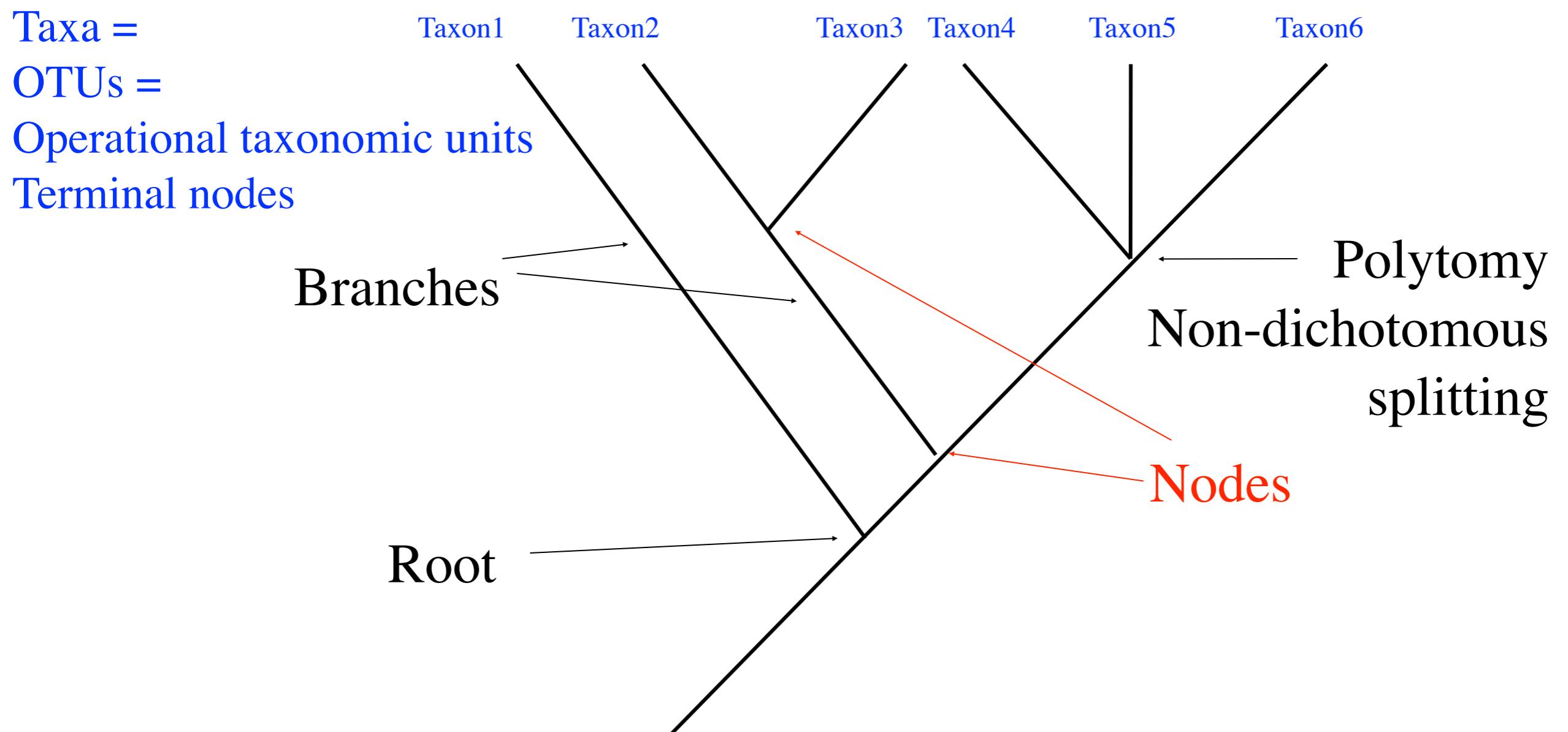


Week 4

Trees and Phylogenetics

Anatomy of a phylogenetic tree



Rooted and unrooted trees

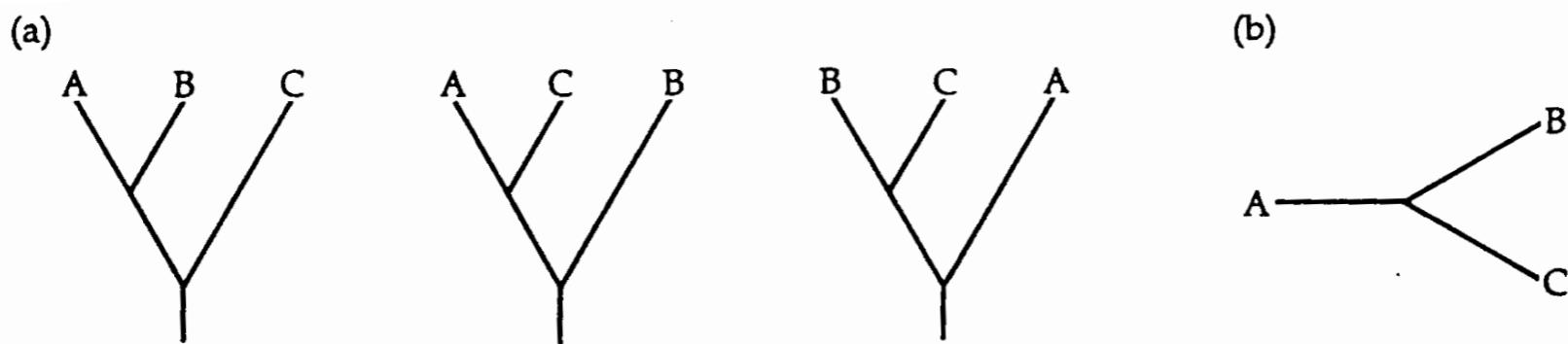


Figure 3. From three OTUs it is possible to construct three different rooted trees (a), but only one unrooted tree (b).

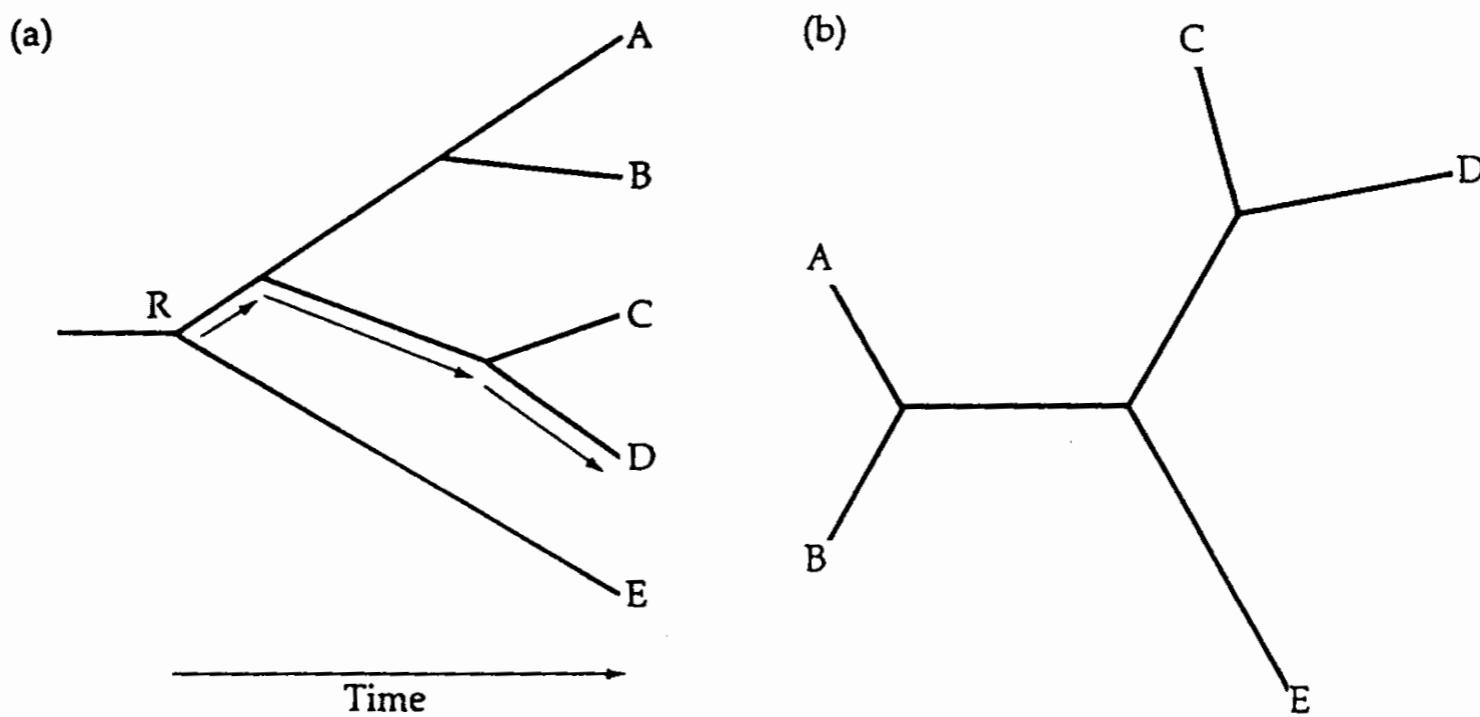


Figure 2. (a) Rooted and (b) unrooted phylogenetic trees. Arrows indicate the unique path leading from the root (R) to OTU D.

