Métodos de reconstrucción filogenética

- **Basados en matrices de distancia:**
  - Distancias por pares → Algoritmo de agrupamiento → Arbol
  - UPGMA y NJ

- **Máxima parsimonia:**
  - Caracteres (nucleótidos o Aa) en cada posición → Ruta más corta que da lugar a estos caracteres → Arbol más parsimonioso

- **Máxima verosimilitud:**
  - Se determina la verosimilitud de todas las configuraciones de caracteres para todos los árboles posibles y se elige el que da el valor máximo

UPGMA: Unweighted Pair-Group Method with Arithmetic mean

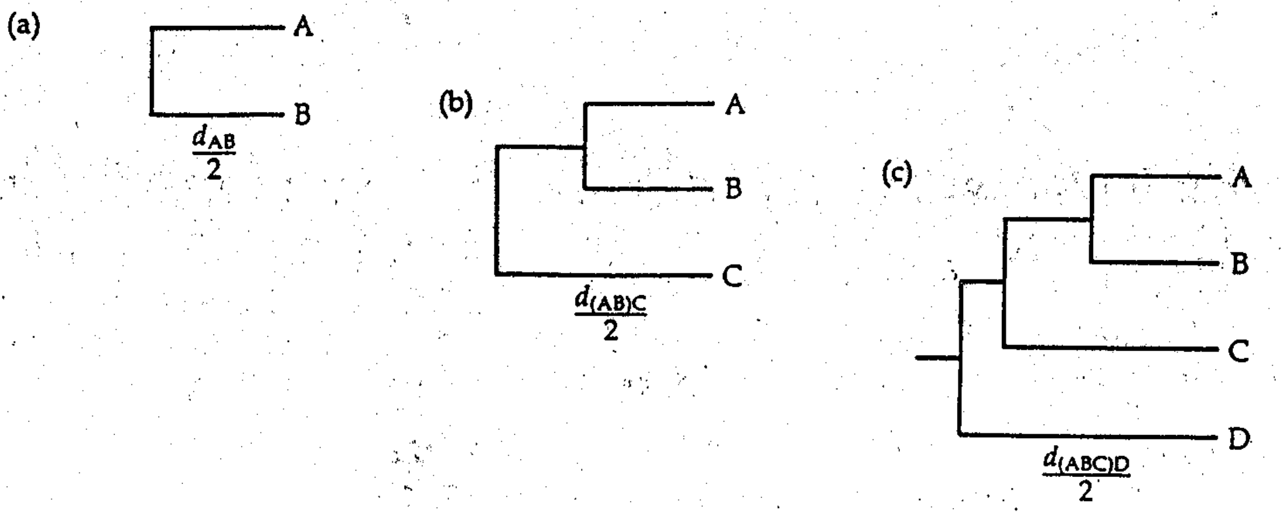
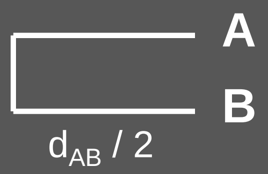


Figure 6. Diagram illustrating the stepwise construction of a phylogenetic tree for four OTUs according to the UPGMA method (see text).

Consideremos 4 OTUs; se calculan las distancias por pares:

OTU	A	B	C
B	d_{AB}		
C	d_{AC}	d_{BC}	
D	d_{AD}	d_{BD}	d_{CD}

Se elige la distancia más pequeña.
Supongamos que sea d_{AB}



OTU	(AB)	C
C	$d_{(AB)C}$	
D	$d_{(AB)D}$	d_{CD}

donde:

$$d_{(AB)C} = (d_{AC} + d_{BC}) / 2$$
$$d_{(AB)D} = (d_{AD} + d_{BD}) / 2$$

Table 11.1 Proportion of different nucleotides (p) (above the diagonal) and estimates of the number of nucleotide substitutions (d) per site and their standard errors (below the diagonal) obtained from nucleotide sequence data for five primate species. Data from Brown et al. (1982).

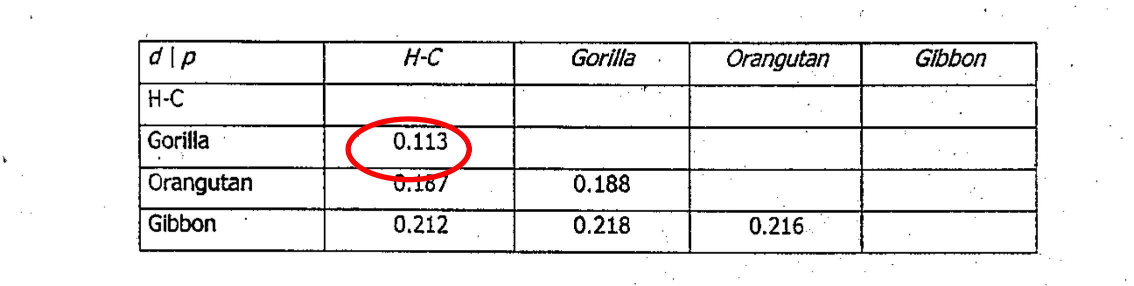
d | p	*Human*	*Chimpanzee*	*Gorilla*	*Orangutan*	*Gibbon*
Human		0.088	0.103	0.160	0.181
Chimpanzee	0.094 ± 0.011		0.106	0.170	0.189
Gorilla	0.111 ± 0.012	0.115 ± 0.012		0.166	0.189
Orangutan	0.180 ± 0.016	0.194 ± 0.016	0.188 ± 0.016		0.188
Gibbon	0.207 ± 0.017	0.218 ± 0.017	0.218 ± 0.017	0.216 ± 0.017	

