

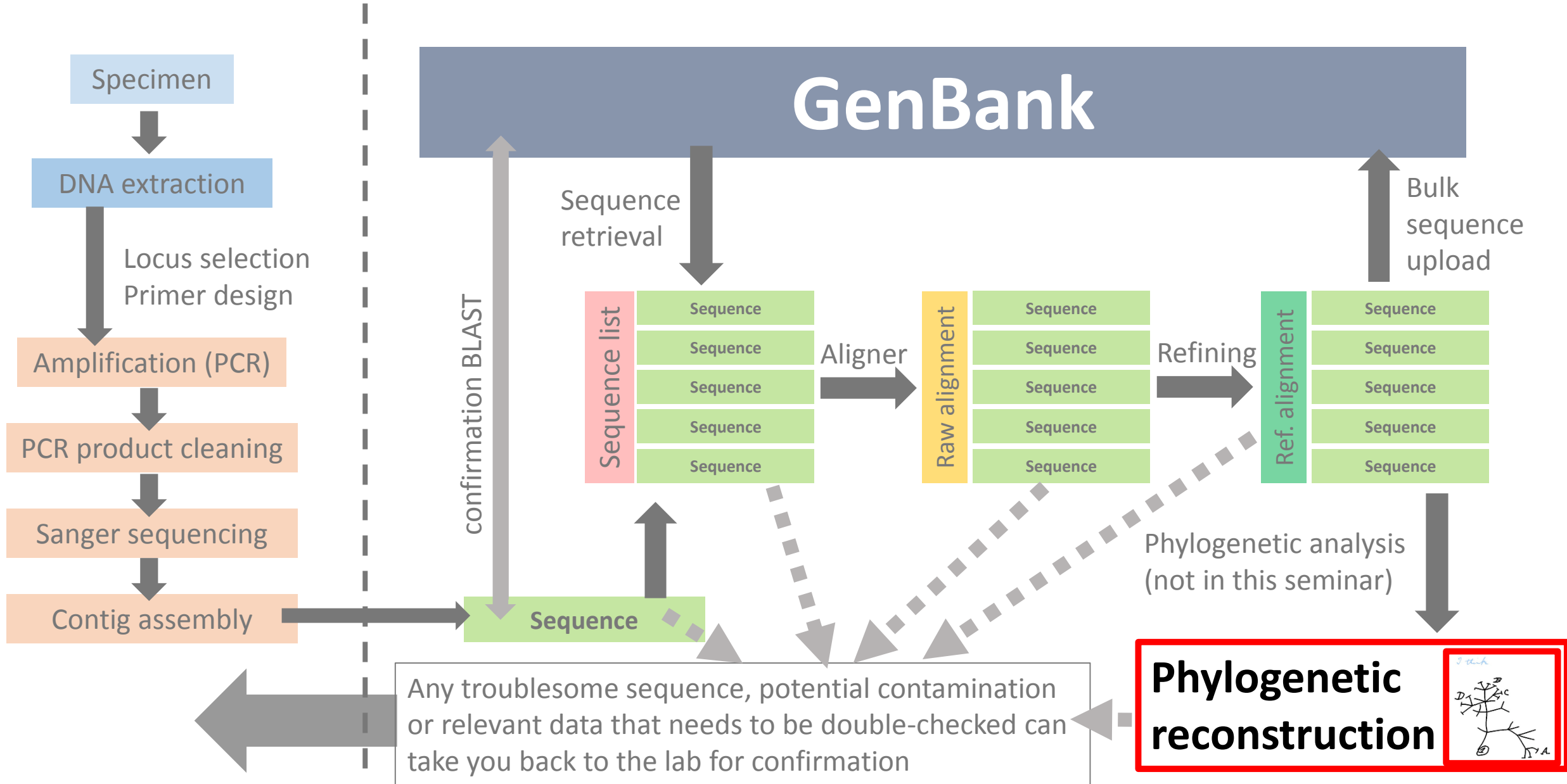
Generation of distance-based phylogenetic trees



Rafael Medina (rafael.medina.bry@gmail.com)

Yang Liu (yang.liu@uconn.edu)

Generation of phylogenetic trees



Some usual ways to infer a phylogenetic tree

- | | |
|---------------------------|--|
| Distance methods | Clustering of the sequences depending on the number of sites they differ |
| Maximum Parsimony | Seeks for the tree(s) with the least number of changes |
| Maximum Likelihood | Infers the tree(s) that makes most probable to observe the data (sequences) |
| Bayesian Inference | Provides the tree(s) with highest probability to be true assuming the data are correct |

Some usual ways to infer a phylogenetic tree

Distance methods



They result in a single tree, but do not explore intensively the topological space

Maximum Parsimony



Maximum Likelihood

Bayesian Inference

They use powerful, heuristic methods to explore the topological space and will result in a high number of trees that can be synthesized in a consensus tree

Some usual ways to infer a phylogenetic tree

Distance methods

Maximum Parsimony



They do not require any nucleotide substitution model to be used

Maximum Likelihood

Bayesian Inference



They require the use of nucleotide substitution models

Some usual ways to infer a phylogenetic tree

The details of phylogenetic inference under most of these methods lie out of the scope of this seminar, but we will learn how to make distance trees according two different distance methods:

UPGMA (Unweighted Pair-Group Method with Arithmetic mean)

NJ (Neighbor Joining)

UPGMA: how to make a tree, step by step

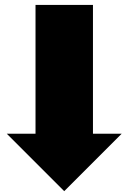
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | T | T | T | C | C | G | T | A | G | T | T | G | C | A | A | G | T | T | A | C | C | G | A | T | T | A | T |
| B | T | A | A | C | C | G | T | A | A | T | T | G | C | T | A | G | T | T | A | C | C | G | A | T | T | T | T |
| C | T | A | A | C | C | G | T | A | T | T | T | G | C | T | A | G | T | T | A | C | C | G | A | T | T | T | T |
| D | T | T | A | C | C | G | T | A | T | T | T | G | C | A | A | G | T | T | T | C | C | G | A | T | T | A | T |
| E | T | A | A | C | G | C | T | A | T | T | T | G | C | T | A | C | T | T | A | C | C | G | A | T | T | T | T |

Alignment

UPGMA: how to make a tree, step by step

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | T | T | C | C | G | T | A | G | T | T | G | C | A | A | G | T | T | A | C | C | G | A | T | T | A | T | |
| B | T | A | A | C | C | G | T | A | A | T | T | G | C | T | A | G | T | T | A | C | C | G | A | T | T | T | T |
| C | T | A | A | C | C | G | T | A | T | T | T | G | C | T | A | G | T | T | A | C | C | G | A | T | T | T | T |
| D | T | T | A | C | C | G | T | A | T | T | T | G | C | A | A | G | T | T | T | C | C | G | A | T | T | A | T |
| E | T | A | A | C | G | C | T | A | T | T | T | G | C | T | A | C | T | T | A | C | C | G | A | T | T | T | T |

