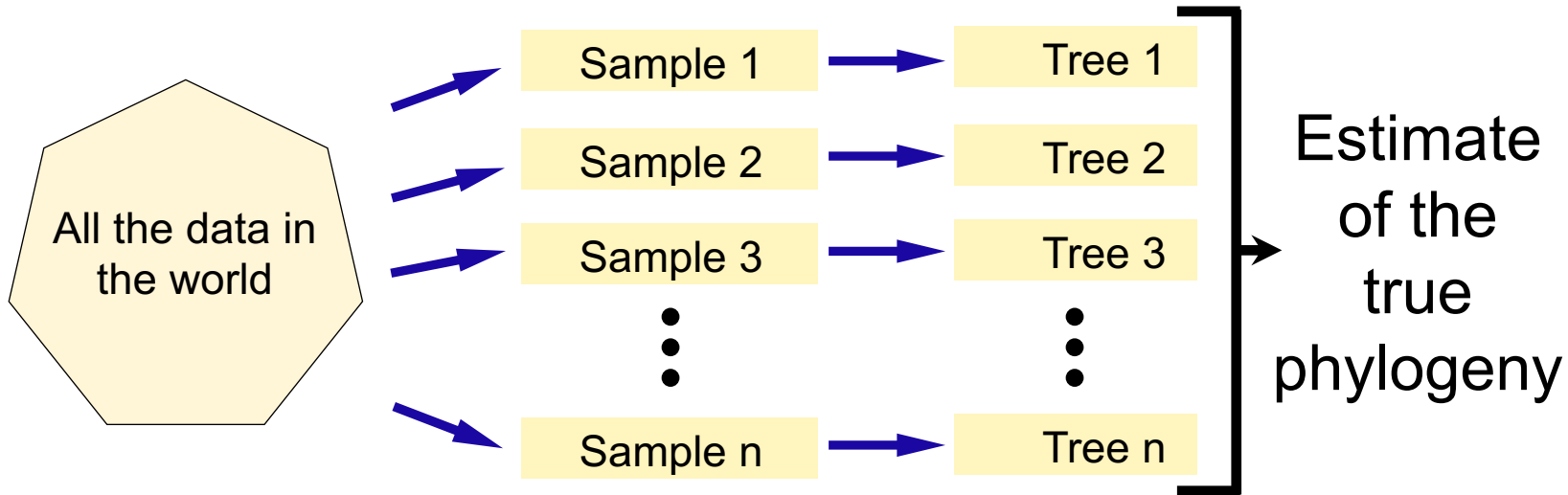
- Make sequence IDs different in the first ten characters
- Only letters, numbers, and “_” in sequence IDs
- Make sure all sequences overlap each other

How reliable are my trees?

Bootstrapping (nonparametric)

