

Filogenómica molecular

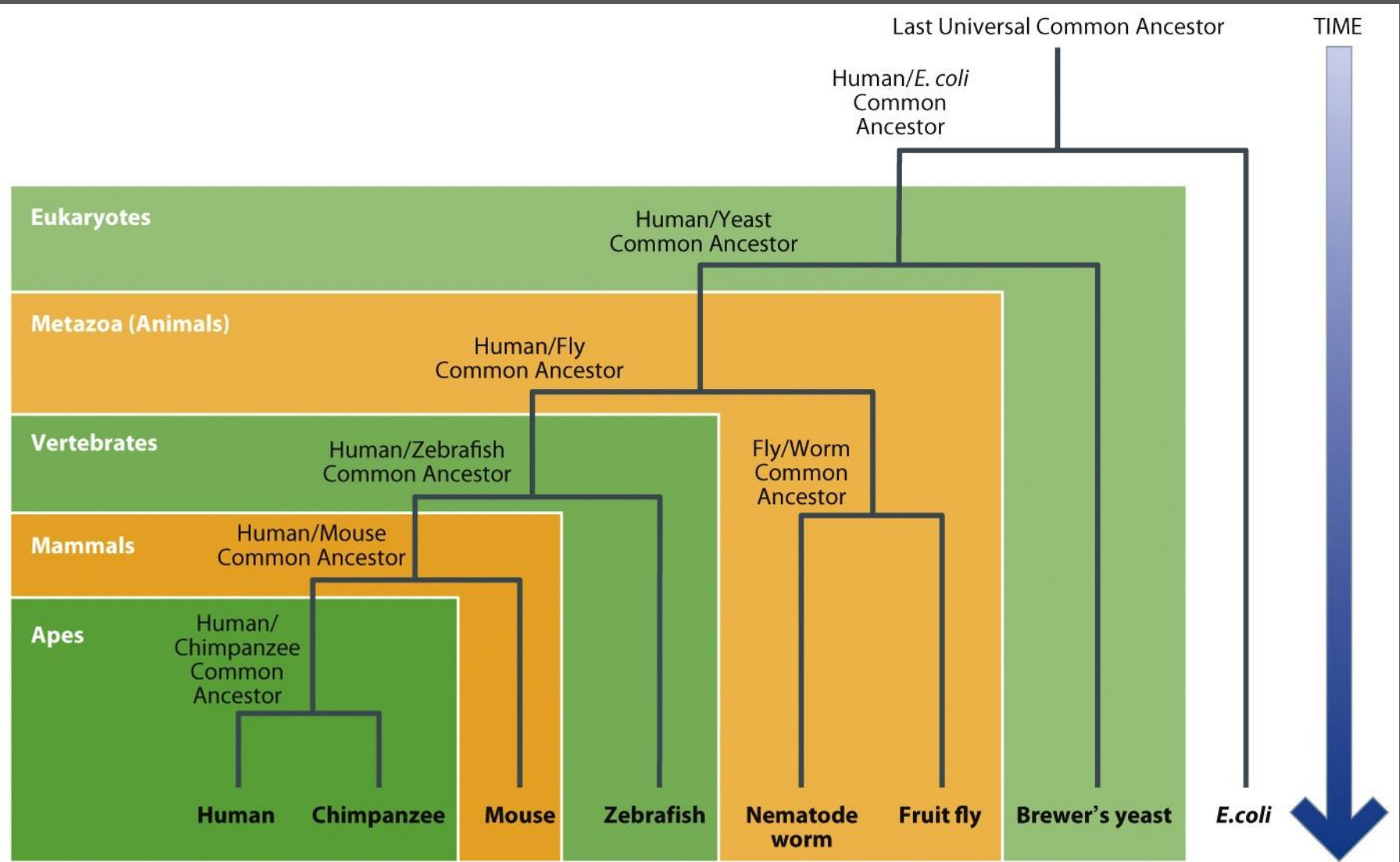
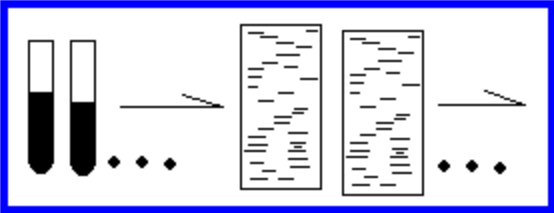
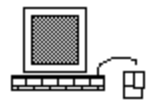


Figure 1.2 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)



ACTGCGTAGACAGTGGTTACAACCACCACTGA
 ACTGGCTAGACGATGGTTACAAGCACCAGTGA
 ACTGGTAGACGATCCTATCAAGTTTTCCATGTA
 ...