Tree shape vs. tree topology

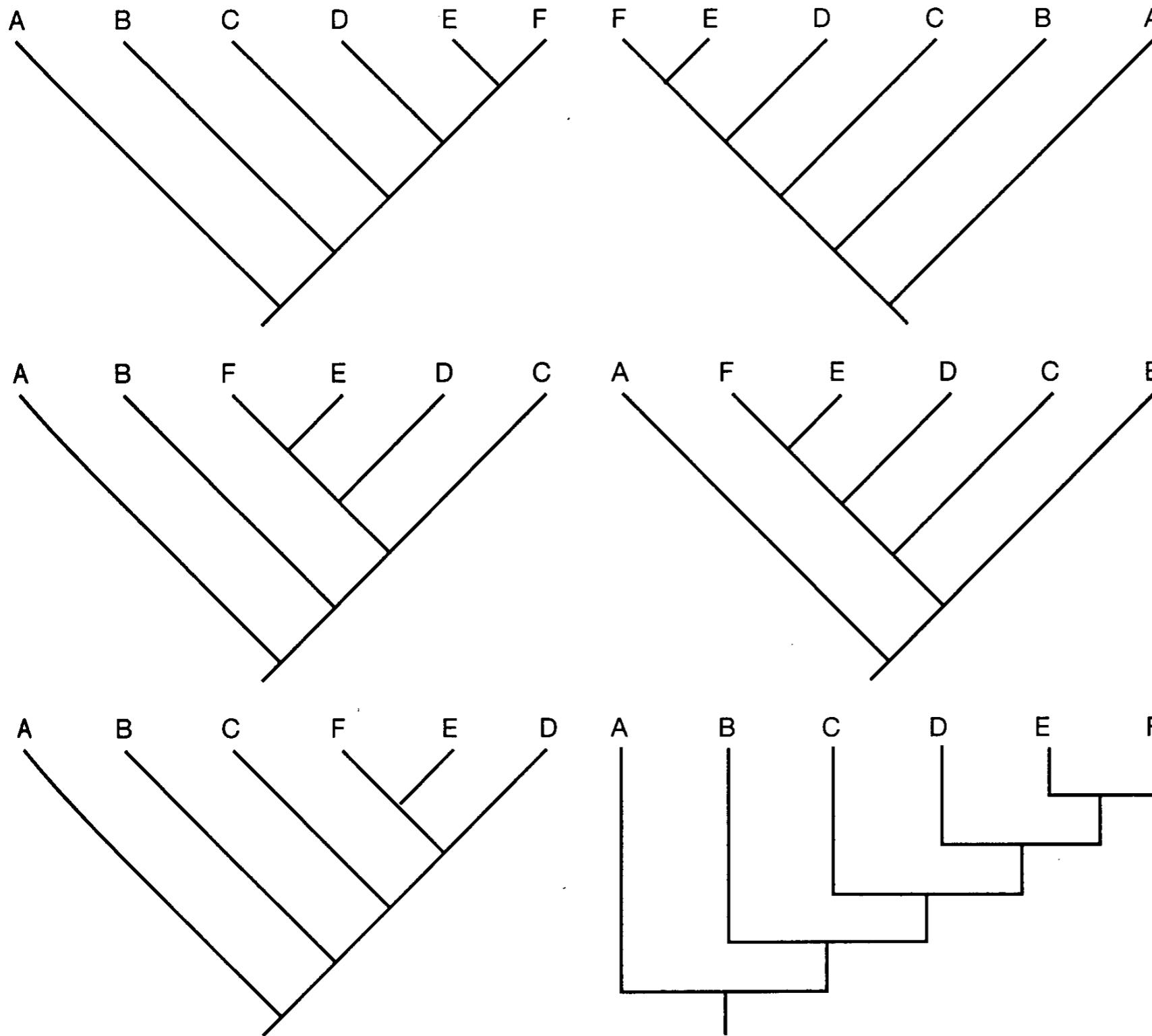


Figure 10.2 Alternative ways of drawing the same tree

These graphs show six ways to illustrate the same evolutionary relationships. These trees all happen to be oriented vertically, so that the basal taxon is at the bottom and derived groups are toward the top, but they could also be tipped 90° and presented horizontally. From Mayden and Wiley (1992).

Phylogenetic reconstruction: many possible trees from several taxa

$$N_R = \frac{(2n - 3)!}{2^{n-2}(n - 2)!} \quad (5.1)$$

The numbers of phylogenetic trees expands rapidly with the number of taxa (OTUs)

Note that rooted trees have the equivalent of one additional taxon, defined by the root

$$N_U = \frac{(2n - 5)!}{2^{n-3}(n - 3)!} \quad (5.2)$$

Table 1. Numbers of possible rooted and unrooted trees for 1–10 OTUs.

Number of OTUs	Number of rooted trees	Number of unrooted trees
2	1	1
3	3	1
4	15	3
5	105	15
6	954	105
7	10,395	954
8	135,135	10,395
9	2,027,025	135,135
10	34,459,425	2,027,025

From Felsenstein (1978).



Lemurs



Spider monkey



macaque

New
World
monkeys

Old
World
monkeys



Gibbon



Orang



Gorilla



Chimp



Bonobo



Australo-
pithecus



Homo

Primate Phylogeny

40 MYA

63 MYA

~6 MYA
~7 MYA Hominins
12-15 MYA Hominids
15-18 MYA Hominoids
25 MYA Catarhini

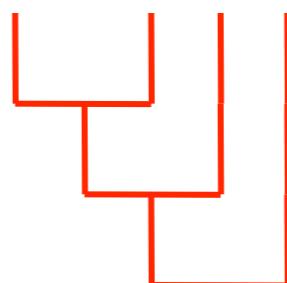
distance (phenetic) method of reconstructing a phylogenetic tree

- Collect as many characters as possible
- Construct a pair-wise distance between each taxon in your study
 - Euclidian (morphological)
 - Sequence divergence (molecular)
- Cluster taxa based on greatest overall similarity (smallest distance)
- Assumes homology outweighs analogy

Distance matrix

	Taxon1	Taxon2	Taxon3	Taxon4
--	--------	--------	--------	--------

Taxon1	0.0			
Taxon2	0.1	0.0		
Taxon3	0.2	0.3	0.0	
Taxon4	0.5	0.5	0.5	0.0



Distance phylogram

The distance method of UPGMA (unweighted pair group method with averages)

Pairs of taxa with the smallest distance score are clustered, and the remaining values are averaged to create a smaller table with one fewer columns e.g., human and chimp are paired, and the new distance of Gorilla to the average of human and chimp (HC) is 1.54 (= average of 1.51 & 1.57).

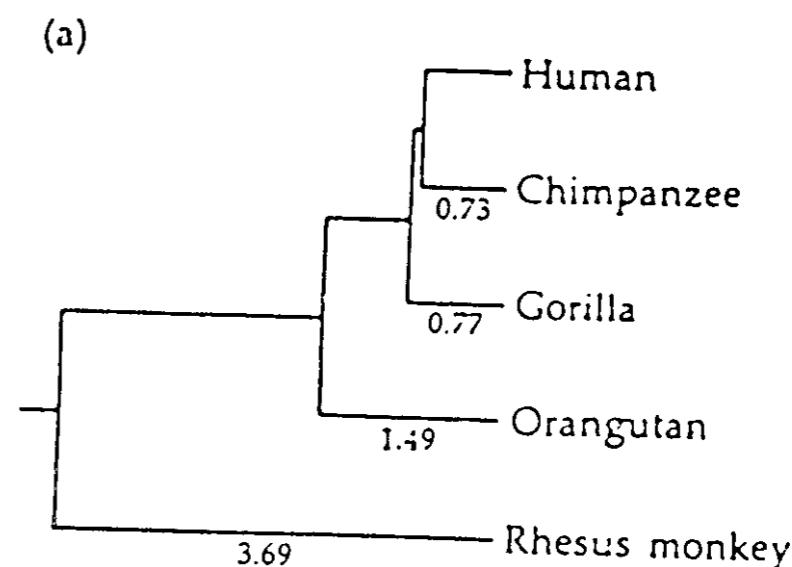
Table 2. Mean (below diagonal) and standard error (above diagonal) of the number of nucleotide substitutions per 100 sites between OTUs.^a

OTU	OTU				
	Human	Chimpanzee	Gorilla	Orangutan	Rhesus monkey
Human		0.17	0.18	0.25	0.41
Chimpanzee	1.45		0.18	0.25	0.42
Gorilla	1.51	1.57		0.26	0.41
Orangutan	2.98	2.94	3.04		0.40
Rhesus monkey	7.51	7.55	7.39	7.10	

From Li et al. (1987b).

^a The sequence data used are 5.3 kb of noncoding DNA, which is made up of two separate regions: (1) the η -globin locus (2.2 kb) described by Koop et al. (1986b) and (2) 3.1 kb of the η - δ globin intergenic region sequenced by Maeda et al. (1983, 1988).