Bootstrapping

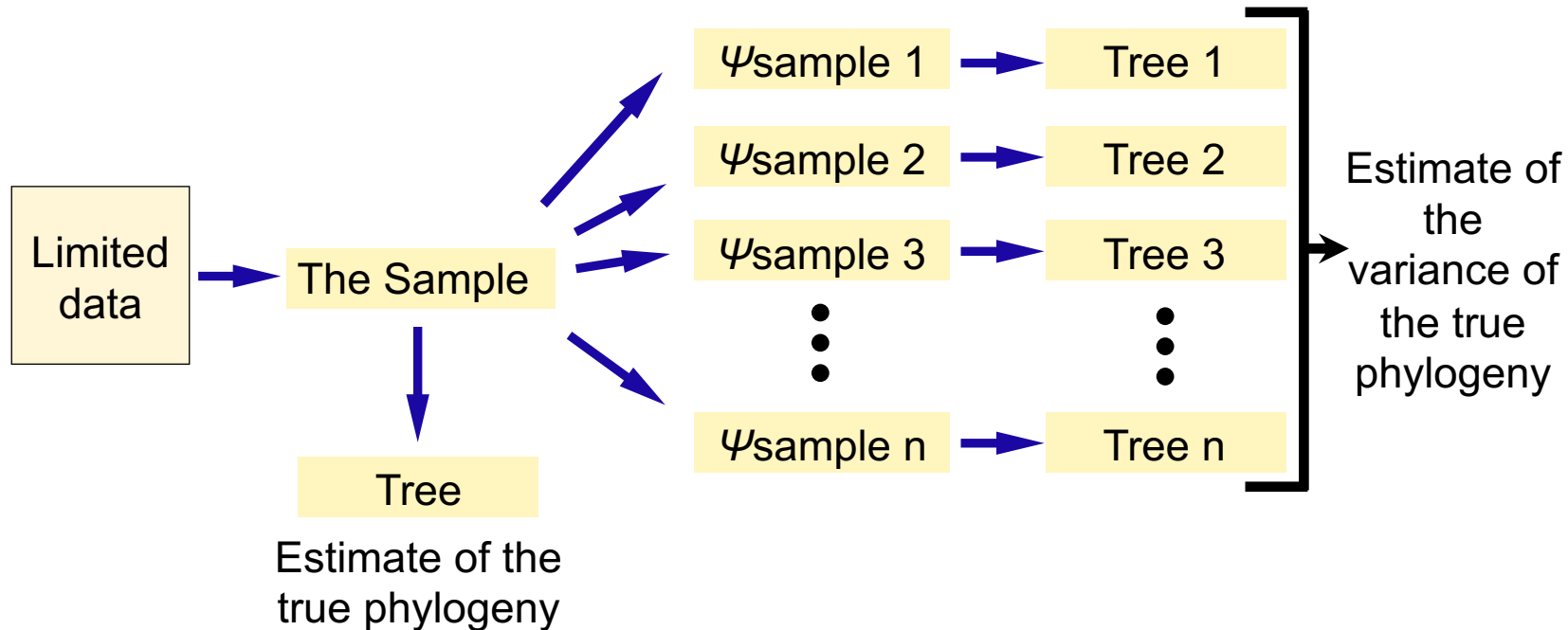
the ideal world



Build replicates by resampling from unlimited data

Bootstrapping

the real world



Build pseudoreplicates of unlimited data by sampling with replacement from limited data

Let's build some trees!

Phylogenetic programs (distance and parsimony)

PAUP* – Phylogenetic Analysis Using Parsimony

*and other methods

PHYLIP – a suite of phylogenetic programs

MEGA – An integrated phylogenetic analysis package

Let's build some trees!



Molecular Evolutionary
Genetics Analysis

tutorial ▾ features documentation ▾ feedback



◀ **Redesigned Tree Explorer toolbar** ▶

The Tree Explorer toolbar has been updated to be more intuitive and accessible.

○ ○ ○ ○ ○ ●

[Info on Log4j](#)

macOS ▾
Graphical (GUI) ▾
MEGA 11 (64-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

Site Links

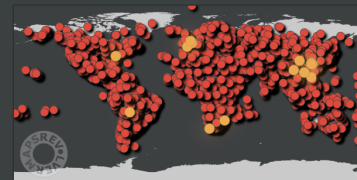
- Home
- Videos
- Walk through
- Books / Articles
- Features
- Publications
- Feedback

Documentation

- Online Manual
- MEGA 1.0 Manual PDF / HTML
- Example Data
- FAQ
- Update History
- Known Issues
- End User Agreement

Downloads

- Windows GUI / CC
- Mac OS X GUI / CC
- Ubuntu/Debian GUI / CC
- RedHat/Fedora GUI / CC
- Other Linux (CC) tar
- Older Versions



2,900,161 Downloads



NIH AID

MEGA11

Imported sequence alignment: Alignment Explorer

The screenshot displays the MEGA11 software interface. The main window has a menu bar with 'MEGA11', 'File', 'Analysis', and 'Help'. Below the menu bar is a toolbar with various icons for different analysis functions. The 'Edit/Build Alignment' icon is circled in black, and its corresponding menu is open, showing options like 'Edit/View Sequence Files (Trace)...', 'Open Saved Alignment Session...', 'Show Web Browser', 'Query Databanks', and 'Do BLAST Search'. Two dialog boxes are overlaid on the main window. The first dialog, titled 'M11: Alignment Editor', asks the user to 'Select An Option' and provides two choices: 'Create a new alignment' and 'Open a saved alignment session'. The second dialog, titled 'Data Type for Alignment', asks 'Are you building a DNA or protein sequence alignment?' and offers three buttons: 'DNA', 'Protein', and 'Cancel'. At the bottom of the interface, there is a 'RECENT PUBLICATIONS' section and a footer with navigation links: 'HELP DOCS', 'EXAMPLES', 'CITATION', 'REPORT BUG', 'UPDATES', 'MEGA LINKS', 'TOOLBAR', and 'PREFERENCES'. There are also buttons for 'ANALYZE', 'PROTOTYPE', and a large 'M1' logo.

MEGA11

Imported sequence alignment: Alignment Explorer

MEGA11 File Analysis Help
Molecular Evolutionary Genetics Analysis

M11: Alignment Explorer (MEGA11demo_IL1A_RawNMSeq.fas)