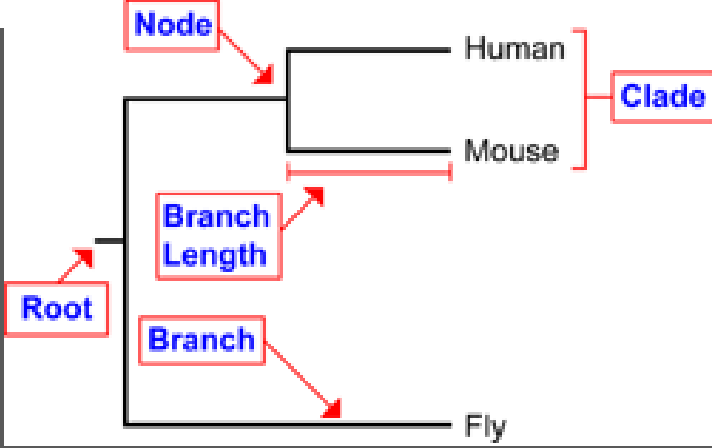
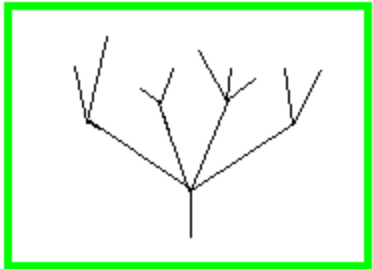


ACTGCGTAGACAGTGGTTACAACCACCACTGA
 ACTGGCTAGACGATGGTTACAAGCACCAGTGA
 ACTGG-TAGACGATCCTATCAAGTTTTCCATGTA



ACTGCGTAGACAGTGGT-TACAAC---CACCACTGA
 ACTGGCTAGACGATGGT-TACAAG---CACCAGTGA
 ACTGG-TAGACGATCCTAT-CAAGTTT--CCATGTA

80% 90% 75%...
 88% 66% 77%
 90% 78% ..
 76%..
 ...



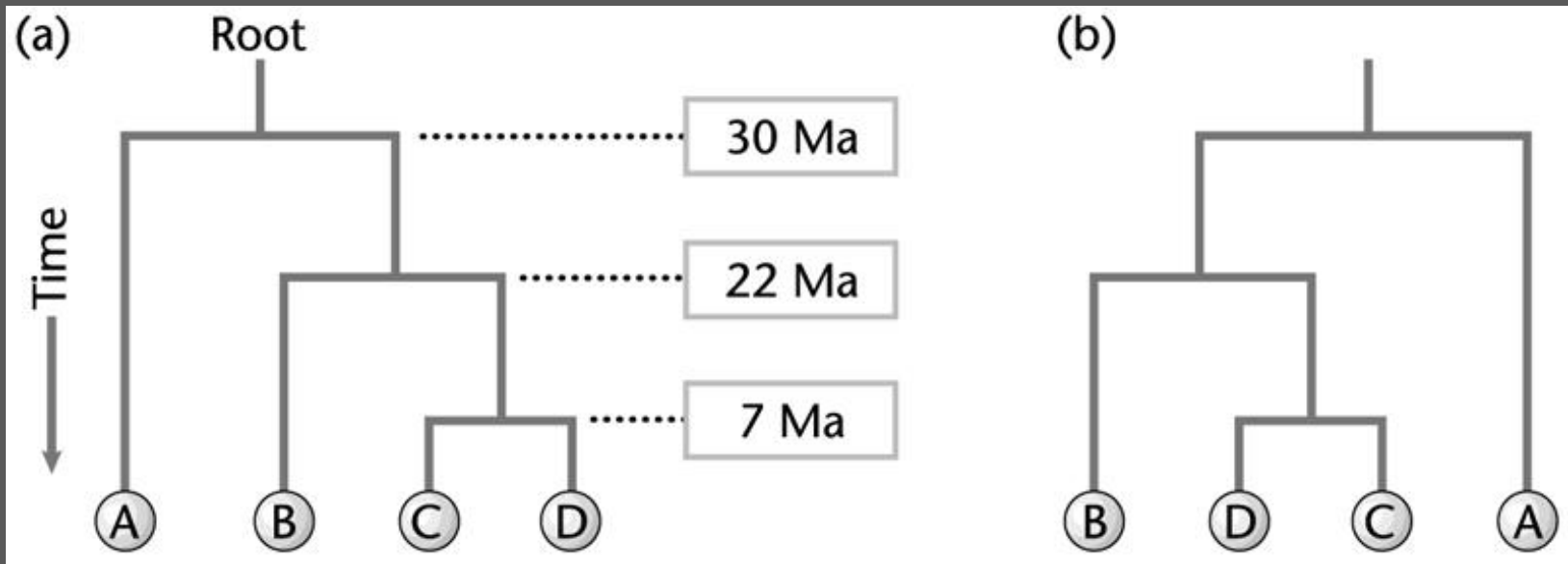


Fig. 8.1 Rooted trees with a time axis. Tree (a) can be converted to tree (b) by swinging around the horizontal branches like mobiles. Hence (a) and (b) are equivalent to one another.

