Filogenia molecular

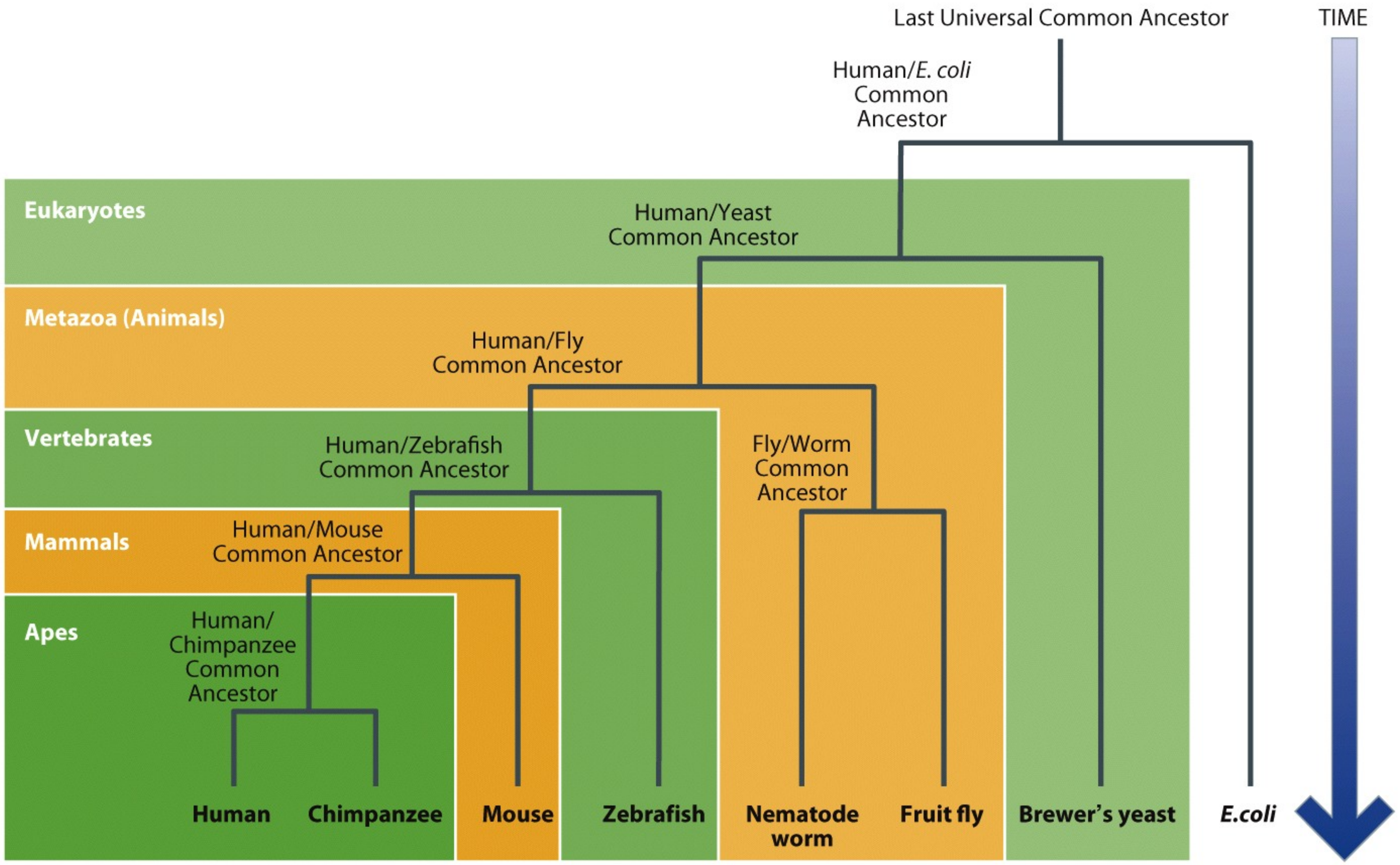
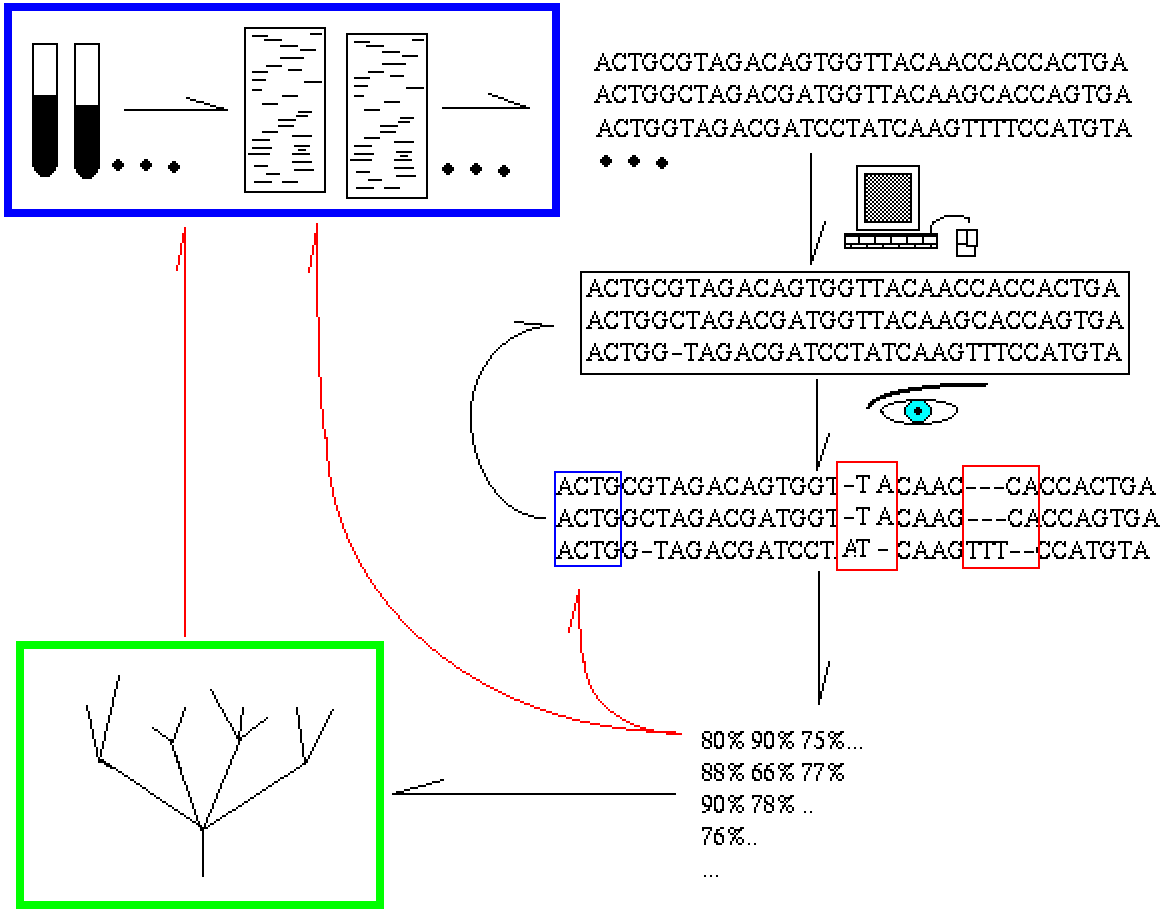