ALIGN

DNA Sequences Translated Protein Sequences

Species/Abbrv	A	G	C	T	G	C	C	A	G	C	C	A	G	A	G	G	G	A	G	T	C	A	T	T	T	C	A	T				
1. NM 000575.5 Homo sapiens interleukin 1 alpha (IL1A) transcript varia	A	G	C	T	G	C	C	A	G	C	C	A	G	A	G	G	G	A	G	T	C	A	T	T	T	C	A	T				
2. NM 001003157.2 Canis lupus familiaris interleukin 1 alpha (IL1A) mRN	A	C	A	A	A	G	G	C	G	A	A	G	T	A	G	T	C	T	G	A	C	T	A	G	G	G	C	T	A	A		
3. NM 001009351.1 Felis catus interleukin 1 alpha (IL1A) mRNA	A	T	G	G	C	C	A	A	A	G	T	T	C	C	T	G	A	C	C	T	C	T	T	T	G	A	A	G	A	C	C	
4. NM 001009808.1 Ovis aries interleukin 1 alpha (IL1A) mRNA	A	T	G	G	C	C	A	A	A	G	T	C	C	T	G	A	C	C	T	C	T	T	T	G	A	A	G	A	C	C		
5. NM 001042757.1 Macaca mulatta interleukin 1 alpha (IL1A) mRNA	A	T	G	G	C	C	A	A	A	G	T	T	C	C	A	G	A	C	A	T	G	T	T	T	G	A	A	G	A	C	C	
6. NM 001082500.2 Equus caballus interleukin 1 alpha (IL1A) mRNA	A	A	G	T	C	T	C	C	A	G	G	C	A	G	A	G	A	G	G	G	A	G	T	C	A	T	T	T	C	A	C	A
7. NM 001101684.1 Oryctolagus cuniculus interleukin 1 alpha (IL1A) mRN	A	G	T	C	A	G	T	T	C	A	T	T	G	G	C	A	C	T	T	G	A	G	T	C	G	G	C	A	A	A	G	A
8. NM 001280664.1 Tursiops truncatus interleukin 1 alpha (IL1A) mRNA	A	T	G	C	C	A	A	A	G	T	C	C	C	G	A	C	C	T	C	T	T	T	T	G	A	A	G	A	C	C		
9. NM 001290904.1 Bubalus bubalis interleukin 1 alpha (IL1A) mRNA	A	T	G	G	C	C	A	A	A	G	T	C	C	T	G	A	C	C	T	C	T	T	T	G	A	A	G	A	C	C		
10. NM 010554.4 Mus musculus interleukin 1 alpha (Il1a) mRNA	A	A	G	T	C	T	C	C	A	G	G	G	C	A	G	A	G	A	G	G	G	A	G	T	C	A	A	C	T	C	A	
11. NM 017019.2 Rattus norvegicus interleukin 1 alpha (Il1a) mRNA	A	A	A	G	T	C	T	C	C	A	G	G	G	C	A	C	A	G	A	G	G	G	A	G	T	C	A	A	C	T	C	
12. NM 174092.1 Bos taurus interleukin 1 alpha (IL1A) mRNA	G	A	C	G	A	G	G	G	A	G	C	C	A	G	T	C	A	T	C	T	C	A	T	T	G	T	T	G	C	T	A	
13. NM 214029.1 Sus scrofa interleukin 1 alpha (IL1A) mRNA	G	T	G	A	A	G	A	T	G	G	C	C	A	A	G	T	C	C	T	G	A	C	C	T	C	T	T	T	T	G	A	
14. XM 002811788.2 PREDICTED: Pongo abelii interleukin 1 alpha (IL1A)	A	G	T	A	A	C	C	A	G	G	C	A	A	C	A	C	C	A	T	T	G	A	A	G	T	C	T	C	A	T	A	
15. XM 003804504.2 PREDICTED: Pan paniscus interleukin 1 alpha (IL1	A	G	T	A	A	C	C	A	G	G	C	A	A	C	A	C	C	A	A	T	G	A	A	G	T	C	T	C	A	T	A	
16. XM 004031622.2 PREDICTED: Gorilla gorilla gorilla interleukin 1 alp	A	G	T	A	A	C	C	A	G	G	C	A	A	C	A	C	C	A	T	T	G	A	A	G	T	C	T	C	G	T	A	
17. XM 030827398.1 PREDICTED: Nomascus leucogenys interleukin 1 a	A	G	T	A	A	C	C	A	G	G	C	A	A	C	A	C	C	A	T	T	G	A	A	G	T	C	T	C	A	T	A	
18. XM 032177081.1 PREDICTED: Hylobates moloch interleukin 1 alpha	A	G	T	A	A	C	C	A	G	G	C	A	A	C	A	C	C	A	T	T	G	A	A	G	T	C	T	C	A	T	A	
19. XM 525866.2 PREDICTED: Pan troglodytes interleukin 1 alpha (IL1A)	A	G	T	A	A	C	C	A	G	G	C	A	A	C	A	C	C	A	A	T	G	A	A	G	T	C	T	C	A	T	A	

RECENT P...

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

HELP DOCS EXAMPLES CITATION REPORT BUG UPDATES MEGA LINKS TOOLBAR PREFERENCES

PROTOTYPE

MEGA11

Imported sequence alignment: Substitution Model

The screenshot shows the MEGA11 software interface. The main window title is 'MEGA11 Molecular Evolutionary Genetics Analysis'. The menu bar includes 'File', 'Analysis', and 'Help'. The toolbar contains icons for 'ALIGN', 'DATA', and 'ANALYZE'. The 'ANALYZE' icon is circled in red. A dialog box titled 'M11: Analysis Preferences' is open, showing the 'Model Selection (ML)' tab. The dialog box is divided into sections: 'ANALYSIS', 'SUBSTITUTION MODEL', 'DATA SUBSET TO USE', and 'SYSTEM RESOURCE USAGE'. The 'SUBSTITUTION MODEL' section is highlighted with a red oval. The 'SUBSTITUTION MODEL' section includes the following options:

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

The 'DATA SUBSET TO USE' section includes the following options:

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

The 'SYSTEM RESOURCE USAGE' section includes the following options:

Number of Threads	8
-------------------	---

The dialog box has 'Help', 'Cancel', and 'OK' buttons at the bottom. The 'OK' button is highlighted in red. The bottom of the screen shows a 'RECENT PUBLICATIONS' section and a toolbar with icons for 'HELP DOCS', 'EXAMPLES', 'CITATION', 'REPORT BUG', 'UPDATES', 'MEGA LINKS', 'TOOLBAR', and 'PREFERENCES'. The 'ANALYZE' button is highlighted in red, and the 'PROTOTYPE' button is also visible. The 'M11' logo is in the bottom right corner.