OTU	OTU		
	(HC)	G	O
G	1.54		
O	2.96	3.04	
R	7.53	7.39	7.10



UPGMA

	Human	Chimp	Gorilla	Orangutan	Rhesus
Human	-				
Chimp	1.45	-			
Gorilla	1.51	1.57	-		
Orangutan	2.98	2.94	3.04	-	
Rhesus	7.51	7.55	7.39	7.10	-

- 1) Find Smallest Cell in distance matrix
- 2) Collapse those taxa into node
- 3) Recompute distance matrix
- 4) Repeat until one node left

distances between new node n and existing node i

$$d(n, i) = \frac{(d(a, i) + d(b, i))}{2}$$

UPGMA

	Human	Chimp	Gorilla	Orangutan	Rhesus
Human	-				
Chimp	1.45	-			
Gorilla	1.51	1.57	-		
Orangutan	2.98	2.94	3.04	-	
Rhesus	7.51	7.55	7.39	7.10	-

- 1) Find Smallest Cell in distance matrix
- 2) Collapse those taxa into node
- 3) Recompute distance matrix
- 4) Repeat until one node left

distances between new node n and existing node i

$$d(n, i) = \frac{(d(a, i) + d(b, i))}{2}$$

UPGMA

	(H,C)	Gorilla	Orangutan	Rhesus
(H,C)	-			
Gorilla	1.54	-		
Orangutan	2.96	3.04	-	
Rhesus	7.53	7.39	7.10	-

- 1) Find Smallest Cell in distance matrix
- 2) Collapse those taxa into node
- 3) Recompute distance matrix
- 4) Repeat until one node left

distances between new node n and existing node i

$$d(n, i) = \frac{(d(a, i) + d(b, i))}{2}$$

UPGMA

	((H,C),G)	Orangutan	Rhesus
((H,C),G)	-		
Orangutan	3.0	-	
Rhesus	7.46	7.10	-

- 1) Find Smallest Cell in distance matrix
- 2) Collapse those taxa into node
- 3) Recompute distance matrix
- 4) Repeat until one node left

distances between new node n and existing node i

$$d(n, i) = \frac{(d(a, i) + d(b, i))}{2}$$

UPGMA

	((H,C),G),O	Rhesus
((H,C),G),O	-	
Rhesus	7.28	-

- 1) Find Smallest Cell in distance matrix
- 2) Collapse those taxa into node
- 3) Recompute distance matrix
- 4) Repeat until one node left

distances between new node n and existing node i

$$d(n, i) = \frac{(d(a, i) + d(b, i))}{2}$$

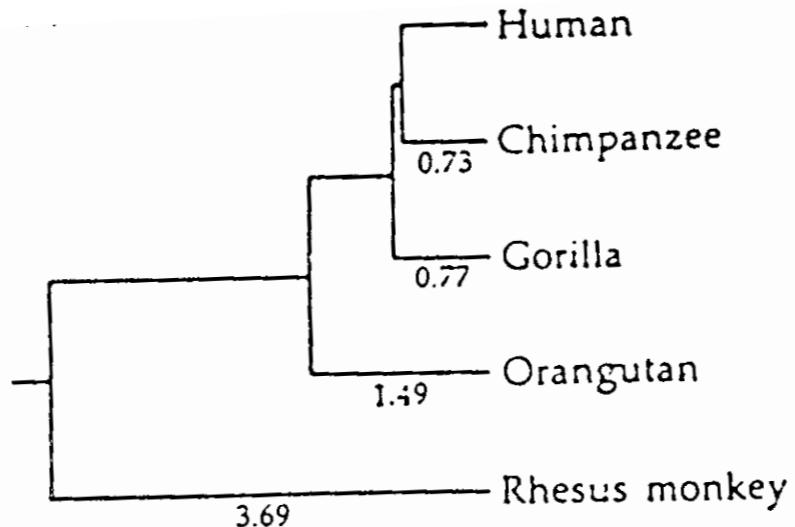
UPGMA

	((H,C),G),O	Rhesus
((H,C),G),O	-	
Rhesus	7.28	-

- 1) Find Smallest Cell in distance matrix
- 2) Collapse those taxa into node
- 3) Recompute distance matrix
- 4) Repeat until one node left

So final tree:

$((H,C),G),O$



Neighbor Joining

	Human	Chimp	Gorilla	Orangutan	Rhesus
Human	-				
Chimp	1.45	-			
Gorilla	1.51	1.57	-		
Orangutan	2.98	2.94	3.04	-	
Rhesus	7.51	7.55	7.39	7.10	-

- 1) Find Smallest Cell in **transformed** distance matrix
- 2) Collapse those taxa into node
- 3) Recompute distance matrix
- 4) Repeat until one node left

distances between new node n and existing node i

$$d(n, i) = \frac{(d(a, i) + d(b, i) - d(a, b)}{2}$$

Neighbor Joining

	Human	Chimp	Gorilla	Orangutan	Rhesus
Human	-				
Chimp	1.45	-			
Gorilla	1.51	1.57	-		
Orangutan	2.98	2.94	3.04	-	
Rhesus	7.51	7.55	7.39	7.10	-

- 1) Find Smallest Cell in **transformed** distance matrix

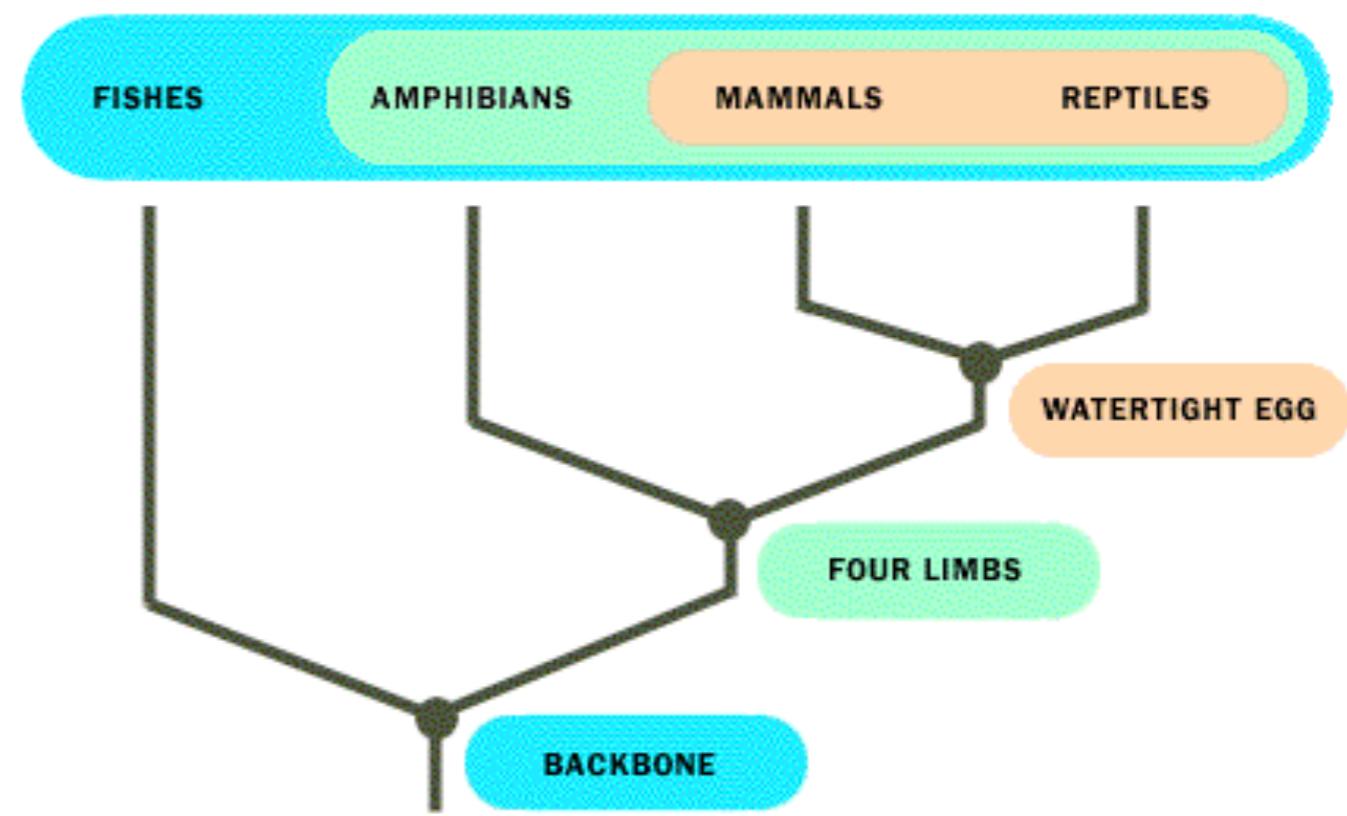
Major advantage is don't assume rates same across branches

$$Q(i, j) = (n - 2) \times d(i, j) - \sum_{l=1}^k d(i, l) - \sum_{l=1}^k d(j, l)$$