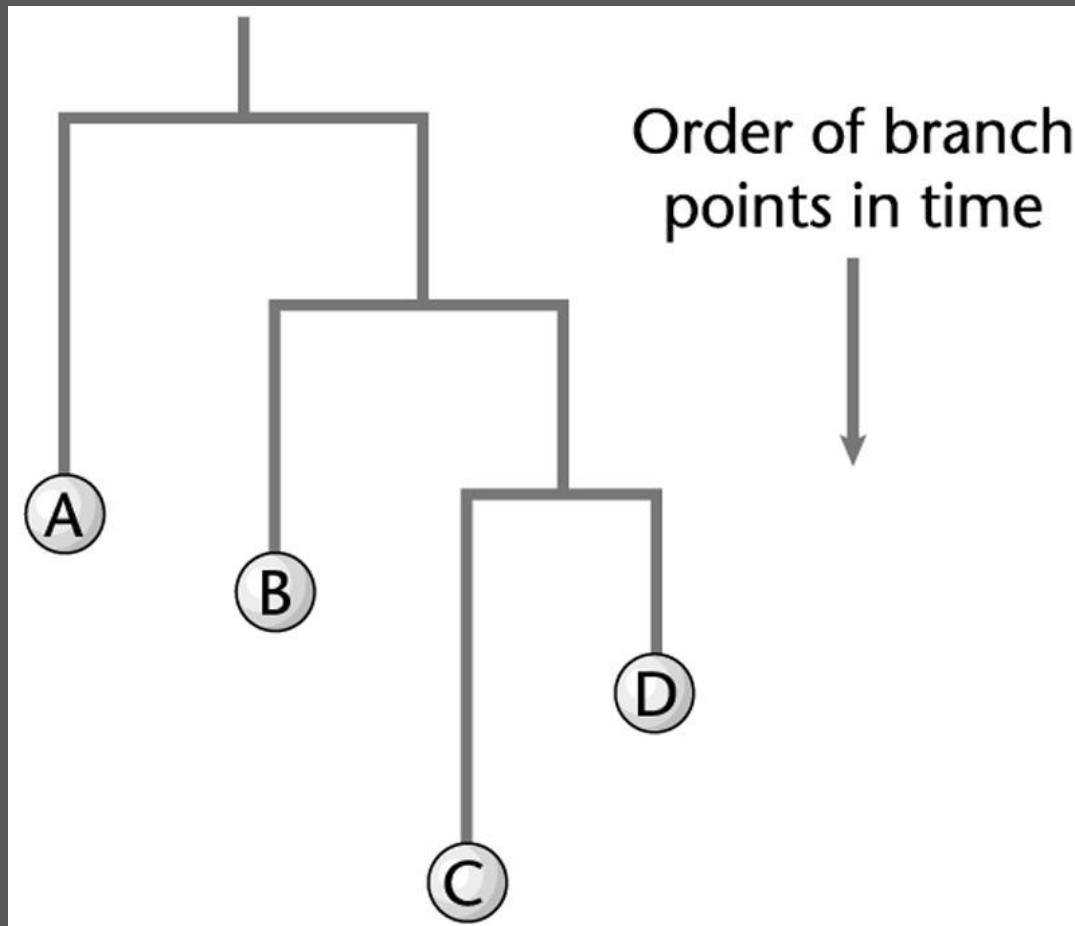


Fig. 8.2 A rooted tree with branches scaled according to the amount of evolutionary change.

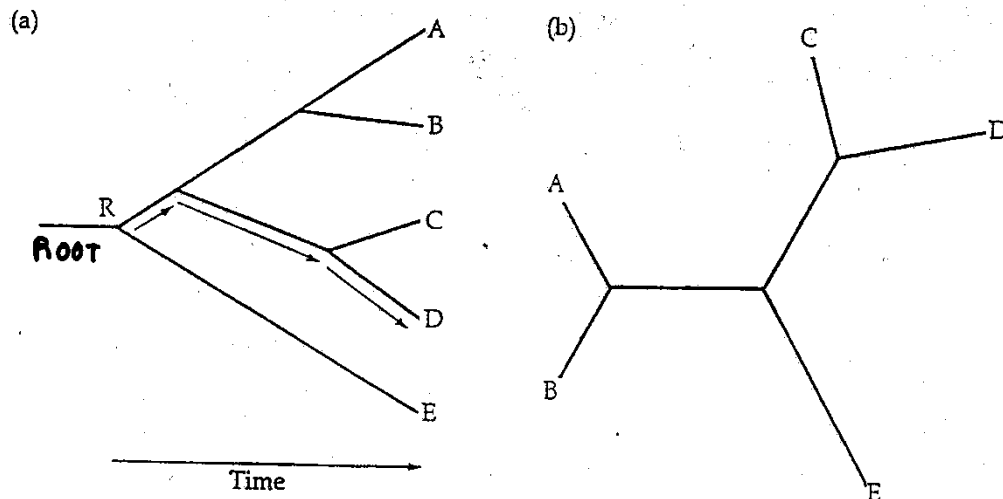


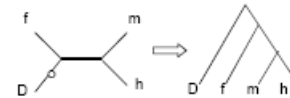
Figure 2. (a) Rooted and (b) unrooted phylogenetic trees. Arrows indicate the unique path leading from the root (R) to OTU D.