$d_{HC} = 0.094 / 2 = 0.047$ →

d | *p*	*Human*	*Chimpanzee*	*Gorilla*	*Orangutan*	*Gibbon*
Human		**0.088**	**0.103**	**0.160**	**0.181**
Chimpanzee	**0.094 ± 0.011**		**0.106**	**0.170**	**0.189**
Gorilla	**0.111 ± 0.012**	**0.115 ± 0.012**		**0.166**	**0.189**
Orangutan	**0.180 ± 0.016**	**0.194 ± 0.016**	**0.188 ± 0.016**		**0.188**
Gibbon	**0.207 ± 0.017**	**0.218 ± 0.017**	**0.218 ± 0.017**	**0.216 ± 0.017**	

Las distancias de HC a las otras especies se calculan como:

HC – Gorilla: (0.111 + 0.115) / 2 = 0.113
HC – Orangutan: (0.180 + 0.194) / 2 = 0.187
HC – Gibbon: (0.207 + 0.218) / 2 = 0.212



d | *p*	*H-C*	*Gorilla*	*Orangutan*	*Gibbon*
H-C				
Gorilla	0.113			
Orangutan	0.187	0.188		
Gibbon	0.212	0.218	0.216	

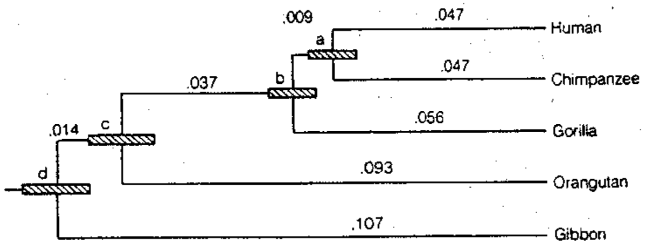


Figure 11.3. Phylogenetic tree reconstructed by UPGMA from the distance matrix in table 11.1. The hatched bar represents one standard error on each side of the branching point.

Filogenia de primates basada en la secuenciación de genomas completos

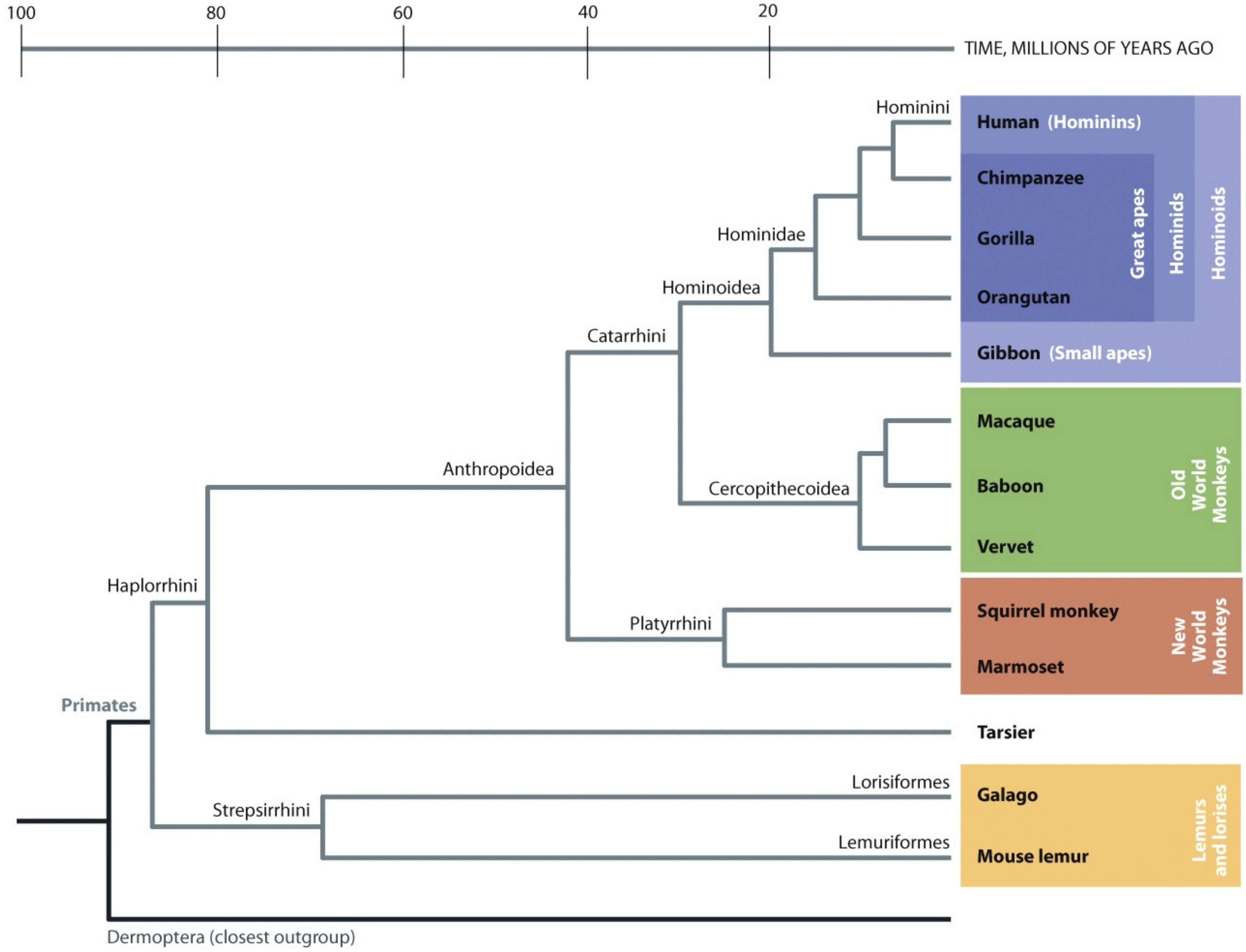


Figure 7.3 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

Combinación de datos de diferentes genes ortólogos para resolver la tricotomía humanos-chimpancé-gorila

