

Parsimony

Sequence Data

Taxa

Characters (=sites in this case)

Agave	G G GCTGGGG C -CGGCCGGTCC C GC-CTCT...
Bougainvillia	G G GGTGGGG T -CGACCGGTCC C GC-CTTG...
Ceratophyllum	G G GTTGGGG C -CGGTCGGTCC G TCACTCT...
Digitalis	G G GTTGGGG T -CGGCCGGTCC C GC-CTTT...
Epilobium	G G GTTGGGG T -CGACCGGTCC C GC-CTTA...
Fagus	G G GTTGGGG C -AGAGCGGTCC C GC-CCCT...
Galax	G G GTTGG ? C -CG ? CCGGTCC C GC-CTAG...

parsimony informative.

*2 states
each in
at least*

2 taxa

constant site

variable site

parsimony informative site

parsimony uninformative site

Discrete Morphology

Taxa

Characters

P._fimbriata	00000000 1 101010110?0 1 ...
P._robusta	00000000 1 10101011000 1 ...
P._articulata	10011101 0 11000000110 0 ...
P._parksii	00011101 2 11000000100 0 ...
P._americana	10011101 0 11110000100 0 ...
P._myriophylla	10011101 0 11100000100 0 ...
P._macrophylla	11011101 2 10000000000 0 ...
P._polygama	11011101 3 10010000000 0 ...
P._gracilis	10111101 3 00000100011 0 ...
P._ciliata	00111011 2 00000100011 0 ...
P._basiramia	00111011 2 00000100011 0 ...

tomentose (hairy)



glabrous (smooth)



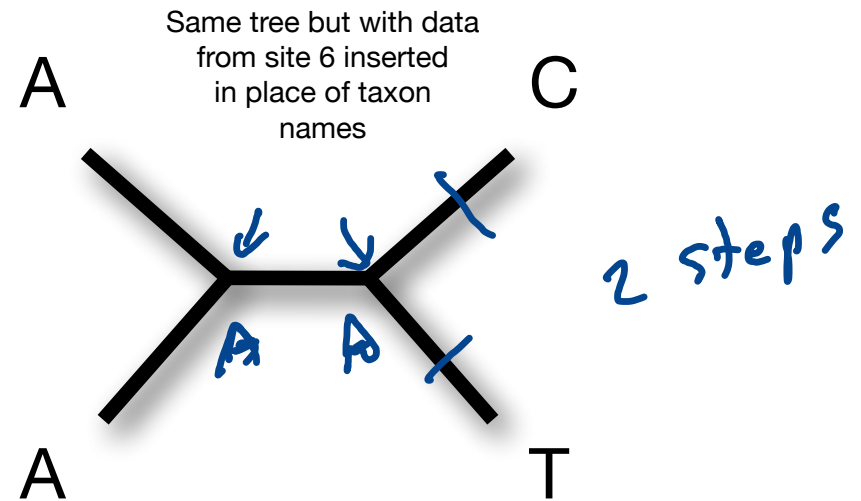
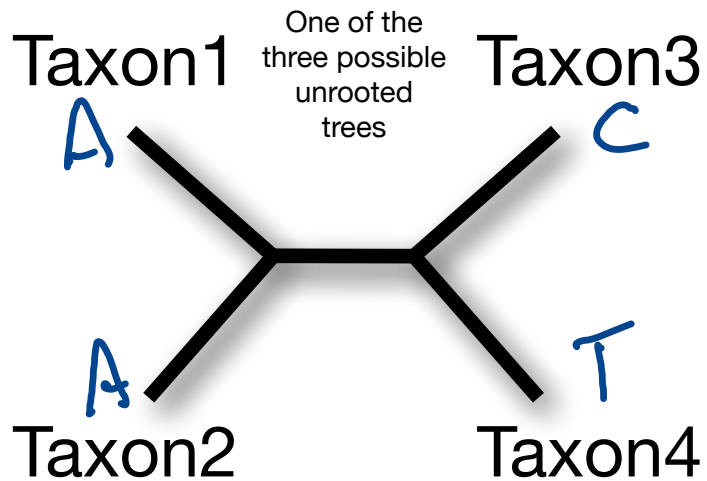
unrelated