Phylogenetic / Molecular Phylogeny

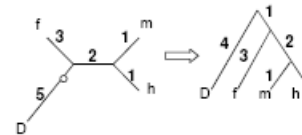
How to root a tree

How to root a tree

- Outgroup – place root between distant sequence and rest group



- Midpoint – place root at midpoint of longest path (sum of branches between any two OTUs)



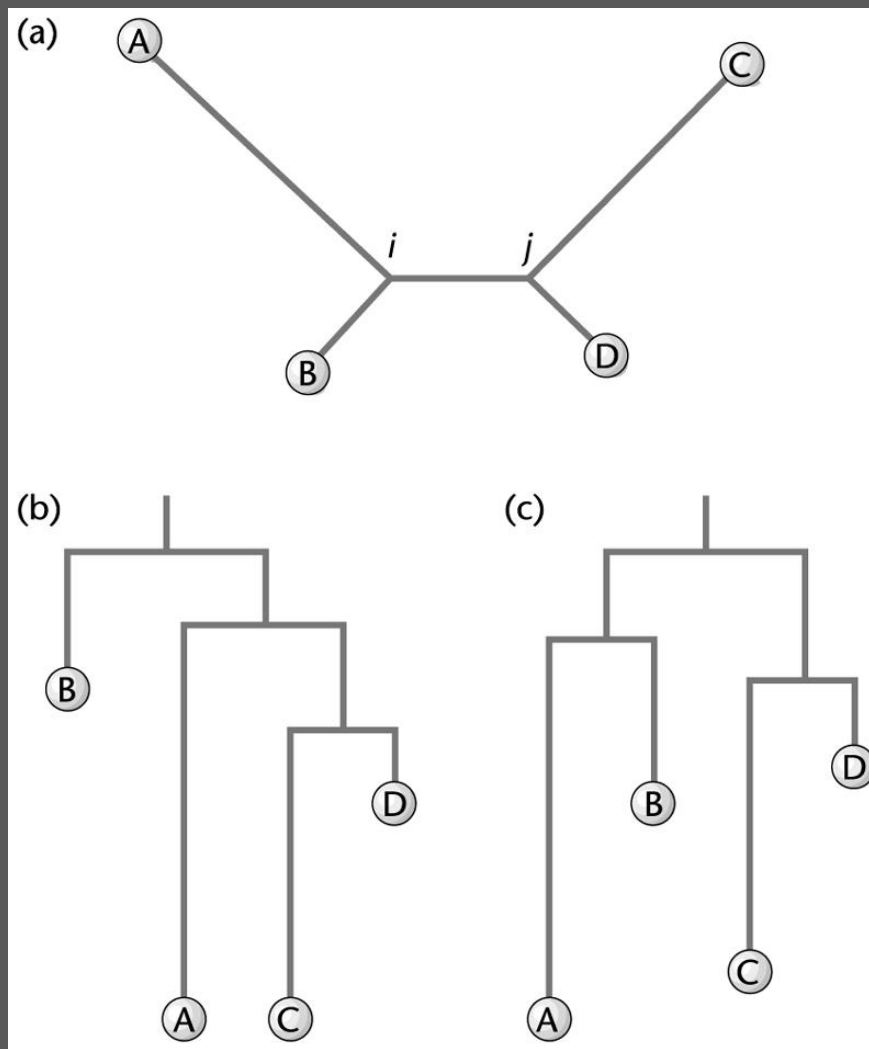


Fig. 8.3 The unrooted tree in (a) can be converted to the rooted trees in (b) and (c) by placing the root in different positions.