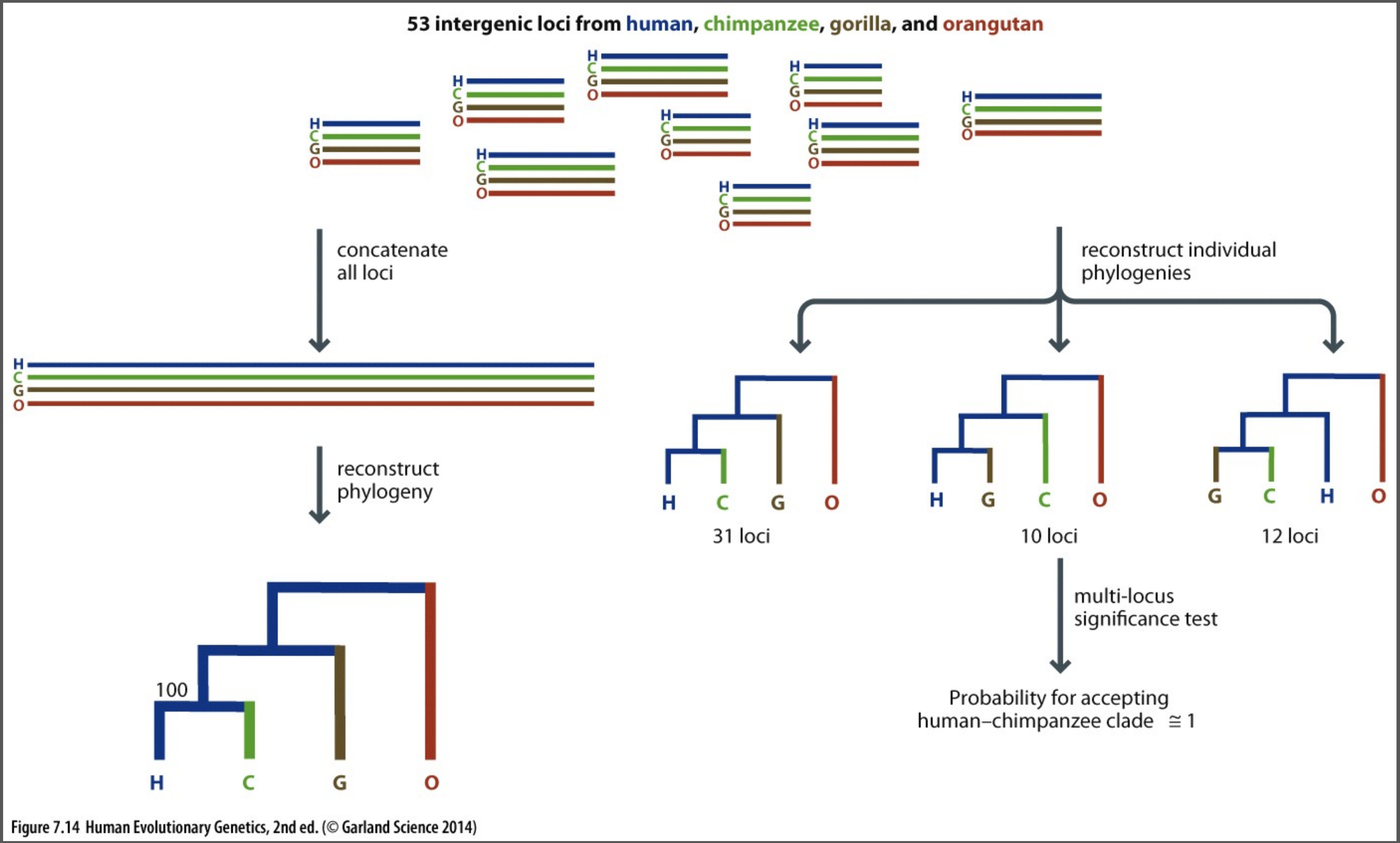


Figure 7.14 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

Neighbor-Joining method (NJ)

Se elige el par de OTUs que minimiza la longitud total del árbol (suma de las ramas)