Figure 1.2 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

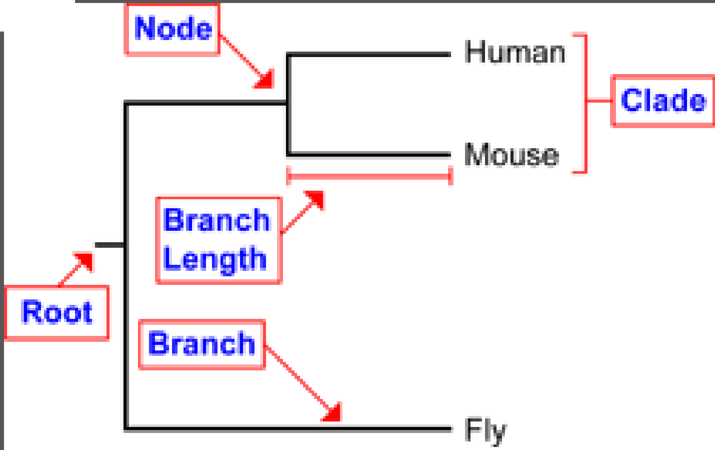


ACTGCGTAGACAGTGGTTACAACCACCACTGA
ACTGGCTAGACGATGGTTACAAGCACCAGTGA
ACTGGTAGACGATCCTATCAAGTTTTCCATGTA
• • •

ACTGCGTAGACAGTGGTTACAACCACCACTGA
ACTGGCTAGACGATGGTTACAAGCACCAGTGA
ACTGG-TAGACGATCCTATCAAGTTTCCATGTA

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

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



Cambio gradual vs. Equilibrio puntuado

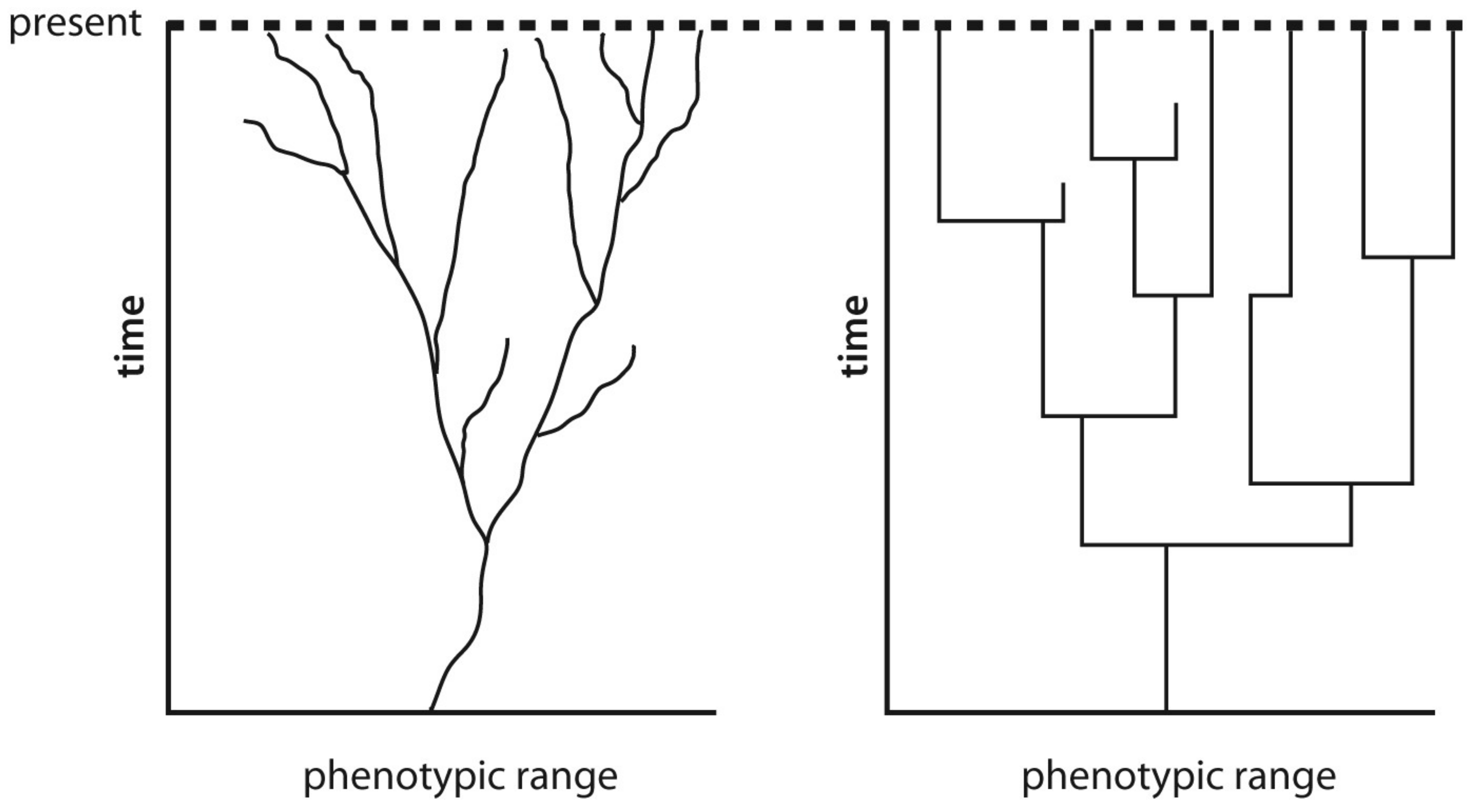


Figure 3.11 Phylogenomics: A Primer (© Garland Science 2013)

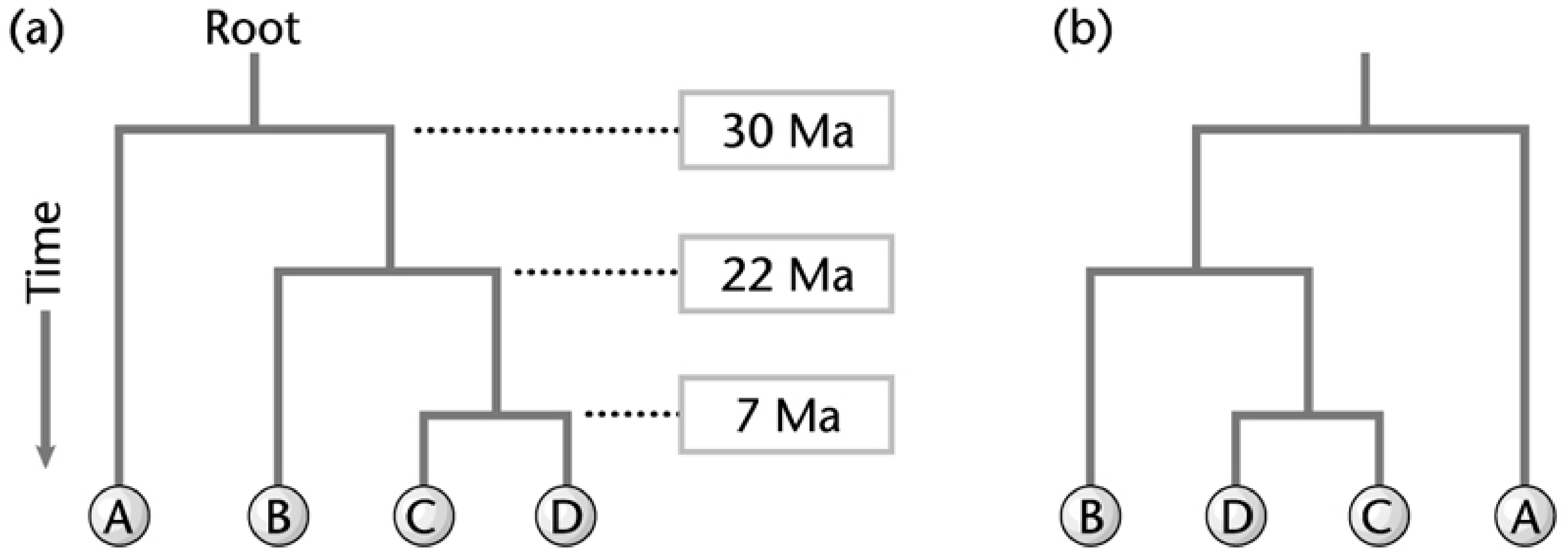


Fig. 8.1 Rooted trees with a time axis. Tree (a) can be converted to tree (b) by swinging around the horizontal branches like mobiles. Hence (a) and (b) are equivalent to one another.

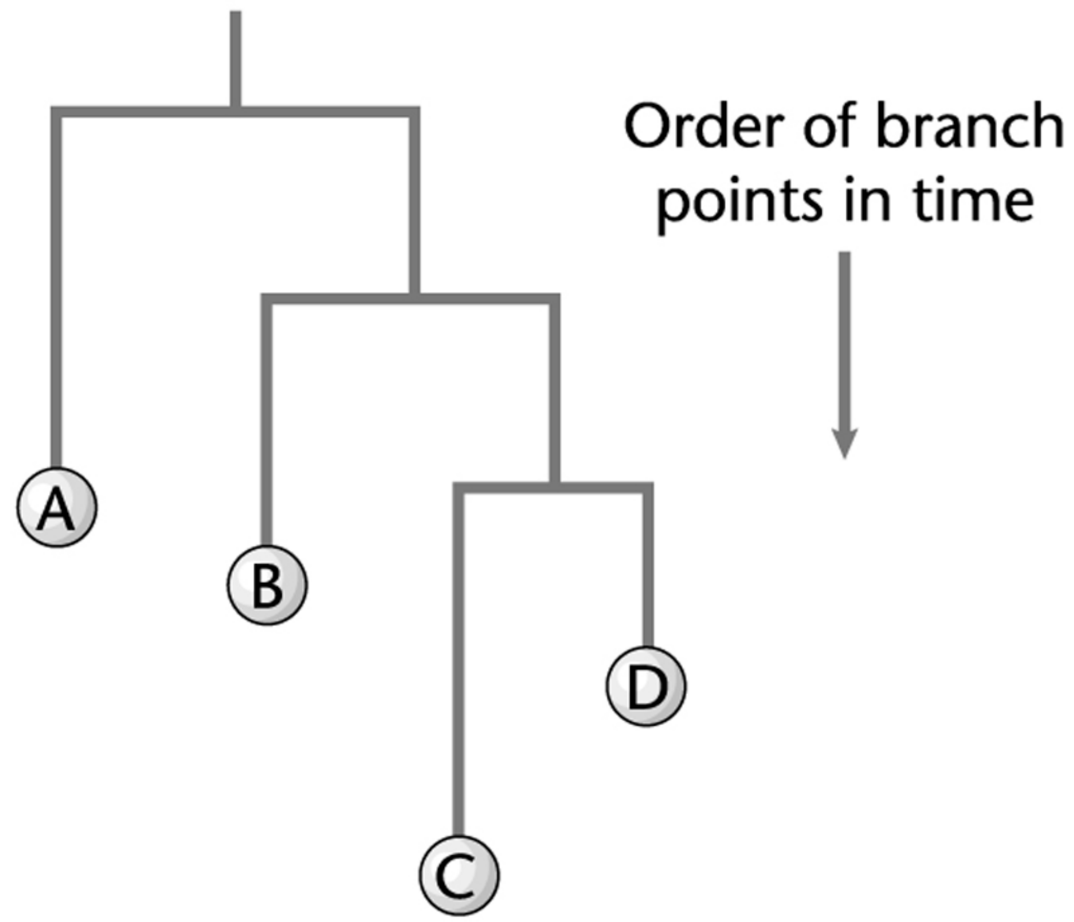


Fig. 8.2 A rooted tree with branches scaled according to the amount of evolutionary change.

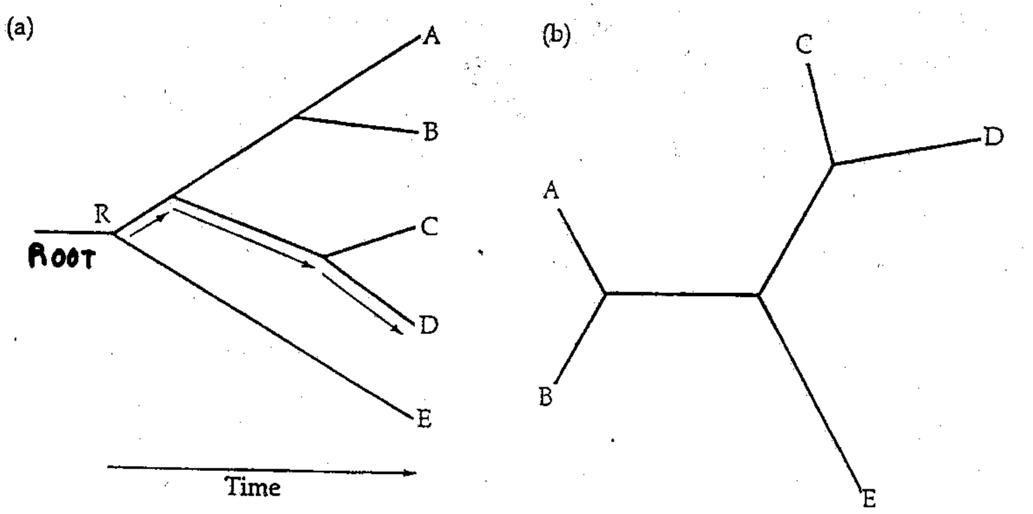
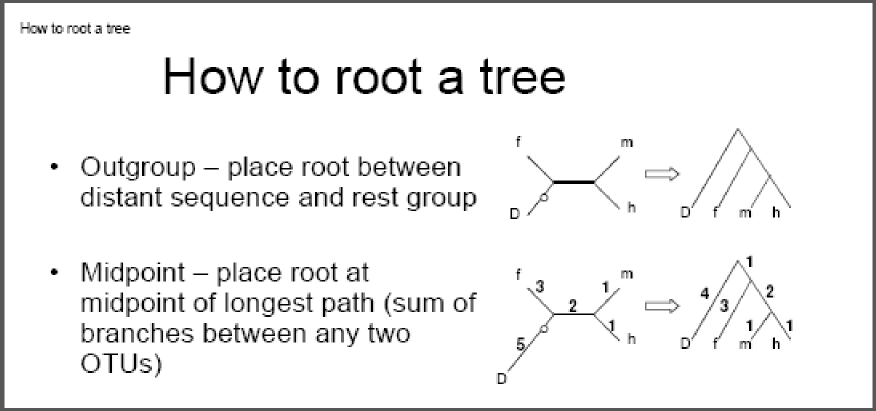


Figure 2. (a) Rooted and (b) unrooted phylogenetic trees. Arrows indicate the unique path leading from the root (R) to OTU D.