Alignment



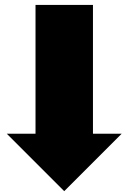
| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | | | | | |
| B | 5 | | | | |
| C | 5 | 1 | | | |
| D | 3 | 5 | 4 | | |
| E | 8 | 4 | 3 | 7 | |

Distance matrix (number of mutations)

UPGMA: how to make a tree, step by step

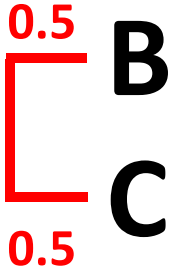
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | T | T | C | C | G | T | A | G | T | T | G | C | A | A | G | T | T | A | C | C | G | A | T | T | A | T | |
| B | T | A | A | C | C | G | T | A | A | T | T | G | C | T | A | G | T | T | A | C | C | G | A | T | T | T | T |
| C | T | A | A | C | C | G | T | A | T | T | T | G | C | T | A | G | T | T | A | C | C | G | A | T | T | T | T |
| D | T | T | A | C | C | G | T | A | T | T | T | G | C | A | A | G | T | T | T | C | C | G | A | T | T | A | T |
| E | T | A | A | C | G | C | T | A | T | T | T | G | C | T | A | C | T | T | A | C | C | G | A | T | T | T | T |

Alignment



| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | | | | | |
| B | 5 | | | | |
| C | 5 | 1 | | | |
| D | 3 | 5 | 4 | | |
| E | 8 | 4 | 3 | 7 | |

Distance matrix (number of mutations)



Scale: 1.0

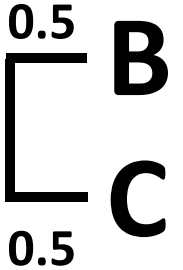
Find the shortest pairwise distance and make the first group.

The length of the branches is half of this value

UPGMA: how to make a tree, step by step

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | | | | | |
| B | 5 | | | | |
| C | 5 | 1 | | | |
| D | 3 | 5 | 4 | | |
| E | 8 | 4 | 3 | 7 | |

| | A | BC | D | E |
|----|---|-----|---|---|
| A | | | | |
| BC | 5 | | | |
| D | | 4.5 | | |
| E | | 3.5 | | |

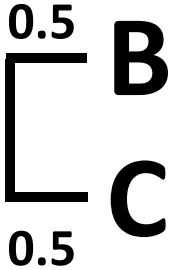


Scale: 1.0

Make a reduced distance matrix using the mean values between each of the remaining taxa and the cluster BC

UPGMA: how to make a tree, step by step

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | | | | | |
| B | 5 | | | | |
| C | 5 | 1 | | | |
| D | 3 | 5 | 4 | | |
| E | 8 | 4 | 3 | 7 | |



| | A | BC | D | E |
|----|---|-----|---|---|
| A | | | | |
| BC | 5 | | | |
| D | 3 | 4.5 | | |
| E | 8 | 3.5 | 7 | |

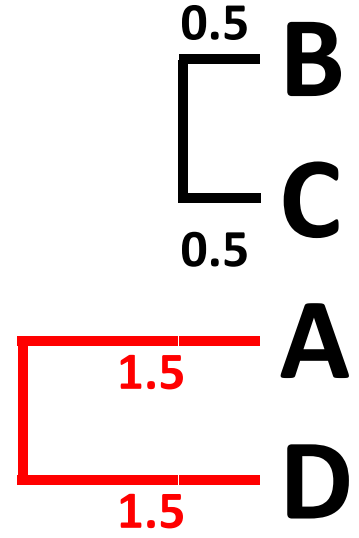
Scale: 1.0

Make a reduced distance matrix using the mean values between each of the remaining taxa and the cluster BC (the rest remains without changes)

UPGMA: how to make a tree, step by step

| | A | BC | D | E |
|----|---|-----|---|---|
| A | | | | |
| BC | 5 | | | |
| D | 3 | 4.5 | | |
| E | 8 | 3.5 | 7 | |

Repeat the first step with the reduced matrix



Scale: 1.0

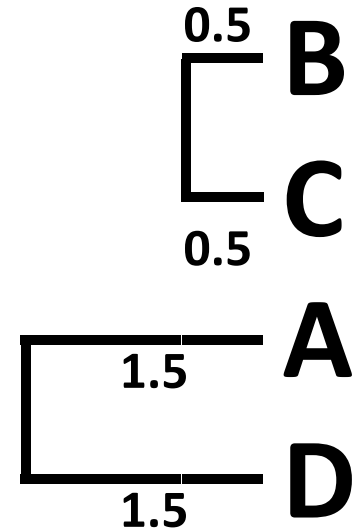
UPGMA: how to make a tree, step by step

| | A | BC | D | E |
|----|---|-----|---|---|
| A | | | | |
| BC | 5 | | | |
| D | 3 | 4.5 | | |
| E | 8 | 3.5 | 7 | |

Repeat the first step with the reduced matrix

| | AD | BC | E |
|----|------|-----|---|
| AD | | | |
| BC | 4.75 | | |
| E | 7.5 | 3.5 | |

And so on...



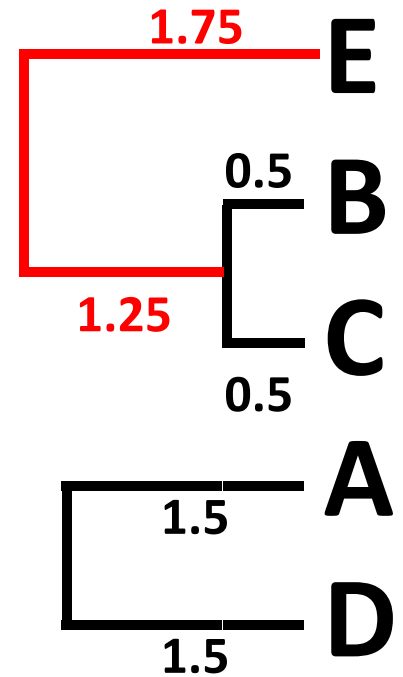
UPGMA: how to make a tree, step by step

| | A | BC | D | E |
|----|---|-----|---|---|
| A | | | | |
| BC | 5 | | | |
| D | 3 | 4.5 | | |
| E | 8 | 3.5 | 7 | |

Repeat the first step with the reduced matrix

| | AD | BC | E |
|----|------|-----|---|
| AD | | | |
| BC | 4.75 | | |
| E | 7.5 | 3.5 | |

And so on...