Calculate UPGMA trees in jupyter notebook

Cladistic method of reconstructing a phylogenetic tree

- Collect characters and character states
- Build a character state matrix
- Cluster organisms according to patterns of shared-derived characters
- Use parsimony to determine best tree



http://www.amnh.org/Exhibition/Fossil_Halls/cladistics.html



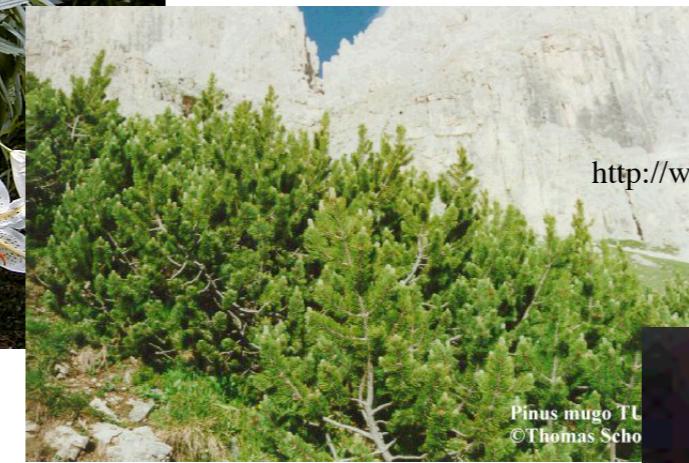
<http://www.plant-pictures.com/>



Dicots

Monocots

Gymnosperms



<http://www.home.aone.net.au/byzantium/ferns/about.html>



Flowers

Seeds

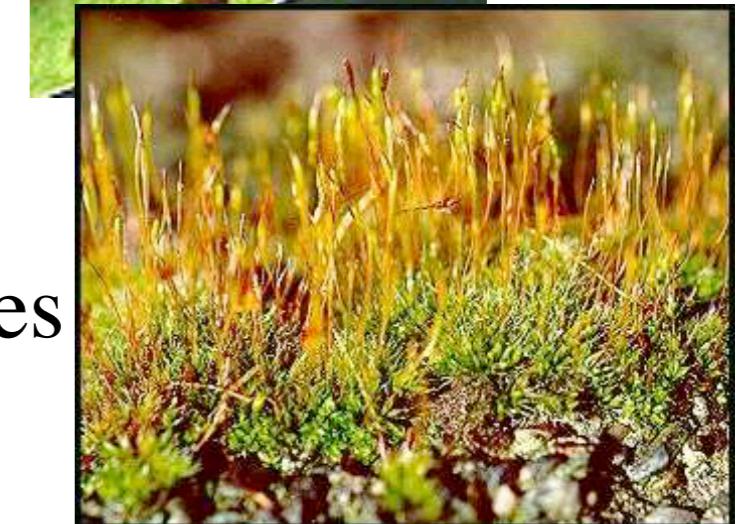
Vascular
Conducting
system

Chlorophyll

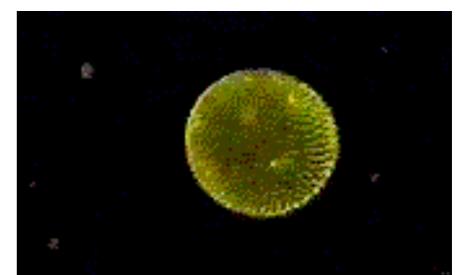
ferns

Bryophytes
(mosses)

Green Algae



<http://home.clara.net/adhale/bryos/phframe.htm>



<http://www.microscopy-uk.org.uk/mag/art97b/volvoxms.html>

Characters and Character states

- Character ~ trait
- Character state = alternative forms of a trait
- Eye color = character
- Blue, brown = character states

Uninformative = invariant, or in a single taxon
Conflicting = cluster taxa in a way that conflicts with clustering of other characters

Informative = both states shared by ≥ 2 taxa

Character	A	B	C	D	E
Species 1	1	0	0	0	0
Species 2	1	1	1	0	0
Species 3	1	0	1	1	1
Species 4	<u>1</u>	0	0	<u>1</u>	<u>1</u>

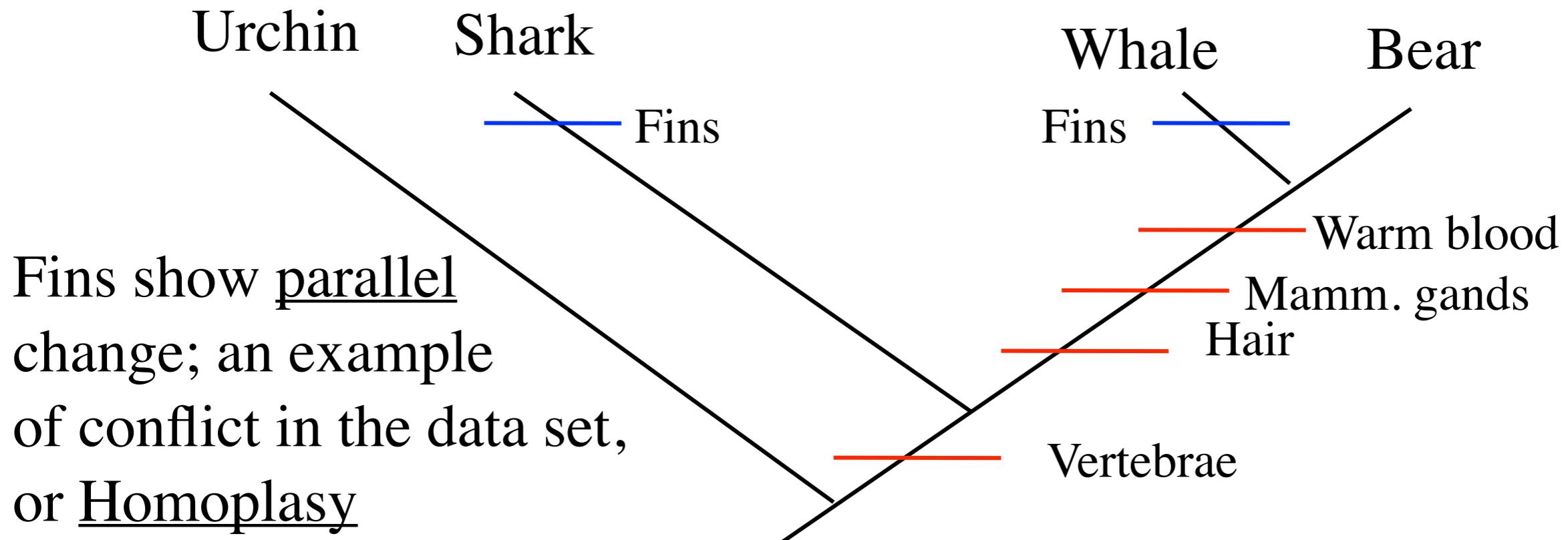
Uninformative characters

conflicting character

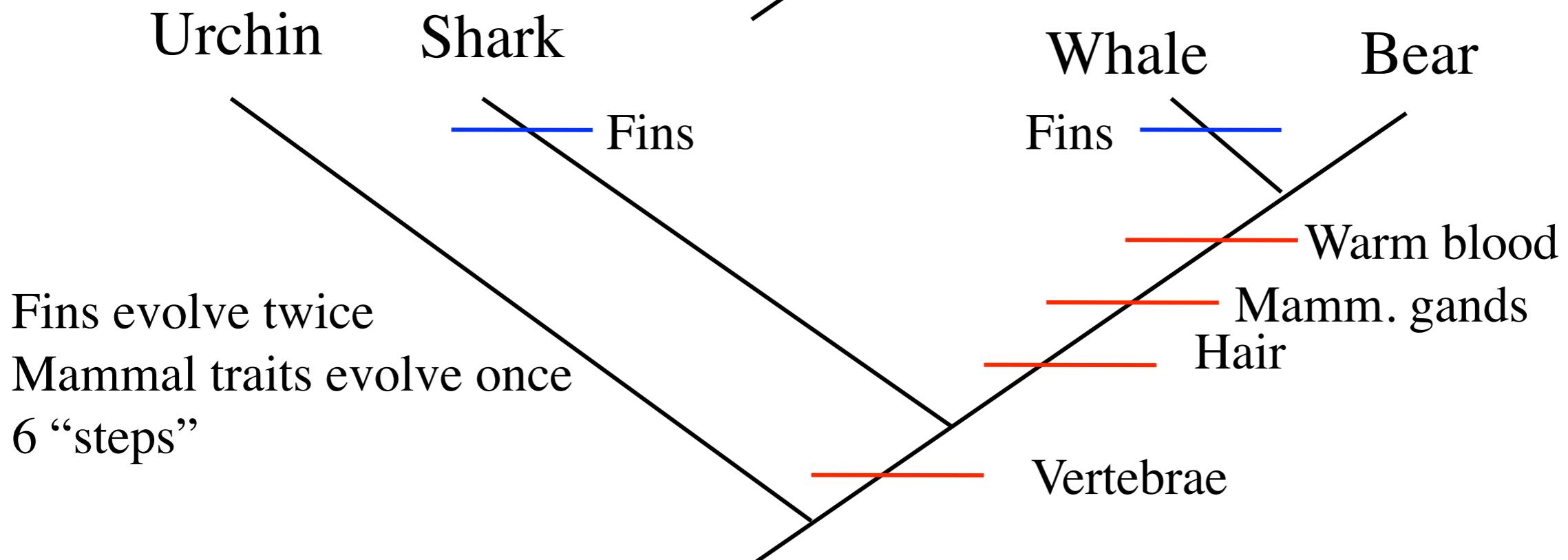
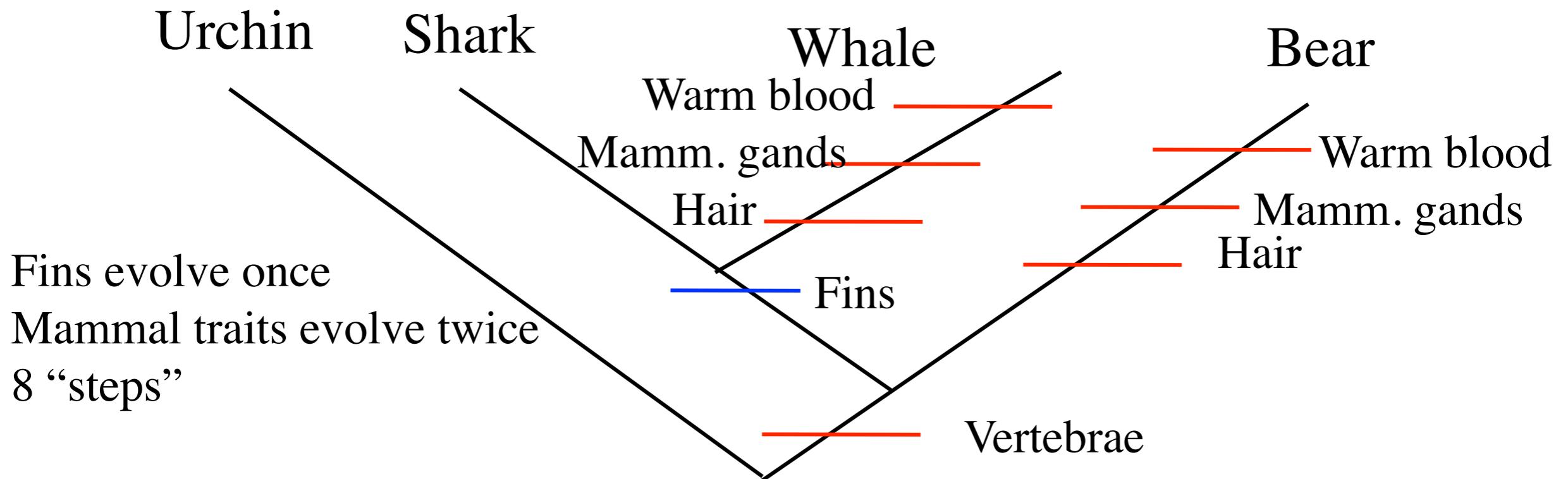
Informative characters