e / 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)

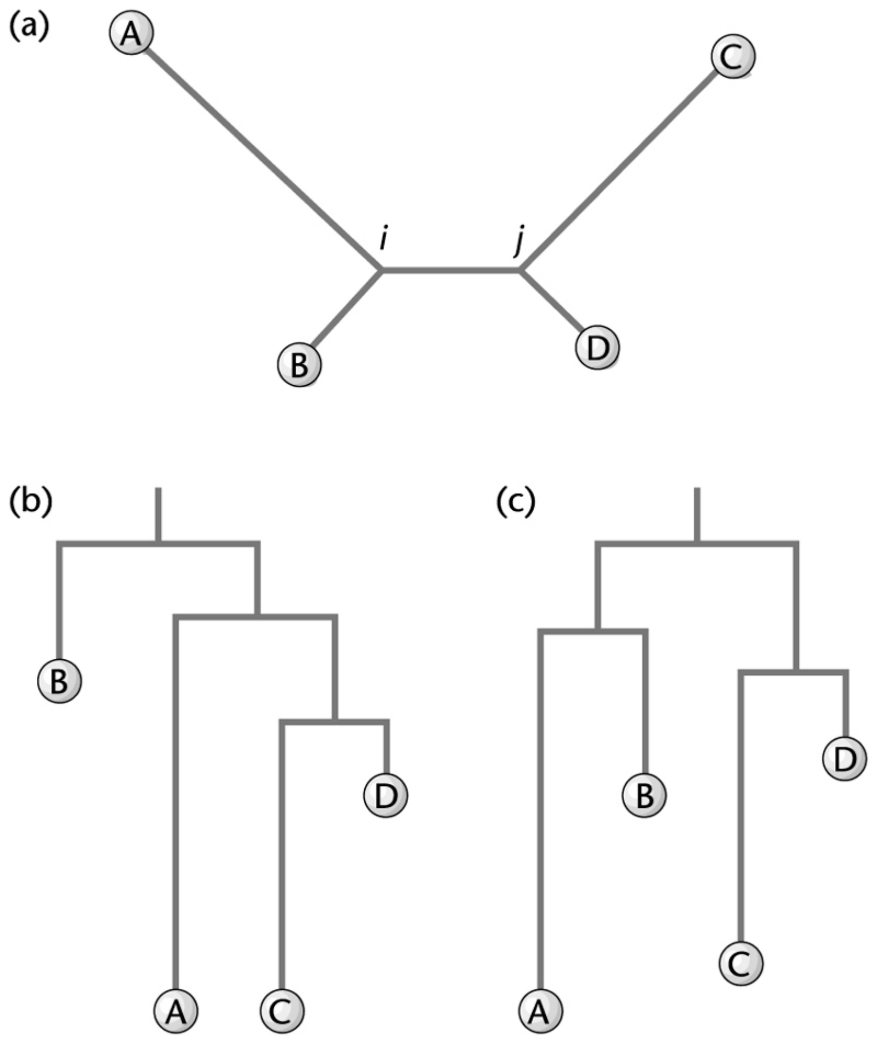


Fig. 8.3 The unrooted tree in (a) can be converted to the rooted trees in (b) and (c) by placing the root in different positions.

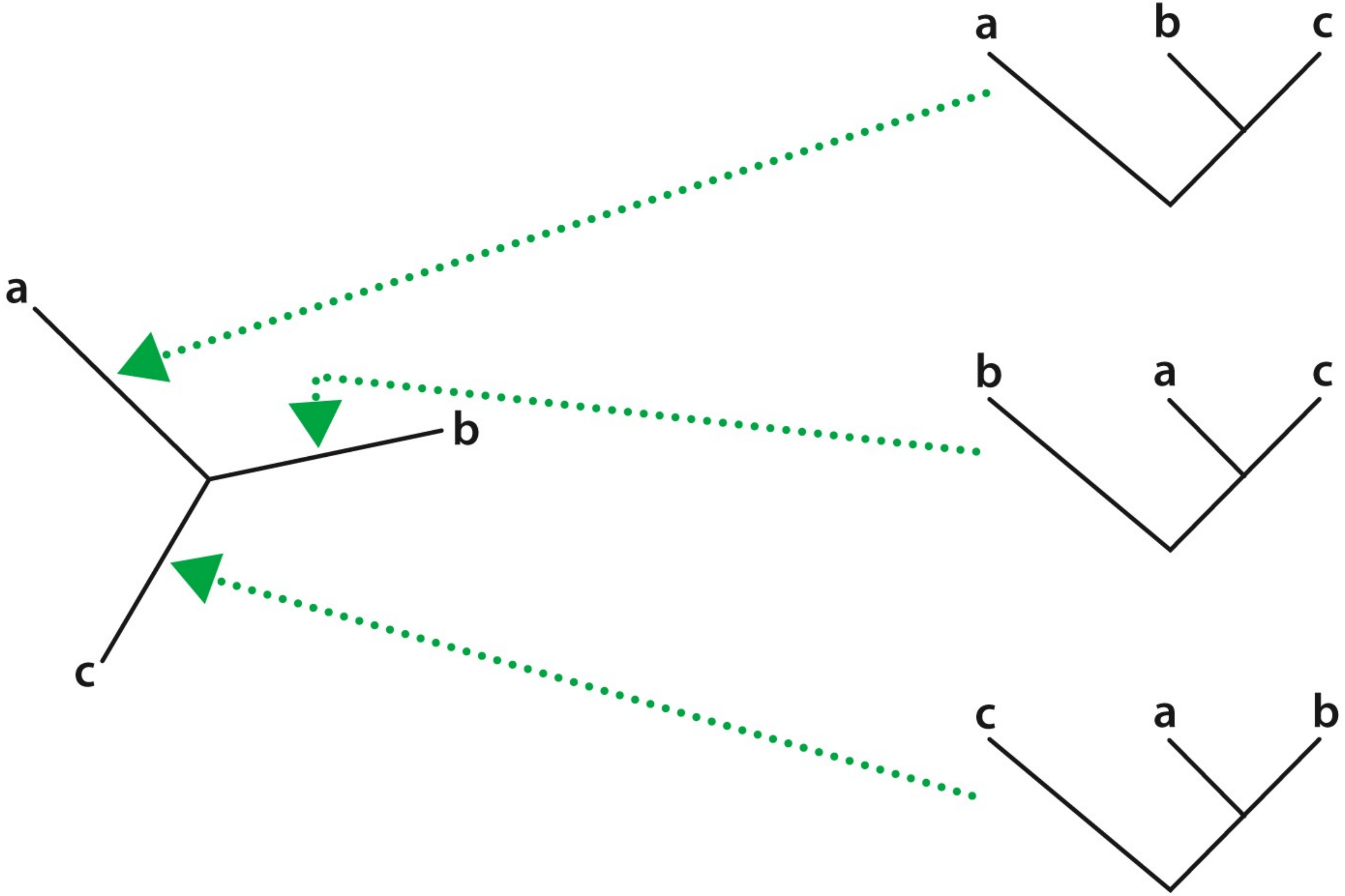


Figure 8.2 Phylogenomics: A Primer (© Garland Science 2013)

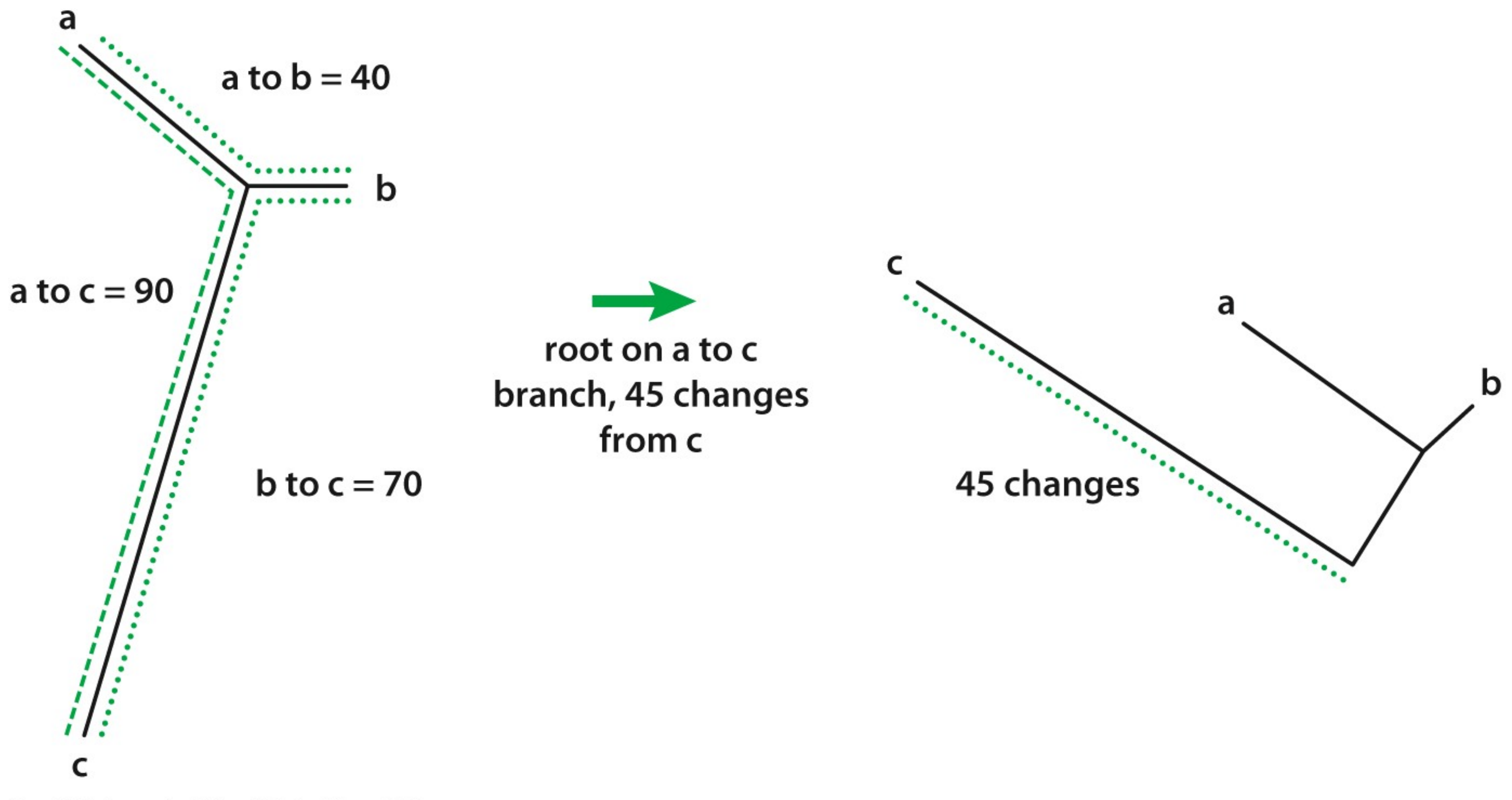


Figure 8.3 Phylogenomics: A Primer (© Garland Science 2013)