binary (2-state) site

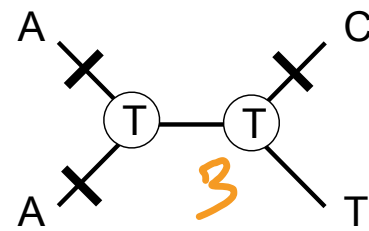
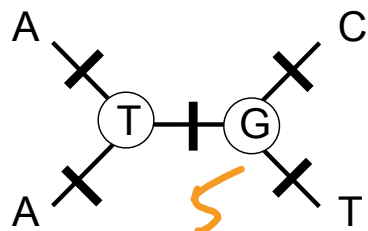
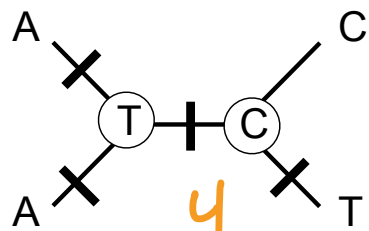
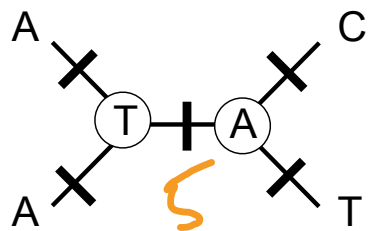
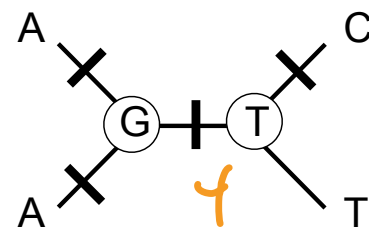
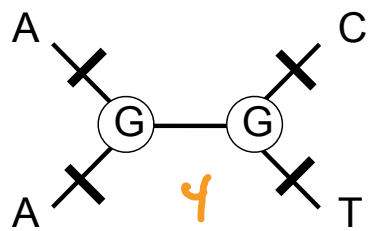
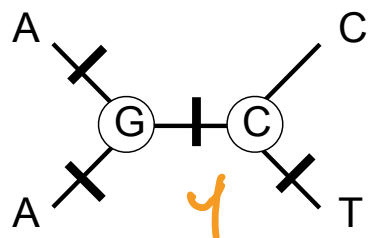
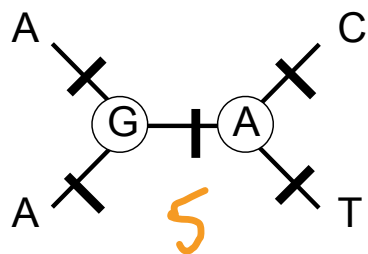
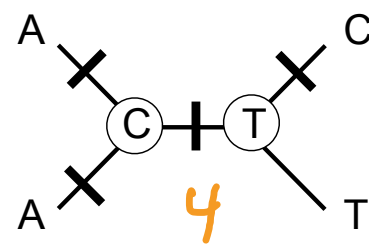
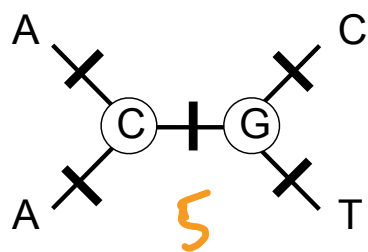
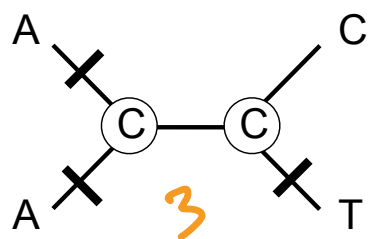
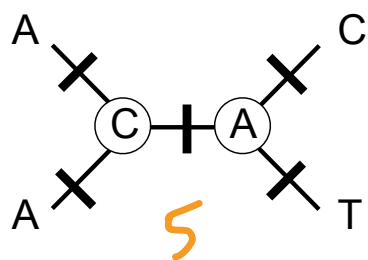
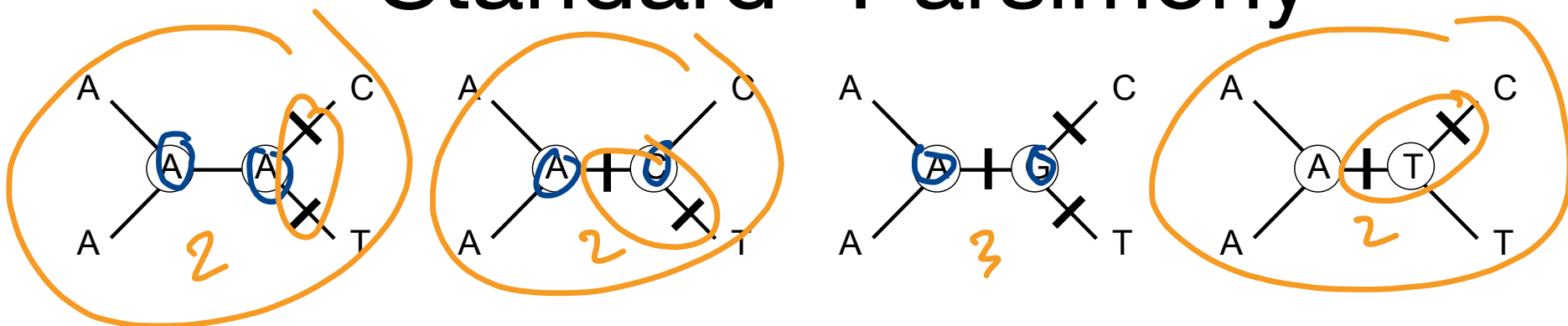
multistate site