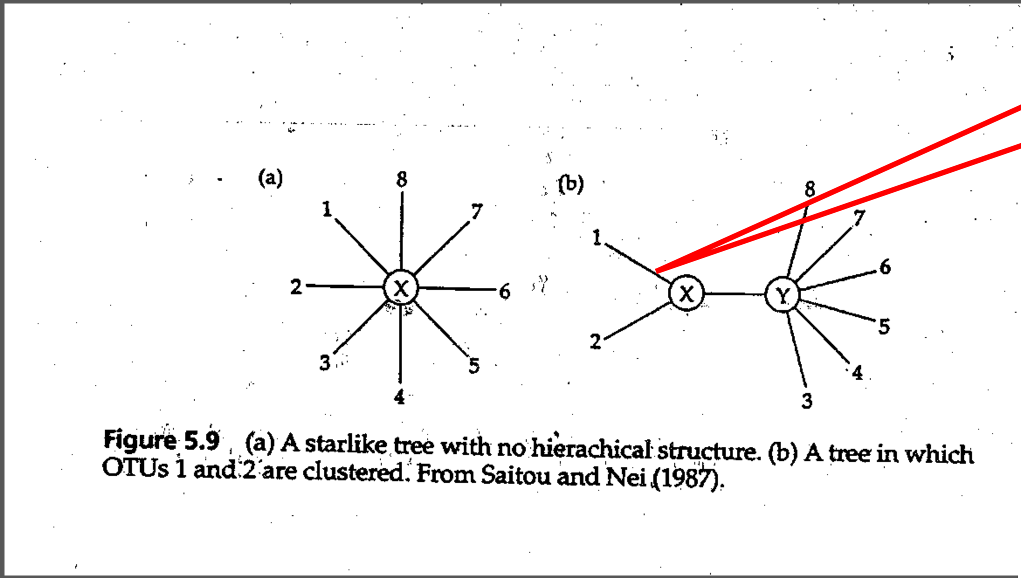
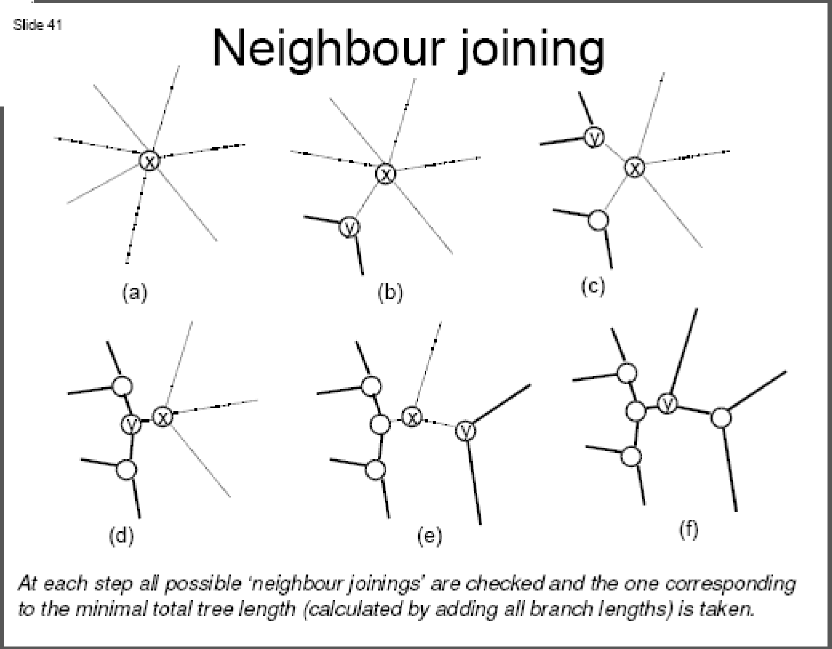


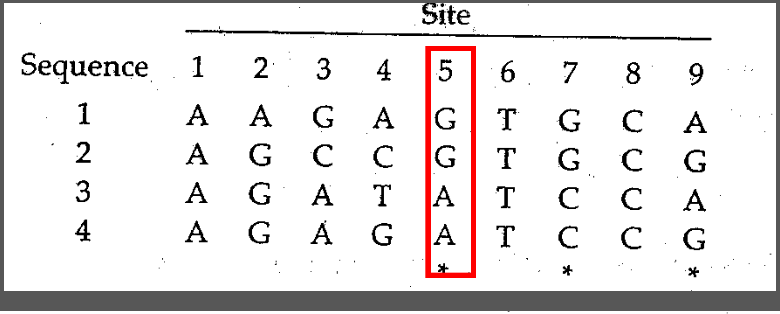
Figure 5.9 (a) A starlike tree with no hierachical structure. (b) A tree in which OTUs 1 and 2 are clustered. From Saitou and Nei (1987).



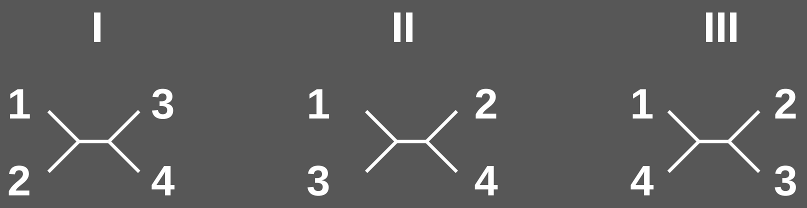
Neighbour joining

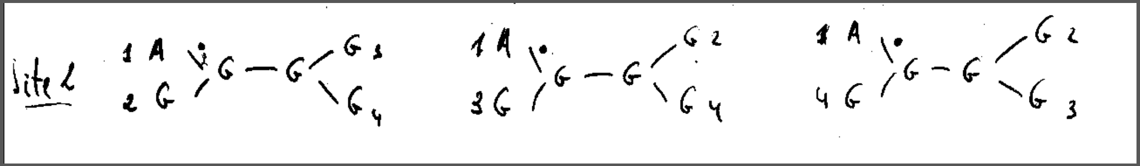
At each step all possible 'neighbour joinings' are checked and the one corresponding to the minimal total tree length (calculated by adding all branch lengths) is taken.