MEGA11

Imported sequence alignment: Model Choice

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

Results

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(CT)	r(GC)
HKY+G	40	10431.265	10125.616	-5022.702	n/a	1.06	1.73	0.319	0.242	0.231	0.208	0.044	0.042	0.133	0.058	0.147	0.038	0.058	0.154	0.038	0.204	0.044	0.042
HKY+G+I	41	10440.257	10126.973	-5022.375	0.14	1.76	1.73	0.319	0.242	0.231	0.208	0.044	0.042	0.133	0.058	0.147	0.038	0.058	0.154	0.038	0.204	0.044	0.042
TN93+G	41	10440.313	10127.028	-5022.403	n/a	1.06	1.73	0.319	0.242	0.231	0.208	0.044	0.042	0.128	0.058	0.154	0.038	0.058	0.161	0.038	0.196	0.044	0.042
K2+G	37	10440.864	10158.125	-5041.971	n/a	1.05	1.72	0.250	0.250	0.250	0.250	0.046	0.046	0.158	0.046	0.158	0.046	0.046	0.158	0.046	0.158	0.046	0.046
HKY+I	40	10442.933	10137.284	-5028.536	0.32	n/a	1.69	0.319	0.242	0.231	0.208	0.044	0.042	0.132	0.059	0.146	0.038	0.059	0.153	0.038	0.202	0.044	0.042
T92+G	38	10443.182	10152.806	-5038.307	n/a	1.06	1.73	0.281	0.281	0.219	0.219	0.051	0.040	0.140	0.051	0.140	0.040	0.051	0.179	0.040	0.179	0.051	0.040
GTR+G	44	10445.557	10109.367	-5010.555	n/a	1.04	1.72	0.319	0.242	0.231	0.208	0.033	0.058	0.127	0.043	0.155	0.023	0.080	0.162	0.046	0.194	0.027	0.051
TN93+G+I	42	10449.242	10128.322	-5022.044	0.14	1.79	1.73	0.319	0.242	0.231	0.208	0.044	0.042	0.127	0.058	0.154	0.038	0.058	0.162	0.038	0.195	0.044	0.042
K2+G+I	38	10450.484	10160.108	-5041.958	0.10	1.50	1.72	0.250	0.250	0.250	0.250	0.046	0.046	0.158	0.046	0.158	0.046	0.046	0.158	0.046	0.158	0.046	0.046
TN93+I	41	10451.673	10138.389	-5028.083	0.32	n/a	1.69	0.319	0.242	0.231	0.208	0.044	0.042	0.126	0.059	0.154	0.038	0.059	0.161	0.038	0.193	0.044	0.042
T92+G+I	39	10452.774	10154.761	-5038.280	0.10	1.50	1.73	0.281	0.281	0.219	0.219	0.051	0.040	0.140	0.051	0.140	0.040	0.051	0.179	0.040	0.179	0.051	0.040
K2+I	37	10454.518	10171.778	-5048.798	0.32	n/a	1.68	0.250	0.250	0.250	0.250	0.047	0.047	0.157	0.047	0.157	0.047	0.047	0.157	0.047	0.157	0.047	0.047
GTR+G+I	45	10454.927	10111.102	-5010.417	0.10	1.46	1.72	0.319	0.242	0.231	0.208	0.033	0.058	0.127	0.043	0.155	0.023	0.080	0.162	0.046	0.194	0.027	0.051
T92+I	38	10457.073	10166.697	-5045.253	0.32	n/a	1.69	0.281	0.281	0.219	0.219	0.052	0.040	0.139	0.052	0.139	0.040	0.052	0.177	0.040	0.177	0.052	0.040
GTR+I	44	10460.300	10124.110	-5017.926	0.32	n/a	1.69	0.319	0.242	0.231	0.208	0.034	0.057	0.125	0.045	0.154	0.025	0.078	0.162	0.046	0.192	0.029	0.051
HKY	39	10545.283	10247.271	-5084.534	n/a	n/a	1.61	0.319	0.242	0.231	0.208	0.046	0.044	0.129	0.060	0.144	0.039	0.060	0.150	0.039	0.199	0.046	0.044
TN93	40	10554.488	10248.839	-5084.313	n/a	n/a	1.61	0.319	0.242	0.231	0.208	0.046	0.044	0.126	0.060	0.148	0.039	0.060	0.155	0.039	0.193	0.046	0.044
K2	36	10555.946	10280.844	-5104.336	n/a	n/a	1.61	0.250	0.250	0.250	0.250	0.048	0.048	0.154	0.048	0.154	0.048	0.048	0.154	0.048	0.154	0.048	0.048
T92	37	10557.604	10274.865	-5100.341	n/a	n/a	1.61	0.281	0.281	0.219	0.219	0.053	0.042	0.136	0.053	0.136	0.042	0.053	0.174	0.042	0.174	0.053	0.042
GTR	43	10563.027	10234.471	-5074.113	n/a	n/a	1.61	0.319	0.242	0.231	0.208	0.036	0.058	0.126	0.048	0.149	0.026	0.080	0.156	0.047	0.193	0.031	0.052
JC+G	36	10695.577	10420.474	-5174.151	n/a	1.14	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G+I	37	10704.586	10421.846	-5173.832	0.14	1.93	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	36	10705.206	10430.104	-5178.965	0.31	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC	35	10798.084	10530.618	-5230.228	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