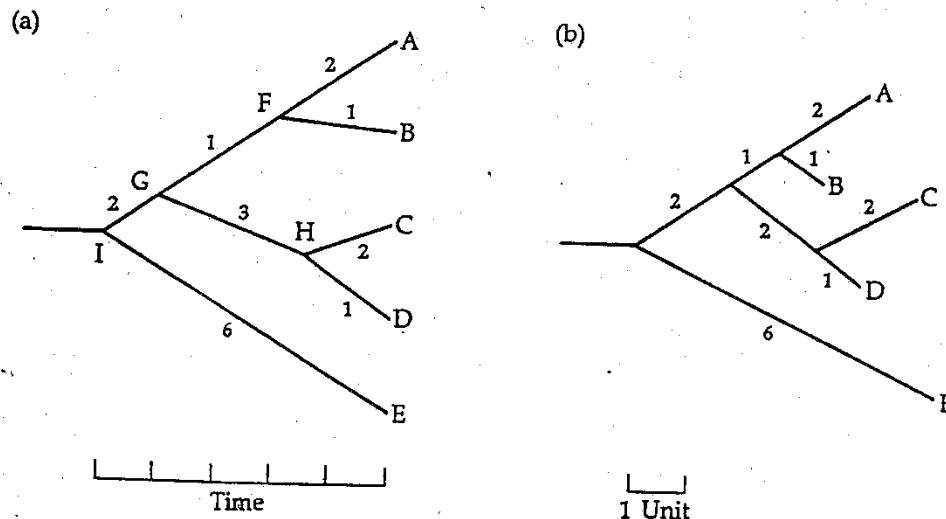
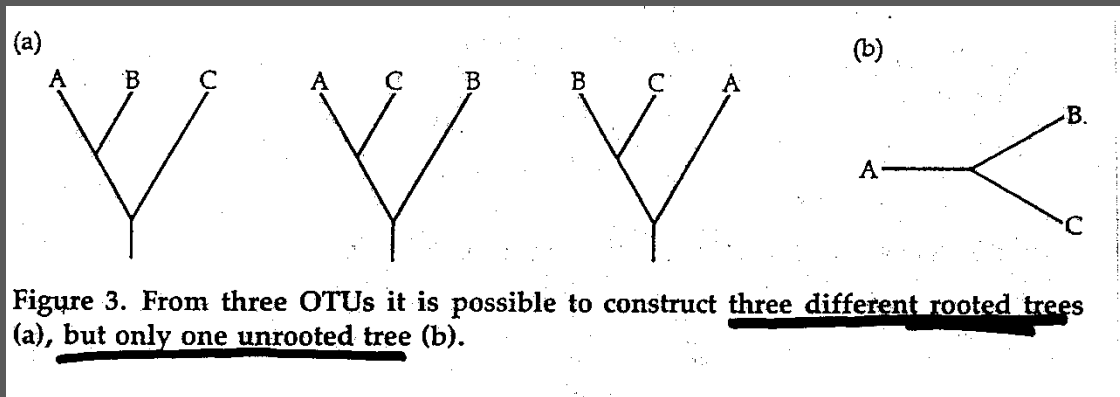


Figure 1. Two alternative representations of a phylogenetic tree for five OTUs. (a) Unscaled branches: extant OTUs are lined up and nodes are positioned proportionally to times of divergence. (b) Scaled branches: lengths of branches are proportional to the numbers of molecular changes.

OTU: Operational taxonomic unit



Para n OTUs ($n \geq 3$):

$$N_U = (2n-5)! / [2^{n-3} (n-3)!]$$

$$N_R = (2n-3)! / [2^{n-2} (n-2)!]$$

Combinatoric explosion

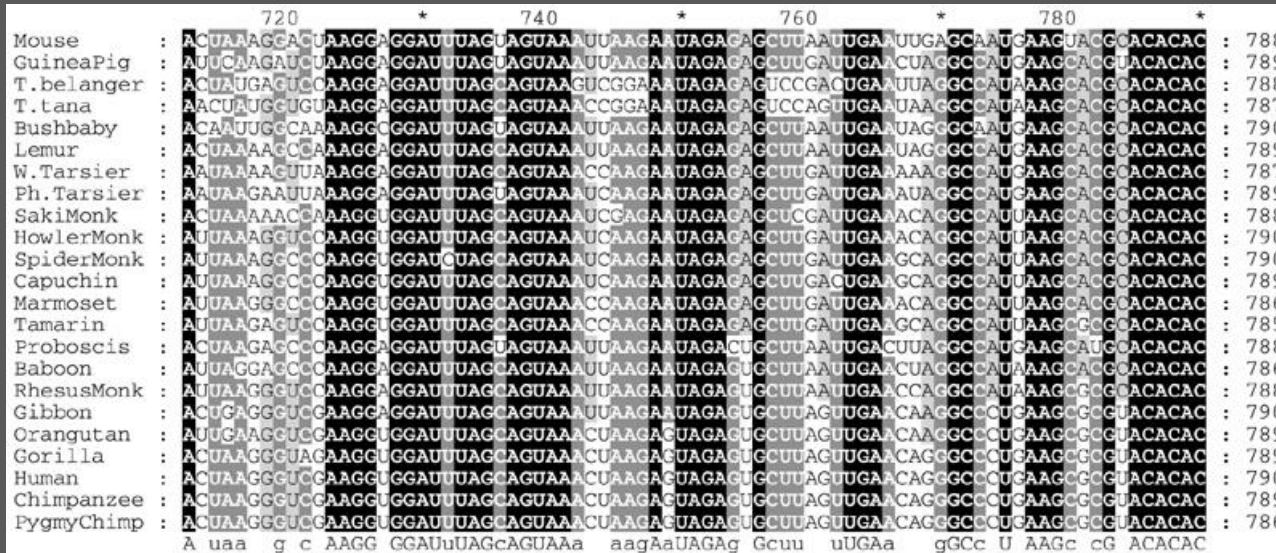
Combinatoric explosion

# sequences	# unrooted trees	# rooted trees
2	1	1
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10,395
8	10,395	135,135
9	135,135	2,027,025
10	2,027,025	34,459,425

- Los árboles filogenéticos deben contener solo **bifurcaciones**
- Las trifurcaciones o multifurcaciones se deben generalmente a falta de resolución del método → deben resolverse con más datos

Selección de las secuencias de partida:

- Ortólogas: ancestro común
- Que estén presentes en todas las especies del grupo de interés
- Ritmo evolutivo adecuado: ni muy alto ni muy bajo



Se descartan los gaps

Fig. 8.4 Part of the alignment of the mitochondrial small subunit rRNA gene from primates, tree shrews, and rodents.