UPGMA: how to make a tree, step by step

| | A | BC | D | E |
|----|---|-----|---|---|
| A | | | | |
| BC | 5 | | | |
| D | 3 | 4.5 | | |
| E | 8 | 3.5 | 7 | |

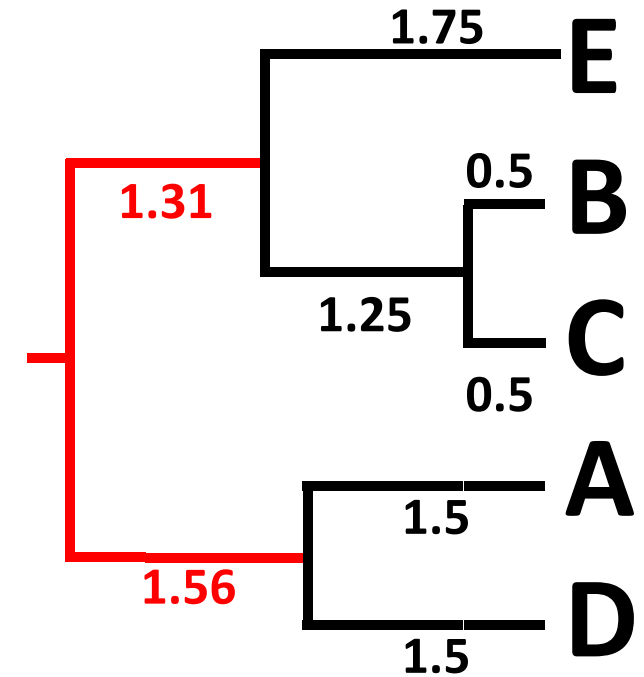
Repeat the first step with the reduced matrix

| | AD | BC | E |
|----|------|-----|---|
| AD | | | |
| BC | 4.75 | | |
| E | 7.5 | 3.5 | |

And so on...

| | AD | BCE |
|-----|-------|-----|
| AD | | |
| BCE | 6.125 | |

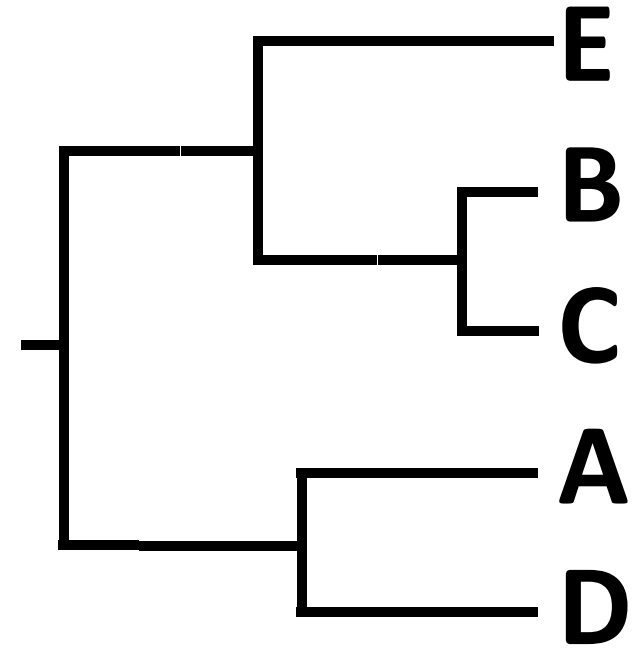
$$1.75 + 1.31 + 1.56 + 1.5 = 6.12$$



Scale: 1.0

UPGMA: how to make a tree, step by step

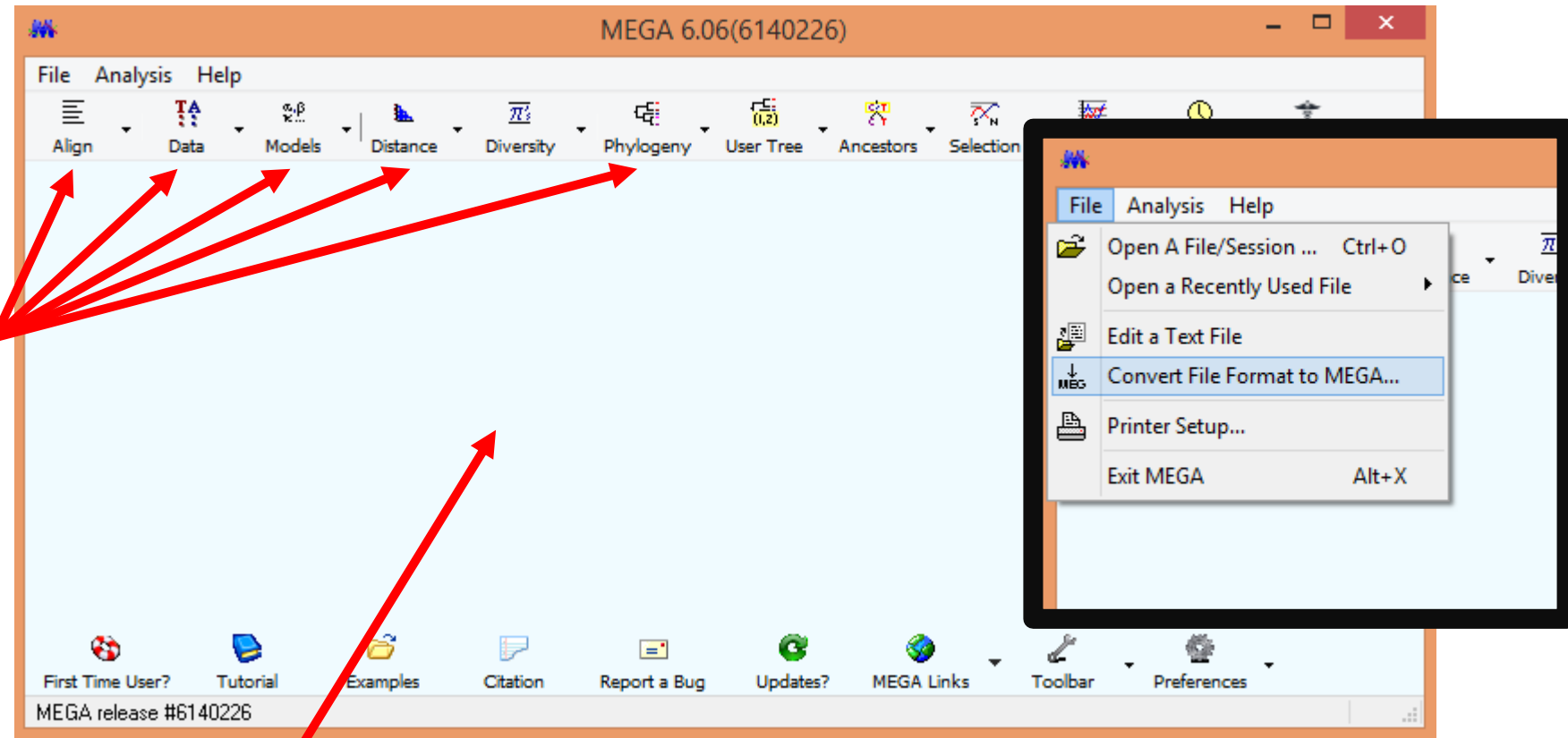
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | T | T | C | G | T | A | G | T | T | G | C | A | A | G | T | T | A | C | C | G | A | T | T | A | T | | |
| B | T | A | A | C | C | G | T | A | A | T | T | G | C | T | A | G | T | T | A | C | C | G | A | T | T | T | |
| C | T | A | A | C | C | G | T | A | T | T | T | G | C | T | A | G | T | T | A | C | C | G | A | T | T | T | |
| D | T | T | A | C | C | G | T | A | T | T | T | G | C | A | A | G | T | T | T | C | C | G | A | T | T | A | T |
| E | T | A | A | C | G | C | T | A | T | T | T | G | C | T | A | C | T | T | A | C | C | G | A | T | T | T | T |