Parsimony

	1	2	3	4	5	6	7	8	9	...
Taxon1	C	G	A	C	C	A	G	G	T	...
Taxon2	C	G	A	C	C	A	G	G	T	...
Taxon3	C	G	G	T	C	C	G	G	T	...
Taxon4	C	G	G	C	C	T	G	G	T	...



"Standard" Parsimony

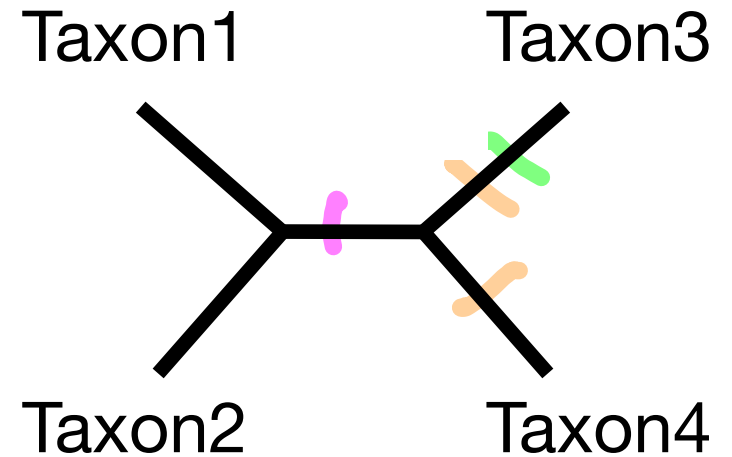


2 parsimony score?

number of most parsimonious reconstructions?

Parsimony Steps

	1	2	3	4	5	6	7	8	9	...
Taxon1	C	G	A	C	C	A	G	G	T	...
Taxon2	C	G	A	C	C	A	G	G	T	...
Taxon3	C	G	G	T	C	C	G	G	T	...
Taxon4	C	G	G	C	T	G	G	T	T	...
Steps	0	0	1	1	0	2	0	0	0	...

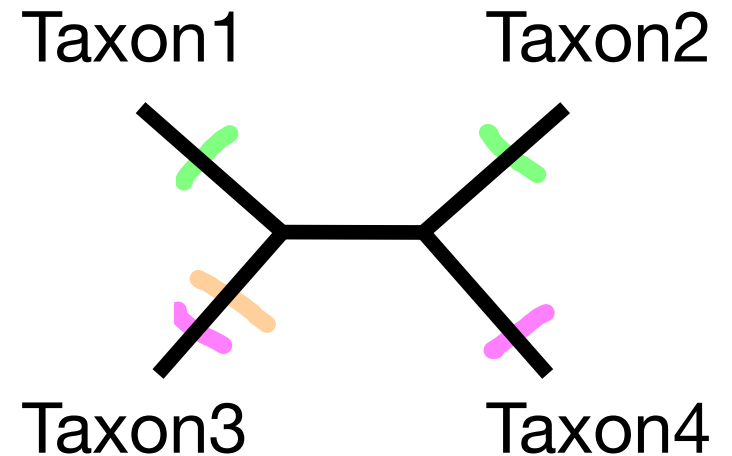


Let's call this tree 1:
(1,2,(3,4))