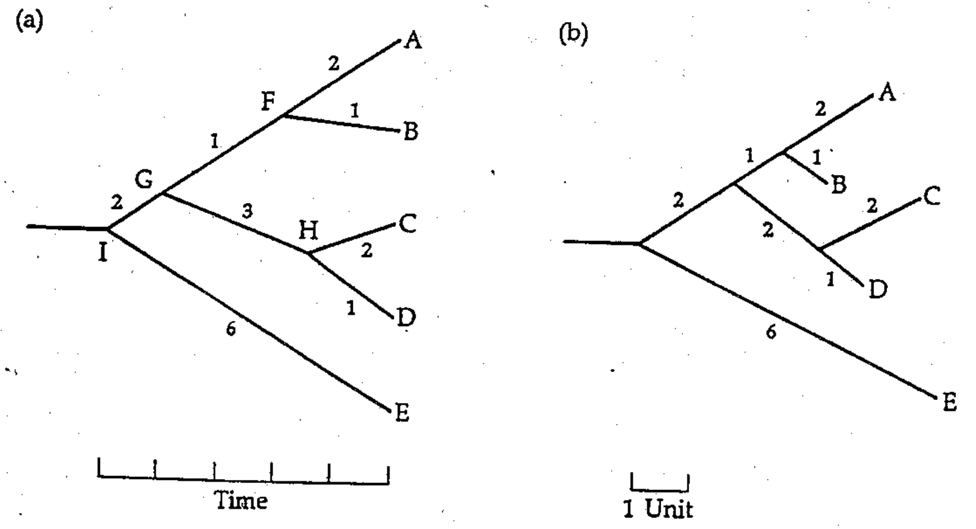


Figure 1. Two alternative representations of a phylogenetic tree for five OTUs. (a) Unscaled branches: extant OTUs are lined up and nodes are positioned proportionally to times of divergence. (b) Scaled branches: lengths of branches are proportional to the numbers of molecular changes.

OTU: Operational taxonomic unit

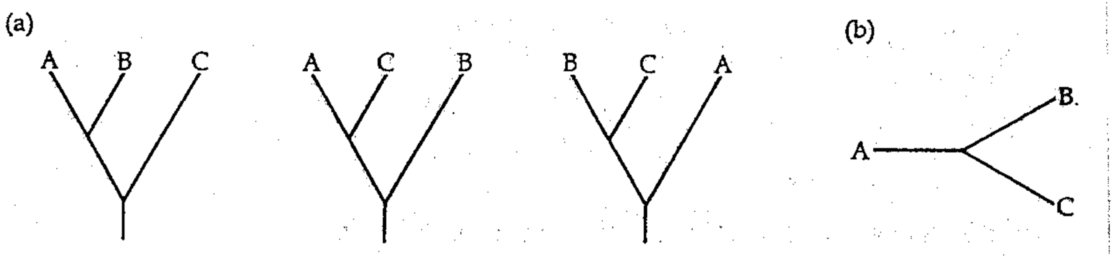
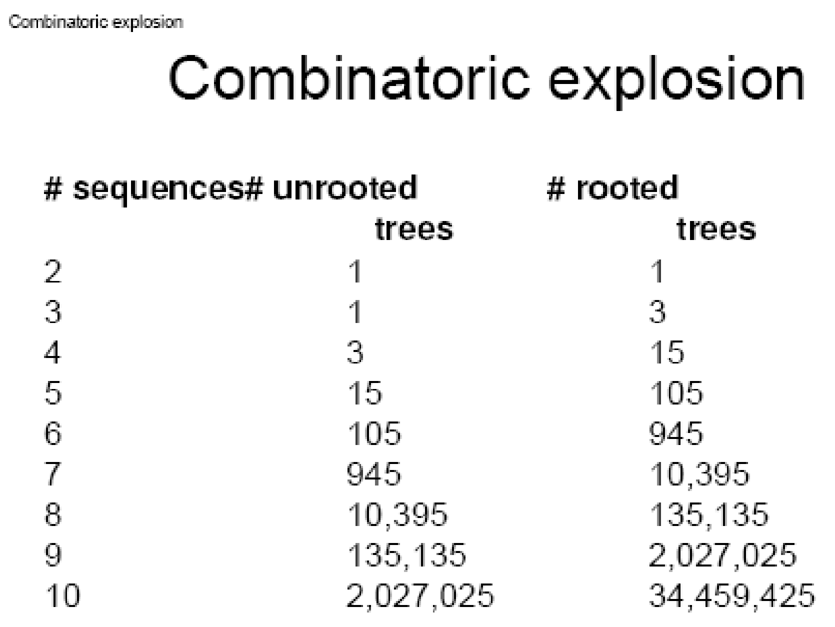


Figure 3. From three OTUs it is possible to construct three different rooted trees (a), but only one unrooted tree (b).

Para n OTUs (n ≥ 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

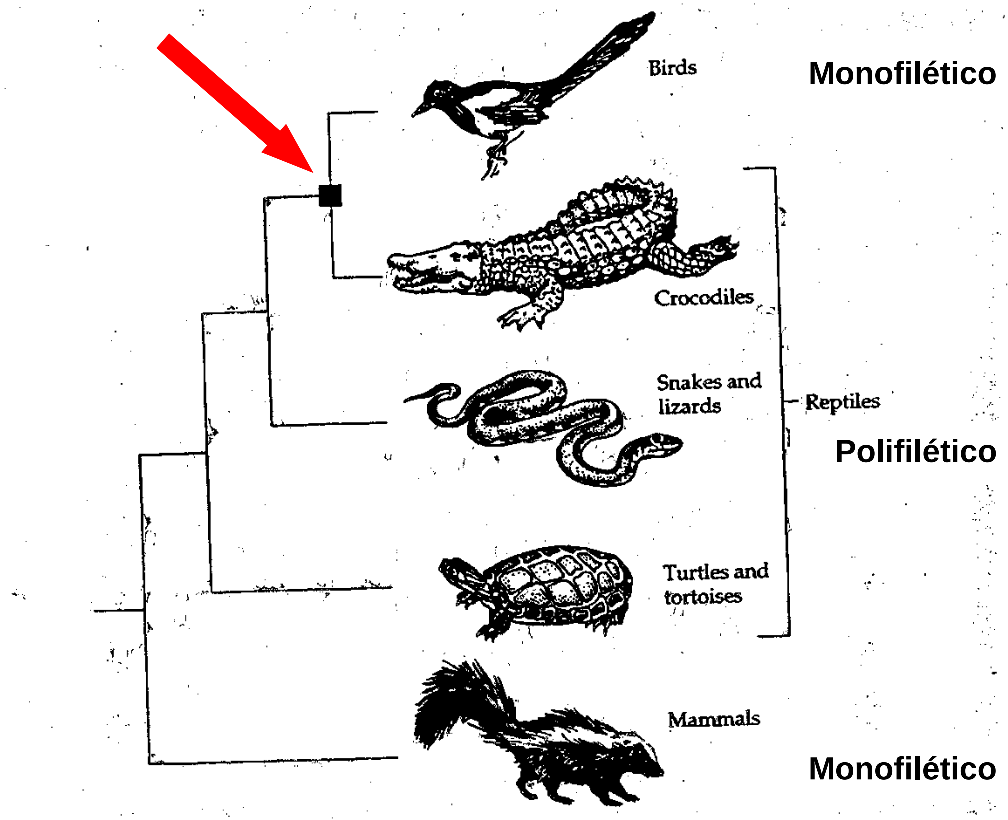


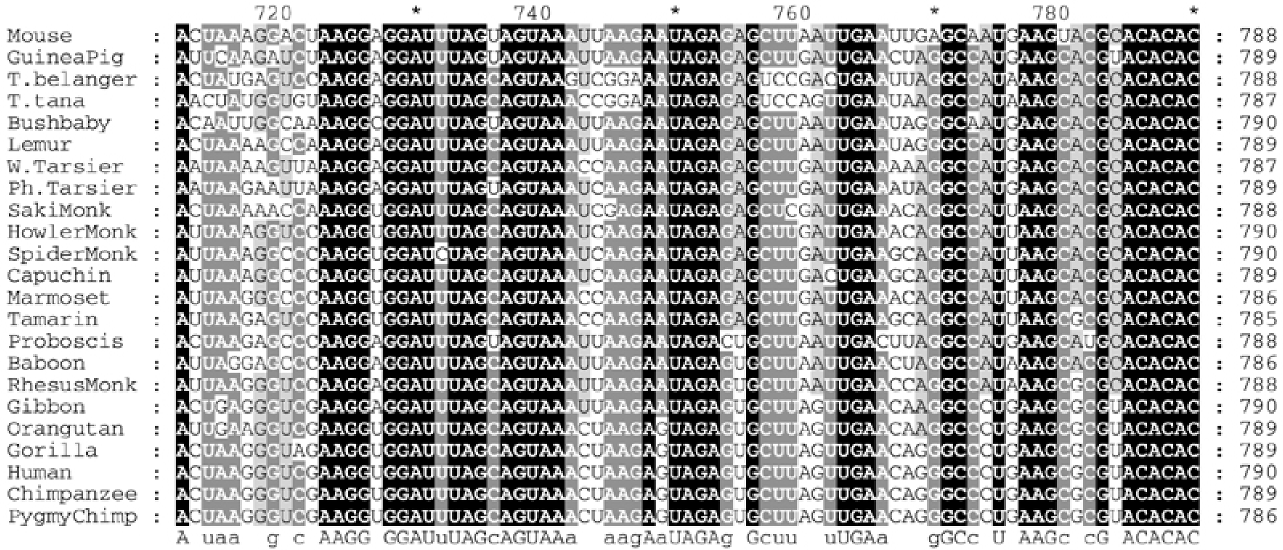
Figure 5.5 Phylogenetic tree of birds, reptiles, and mammals. The reptiles do not constitute a natural clade since they share ancestors with the birds, which are not included in the Reptilia. Birds and crocodiles, on the other hand, constitute a natural clade (Archosauria) since they share a common ancestor (black box) not shared by any other organism. From Li and Graur (1991).

- Un grupo **monofilético** es aquel que comprende una especie ancestral y a todos sus descendientes. Por ejemplo, **las Aves son un grupo monofilético** (incluye a todos los descendientes de la primera ave). Los mamíferos también lo son.

- Un grupo **polifilético** es aquel que comprende una especie ancestral pero no a todos sus descendientes. Un ejemplo son **los Reptiles,** ya que no incluyen a todos los descendientes del primer reptil que existió (las aves quedan fuera del grupo).