Making trees using MEGA6

Download MEGA6 at <http://www.megasoftware.net/>
(Free, versatile and intuitive)

Once you run MEGA you will see this window

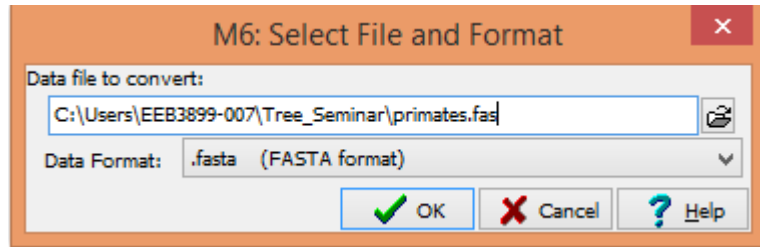


These menus will allow you to do different things with your data

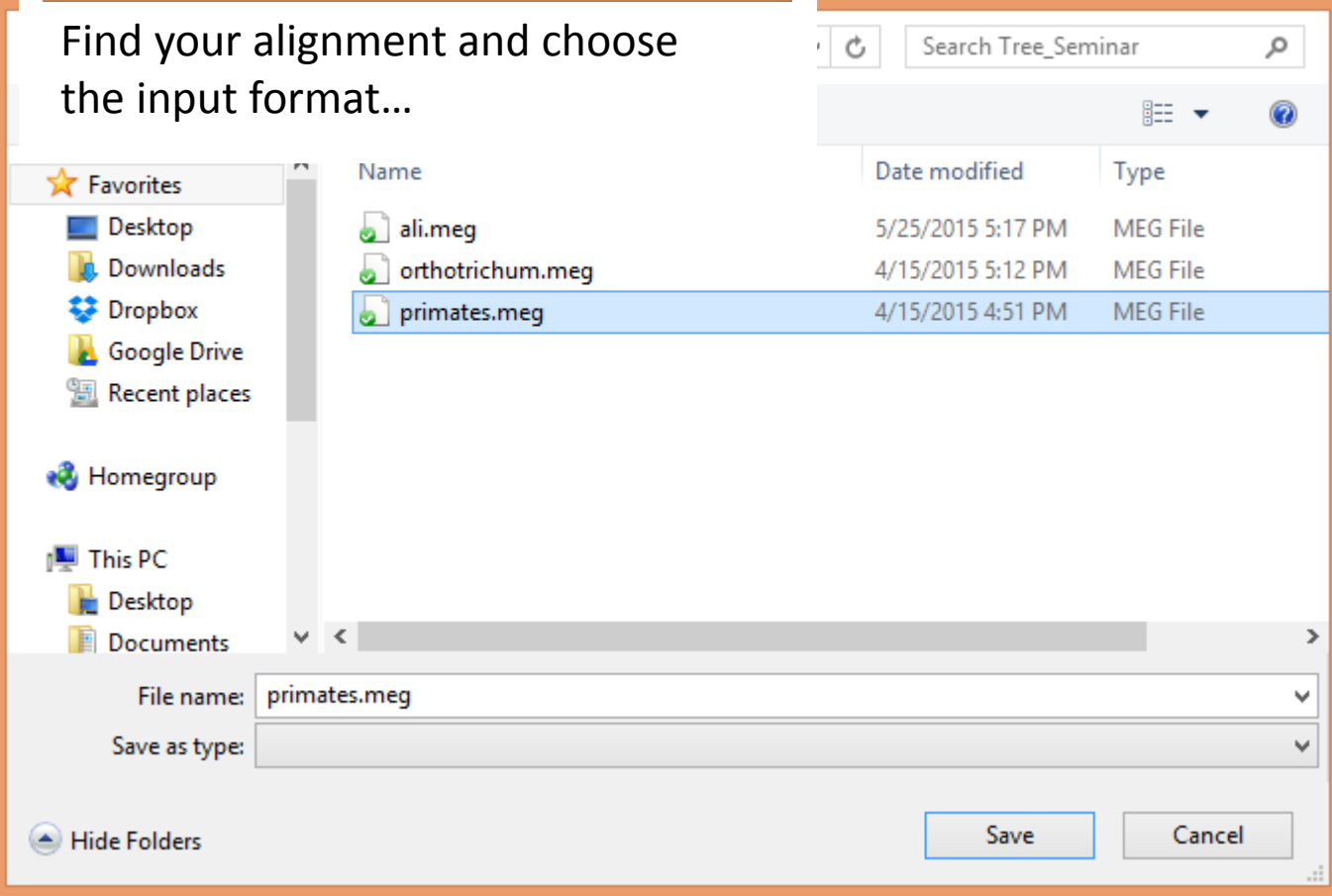
The different datasets, sessions and trees will appear in the central area

MEGA uses its own alignment format, so the first thing you need to do is convert your fasta/nexus file into MEGA