Máxima parsimonia



Sequence	Site 1	2	3	4	5	6	7	8	9
1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G
					*		*		*





No informativo

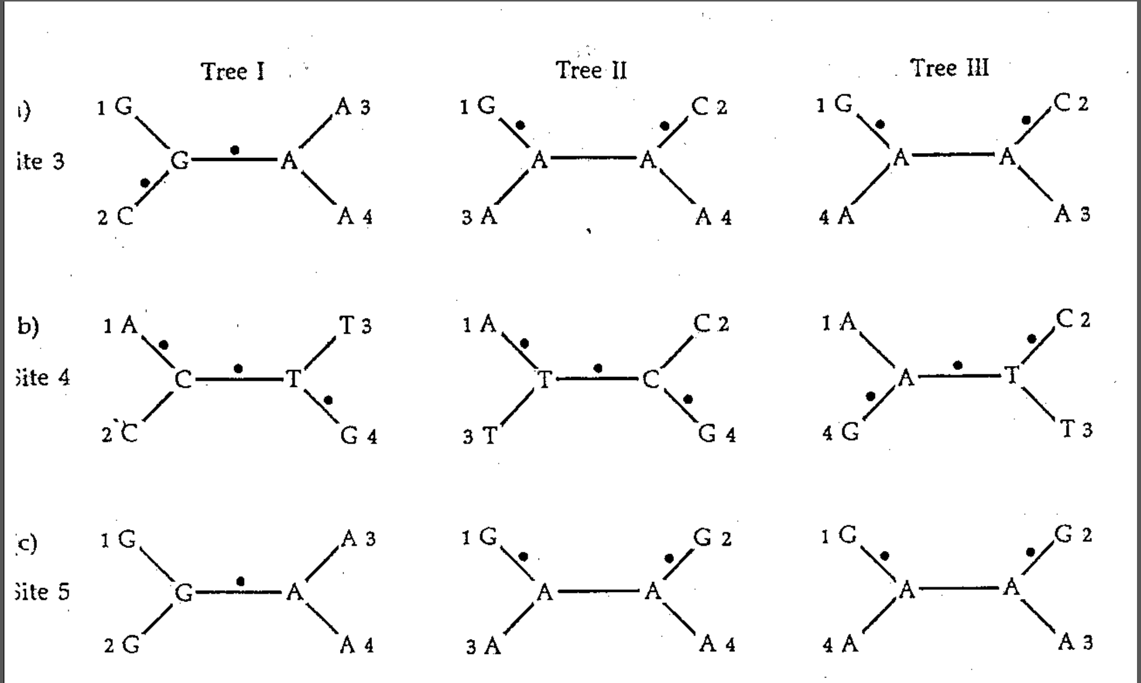


Figure 9. Three possible unrooted trees (I, II, and III) for four DNA sequences (1, 2, 3, and 4) that have been used to choose the most parsimonious tree (see text).

No informativo

No informativo

Informativo: → I

Tree \ Sites	5	7	9	Total
I	1	1	2	4
II	2	2	1	5
III	2	2	2	6

Significación estadística: Bootstrapping

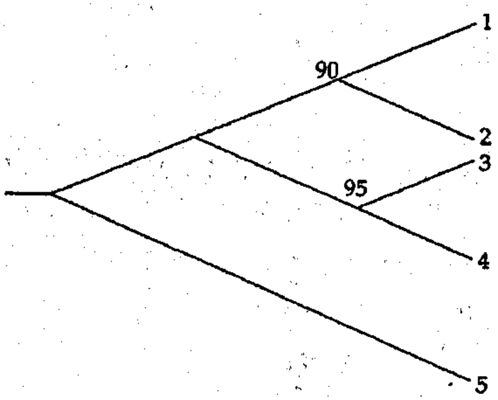
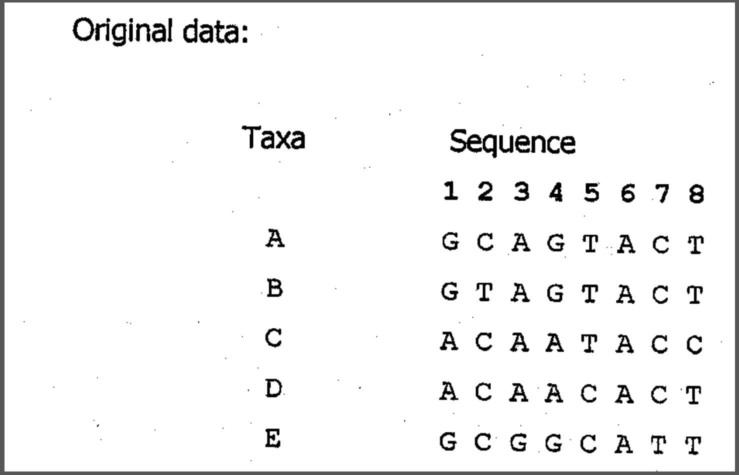
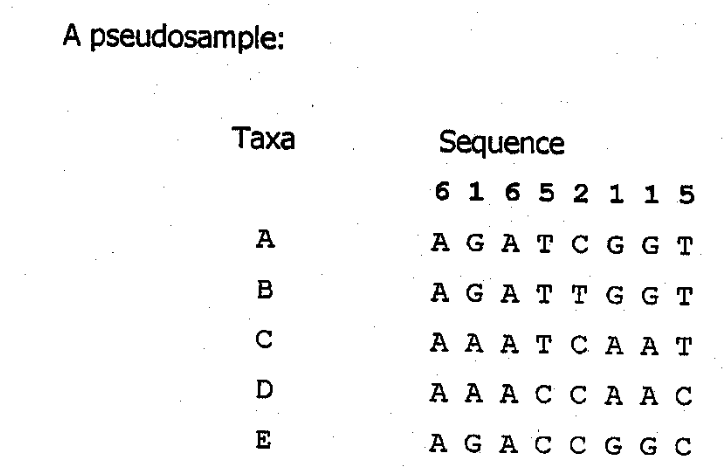


Figure 5.21 Hypothetical tree for illustrating the bootstrap technique. The numbers on the interior nodes are the bootstrap proportions supporting the subsets (1,2) and (3,4).