Parsimony example

	Vertebrae	Hair	Mammary glands	Warm blood	Fins
Urchin	0	0	0	0	0
Shark	1	0	0	0	1
Whale	1	1	1	1	1
Bear	1	1	1	1	0



Different Trees yield different interpreted evolutionary histories



Building Trees with Parsimony

- **Parsimony** involves evaluating all possible tree topologies and giving each a **Length** score based on the total number of evolutionary changes that are needed to explain the observed data.
- The best tree is the one that requires the fewest hypothesized character state changes for all characters to account for all changes derived from a common ancestor.

(A)

a	b	c	d	e	f	g	h	i	j
0	0	1	1	1	0	0	0	0	0

Species 1

a	b	c	d	e	f	g	h	i	j
0	0	1	0	0	1	1	1	1	0

Species 2

a	b	c	d	e	f	g	h	i	j
0	0	1	0	0	1	1	0	0	1

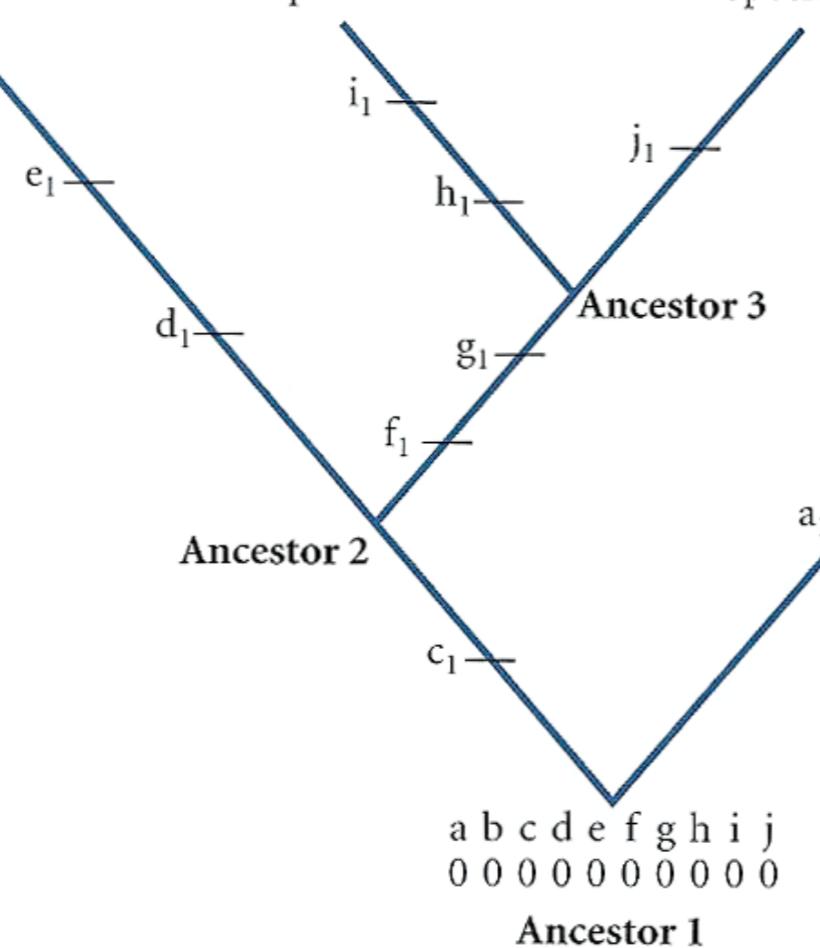
Species 3

a	b	c	d	e	f	g	h	i	j
1	1	0	0	0	0	0	0	0	0

Species 4 (outgroup)

The character state (0=ancestral, 1=derived) is shown for each of 10 characters a-j.

Time ↑



Shared character states

	1	2	3	4
1	-	4	5	5
2	-	7	3	
3		-	4	

Species 1 shares 5 states (f-j) with species 4...

Shared derived character states

	1	2	3	4
1	-	1	1	0
2	-	3	0	
3		-	0	

...but they have no derived states in common.

(B)

a	b	c	d	e	f	g	h	i	j
0	0	1	1	0	0	0	0	0	0

a	b	c	d	e	f	g	h	i	j
0	0	1	0	1	1	1	1	1	1

a	b	c	d	e	f	g	h	i	j
0	0	1	0	1	1	0	0	0	0

a	b	c	d	e	f	g	h	i	j
1	1	0	0	0	0	0	0	0	0

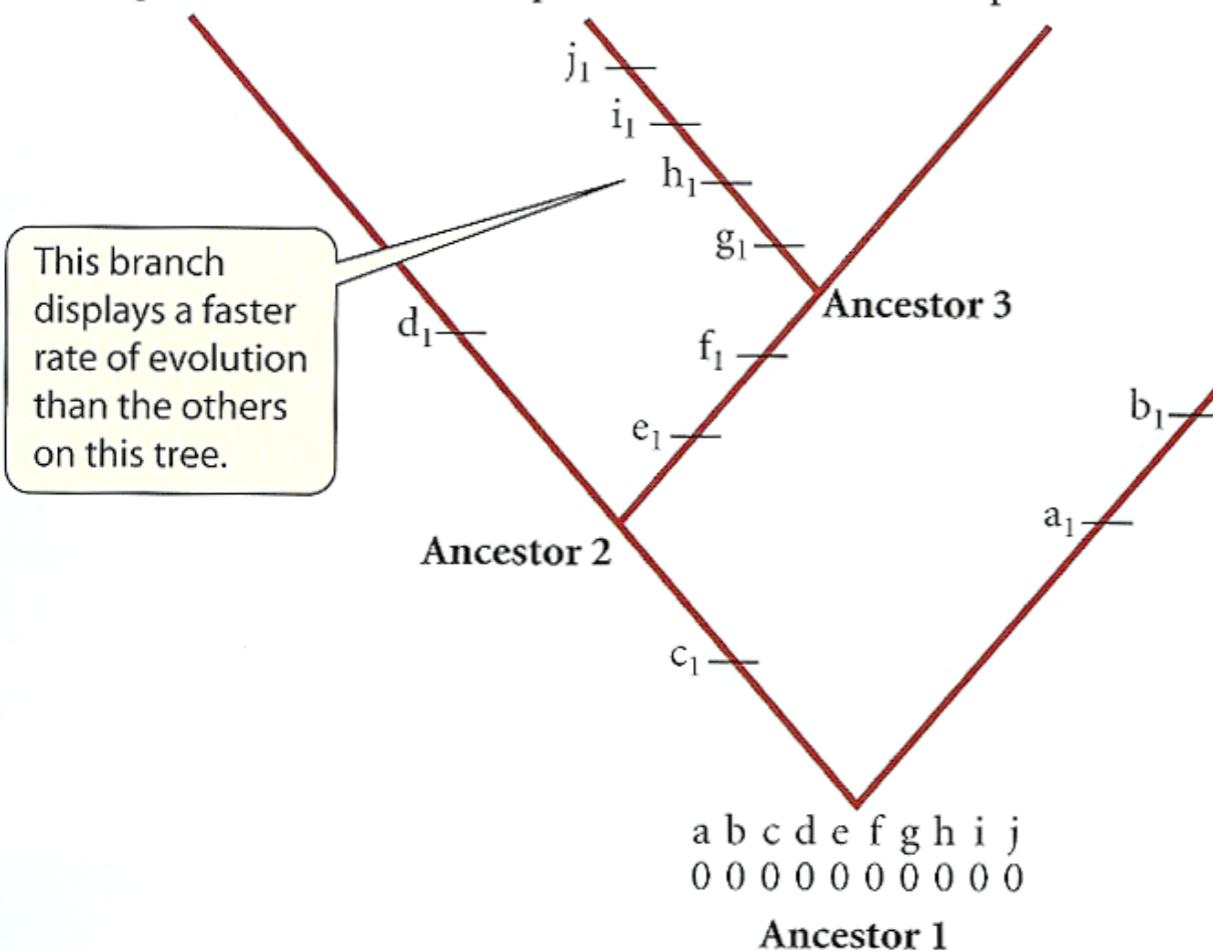
Species 1

Species 2

Species 3

Species 4 (outgroup)

Time ↑



Shared character states

	1	2	3	4
1	-	3	7	6
2		-	6	1
3			-	5

Species 1 and 3 share the most character states...