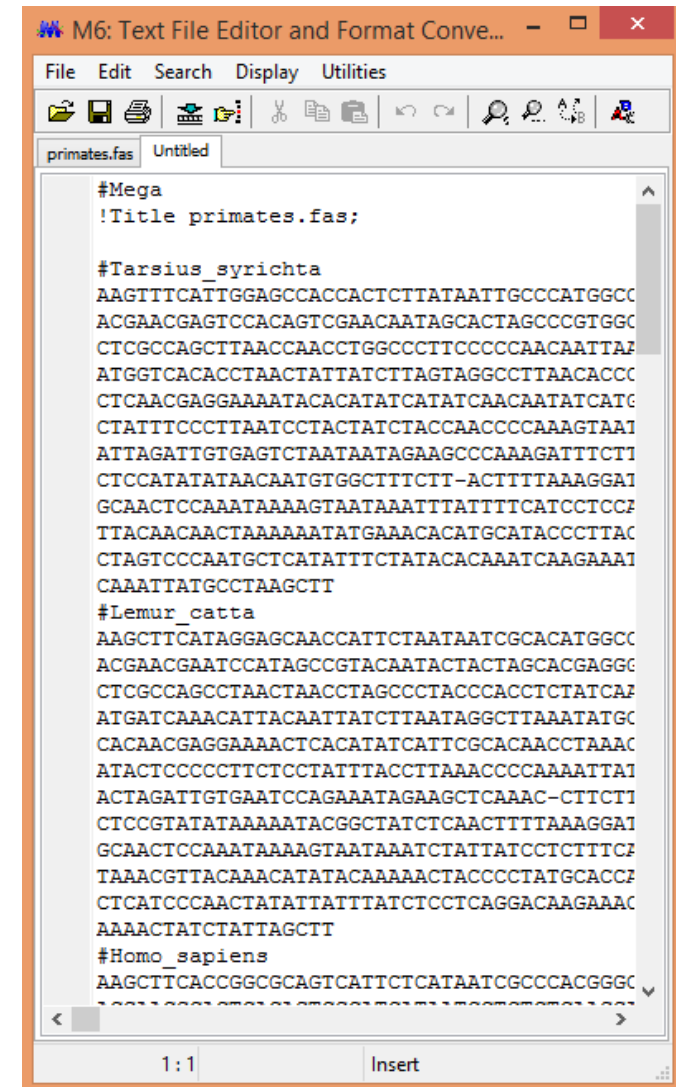
Making trees using MEGA6



Find your alignment and choose the input format...

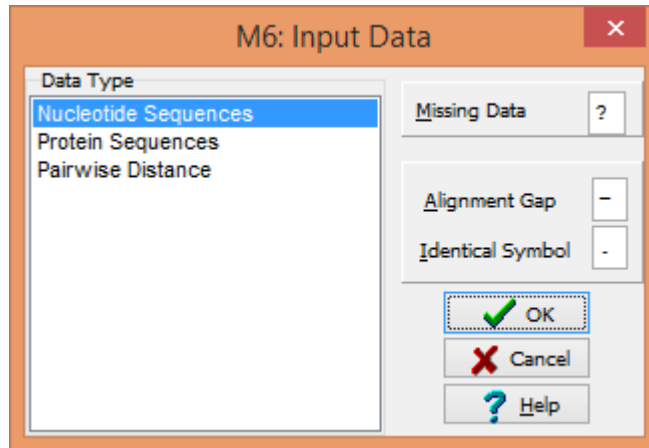


...and save it as a .meg file

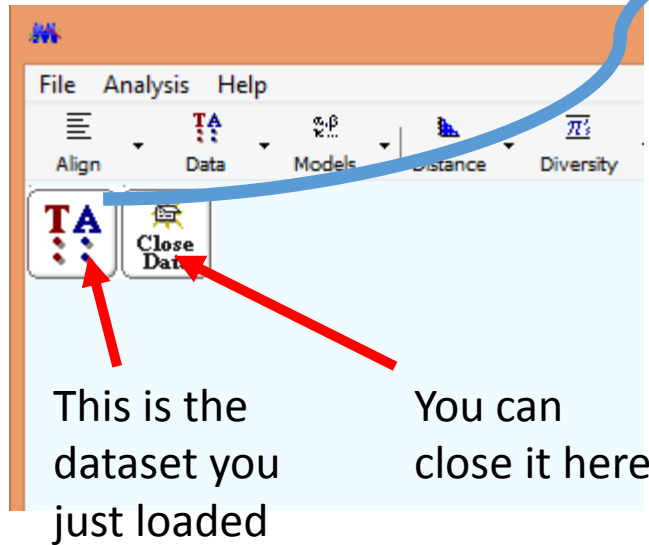


You will also see your alignment through the MEGA text editor, but it is not very useful for us at this point and you can just close it

Making trees using MEGA6

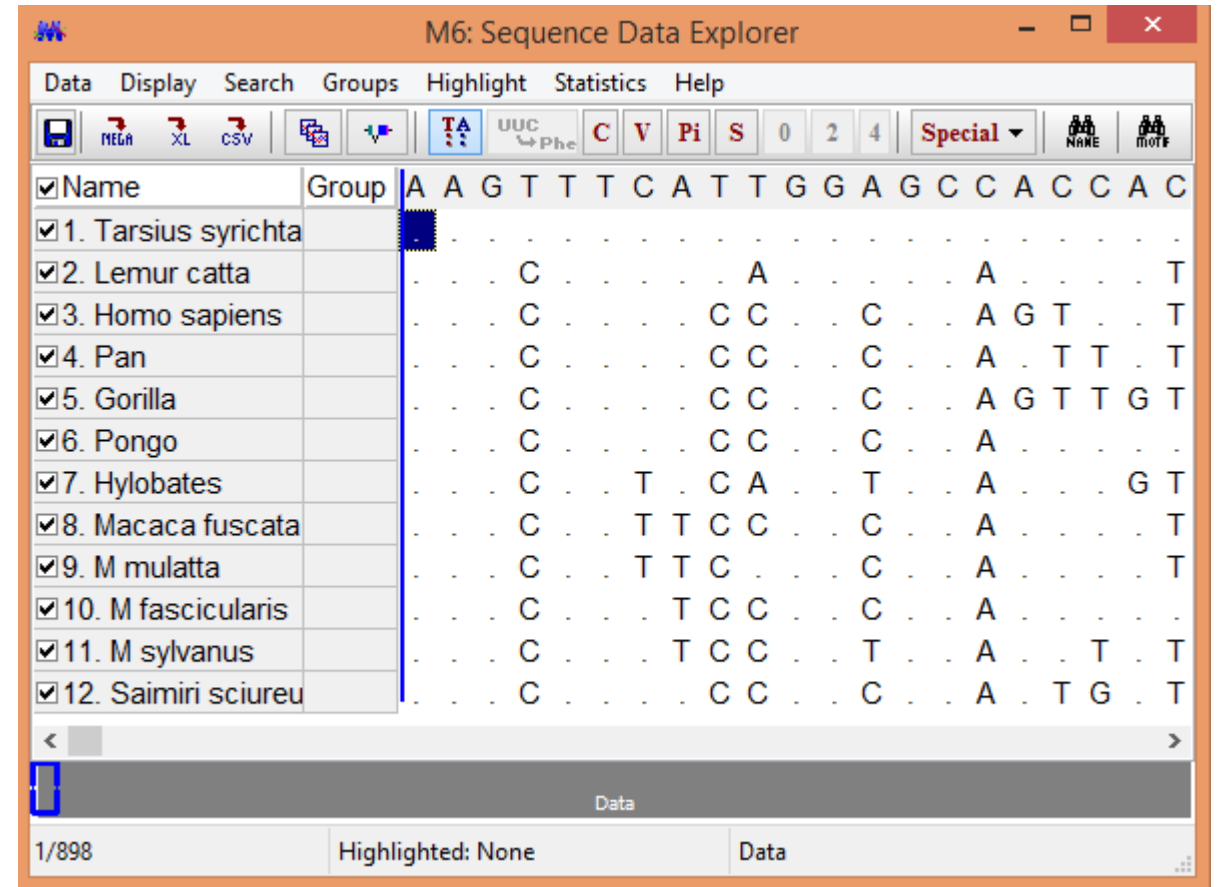


For our purposes we will not consider this sequence protein-coding region.