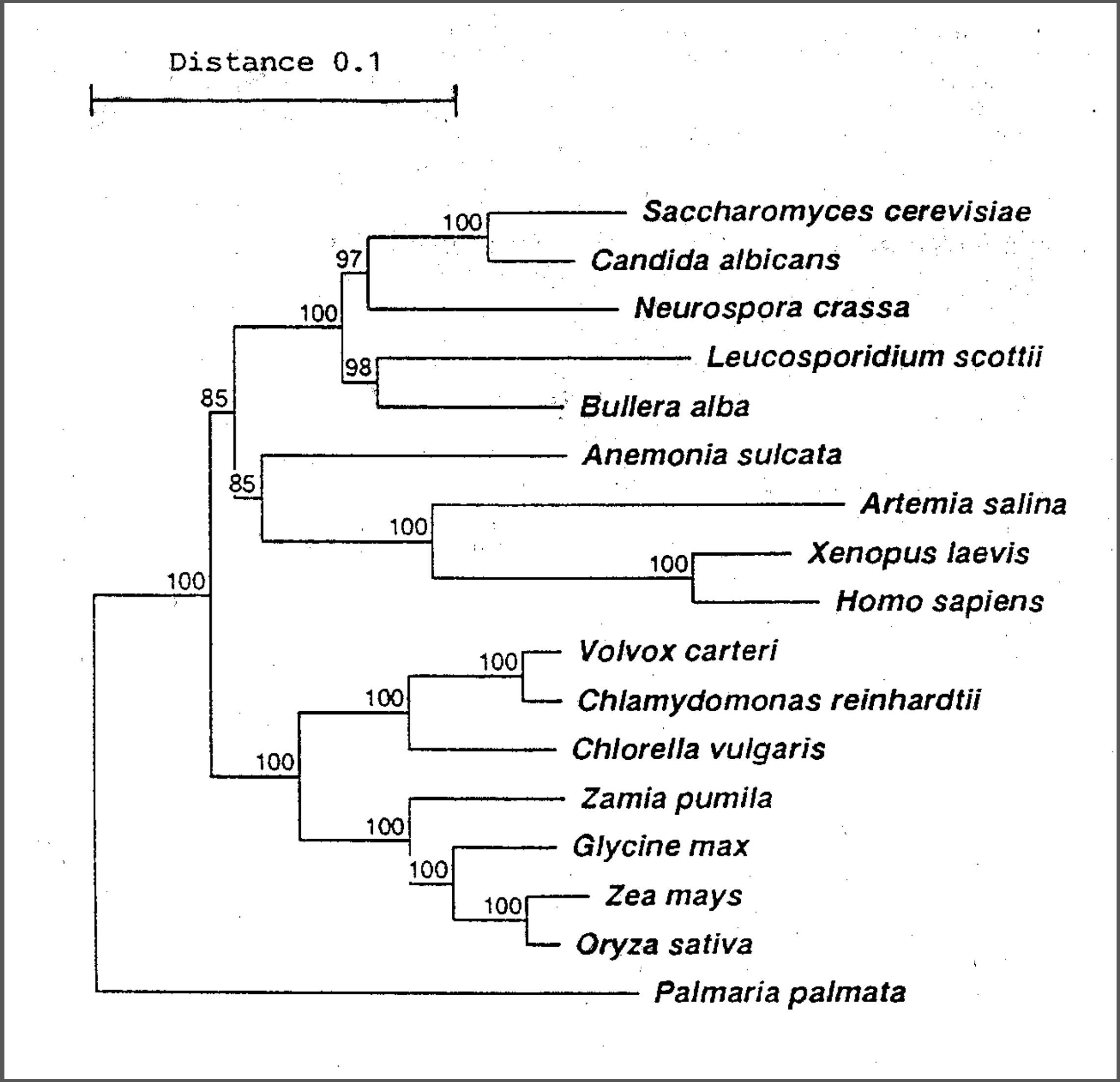
Original data:

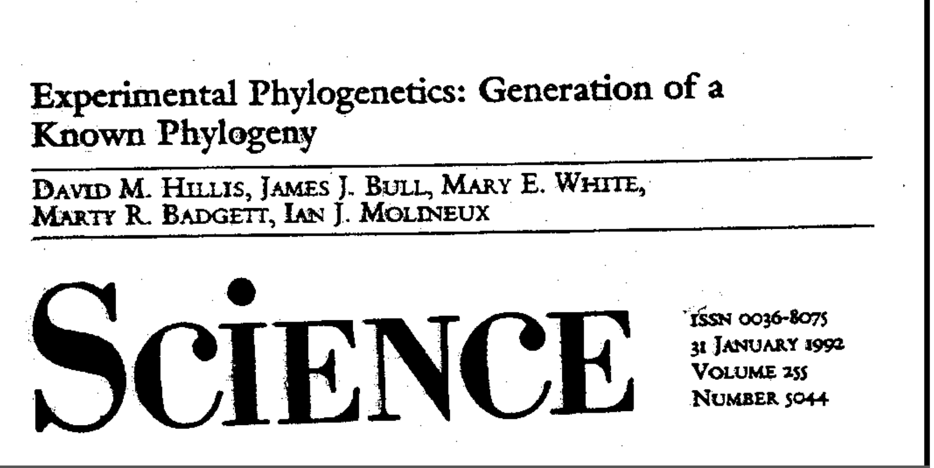
Taxa	Sequence							
	1	2	3	4	5	6	7	8
A	G	C	A	G	T	A	C	T
B	G	T	A	G	T	A	C	T
C	A	C	A	A	T	A	C	C
D	A	C	A	A	C	A	C	T
E	G	C	G	G	C	A	T	T