Shared derived character states

	1	2	3	4
1	-	1	1	0
2		-	3	0
3			-	0

...but species 2 and 3 share the most derived character states, and thus are the most closely related.

(C)

a	b	c	d	e	f	g	h	i	j
0	1	1	0	0	0	1	1	0	1

Species 1

a	b	c	d	e	f	g	h	i	j
0	1	1	1	1	1	0	1	0	0

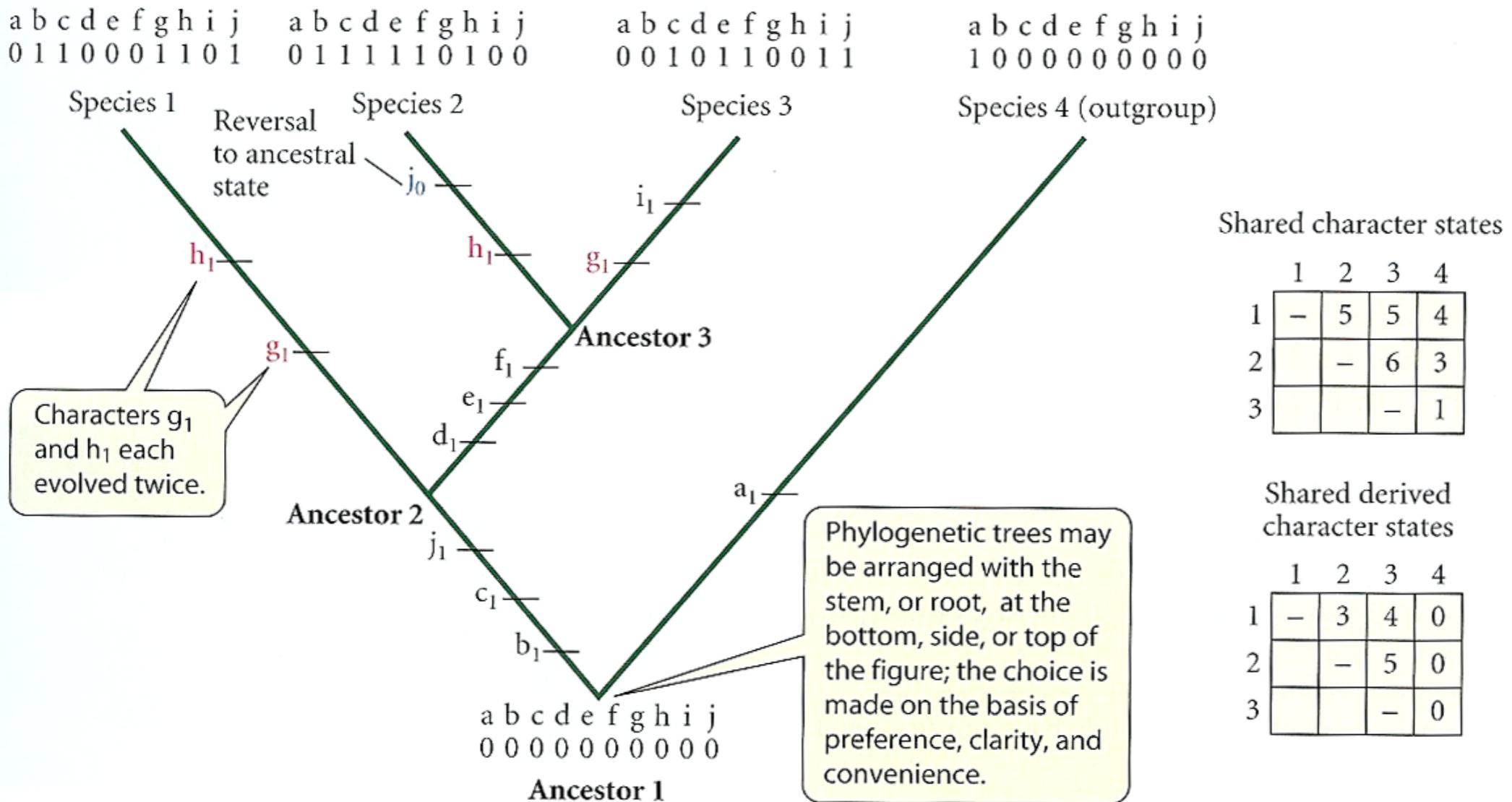
Species 2

a	b	c	d	e	f	g	h	i	j
0	0	1	0	1	1	0	0	1	1

Species 3

a	b	c	d	e	f	g	h	i	j
1	0	0	0	0	0	0	0	0	0

Species 4 (outgroup)



Shared character states

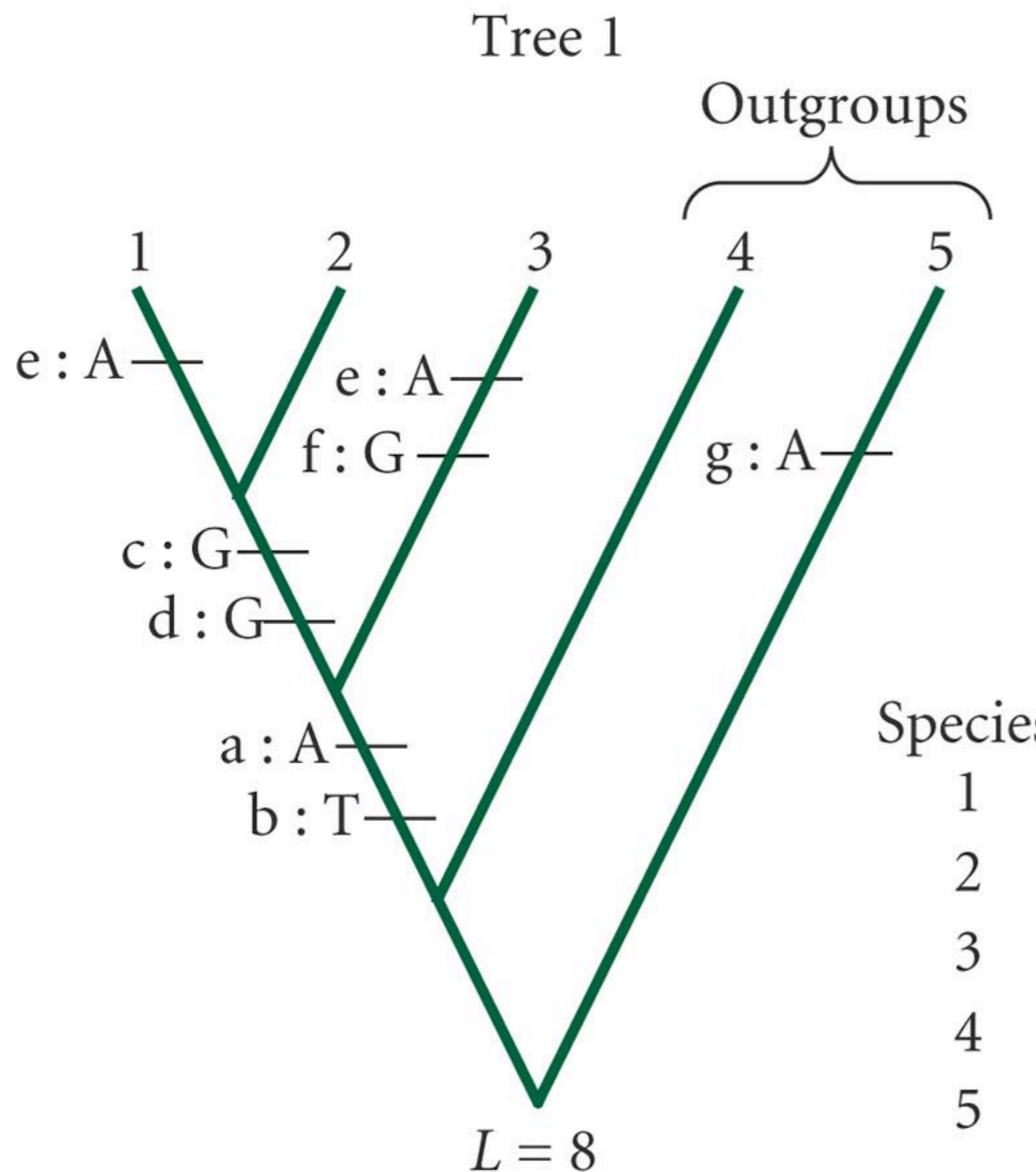
	1	2	3	4
1	-	5	5	4
2		-	6	3
3			-	1

Shared derived character states

	1	2	3	4
1	-	3	4	0
2		-	5	0
3			-	0

How else could characters g, h and j have evolved on this tree, and have the change still be most parsimonious?