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 19 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 828 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2][3]

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

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

MEGA11

Imported sequence alignment: Model Choice

- MEGA11>Models and jModeltest2 use Bayesian Information Criteria (BIC) to evaluate model fit to data
- BIC is a function of likelihood of the model given the data, the number of parameters in the model, and the number of data points
- Lowest BIC = best fit model
- $\Delta\text{BIC} < 2$: no difference in how these models fit the data
- $\Delta\text{BIC} > 2, < 6$: good evidence that the best model is the best fit
- $\Delta\text{BIC} > 6, < 10$: strong evidence that the best model is the best fit
- $\Delta\text{BIC} > 10$: I can't even.

MEGA11

Phylogeny Construction: Neighbor-Joining

The screenshot displays the MEGA11 software interface. At the top, the title bar reads 'MEGA11 Molecular Evolutionary Genetics Analysis'. Below the title bar is a menu bar with 'File', 'Analysis', and 'Help'. A toolbar contains various icons for different functions. A red arrow points to the 'Phylogeny Construction' icon (a tree with a red branch). A context menu is open over this icon, listing several options: '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)', and 'Open Tree Session'. A second red arrow points to the 'Construct/Test Neighbor-Joining Tree...' option. Below the toolbar, there are logos for 'TIMETREE' and 'DATAMONKEY'. At the bottom of the interface, there is a 'RECENT PUBLICATIONS' section and a footer with navigation links: 'HELP DOCS', 'EXAMPLES', 'CITATION', 'REPORT BUG', 'UPDATES', 'MEGA LINKS', 'TOOLBAR', 'PREFERENCES', 'ANALYZE', 'PROTOTYPE', and the MEGA logo.

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Scope	→ All Selected Taxa
Statistical Method	→ Neighbor-joining
PHYLOGENY TEST	
Test of Phylogeny	✓ None
No. of Bootstrap Replications	Bootstrap method
	Interior-branch test
SUBSTITUTION MODEL	
Substitutions Type	→ Nucleotide
Genetic Code Table	→ Not Applicable
Model/Method	→ Maximum Composite Likelihood
Fixed Transition/Transversion Ratio	→ Not Applicable
Substitutions to Include	→ d: Transitions + Transversions
RATES AND PATTERNS	
Rates among Sites	→ Uniform Rates
Gamma Parameter	→ Not Applicable
Pattern among Lineages	→ Same (Homogeneous)
DATA SUBSET TO USE	
Gaps/Missing Data Treatment	→ Pairwise deletion
Site Coverage Cutoff (%)	→ Not Applicable
Select Codon Positions	→ <input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
SYSTEM RESOURCE USAGE	
Number of Threads	→ Not Applicable

? Help X Cancel ✓ OK

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
PHYLOGENY TEST	
Test of Phylogeny	Bootstrap method
No. of Bootstrap Replications	500
SUBSTITUTION MODEL	
Substitutions Type	No. of differences
Genetic Code Table	p-distance
Model/Method	Jukes-Cantor model
Fixed Transition/Transversion Ratio	Kimura 2-parameter model
Substitutions to Include	<input checked="" type="checkbox"/> Tajima-Nei model
	Tamura 3-parameter model
	Tamura-Nei model
	Maximum Composite Likelihood
	LogDet (Tamura-Kumar)
RATES AND PATTERNS	
Rates among Sites	Uniform Rates
Gamma Parameter	Not Applicable
Pattern among Lineages	Same (Homogeneous)
DATA SUBSET TO USE	
Gaps/Missing Data Treatment	Pairwise deletion
Site Coverage Cutoff (%)	Not Applicable
Select Codon Positions	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
SYSTEM RESOURCE USAGE	
Number of Threads	8

? Help X Cancel ✓ OK

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Scope	→ All Selected Taxa
Statistical Method	→ Neighbor-joining
PHYLOGENY TEST	
Test of Phylogeny	→ Bootstrap method
No. of Bootstrap Replications	→ 500
SUBSTITUTION MODEL	
Substitutions Type	→ Nucleotide
Genetic Code Table	→ Not Applicable
Model/Method	→ Tajima-Nei model
Fixed Transition/Transversion Ratio	→ Not Applicable
Substitutions to Include	→ All
RATES AND PATTERNS	
Rates among Sites	→ <input checked="" type="checkbox"/> Uniform Rates
Gamma Parameter	→ Gamma Distributed (G)
Pattern among Lineages	→ Same (Homogeneous)
DATA SUBSET TO USE	
Gaps/Missing Data Treatment	→ Pairwise deletion
Site Coverage Cutoff (%)	→ Not Applicable
Select Codon Positions	→ <input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
SYSTEM RESOURCE USAGE	
Number of Threads	→ 8