Selección de las secuencias de partida:

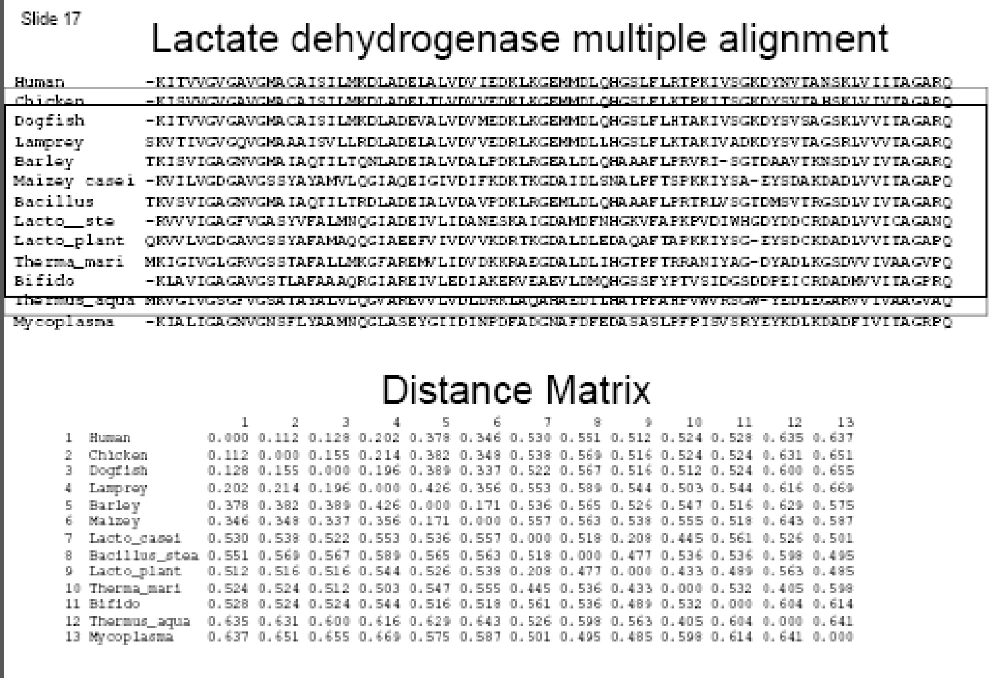
- Ortólogas: ancestro común
- Que estén presentes en todas las especies del grupo de interés
- Ritmo evolutivo adecuado: ni muy alto ni muy bajo



```
                    720        *        740        *        760        *        780        *
Mouse      : ACUAAAGGACUAAGGAGGAUUUAGUAGUAAAUUAAGAAUAGAGAGCUUAAUUGAAUUGAGCAAUGAAGUACGCACACAC : 788
GuineaPig  : AUUCAAGAUCUAAGGAGGAUUUAGUAGUAAAUUAAGAAUAGAGAGCUUGAUUGAACUAGGCCAUGAAGCACGUACACAC : 789
T.belanger : ACUAUGAGUCCAAGGAGGAUUUAGCAGUAAGUCGGAAAUAGAGAGUCCGACUGAAUUAGGCCAUAAAGCACGCACACAC : 788
T.tana     : AACUAUGGUGUAAGGAGGAUUUAGCAGUAAACCGGAAAUAGAGAGUCCAGUUGAAUAAGGCCAUAAAGCACGCACACAC : 787
Bushbaby   : ACAAUUGGCAAAAGGCGGAUUUAGUAGUAAAUUAAGAAUAGAGAGCUUAAUUGAAUAGGGCAAUGAAGCACGCACACAC : 790
Lemur      : ACUAAAAGCCAAAGGAGGAUUUAGCAGUAAAUUAAGAAUAGAGAGCUUAAUUGAAUAGGGCCAUGAAGCACGCACACAC : 789
W.Tarsier  : AAUAAAAGUUAAAGGAGGAUUUAGCAGUAAACCAAGAAUAGAGAGCUUGAUUGAAAAAGGCCAUGAAGCACGCACACAC : 787
Ph.Tarsier : AAUAAGAAUUAAAGGAGGAUUUAGUAGUAAAUCAAGAAUAGAGAGCUUGAUUGAAAUAGGCCAUGAAGCACGCACACAC : 789
SakiMonk   : ACUAAAAACCAAAGGUGGAUUUAGCAGUAAAUCGAGAAUAGAGAGCUCGAUUGAAACAGGCCAUUAAGCACGCACACAC : 788
HowlerMonk : AUUAAAGGUCCAAGGUGGAUUUAGCAGUAAAUCAAGAAUAGAGAGCUUGAUUGAAACAGGCCAUUAAGCACGCACACAC : 790
SpiderMonk : AUUAAAGGCCCAAGGUGGAUCUAGCAGUAAAUCAAGAAUAGAGAGCUUGAUUGAAGCAGGCCAUUAAGCACGCACACAC : 790
Capuchin   : AUUAAAGGCCCAAGGUGGAUUUAGCAGUAAAUCAAGAAUAGAGAGCUUGACUGAAGCAGGCCAUUAAGCACGCACACAC : 789
Marmoset   : AUUAAGGGCCCAAGGUGGAUUUAGCAGUAAACCAAGAAUAGAGAGCUUGAUUGAAACAGGCCAUUAAGCACGCACACAC : 786
Tamarin    : AUUAAGAGUCCAAGGUGGAUUUAGCAGUAAACCAAGAAUAGAGAGCUUGAUUGAAGCAGGCCAUUAAGCGCGCACACAC : 785
Proboscis  : ACUAAGAGCCCAAGGAGGAUUUAGUAGUAAAUUAAGAAUAGACUGCUUAAUUGACUUAGGCCAUGAAGCAUGCACACAC : 788
Baboon     : AUUAGGAGCCCAAGGAGGAUUUAGCAGUAAAUUAAGAAUAGAGUGCUUAAUUGAACUAGGCCAUAAAGCACGCACACAC : 786
RhesusMonk : AUUAAGGGUCCAAGGAGGAUUUAGCAGUAAAUUAAGAAUAGAGUGCUUAAUUGAACCAGGCCAUAAAGCGCGCACACAC : 788
Gibbon     : ACUGAGGGUCGAAGGAGGAUUUAGCAGUAAAUUAAGAAUAGAGUGCUUAGUUGAACAAGGCCCUGAAGCGCGUACACAC : 790
Orangutan  : AUUGAAGGUCGAAGGUGGAUUUAGCAGUAAACUAAGAGUAGAGUGCUUAGUUGAACAAGGCCCUGAAGCGCGUACACAC : 789
Gorilla    : ACUAAGGGUAGAAGGUGGAUUUAGCAGUAAACUAAGAGUAGAGUGCUUAGUUGAACAGGGCCCUGAAGCGCGUACACAC : 789
Human      : ACUAAGGGUCGAAGGUGGAUUUAGCAGUAAACUAAGAGUAGAGUGCUUAGUUGAACAGGGCCCUGAAGCGCGUACACAC : 790
Chimpanzee : ACUAAGGGUCGAAGGUGGAUUUAGCAGUAAACUAAGAGUAGAGUGCUUAGUUGAACAGGGCCCUGAAGCGCGUACACAC : 789
PygmyChimp : ACUAAGGGUCGAAGGUGGAUUUAGCAGUAAACUAAGAGUAGAGUGCUUAGUUGAACAGGGCCCUGAAGCGCGUACACAC : 786
             A uaa  g c AAGG GGAUuUAGcAGUAAa  aagAaUAGAg Gcuu  uUGAa   gGCc U AAGc cG ACACAC
```

Se descartan los gaps

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



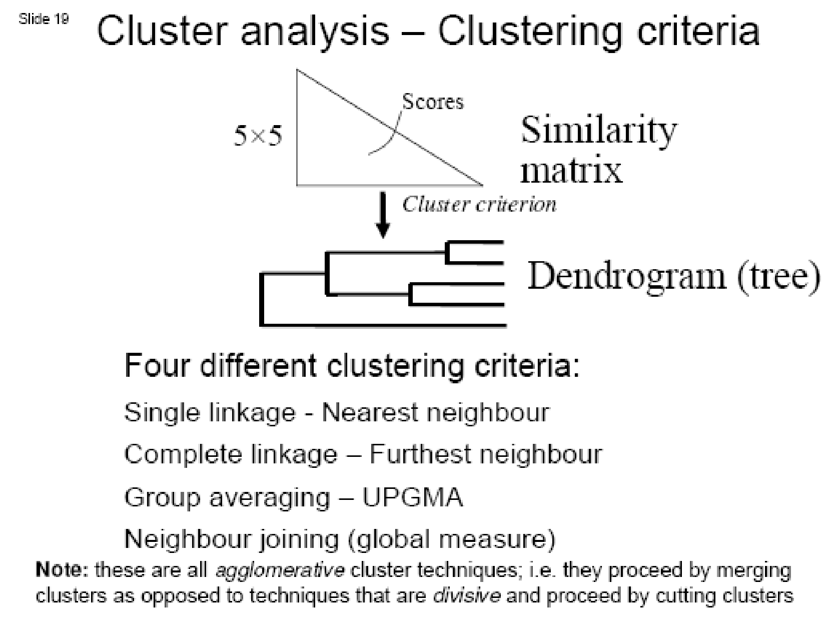
Slide 17

Lactate dehydrogenase multiple alignment

```
Human        -KITVVGVGAVGMACAISILMKDLADEIALVDVIEDKLKGEMMDLQHGSLFLRTPKIVSGKDYNVTANSKLVIIITAGARQ
Chicken      -KISVVGVGAVGMACAISILMKDLADELTLVDVVEDKLKGEMMDLQHGSLFLKTPKITSGKDYSVTAHSKLVIVTAGARQ
Dogfish      -KITVVGVGAVGMACAISILMKDLADEVALVDVMEDKLKGEMMDLQHGSLFLHTAKIVSGKDYSVSAGSKLVVITAGARQ
Lamprey      SKVTIVGVGQVGMAAAISVLLRDLADEIALVDVVEDRLKGEMMDLLHGSLFLKTAKIVADKDYSVTAGSRLVVVTAGARQ
Barley       TKISVIGAGNVGMAIAQTILTQNLADEIALVDALPDKLRGEALDLQHAAAFLPRVRI-SGTDAAVTKNSDLVIVTAGARQ
Maizey casei -KVILVGDGAVGSSYAYAMVLQGIAQEIGIVDIFKDKTKGDAIDLSNALPFTSPKKIYSA-EYSDAKDADLVVITAGAPQ
Bacillus     TKVSVIGAGNVGMAIAQTILTRDLADEIALVDAVPDKLRGEMLDLQHAAAFLPRTRLVSGTDMSVTRGSDLVIVTAGARQ
Lacto__ste   -RVVVIGAGFVGASYVFALMNQGIADEIVLIDANESKAIGDAMDFNHGKVFAPKPVDIWHGDYDDCRDADLVVICAGANQ
Lacto_plant  QKVVLVGDGAVGSSYAFAMAQQGIAEEFVIVDVVKDRTKGDALDLEDAQAFTAPKKIYSG-EYSDCKDADLVVITAGAPQ
Therma_mari  MKIGIVGLGRVGSSTAFALLMKGFAREMVLIDVDKKRAEGDALDLIHGTPFTRRANIYAG-DYADLKGSDVVIVAAGVPQ
Bifido       -KLAVIGAGAVGSTLAFAAAQRGIAREIVLEDIAKERVEAEVLDMQHGSSFYPTVSIDGSDDPEICRDADMVVITAGPRQ
Thermus_aqua MKVGIVGSGFVGSATAYALVLQGVAREVVLVDLDRKLAQAHAEDILHATPFAHPVWVRSGW-YEDLEGARVVIVAAGVAQ
Mycoplasma   -KIALIGAGNVGNSFLYAAMNQGLASEYGIIDINPDFADGNAFDFEDASASLPFPISVSRYEYKDLKDADFIVITAGRPQ
```