This is the dataset you just loaded

You can close it here



The alignment can be opened in the MEGA data explorer, which allows to find easily relevant information in the alignment

Making trees using MEGA6

Visualize:

Conserved Variable Parsimony informative Singletons

Display invariant sites as dots

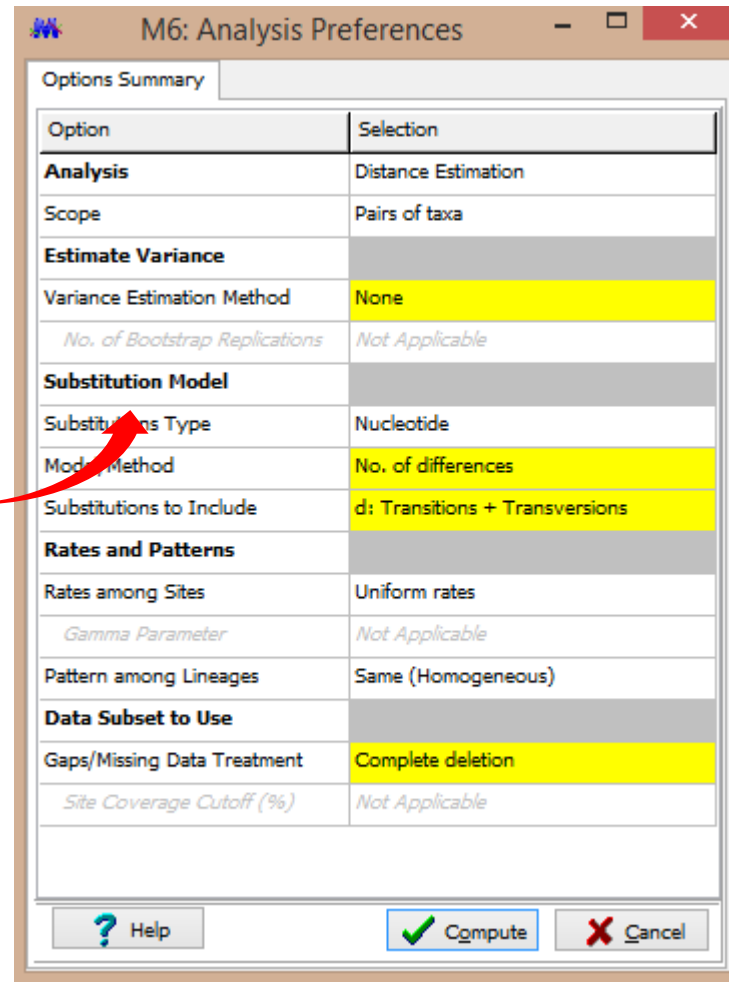
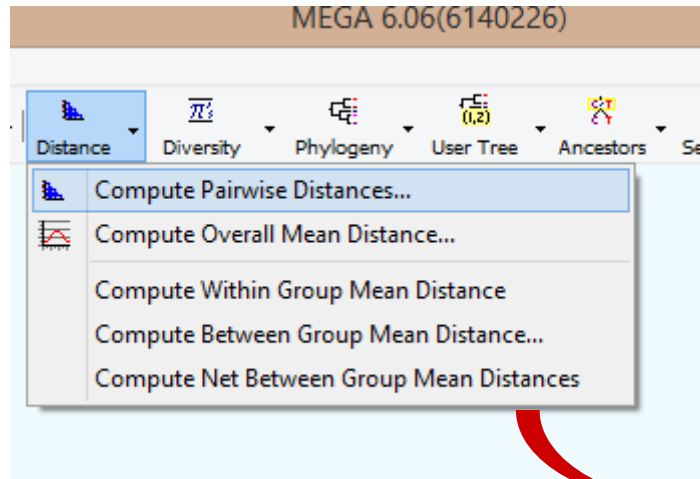
The screenshot shows the MEGA6 Sequence Data Explorer interface. The main window displays a sequence alignment for 12 species: 1. Tarsius syrichta, 2. Lemur catta, 3. Homo sapiens, 4. Pan, 5. Gorilla, 6. Pongo, 7. Hylobates, 8. Macaca fuscata, 9. M. mulatta, 10. M. fascicularis, 11. M. sylvanus, and 12. Saimiri sciureu. The alignment is shown as a grid of nucleotides (A, G, T, C) and dots representing invariant sites. Several columns are highlighted in yellow, indicating parsimony-informative sites. The status bar at the bottom shows '1/898' and 'Parsim-info: 367/898'. A red arrow points from the text 'Highlighted nucleotides / total' to the status bar.

| Name | Group | A | A | G | T | T | C | A | T | G | G | A | G | C | C | A | C | T | C | T | T | A | T | A | A | T | T | G | C | C | C | A | T | G | C | | | | | | | | | | |
|---------------------|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Tarsius syrichta | | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | | | | | | | |
| 2. Lemur catta | | . | . | C | . | . | . | . | A | . | . | . | A | . | . | . | T | . | . | A | . | . | . | . | . | C | . | . | A | . | . | . | . | . | . | . | | | | | | | | | |
| 3. Homo sapiens | | . | . | C | . | . | . | C | C | . | C | . | A | G | T | . | T | . | . | C | . | . | . | . | C | . | C | . | . | . | . | . | . | C | . | . | | | | | | | | | |
| 4. Pan | | . | . | C | . | . | . | C | C | . | C | . | A | . | T | T | T | C | . | C | . | . | . | . | C | . | C | . | . | . | . | . | . | . | C | . | . | | | | | | | | |
| 5. Gorilla | | . | . | C | . | . | . | C | C | . | C | . | A | G | T | T | G | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | | | | | | | |
| 6. Pongo | | . | . | C | . | . | . | C | C | . | C | . | A | . | . | . | . | C | . | C | . | . | . | C | G | . | . | . | . | . | . | . | . | . | . | . | . | | | | | | | | |
| 7. Hylobates | | . | . | C | . | . | T | C | A | . | T | . | A | . | . | G | T | C | . | C | . | . | . | . | . | C | . | . | C | . | . | . | . | . | . | . | C | . | . | | | | | | |
| 8. Macaca fuscata | | . | . | C | . | . | T | T | C | C | . | C | . | A | . | . | . | T | C | . | . | . | . | . | G | . | C | . | . | T | . | . | . | . | . | . | C | . | . | | | | | | |
| 9. M. mulatta | | . | . | C | . | . | T | T | C | . | C | . | A | . | . | . | . | T | C | . | C | . | . | G | . | . | . | . | T | . | . | . | . | . | . | . | . | C | . | . | | | | | |
| 10. M. fascicularis | | . | . | C | . | . | . | T | C | C | . | C | . | A | . | . | . | . | C | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | | | | |
| 11. M. sylvanus | | . | . | C | . | . | . | T | C | C | . | T | . | A | . | T | . | T | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | C | . | . | | | |
| 12. Saimiri sciureu | | . | . | C | . | . | . | . | C | C | . | C | . | A | . | T | G | . | T | C | . | . | . | A | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | T | . | . | C | . | . |