? Help X Cancel ✓ OK

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Scope	→ All Selected Taxa
Statistical Method	→ Neighbor-joining
PHYLOGENY TEST	
Test of Phylogeny	→ Bootstrap method
No. of Bootstrap Replications	→ 500
SUBSTITUTION MODEL	
Substitutions Type	→ Nucleotide
Genetic Code Table	→ Not Applicable
Model/Method	→ Tajima-Nei model
Fixed Transition/Transversion Ratio	→ Not Applicable
Substitutions to Include	→ All
RATES AND PATTERNS	
Rates among Sites	→ Gamma Distributed (G)
Gamma Parameter	→ 1.00
Pattern among Lineages	→ Same (Homogeneous)
DATA SUBSET TO USE	
Gaps/Missing Data Treatment	→ <input checked="" type="checkbox"/> Pairwise deletion
Site Coverage Cutoff (%)	→ <input type="checkbox"/> Complete deletion <input type="checkbox"/> Partial deletion
Select Codon Positions	→ <input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
SYSTEM RESOURCE USAGE	
Number of Threads	→ 8

? Help X Cancel ✓ OK

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Tree Explorer (MEGA11demo_PhyloAnalysis.mtsx)

Original Tree Bootstrap Tree

76
75
99
100
99
85
100
97
67
96
85
100
79
100

NM 000575.5 Homo sapiens interleukin 1 alpha (IL1A) transcript variant 1 mRNA
XM 525866.2 PREDICTED: Pan troglodytes interleukin 1 alpha (IL1A) mRNA
XM 003804504.2 PREDICTED: Pan paniscus interleukin 1 alpha (IL1A) mRNA
XM 004031622.2 PREDICTED: Gorilla gorilla gorilla interleukin 1 alpha (IL1A) transcript variant X1 mRNA
XM 002811788.2 PREDICTED: Pongo abelii interleukin 1 alpha (IL1A) mRNA
XM 030827398.1 PREDICTED: Nomascus leucogenys interleukin 1 alpha (IL1A) transcript variant X1 mRNA
XM 032177081.1 PREDICTED: Hylobates moloch interleukin 1 alpha (IL1A) transcript variant X1 mRNA
NM 001042757.1 Macaca mulatta interleukin 1 alpha (IL1A) mRNA
NM 001003157.2 Canis lupus familiaris interleukin 1 alpha (IL1A) mRNA
NM 001009351.1 Felis catus interleukin 1 alpha (IL1A) mRNA
NM 001082500.2 Equus caballus interleukin 1 alpha (IL1A) mRNA
NM 214029.1 Sus scrofa interleukin 1 alpha (IL1A) mRNA
NM 001280664.1 Tursiops truncatus interleukin 1 alpha (IL1A) mRNA
NM 001009808.1 Ovis aries interleukin 1 alpha (IL1A) mRNA
NM 001290904.1 Bubalus bubalis interleukin 1 alpha (IL1A) mRNA
NM 174092.1 Bos taurus interleukin 1 alpha (IL1A) mRNA
NM 001101684.1 Oryctolagus cuniculus interleukin 1 alpha (IL1A) mRNA
NM 010554.4 Mus musculus interleukin 1 alpha (Il1a) mRNA
NM 017019.2 Rattus norvegicus interleukin 1 alpha (Il1a) mRNA

0.050

Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tajima-Nei method [3] and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). This analysis involved 19 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 828 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [4][5]

1. Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
2. Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
3. Tajima F. and Nei M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* 1:269-285.
4. Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Tree Explorer (MEGA11demo_PhyloAnalysis.mtsx)

Original Tree Bootstrap Tree

Layout

- Taxon Names
- 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
- Collapse/Expand Lineages
- Compute
- Display Caption

Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tajima-Nei method [3] and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). This analysis involved 19 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 828 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [4][5]

- Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
- Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
- Tajima F. and Nei M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* 1:269-285.
- 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>.

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Tree Explorer (MEGA11demo_PhyloAnalysis.mtsx)