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_stea	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



Slide 19

Cluster analysis – Clustering criteria

Four different clustering criteria:

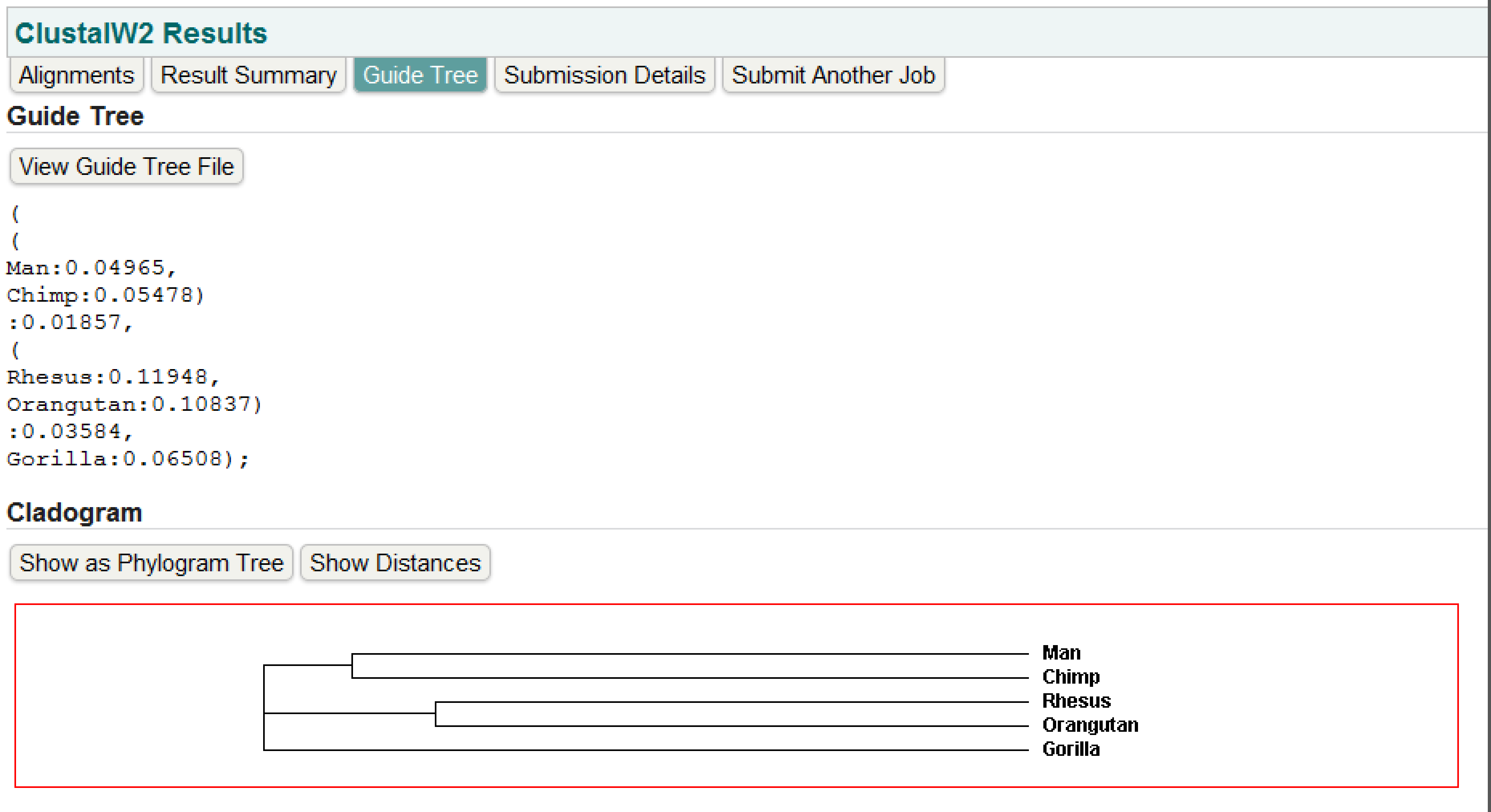
Single linkage - Nearest neighbour

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

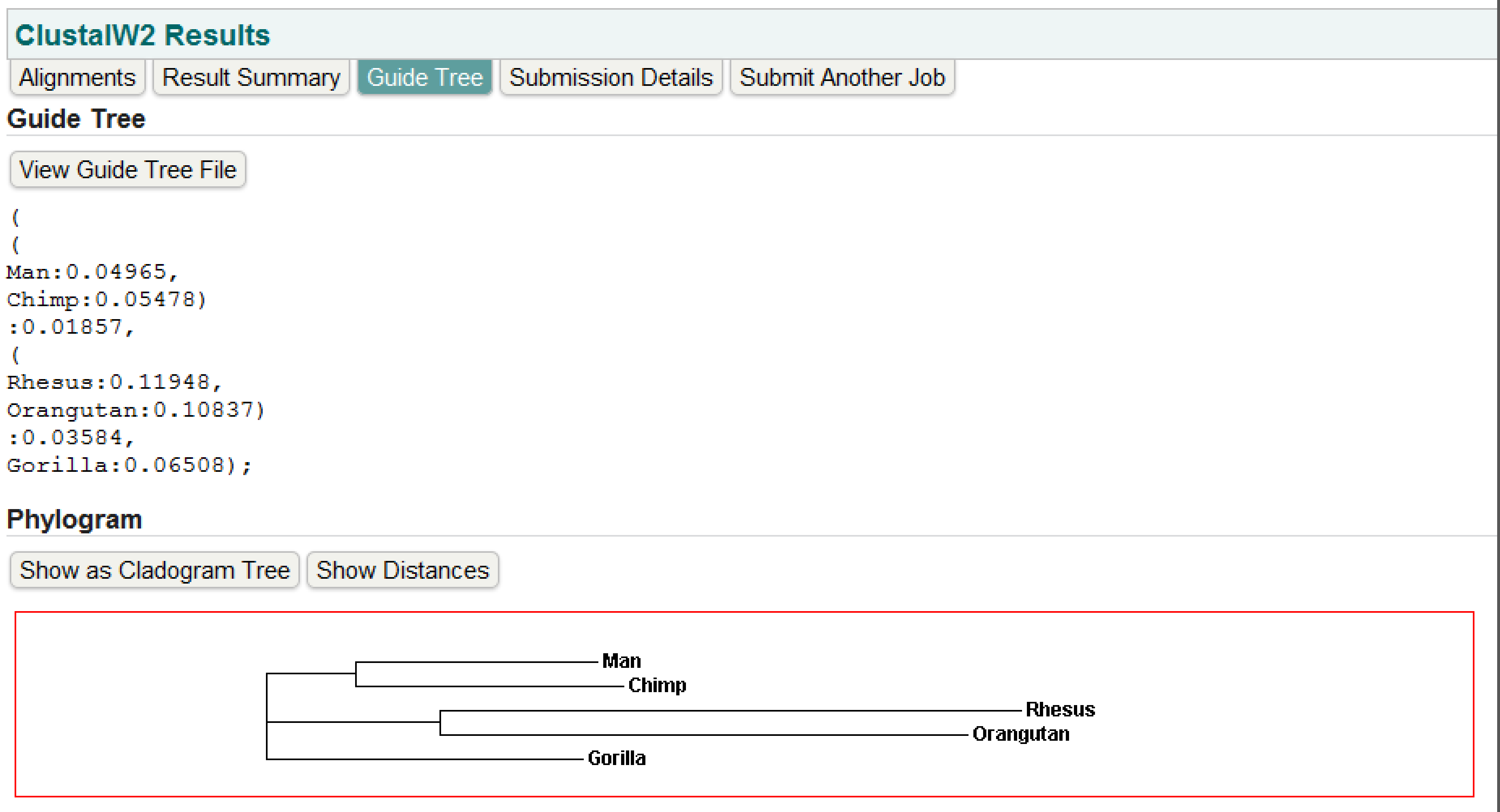


ClustalW2 Results

Guide Tree

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

Cladogram



ClustalW2 Results