Lactate dehydrogenase multiple alignment

```

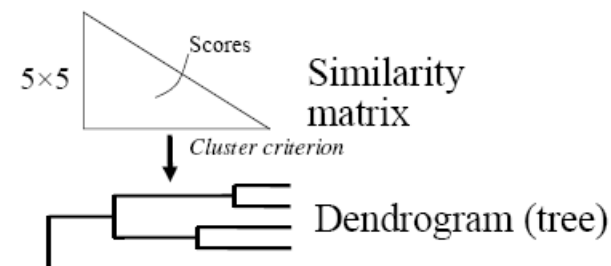
Human      -KIIVVGVGAVGMRCAISILMKDLADELALVDVIEDKLRGEMMDLQHGSLFLRTPKIVSGKDYVNTANSKIVIIITAGARQ
chicken    -KISVGVGVGAVGMRCAISILMKDLADELALVDVIEDKLRGEMMDLQHGSLFLRTPKIVSGKDYVNTANSKIVIIITAGARQ
Dogfish    -KIIVVGVGAVGMRCAISILMKDLADELALVDVIEDKLRGEMMDLQHGSLFLRHTAKIVSGKDYVNTANSKIVIIITAGARQ
Lamprey    SKVITVGVGQVGMRAAISVLLRDLADELALVDVVEDRLRGEMMDLHGSLFLKHTAKIVADKDYVNTANSKIVIIITAGARQ
Barley     TKISVIGAQNVGMR IAQTILTONLADEIALVDA LFPDKLRGEALDLQRAAFLFRVRI-SGTDAAVTKNSDIVIVTAGARQ
Maizey casei -KVILVGDGAVGSSYAYAMVLQGI AQEIGIVDFKDKTRGDAIDLQNALFFTSFKKIYSA-EYSDAKDADLVVITAGAPQ
Bacillus   TKVSVIGAQNVGMR IAQTILTRDLADEIALVDAVFDKLRGEMMDLQRAAFLFRTRILVSGTDMSEVIRGSDIVIVTAGARQ
Lacto__ste -RIVVIGAQFVGA SYVFA LNMNQGIADEIVLIDANESKAIGDAMD FNHGKVFAPKPVDIWHGQYDDCRDADLVVICAGANQ
Lacto_plant QKVVLVGDGAVGSSYAFAMAQQGIAEEFVIVDVVKDRTRGDALDLEDAQAFIAPKKIYSG-EYSDCKDADLVVITAGAPQ
Therma_mari MKIGIVGLGRVGSSTAFALLMKGFAREN/LIIVDKKRAEGDALDLHGCTPFFRRANIIYAG-DYADLKGSDVVIVVAGVQPQ
Bifido     -KLAIVIGAQAVGSLTAAFAAQRGIAREIVLEDI AKERV EAEVLDLMQHGSSFYFTVSDIGSDDFEICRDADMVIVITAGFRQ
Thermus_aqua SKVGVGSSG VGSATATAVLDVGGVAREVVLDLDRRLQAQRAEDILRRITPFAKFWVVRSGWF-YEDLEGARVVIVAGVAVQ
Mycoplasma -KIALIGAQNVGNSFLVAAAMNQLASEYGIIDINPFDADGNAFDPEQASASLFFPIVSRVRYEKDLKDADEIVITAGRPQ

```

Distance Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Human	0.000	0.112	0.128	0.202	0.378	0.346	0.530	0.551	0.512	0.524	0.528	0.635	0.637
2 Chicken	0.112	0.000	0.155	0.214	0.382	0.348	0.538	0.569	0.516	0.524	0.524	0.631	0.651
3 Dogfish	0.128	0.155	0.000	0.196	0.389	0.337	0.522	0.567	0.516	0.512	0.524	0.600	0.655
4 Lamprey	0.202	0.214	0.196	0.000	0.426	0.356	0.553	0.589	0.544	0.503	0.544	0.616	0.669
5 Barley	0.378	0.382	0.389	0.426	0.000	0.171	0.536	0.565	0.526	0.547	0.516	0.629	0.575
6 Maizey	0.346	0.348	0.337	0.356	0.171	0.000	0.557	0.563	0.538	0.555	0.518	0.643	0.587
7 Lacto_casei	0.530	0.538	0.522	0.553	0.536	0.557	0.000	0.518	0.208	0.445	0.561	0.526	0.501
8 Bacillus_ste	0.551	0.569	0.567	0.589	0.565	0.563	0.518	0.000	0.477	0.536	0.536	0.598	0.495
9 Lacto_plant	0.512	0.516	0.516	0.544	0.526	0.538	0.208	0.477	0.000	0.433	0.489	0.563	0.485
10 Therma_mari	0.524	0.524	0.512	0.503	0.547	0.555	0.445	0.536	0.433	0.000	0.532	0.405	0.598
11 Bifido	0.528	0.524	0.524	0.544	0.516	0.518	0.561	0.536	0.489	0.532	0.000	0.604	0.614
12 Thermus_aqua	0.635	0.631	0.600	0.616	0.629	0.643	0.526	0.598	0.563	0.405	0.604	0.000	0.641
13 Mycoplasma	0.637	0.651	0.655	0.669	0.575	0.587	0.501	0.495	0.485	0.598	0.614	0.641	0.000