Original Tree Bootstrap Tree

76 NM 000575.5 Homo sapiens interleukin 1 alpha (IL1A) transcript variant 1 mRNA
75 XM 525866.2 PREDICTED: Pan troglodytes interleukin 1 alpha (IL1A) mRNA
99 XM 003804504.2 PREDICTED: Pan paniscus interleukin 1 alpha (IL1A) mRNA
99 XM 004031622.2 PREDICTED: Gorilla gorilla gorilla interleukin 1 alpha (IL1A) transcript variant X1 mRNA
100 XM 002811788.2 PREDICTED: Pongo abelii interleukin 1 alpha (IL1A) mRNA
99 XM 030827398.1 PREDICTED: Nomascus leucogenys interleukin 1 alpha (IL1A) transcript variant X1 mRNA
99 XM 032177081.1 PREDICTED: Hylobates moloch interleukin 1 alpha (IL1A) transcript variant X1 mRNA
NM 001042757.1 Macaca mulatta interleukin 1 alpha (IL1A) mRNA
85 NM 001101684.1 Oryctolagus cuniculus interleukin 1 alpha (IL1A) mRNA
100 NM 010554.4 Mus musculus interleukin 1 alpha (Il1a) mRNA
100 NM 017019.2 Rattus norvegicus interleukin 1 alpha (Il1a) mRNA
100 NM 001003157.2 Canis lupus familiaris interleukin 1 alpha (IL1A) mRNA
97 NM 001009351.1 Felis catus interleukin 1 alpha (IL1A) mRNA
67 NM 001082500.2 Equus caballus interleukin 1 alpha (IL1A) mRNA
96 NM 214029.1 Sus scrofa interleukin 1 alpha (IL1A) mRNA
85 NM 001280664.1 Tursiops truncatus interleukin 1 alpha (IL1A) mRNA
100 NM 001009808.1 Ovis aries interleukin 1 alpha (IL1A) mRNA
79 100 NM 001290904.1 Bubalus bubalis interleukin 1 alpha (IL1A) mRNA
NM 174092.1 Bos taurus interleukin 1 alpha (IL1A) mRNA

0.050

Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tajima-Nei method [3] and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). This analysis involved 19 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 828 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [4][5]

- Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
- Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
- Tajima F. and Nei M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* 1:269-285.
- Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*

SBL = 1.19530839 Ready

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Tree Explorer (MEGA11demo_PhyloAnalysis.mtsx)

Original Tree Bootstrap Tree

76 NM 000575.5 Homo sapiens interleukin 1 alpha (IL1A) transcript variant 1 mRNA
75 XM 525866.2 PREDICTED: Pan troglodytes interleukin 1 alpha (IL1A) mRNA
75 XM 003804504.2 PREDICTED: Pan paniscus interleukin 1 alpha (IL1A) mRNA
99 XM 004031622.2 PREDICTED: Gorilla gorilla gorilla interleukin 1 alpha (IL1A) transcript variant X1 mRNA
99 XM 002811788.2 PREDICTED: Pongo abelii interleukin 1 alpha (IL1A) mRNA
100 XM 030827398.1 PREDICTED: Nomascus leucogenys interleukin 1 alpha (IL1A) transcript variant X1 mRNA
99 XM 032177081.1 PREDICTED: Hylobates moloch interleukin 1 alpha (IL1A) transcript variant X1 mRNA
NM 001042757.1 Macaca mulatta interleukin 1 alpha (IL1A) mRNA
NM 001101684.1 Oryctolagus cuniculus interleukin 1 alpha (IL1A) mRNA
85 NM 010554.4 Mus musculus interleukin 1 alpha (Il1a) mRNA
100 NM 017019.2 Rattus norvegicus interleukin 1 alpha (Il1a) mRNA
100 NM 001003157.2 Canis lupus familiaris interleukin 1 alpha (IL1A) mRNA
NM 001009351.1 Felis catus interleukin 1 alpha (IL1A) mRNA
97 NM 001082500.2 Equus caballus interleukin 1 alpha (IL1A) mRNA
67 NM 214029.1 Sus scrofa interleukin 1 alpha (IL1A) mRNA
96 NM 001280664.1 Tursiops truncatus interleukin 1 alpha (IL1A) mRNA
85 NM 001009808.1 Ovis aries interleukin 1 alpha (IL1A) mRNA
100 NM 001290904.1 Bubalus bubalis interleukin 1 alpha (IL1A) mRNA
79 NM 174092.1 Bos taurus interleukin 1 alpha (IL1A) mRNA

0.050

Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tajima-Nei method [3] and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). This analysis involved 19 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 828 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [4][5]

- Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
- Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
- Tajima F. and Nei M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* 1:269-285.
- Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*

SBL = 1.19530839

Ready

MEGA11

Phylogeny Construction: Neighbor-Joining

M11: Tree Explorer (MEGA11demo_PhyloAnalysis.mtsx)