Tree 1's "length" for first 9 sites = 4

Parsimony Steps

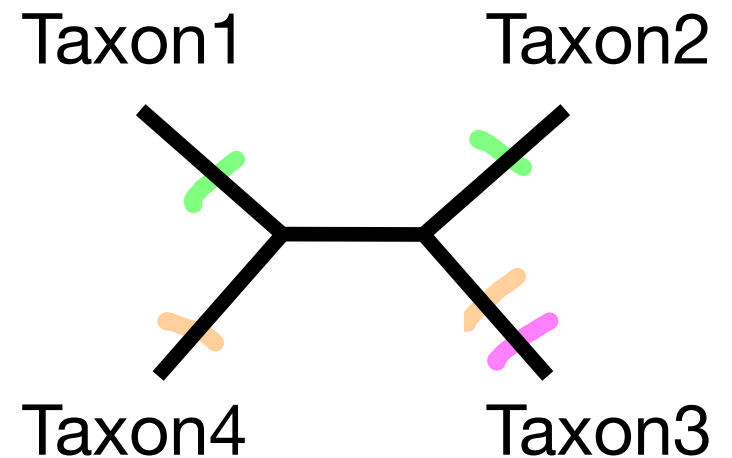
	1	2	3	4	5	6	7	8	9	...
Taxon1	C	G	A	C	A	G	G	T	...	
Taxon2	C	G	A	C	A	G	G	T	...	
Taxon3	C	G	G	T	C	G	G	T	...	
Taxon4	C	G	G	C	T	G	G	T	...	
Steps	0	0	2	1	0	2	0	0	0	...



Tree 2's length for first 9 sites = **5**

Parsimony Steps

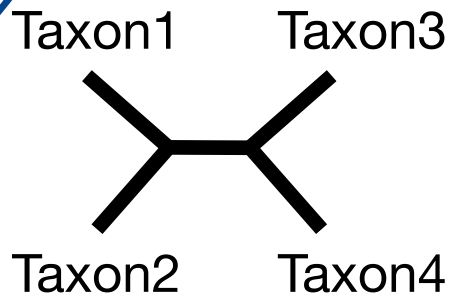
	1	2	3	4	5	6	7	8	9	...
Taxon1	C	G	A	C	A	G	G	T	...	
Taxon2	C	G	A	C	A	G	G	T	...	
Taxon3	C	G	G	T	C	C	G	G	T	...
Taxon4	C	G	G	C	T	G	G	T	...	
<hr/> Steps	0	0	2	1	0	2	0	0	0	...



Tree 3: (1,4,(2,3))

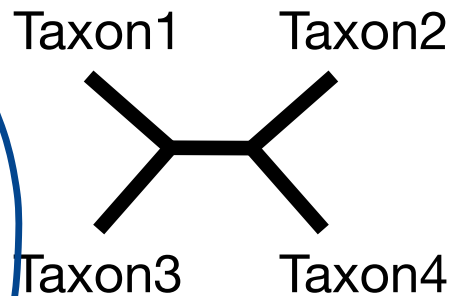
Tree 3's length for first 9 sites = 5

Parsimony (using only 9 sites)

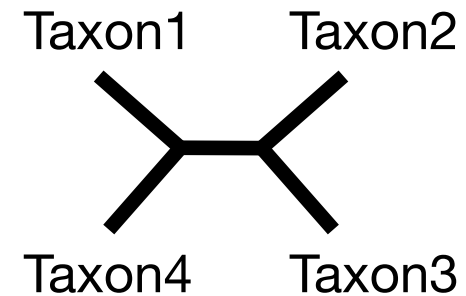


TREE LENGTH
= 4 steps

most
parsimonious



5 steps



5 steps

This is the simplest explanation of the data for the first 9 sites according to the parsimony criterion.