Cluster analysis – Clustering criteria



Four different clustering criteria:

- Single linkage - Nearest neighbour
- Complete linkage – Furthest neighbour
- Group averaging – UPGMA
- Neighbour joining (global measure)

Note: these are all *agglomerative* cluster techniques; i.e. they proceed by merging clusters as opposed to techniques that are *divisive* and proceed by cutting clusters

ClustalW2 Results

[Alignments](#) [Result Summary](#) [Guide Tree](#) [Submission Details](#) [Submit Another Job](#)

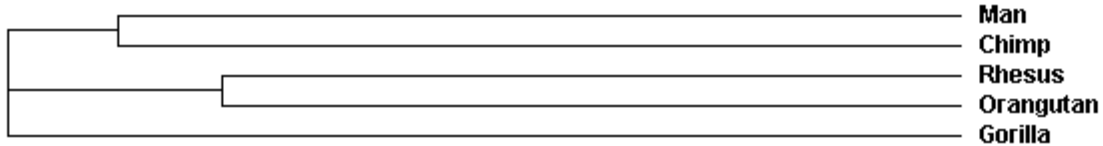
Guide Tree

[View Guide Tree File](#)

```
(  
(  
Man:0.04965,  
Chimp:0.05478)  
:0.01857,  
(  
Rhesus:0.11948,  
Orangutan:0.10837)  
:0.03584,  
Gorilla:0.06508);
```

Cladogram

[Show as Phylogram Tree](#) [Show Distances](#)



ClustalW2 Results

[Alignments](#) [Result Summary](#) [Guide Tree](#) [Submission Details](#) [Submit Another Job](#)

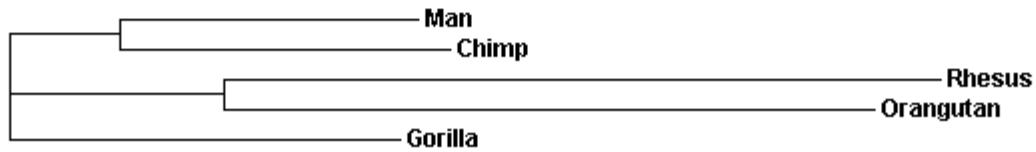
Guide Tree


[View Guide Tree File](#)

```
(  
(  
Man:0.04965,  
Chimp:0.05478)  
:0.01857,  
(  
Rhesus:0.11948,  
Orangutan:0.10837)  
:0.03584,  
Gorilla:0.06508);
```

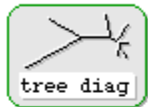
Phylogram

[Show as Cladogram Tree](#) [Show Distances](#)



Phylodendron  Phylogenetic tree printer

Tree styles



Tree data (newick / nh format)

Upload tree file: or paste data or URL in box below

Title:

See sample [data 1](#) and [data 2](#)

Guide Tree

Show as Cladogram Tree

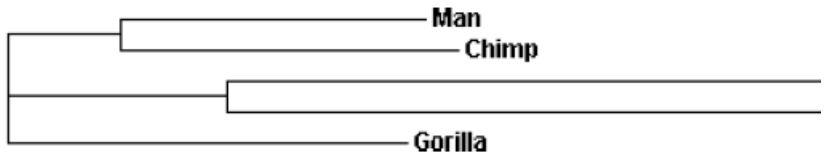
Show Distances

View DND File

```
(  
(  
Man:0.04965,  
Chimp:0.05478)  
:0.01857,  
(  
Rhesus:0.11948,  
Orangutan:0.10837)  
:0.03584,  
Gorilla:0.06508);
```