Original Tree Bootstrap Tree

76 NM 000575.5 Homo sapiens interleukin 1 alpha (IL1A) transcript variant 1 mRNA
75 XM 525866.2 PREDICTED: Pan troglodytes interleukin 1 alpha (IL1A) mRNA
75 XM 003804504.2 PREDICTED: Pan paniscus interleukin 1 alpha (IL1A) mRNA
99 XM 004031622.2 PREDICTED: Gorilla gorilla gorilla interleukin 1 alpha (IL1A) transcript variant X1 mRNA
99 XM 002811788.2 PREDICTED: Pongo abelii interleukin 1 alpha (IL1A) mRNA
100 XM 030827398.1 PREDICTED: Nomascus leucogenys interleukin 1 alpha (IL1A) transcript variant X1 mRNA
99 XM 032177081.1 PREDICTED: Hylobates moloch interleukin 1 alpha (IL1A) transcript variant X1 mRNA
100 NM 001042757.1 Macaca mulatta interleukin 1 alpha (IL1A) mRNA
85 NM 001101684.1 Oryctolagus cuniculus interleukin 1 alpha (IL1A) mRNA
100 NM 010554.4 Mus musculus interleukin 1 alpha (Il1a) mRNA
100 NM 017019.2 Rattus norvegicus interleukin 1 alpha (Il1a) mRNA
100 NM 001003157.2 Canis lupus familiaris interleukin 1 alpha (IL1A) mRNA
97 NM 001009351.1 Felis catus interleukin 1 alpha (IL1A) mRNA
67 NM 001082500.2 Equus caballus interleukin 1 alpha (IL1A) mRNA
96 NM 214029.1 Sus scrofa interleukin 1 alpha (IL1A) mRNA
85 NM 001280664.1 Tursiops truncatus interleukin 1 alpha (IL1A) mRNA
100 NM 001009808.1 Ovis aries interleukin 1 alpha (IL1A) mRNA
79 NM 001290904.1 Bubalus bubalis interleukin 1 alpha (IL1A) mRNA
79 NM 174092.1 Bos taurus interleukin 1 alpha (IL1A) mRNA

0.050

Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tajima-Nei method [3] and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). This analysis involved 19 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 828 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [4][5]

1. Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
2. Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
3. Tajima F. and Nei M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* 1:269-285.
4. Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*

SBL = 1.19530839 Ready

MEGA11

Phylogeny Construction: The Caption Window

The screenshot displays the MEGA11 software interface. On the left is a toolbar with options: Display Caption, Zoom In, Zoom Out, Print, Copy to Clip Board, Save as Text File, and Show in New Window. The main window shows a phylogenetic tree of interleukin 1 alpha (IL1A) mRNA sequences from various species. The tree is rooted at the top left with a diamond symbol. Bootstrap values are shown at the nodes: 100 for the root, 100 for the Rattus norvegicus branch, 97 for the Equus caballus branch, 67 for the Sus scrofa branch, 96 for the Tursiops truncatus branch, 85 for the Ovis aries branch, 100 for the Bubalus bubalis branch, and 79 for the Bos taurus branch. A scale bar at the bottom left indicates 0.050 substitutions per site.

Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tajima-Nei method [3] and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). This analysis involved 19 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 828 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [4][5]

1. Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
2. Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
3. Tajima F. and Nei M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* 1:269-285.
4. Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*

SBL = 1.19530839 Ready

Recapitulation

Where have we been? What have we done?

- What is a phylogenetic tree?
- How to build trees
 - Distance
 - Parsimony
- Calculated bootstrap support
- Calculated a neighbor-joining tree with MEGA11

Seminar Follow-Up Site

<https://bioinformatics.niaid.nih.gov>

The screenshot displays the homepage of the bioinformatics.niaid.nih.gov website. At the top left is the NIH bioinformatics NIAID @NIAID logo. A navigation menu at the top right includes links for Applications, Events, Training Resources, Services, and Code, along with a search bar. The main content area is divided into several sections: 'Applications' featuring 'Nephele' with a workflow diagram; 'Upcoming Events' listing 'Becoming a Reproducible Scientist (Part 1)'; 'Training Resources' which is circled in red; 'Code'; and 'Services'. Each section includes a 'VIEW ALL' button.

Seminar Follow-Up Site

<https://bioinformatics.niaid.nih.gov>

- ▶ 3D Printing
- ▶ Biostatistics
- ▶ General Bioinformatics
- ▶ **Next Generation Sequencing**
- ▶ Phylogenetics and Similarity
- ▶ Structural Biology
- ▶ Scientific Programming
- ▶ Reproducible Science
- ▶ Systems Biology

Welcome to the Training Resources section!

Here you can find training materials on a wide variety of topics from next generation sequencing (NGS) to molecular dynamics. Navigate through various topics on the left side of this page to find an appropriate course. Course content ranges from short videos to PDF and PowerPoint files.

Looking for a particular topic that is not found here? [Contact us](#) and we will point you in the right direction.



Related Events

[Connectivity Map Workshop](#)

Date: Tuesday, December 04, 2018
Location: Broad Institute, 415 Main St., Cambridge, MA 02142
Classification: Workshop

[BioC 2019](#)

Date: Tuesday, June 25, 2019
Location: Rockefeller University
Classification: Workshop

Related Code

[Blender_PNG_from_X3D.py](#)

Description: Imports a monochrome .x3d model into Blender and automatically generates a snapshot in .png format and exports a model in .stl format.

[Blender_STL_Cleanup.py](#)

Description: Imports a .wrl file into Blender, cleans the mesh, and exports .stl, .x3d, and .blend formats.

[Blender_VRML_cleanup_ribbon.py](#)

Description: Imports a .wrl model into Blender and cleans the mesh, exports a .stl model, and

Questions?

Email us!

bioinformatics@niaid.nih.gov



Next

Wednesday, 13 March

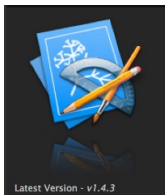


jModeltest: Fitting analysis parameters to your data

IQtree: Maximum Likelihood



MrBayes: Robust statistical phylogeny inference



FigTree: Making publication-quality tree figures

R/ggtree2: Making annotated tree figures in R