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



Experimental Phylogenetics: Generation of a Known Phylogeny

DAVID M. HILLIS, JAMES J. BULL, MARY E. WHITE,
MARTY R. BADGETT, IAN J. MOLINEUX

SCIENCE

ISSN 0036-8075
31 JANUARY 1992
VOLUME 255
NUMBER 5044

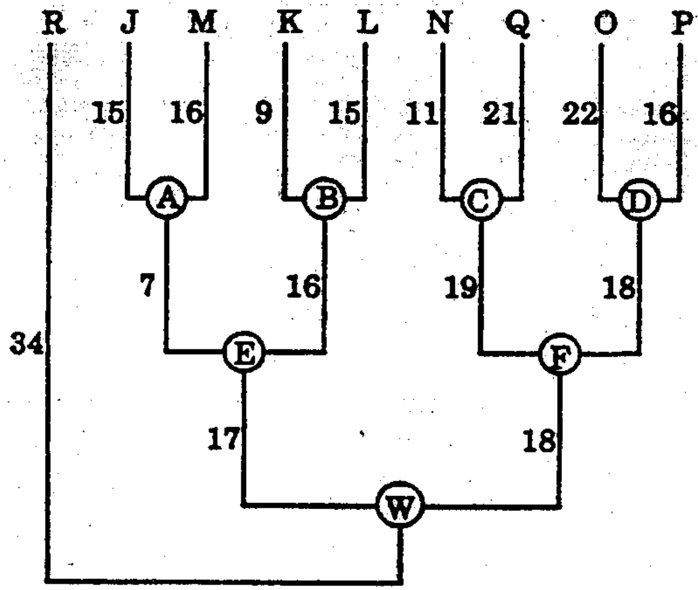
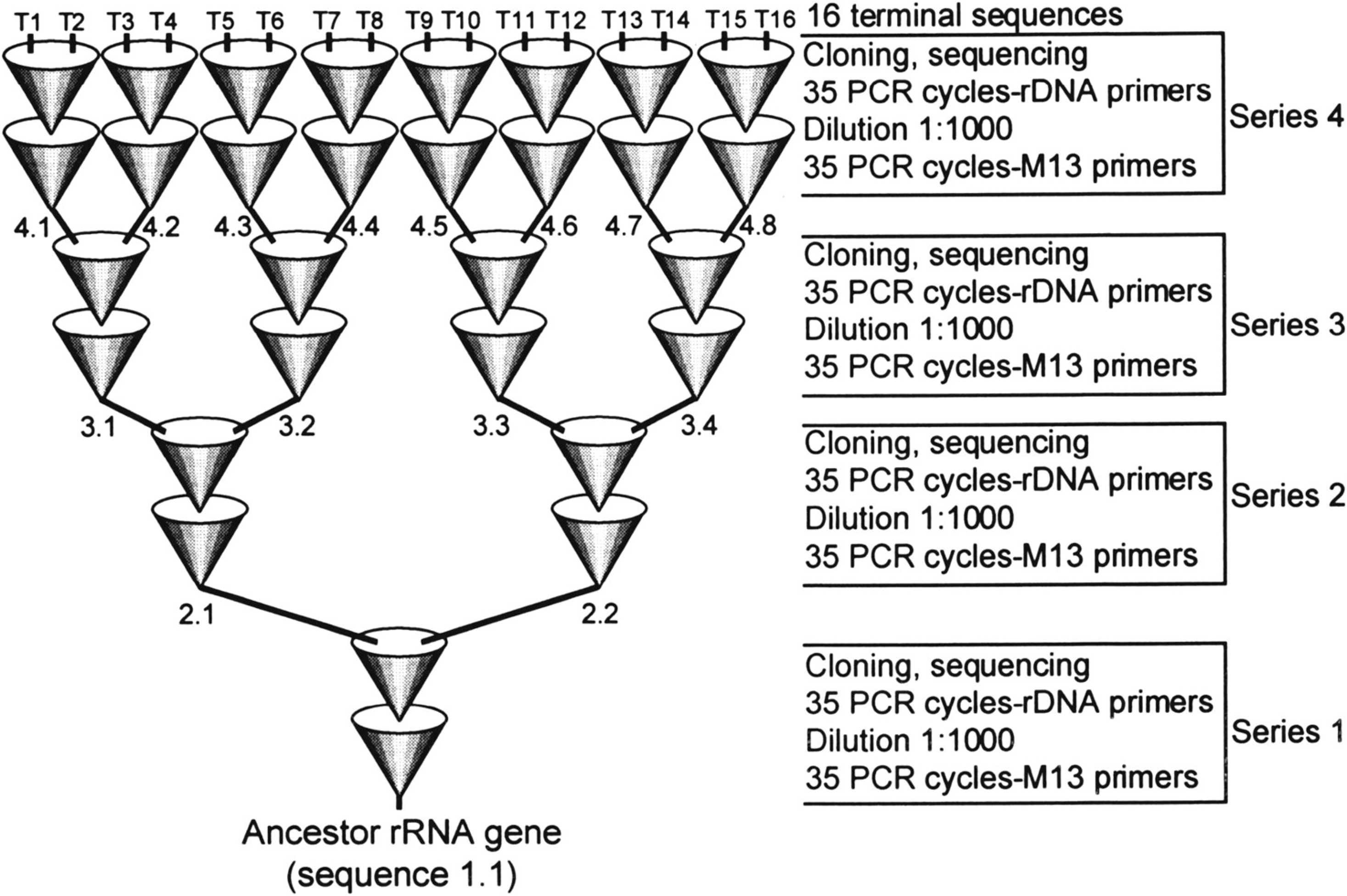