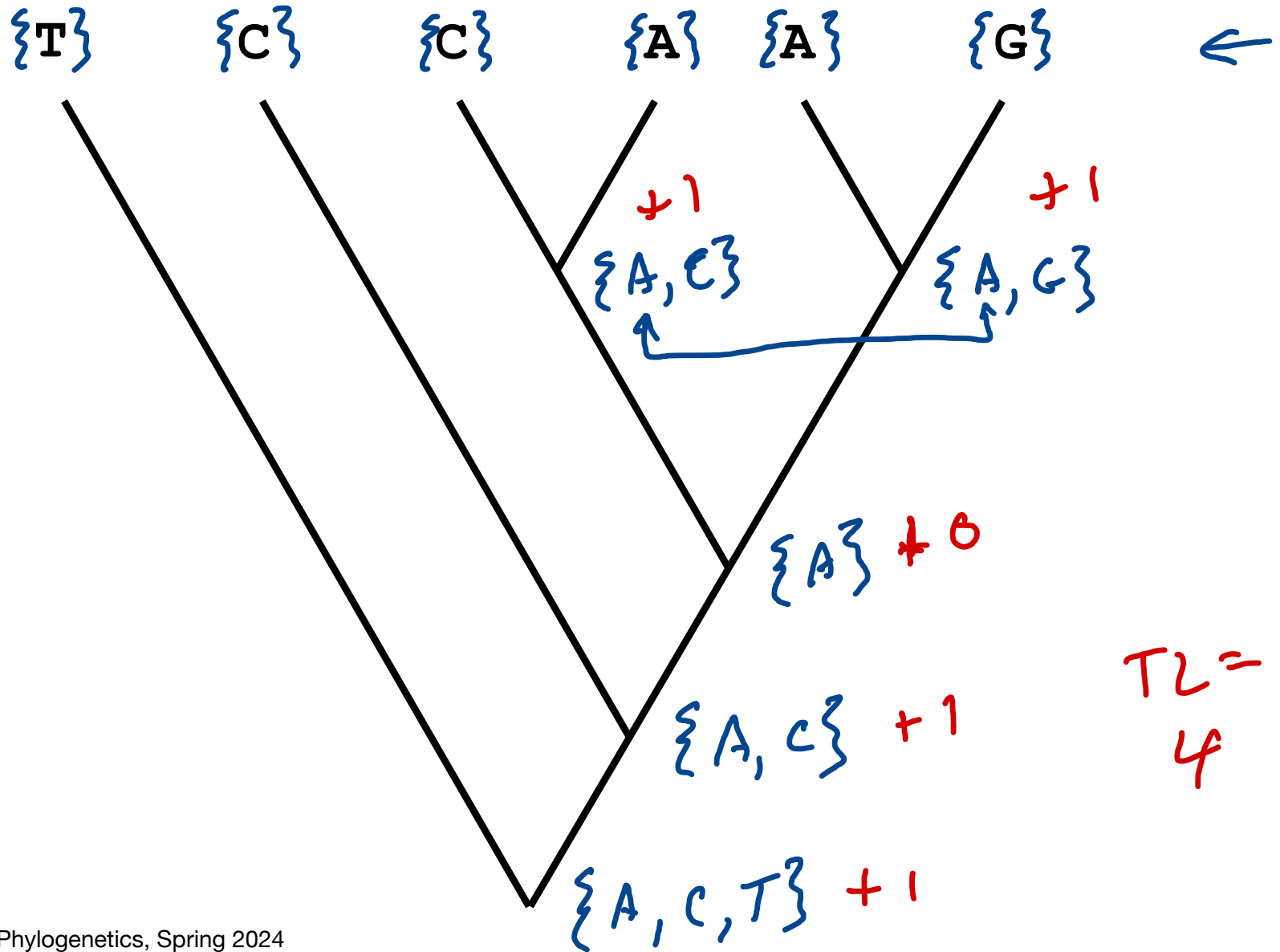
Ockham's Razor

"Essentia non sunt multiplicanda praeter necessitatem"
(Entities should not be multiplied unnecessarily)
— William of Ockham, 14th. century