Highlighted nucleotides / total

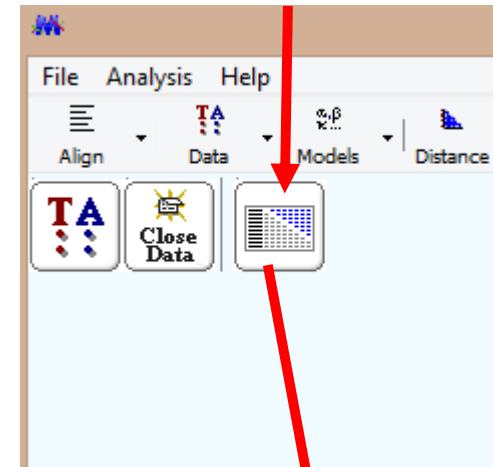
Making trees using MEGA6

Distance matrix:



Choose among different options in the yellow drop-menus

The distance matrix will appear among the open items



File Display Average Caption Help

(A,B) 0.0 0.00

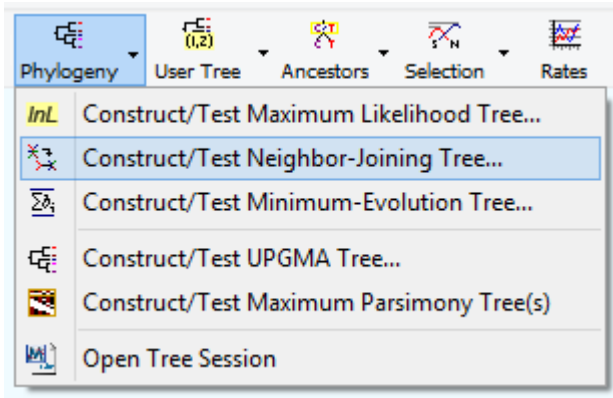
XL CSV MEGA TXT Caption

| | 1 | 2 | 3 | 4 | |
|---------------------|---------|---------|---------|---------|----|
| 1. Tarsius syrichta | | | | | |
| 2. Lemur catta | 225.000 | | | | |
| 3. Homo sapiens | 285.000 | 272.000 | | | |
| 4. Pan | 284.000 | 273.000 | 80.000 | | |
| 5. Gorilla | 279.000 | 260.000 | 92.000 | 94.000 | |
| 6. Pongo | 268.000 | 258.000 | 143.000 | 152.000 | 14 |
| 7. Hylobates | 272.000 | 260.000 | 161.000 | 167.000 | 16 |
| 8. Macaca fuscata | 277.000 | 249.000 | 205.000 | 215.000 | 20 |
| 9. M mulatta | 280.000 | 256.000 | 206.000 | 222.000 | 20 |
| 10. M fascicularis | 275.000 | 264.000 | 219.000 | 236.000 | 23 |

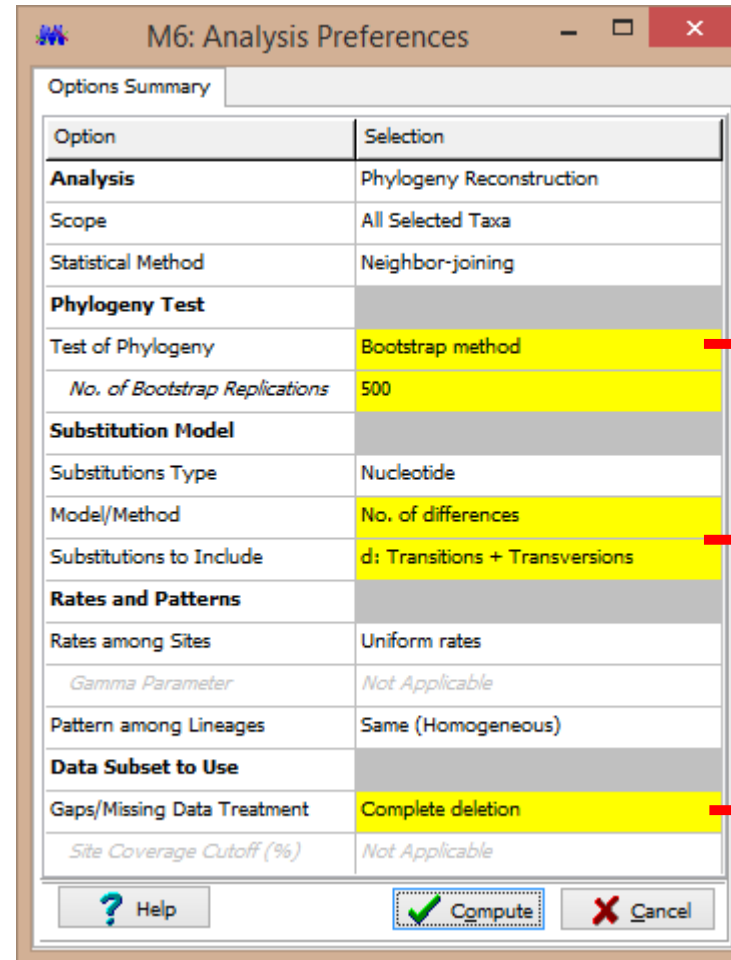
[1,1] (Tarsius syrichta-Tarsius syrichta) / Nucleotide: Nt

Making trees using MEGA6

Generation of a tree:



Pick the desired inference



Setup parameters in the drop menus:

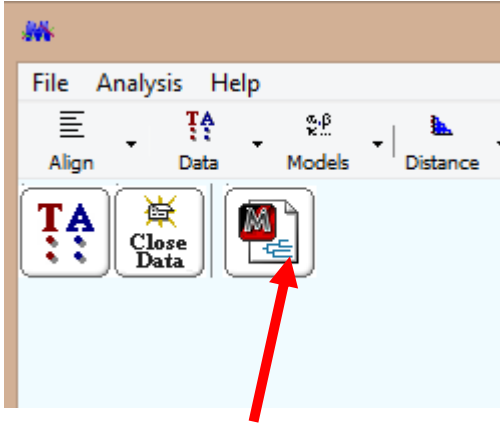
Support test?
Number of pseudoreplicates

How the changes are considered?

How gaps/missing data are considered?