Formato Newick

Phylogram

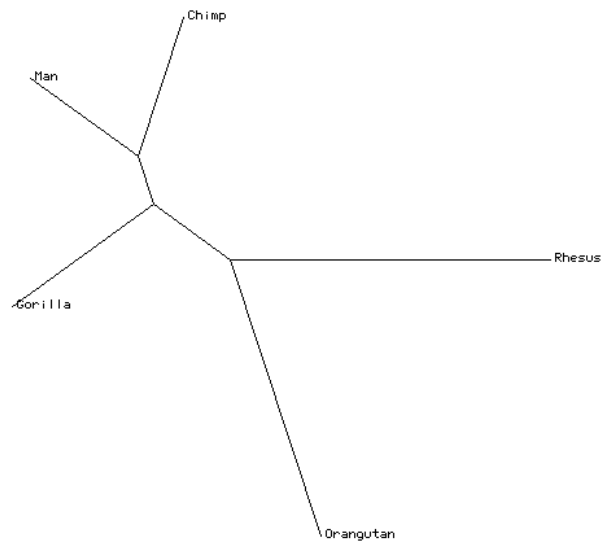


Show as Cladogram Tree

Show Distances

View DND File

Phylogenetic tree



0.1

Filogenia de una amplia variedad de organismos. Se muestra el número de proyectos genómicos acabados que se ha usado para cada grupo de taxones.

Los círculos y óvalos blancos indican proyectos cromosómicos.

Los círculos y óvalos verdes indican proyectos de orgánulos (mitocondrias y cloroplastos).

Fuente: Phylogenomics, a primer, Fig. 7.4

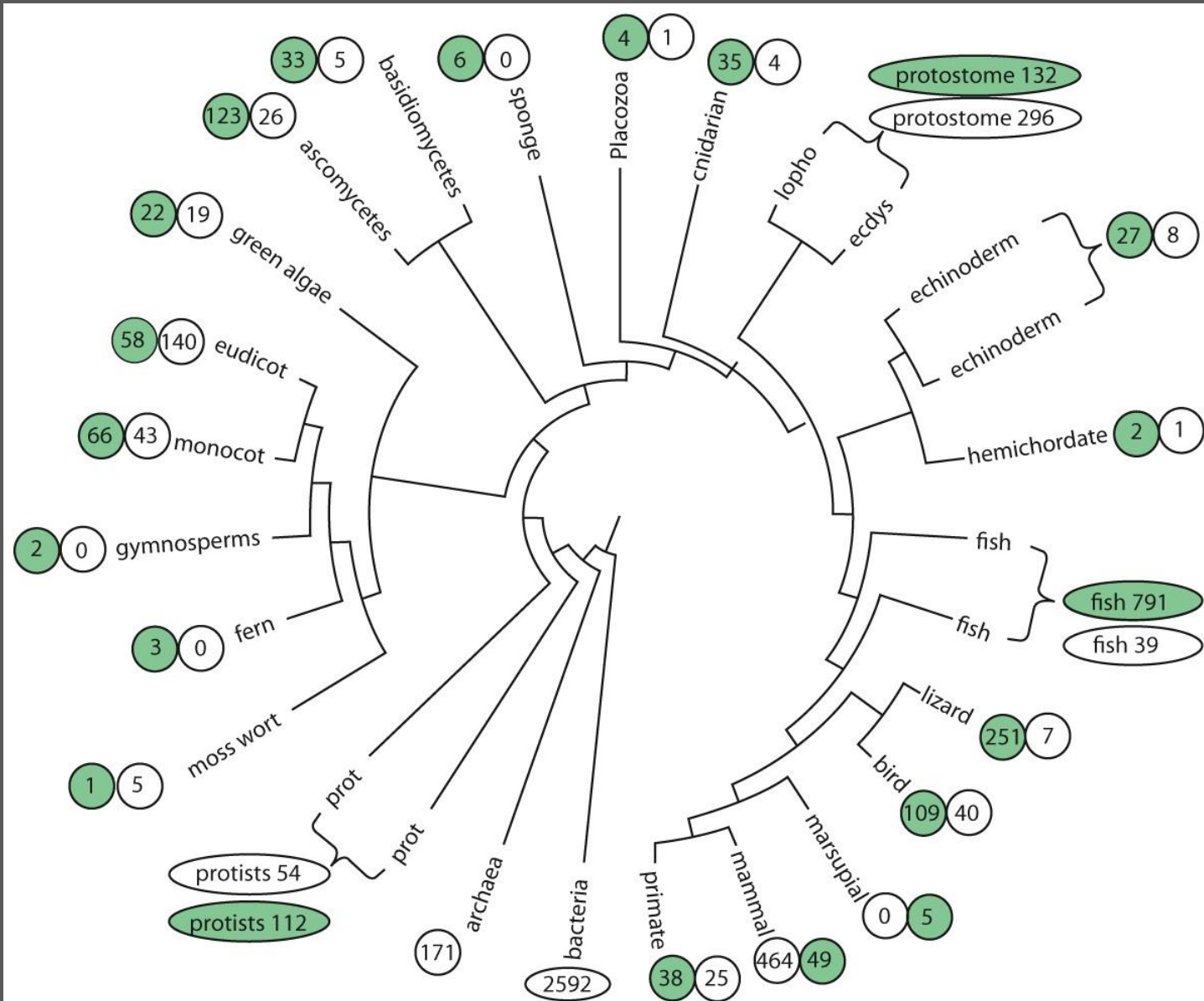


Figure 7.4 Phylogenomics: A Primer (© Garland Science 2013)