Fig. 1. True phylogeny for the experimental lineages of bacteriophage T7. The ancestors at each node are labeled with letters A through F and W (the latter represents wild-type T7). The numbers represent the number of restriction-site differences scored between the phages at each node of the phylogeny.

1. Maxima parsimonia
2. Fitch-Margoliash
3. Cavalli-Sforza
4. NJ
5. UPGMA

- Todos los métodos predicen la topología correcta
- Ninguno predice la longitud de TODAS las ramas
- El mejor es el de máxima parsimonia y el peor UPGMA

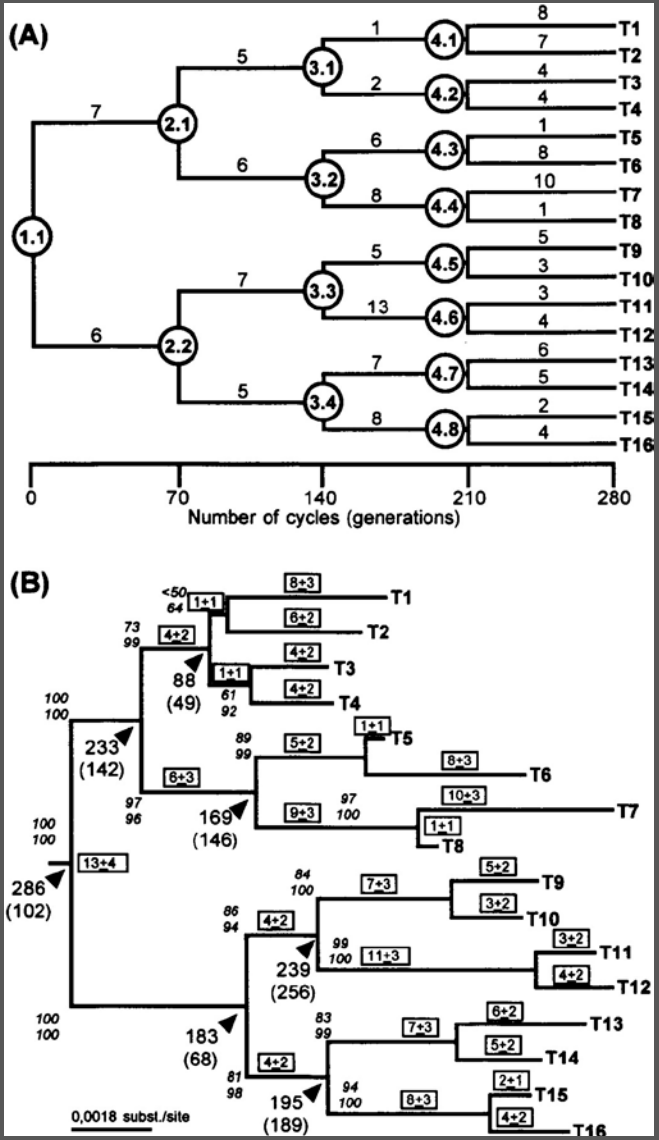
Evolution of DNA sequences by a series of bifurcate PCRs. An ancestor SSU rDNA cloned in pBluescript was used as template for series 1 of 70 nested PCR cycles with M13 primers. After the initial 35 cycles, reaction products were diluted 1:1,000 and used as templates for the subsequent 35 cycles, with rDNA primers RIBA and RIBB. After 70 cycles amplicons were cloned, and two clones were picked randomly and used as templates for the next series of nested PCR cycles. Lineages are propagated at random, and therefore the evolution is neutral and behaves as a stochastic process. Tree nodes T1 to T16 indicate terminal sequences, and 1.1 to 4.8, internal ancestors



Experimental Phylogeny of Neutrally Evolving DNA Sequences Generated by a Bifurcate Series of Nested Polymerase Chain Reactions
Gerdine F. O. Sanson, Silvia Y. Kawashita, Adriana Brunstein, and Marcelo R. S. Briones. Mol. Biol. Evol. 19(2):170–178. 2002

Comparison of real phylogeny with inferred maximum likelihood phylogeny. The serial PCR in vitro evolution resulted in the topology depicted (A).

The inferred phylogeny (B) has a topology identical to the real tree (A) and 9 out of 30 branch lengths were estimated correctly. It was inferred using a submodel of the General Stochastic-Model of Nucleotide Substitution (Rodríguez, Oliver, Marín & Medina. 1990), with parameter determination by hierarchical likelihood ratio tests.



Experimental Phylogeny of Neutrally Evolving DNA Sequences Generated by a Bifurcate Series of Nested Polymerase Chain Reactions
Gerdine F. O. Sanson, Silvia Y. Kawashita, Adriana Brunstein, and Marcelo R. S. Briones. Mol. Biol. Evol. 19(2):170–178. 2002