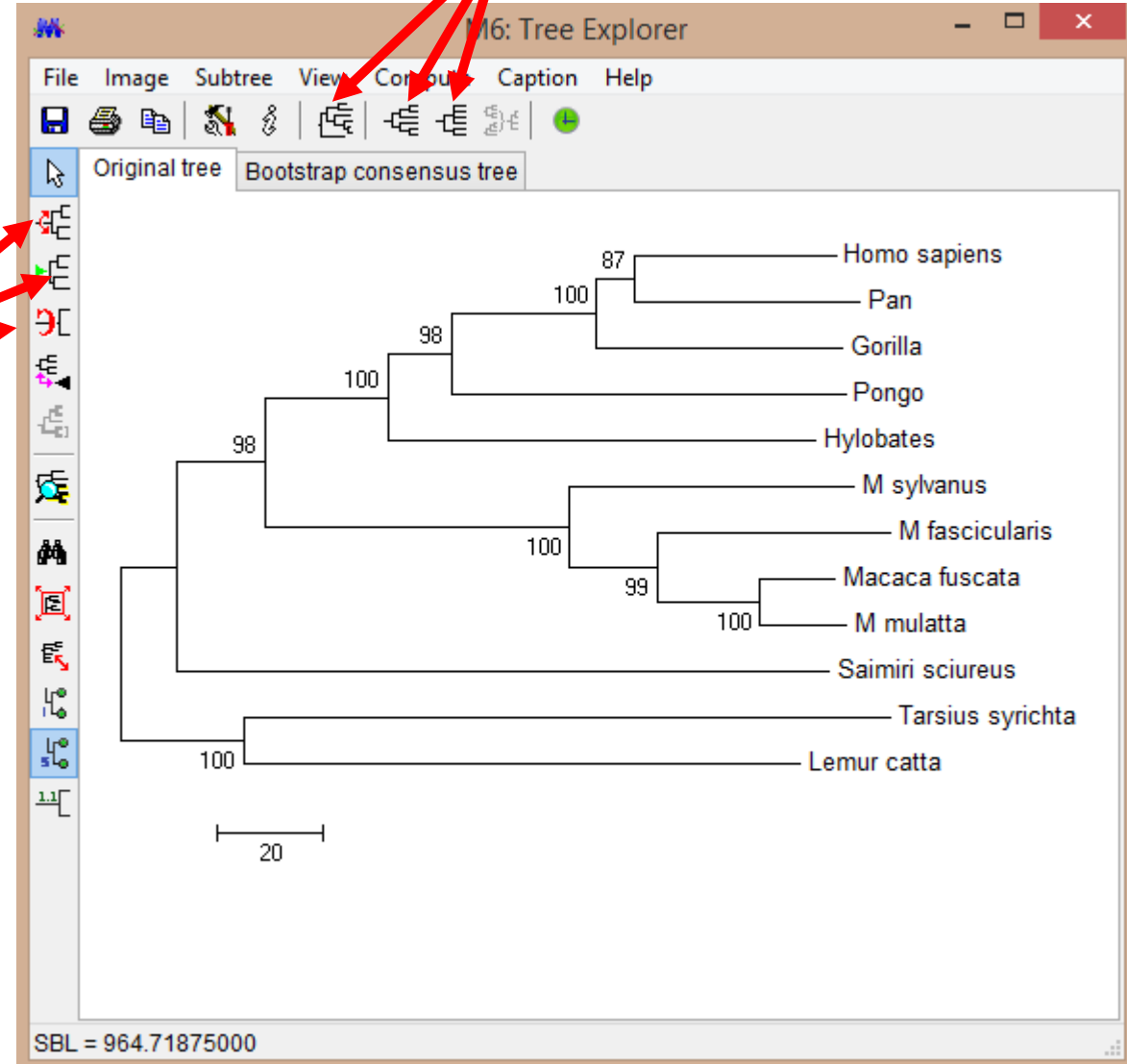
Making trees using MEGA6

Choose among different “format styles”



The new tree will appear among the open items

Change the root, flip the nodes etc



Tree explorer window

Making trees using MEGA6

Use the different tools to polish the displayed figure

The image displays the MEGA6 software interface. The main window, titled "M6: Tree Explorer", shows a phylogenetic tree with the following taxa: Homo sapiens, Pan, Gorilla, Pongo, Hylobates, M sylvanus, M fascicularis, Macaca fuscata, M mulatta, Saimiri sciureus, Tarsius syrich, and Lemur catta. A scale bar of 20 is visible at the bottom left of the tree. The "Tree Options" panel is open on the right, showing the "Tree" tab with "Rectangular Tree" selected. The panel includes a diagram illustrating the parameters: Taxon Separation, Branch Length, and Tree Width. The current settings are: Taxon Separation: 24 Pixels, Branch Length: 300 Pixels / 100, and Tree Width: 432 Pixels. A red arrow points from the "Tree" tab in the options panel to the tree in the main window.

M6: Tree Explorer

File Image Subtree View Compute Caption Help

Original tree Bootstrap consensus tree Timetree

Homo sapiens
Pan
Gorilla
Pongo
Hylobates
M sylvanus
M fascicularis
Macaca fuscata
M mulatta
Saimiri sciureus
Tarsius syrich
Lemur catta

20

SBL = 964.71875000

M6: Tree Options

Tree Branch Labels Scale Cutoff

Rectangular Tree Circle Tree Radiati

Taxon Separation
Branch Length
Tree Width

Taxon Separation 24 Pixels
Branch Length 300 Pixels / 100
Tree Width 432 Pixels

Help OK Cancel

Making trees using MEGA6

Then export the Newick tree, or print as pdf

The screenshot displays the MEGA6 software interface. The main window, titled "M6: Tree Explorer", shows a phylogenetic tree with the following taxa and bootstrap values:

- Homo sapiens (37.703)
- Pan (42.297)
- Gorilla (45.969)
- Pongo (73.563)
- Hylobates (79.844)
- Saimiri sciureus (56.925)
- Tarsius syrichta (120.946)
- Lemur catta (104.054)

A red bracket on the right side of the tree groups Homo sapiens, Pan, and Gorilla under the label "Hominidae". A green box highlights the branch leading to the Hominidae clade, with a bootstrap value of 27.094. A scale bar at the bottom left indicates a distance of 20. The status bar at the bottom shows "SBL = 964.71875000".

The "File" menu is open, showing the following options:

- Save Current Session (Ctrl+S)
- Export Current Tree (Newick)
- Export Current Tree (Timetree)
- Export All Trees (Newick)
- Write Tree in a Table format
- Export Group Names
- Import Group Names
- Show Information (Ctrl+I)
- Print
- Print in a sheet
- Printer setup...
- Exit Tree Explorer (Ctrl+Q)

The background window shows the "File" menu of another MEGA6 instance, with the "Export Current Tree (Newick)" option highlighted.