Guide Tree

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

Phylogram

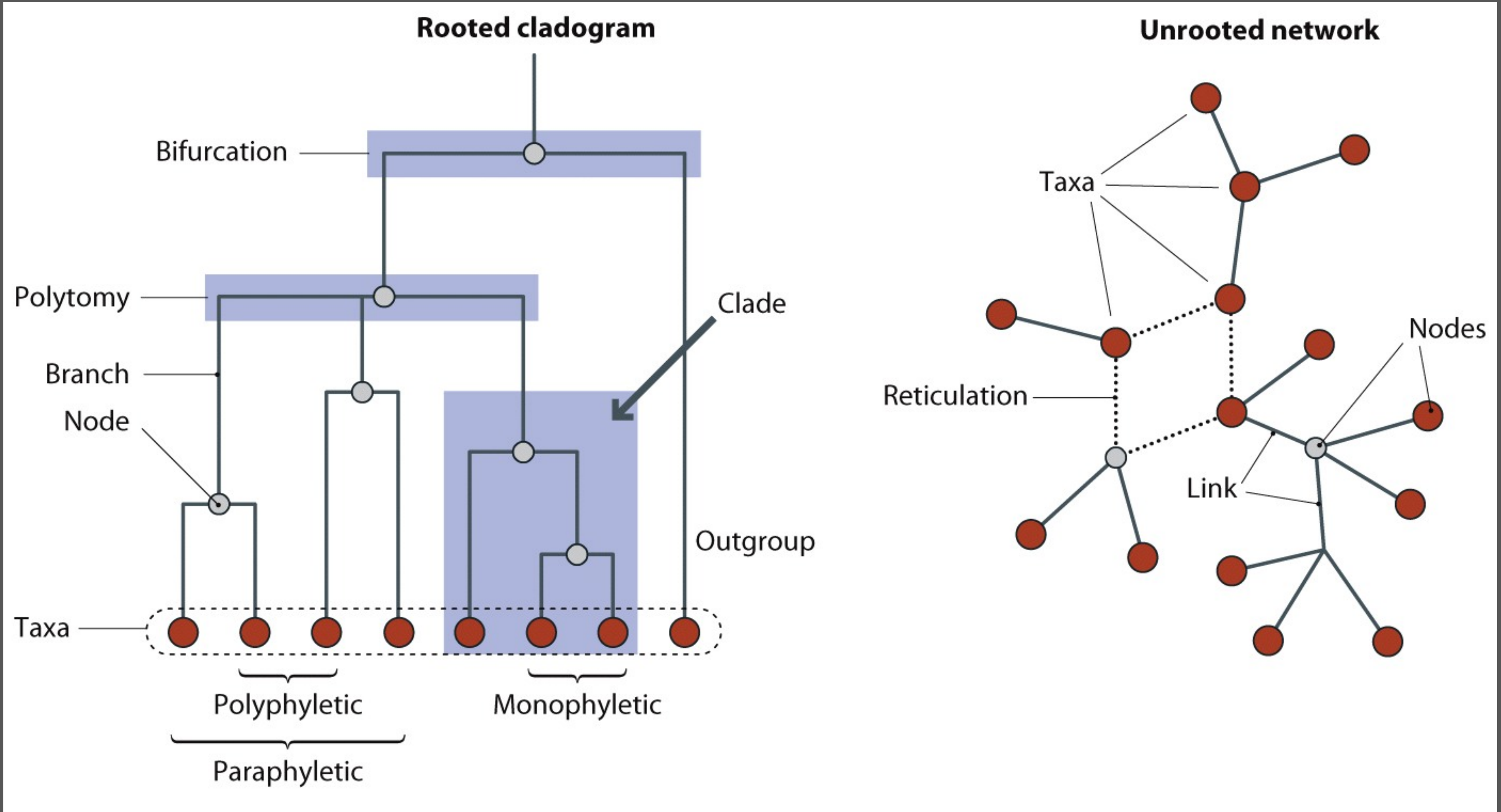
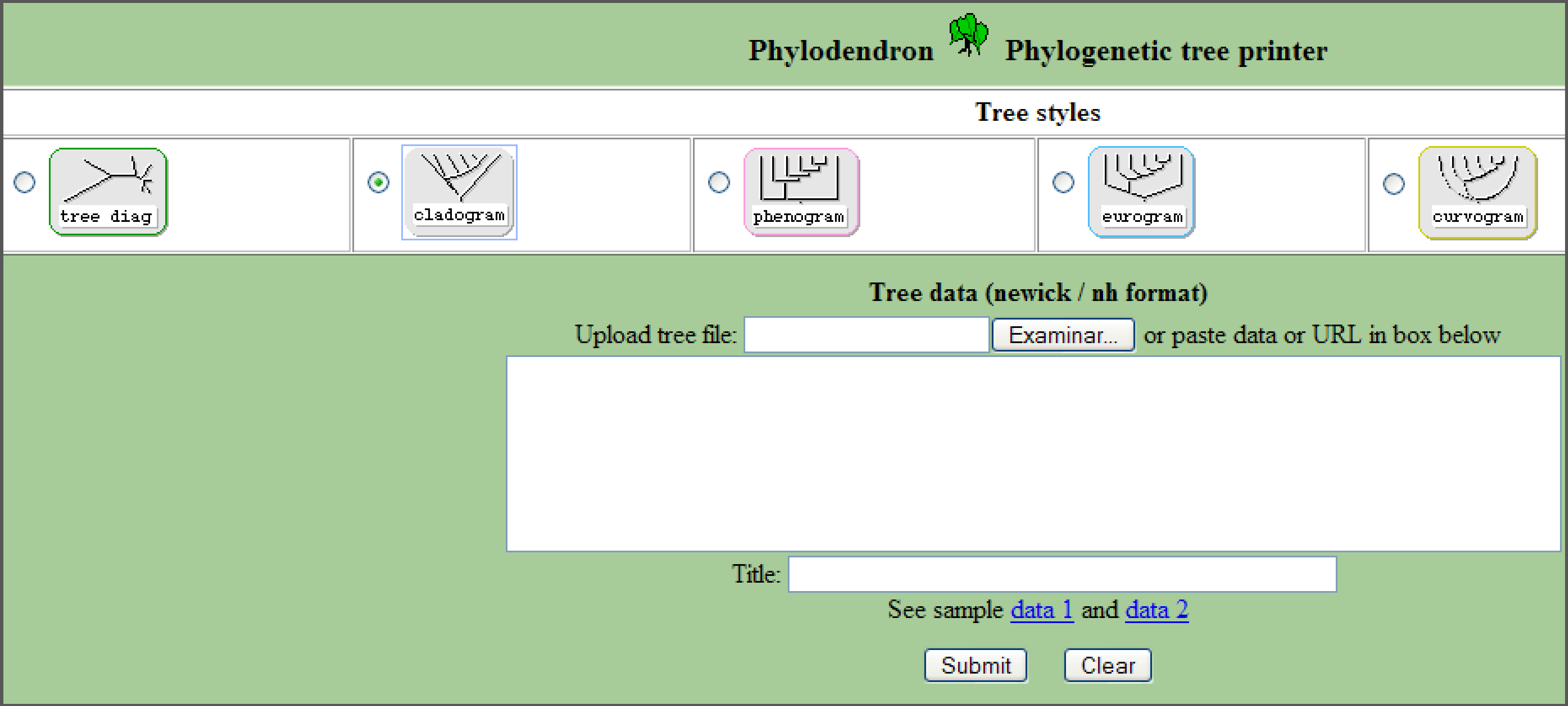


Figure 6.8 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)



Phylodendron **Phylogenetic tree printer**

Tree styles

○ tree diag

● cladogram

○ phenogram

○ eurogram

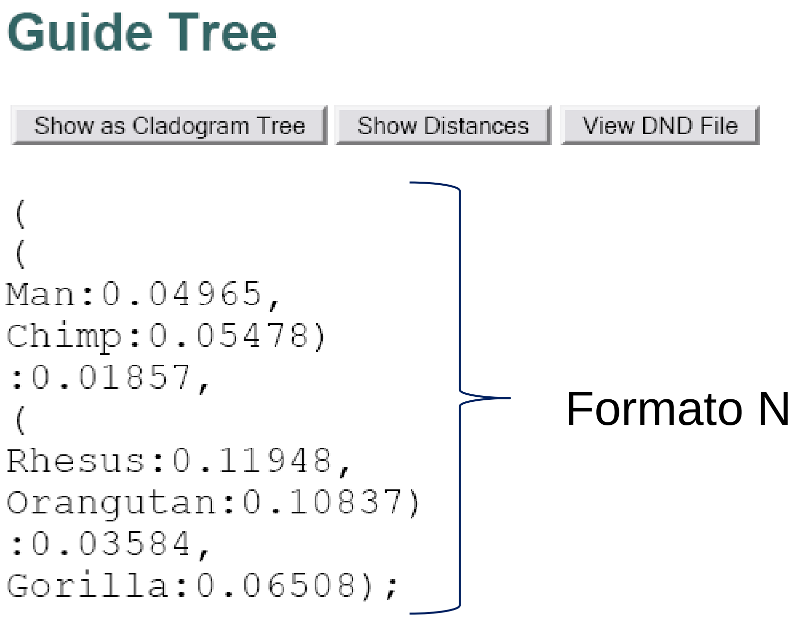
○ curvogram

Tree data (newick / nh format)

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

Title:

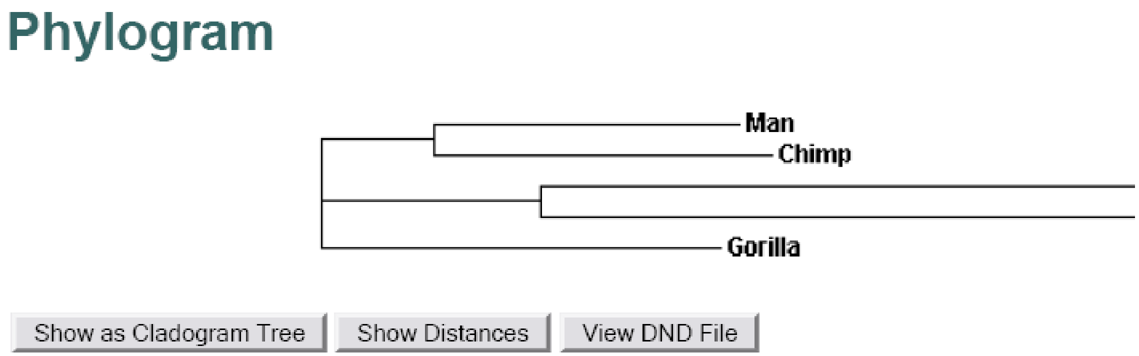
See sample data 1 and data 2



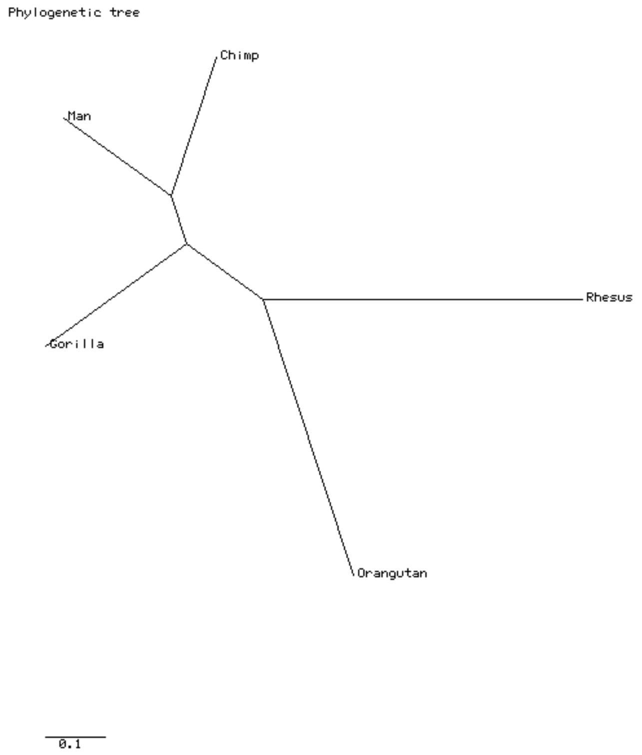
Guide Tree

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

Formato Newick



Phylogram



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

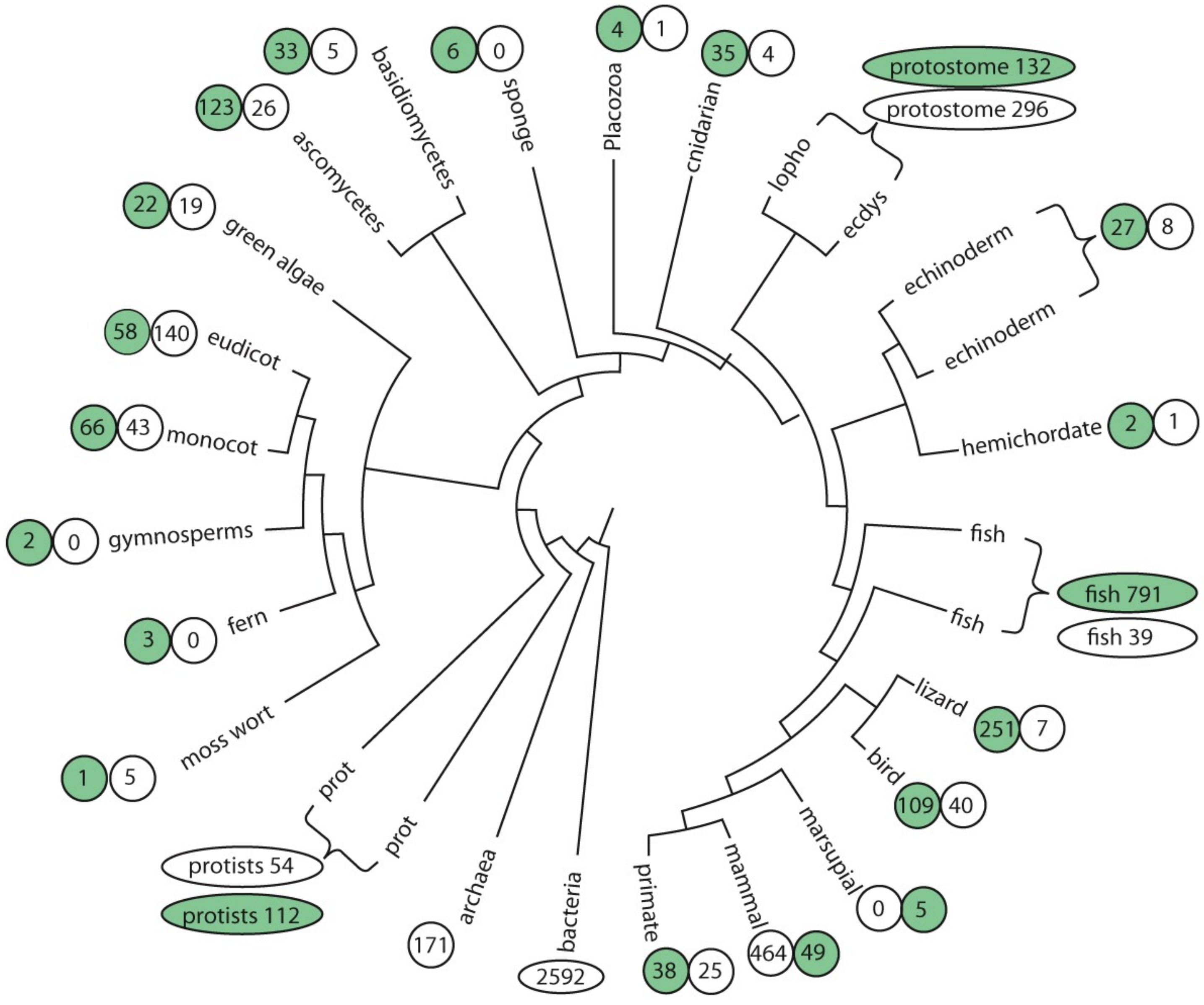


Figure 7.4 Phylogenomics: A Primer (© Garland Science 2013)

Bootstrapping

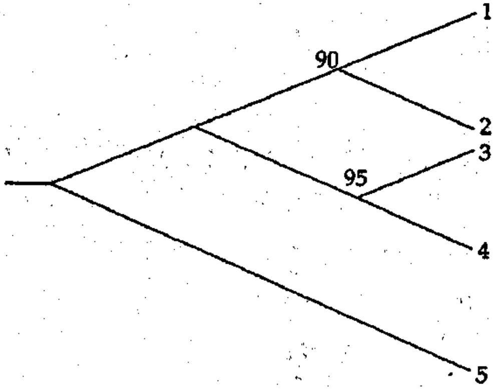
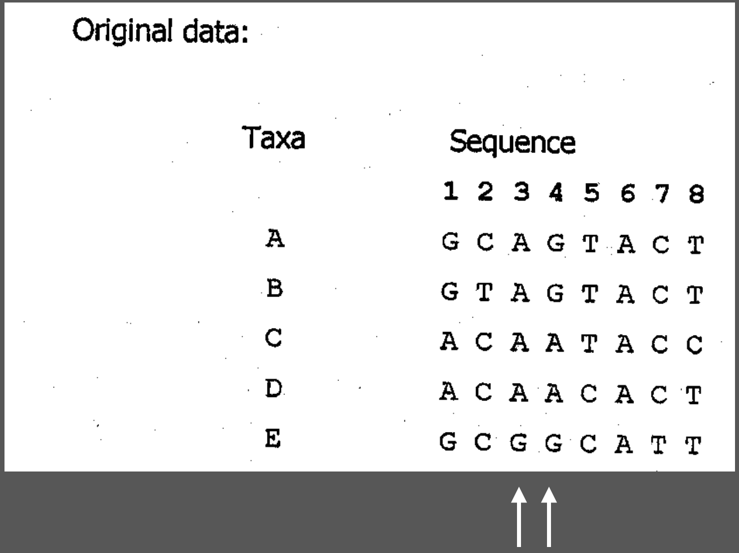
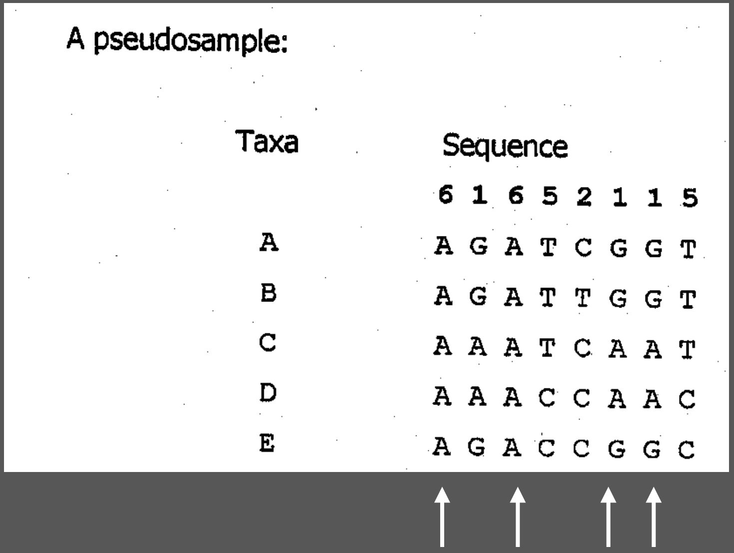


Figure 5.21 Hypothetical tree for illustrating the bootstrap technique. The numbers on the interior nodes are the bootstrap proportions supporting the subsets (1,2) and (3,4).



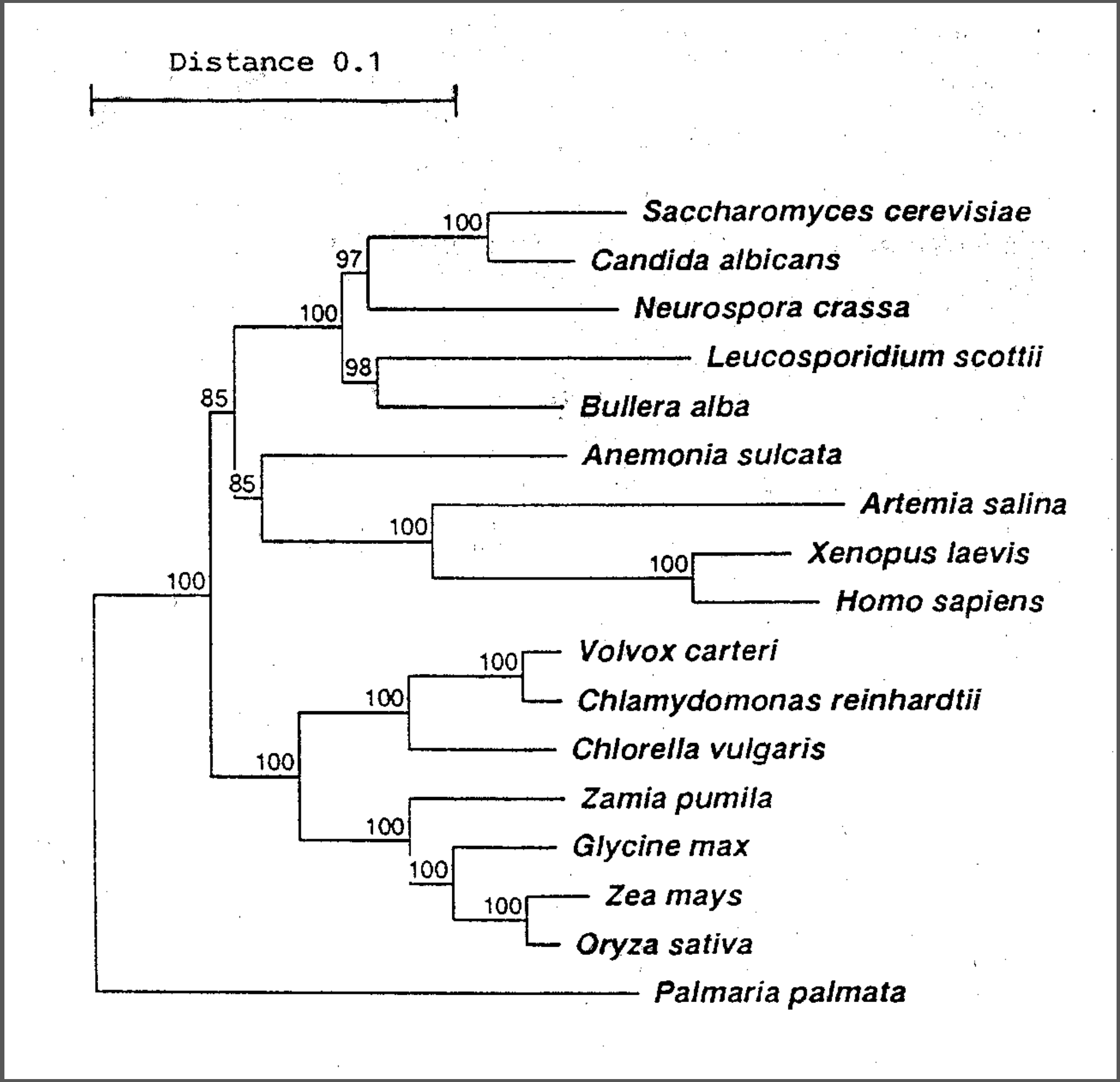
Original data:

Taxa	Sequence							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	G	C	A	G	T	A	C	T
B	G	T	A	G	T	A	C	T
C	A	C	A	A	T	A	C	C
D	A	C	A	A	C	A	C	T
E	G	C	G	G	C	A	T	T



A pseudosample:

Taxa	Sequence							
	6	**1**	**6**	**5**	**2**	**1**	**1**	**5**
A	A	G	A	T	C	G	G	T
B	A	G	A	T	T	G	G	T
C	A	A	A	T	C	A	A	T
D	A	A	A	C	C	A	A	C
E	A	G	A	C	C	G	G	C


Distance 0.1
100
97
100
98
85
85
100
100
100
100
100
100
100
100
100
100
Saccharomyces cerevisiae
Candida albicans
Neurospora crassa
Leucosporidium scottii
Bullera alba
Anemonia sulcata
Artemia salina
Xenopus laevis
Homo sapiens
Volvox carteri
Chlamydomonas reinhardtii
Chlorella vulgaris
Zamia pumila
Glycine max
Zea mays
Oryza sativa
Palmaria palmata