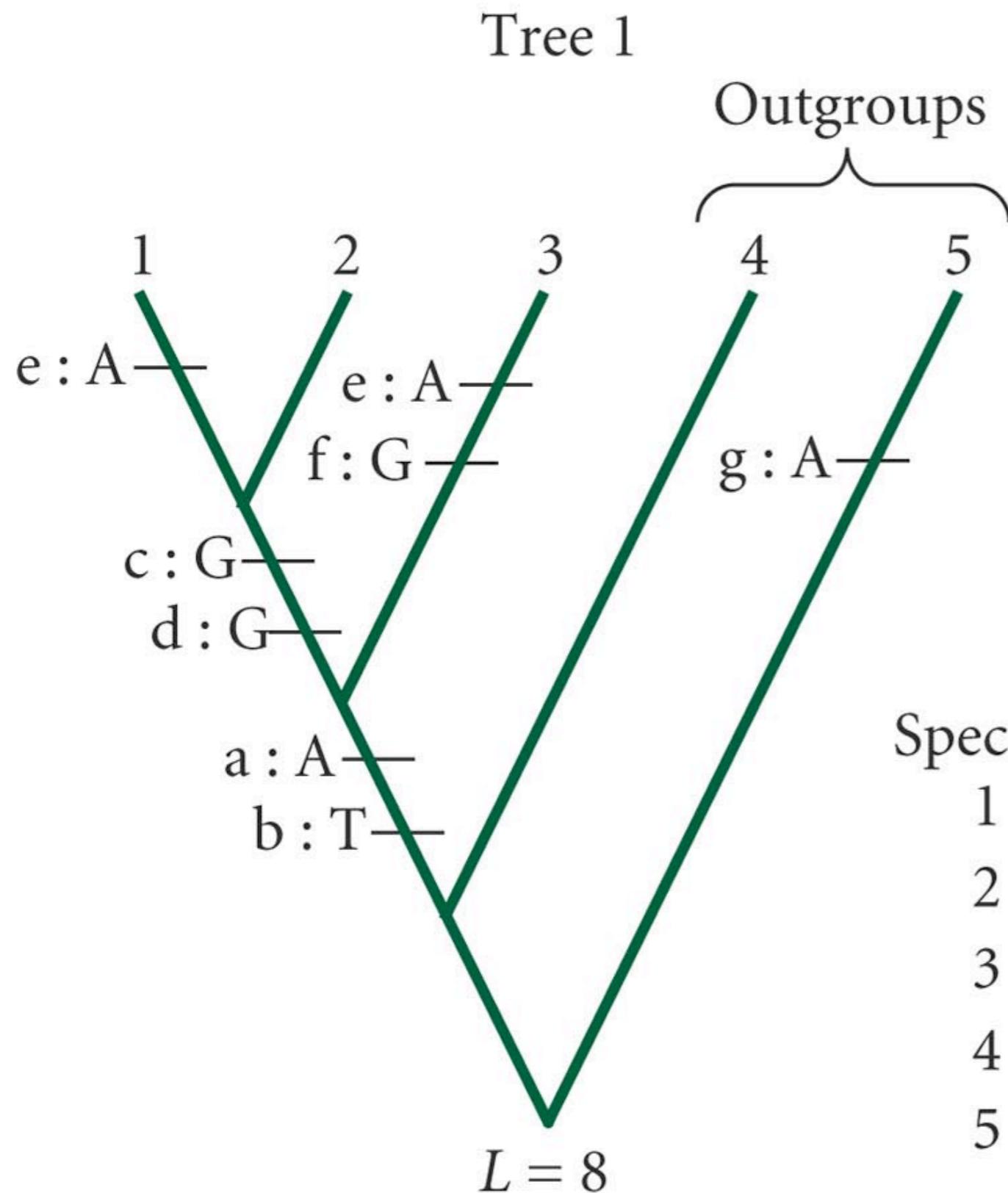
Inferring a phylogeny by the method of maximum parsimony



First an **outgroup** must be defined. The outgroup character states are assumed to be ancestral for the **ingroup** in determining which are shared, derived character states.

Species	a	b	c	d	e	f	g
1	A	T	G	G	A	C	T
2	A	T	G	G	T	C	T
3	A	T	T	C	A	G	T
4	C	G	T	C	T	C	T
5	C	G	T	C	T	C	A

Inferring a phylogeny by the method of maximum parsimony

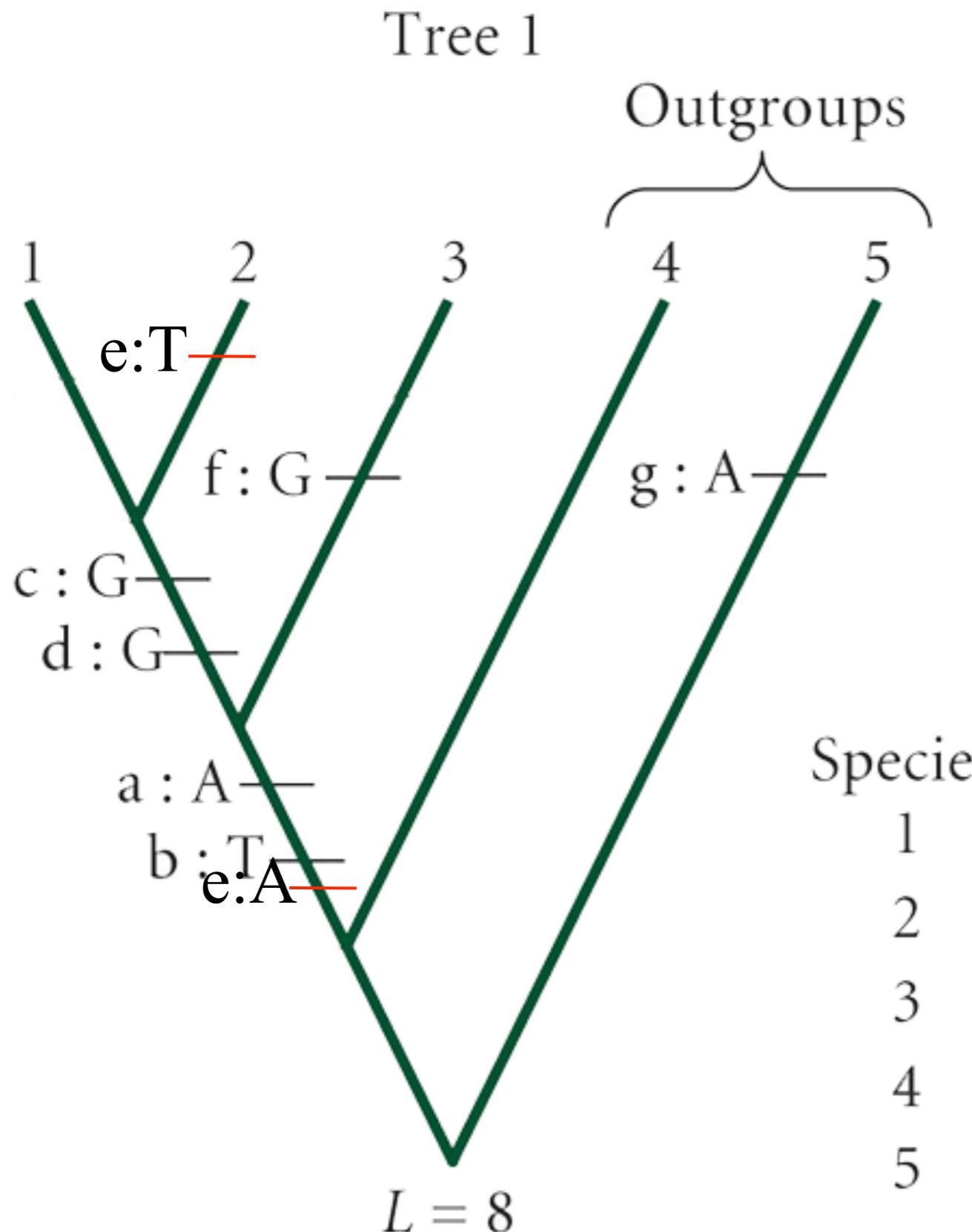


Note that two evolutionary events must be inferred for this homoplasious character on this tree.

Character

Species	a	b	c	d	e	f	g
1	A	T	G	G	A	C	T
2	A	T	G	G	T	C	T
3	A	T	T	C	A	G	T
4	C	G	T	C	T	C	T
5	C	G	T	C	T	C	A

Inferring a phylogeny by the method of maximum parsimony

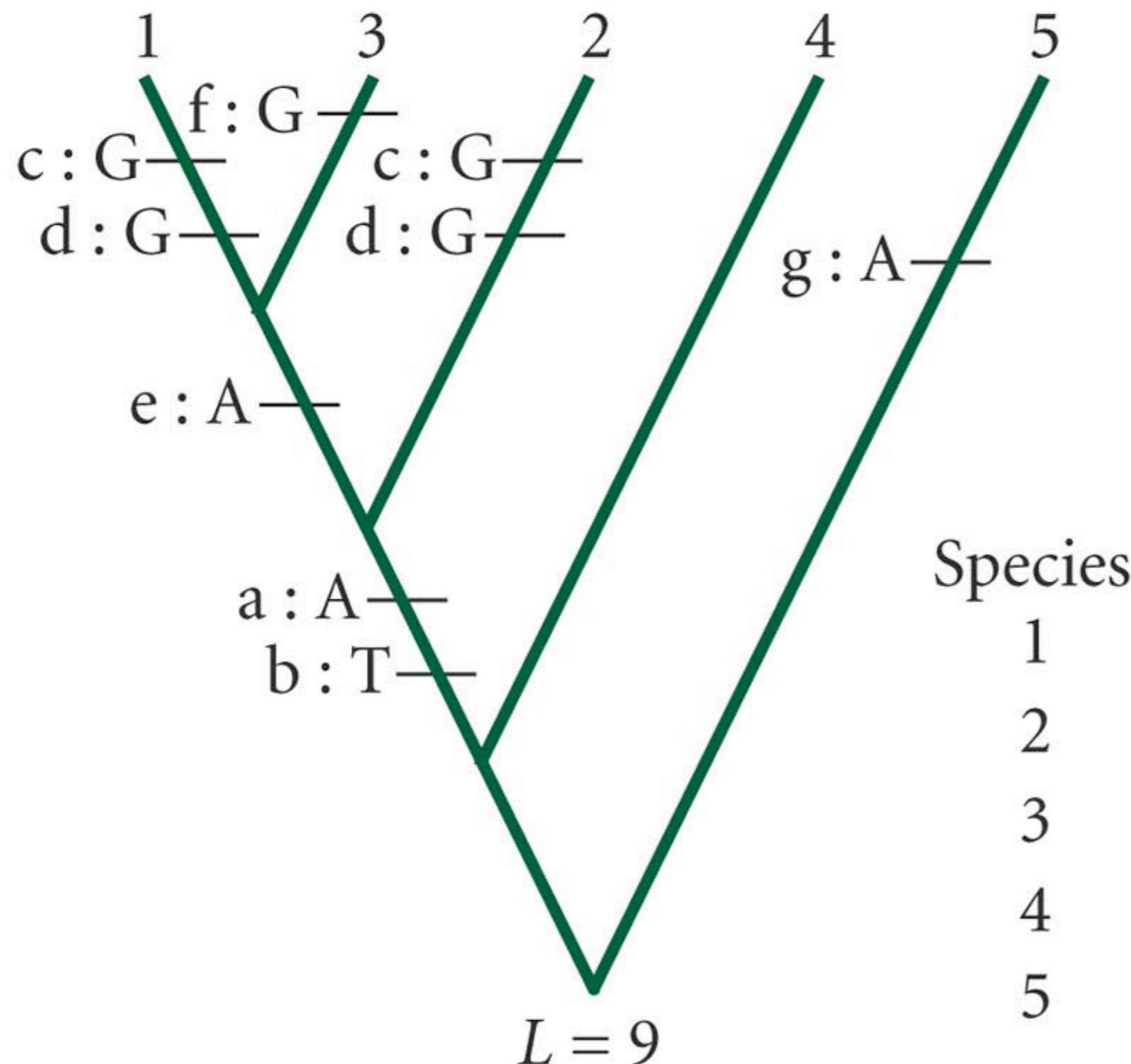


But an equally parsimonious evolutionary scenario would be a gain and loss.
Either way, two character state changes must be inferred.

Species	a	b	c	d	e	f	g
1	A	T	G	G	A	C	T
2	A	T	G	G	T	C	T
3	A	T	T	C	A	G	T
4	C	G	T	C	T	C	T
5	C	G	T	C	T	C	A

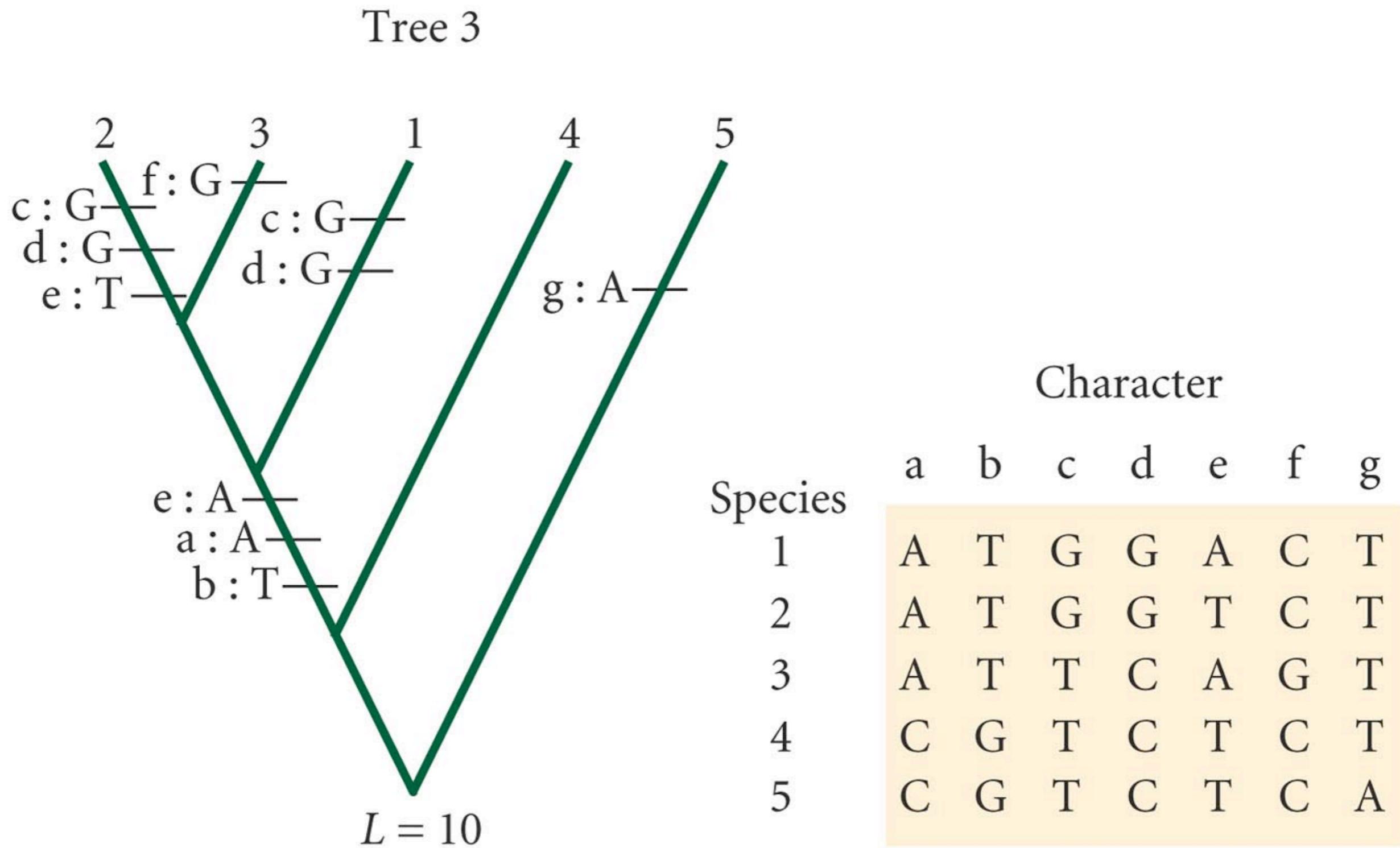
Inferring a phylogeny by the method of maximum parsimony

Tree 2

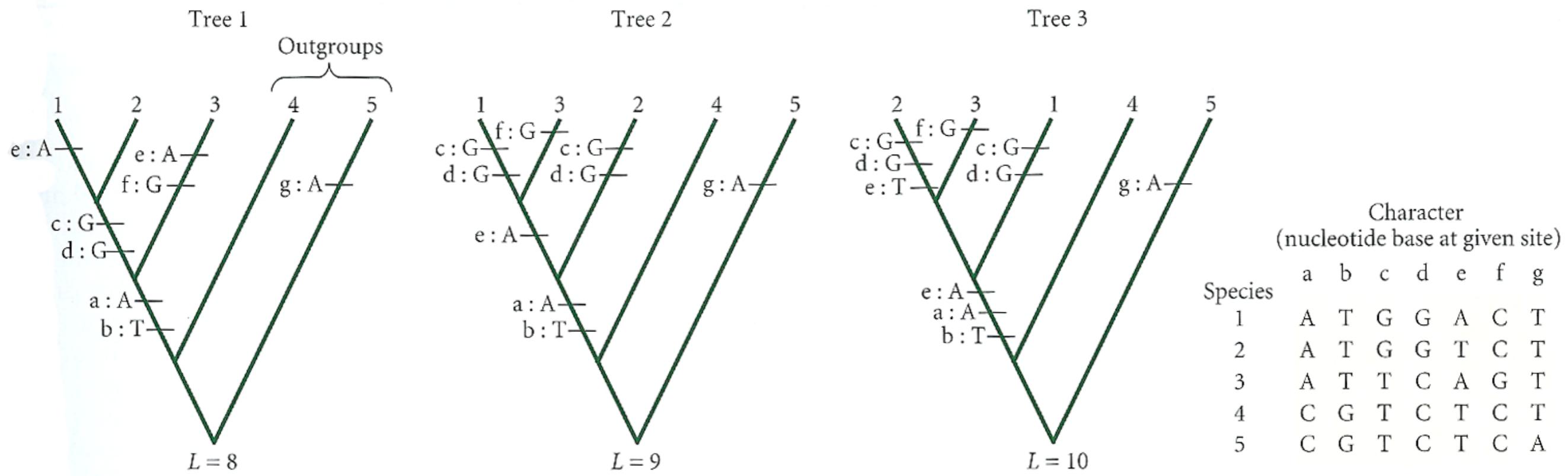


Species	Character						
	a	b	c	d	e	f	g
1	A	T	G	G	A	C	T
2	A	T	G	G	T	C	T
3	A	T	T	C	A	G	T
4	C	G	T	C	T	C	T
5	C	G	T	C	T	C	A

Inferring a phylogeny by the method of maximum parsimony



Inferring a phylogeny by the method of maximum parsimony



For these data, Tree 1 is **Most Parsimonious** because the fewest character state changes must be hypothesized.