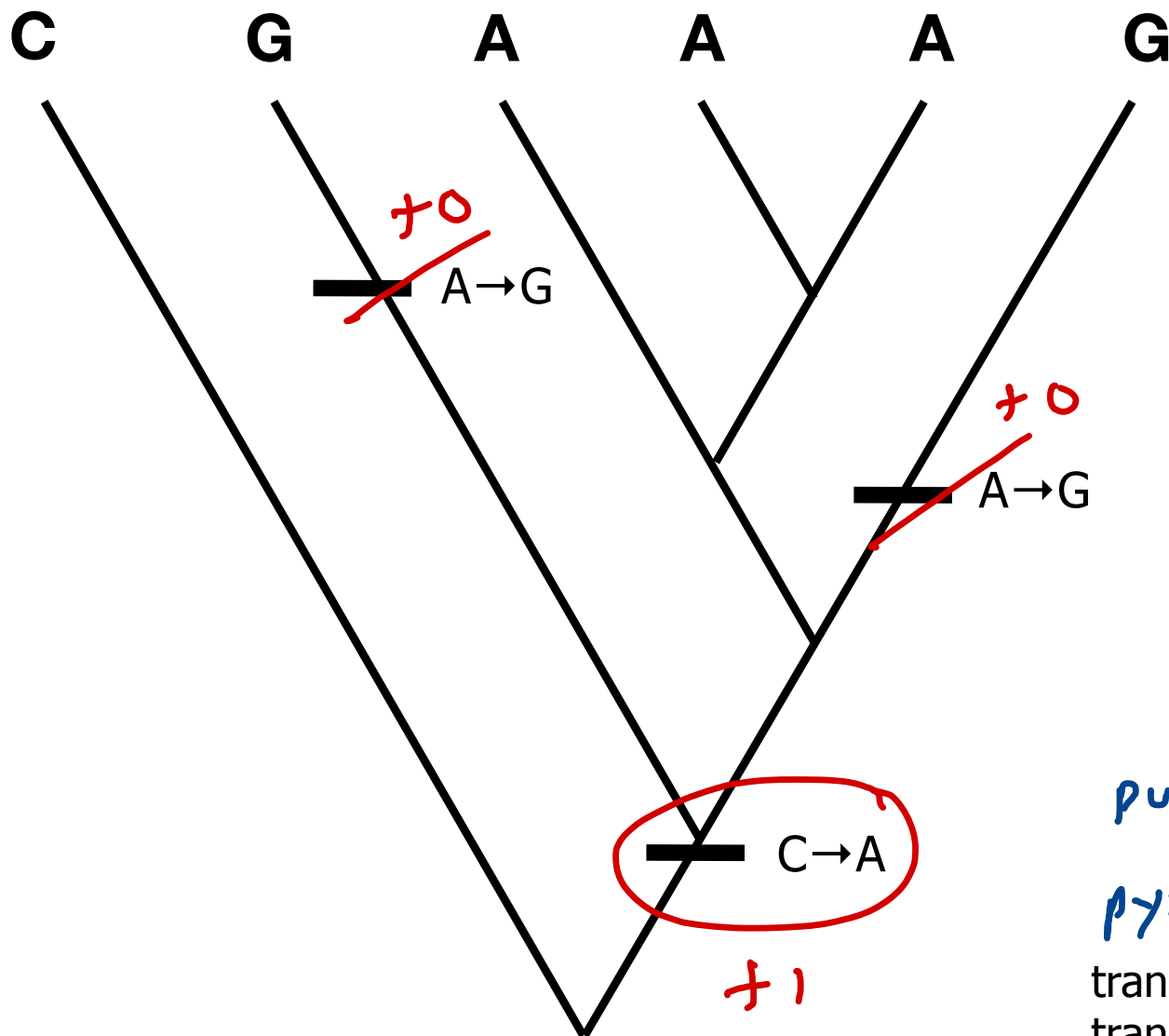
In **science**, this is the **principle of parsimony**: all other things equal, go with the simplest, least complicated hypothesis

In **phylogenetics**, it has become the **parsimony criterion**: all other things equal, go with the tree that requires the fewest inferred character state changes

Counting steps with a minimum of effort



Parsimony variants



TRANSITIONS

A ↔ G

C ↔ T

TRANSVERSION

A ↔ C

A ↔ T

G ↔ C

G ↔ T

purine (R) A, G

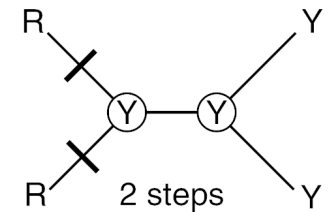
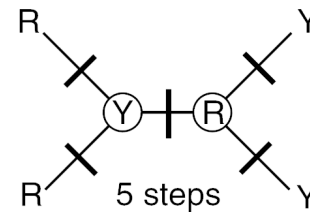
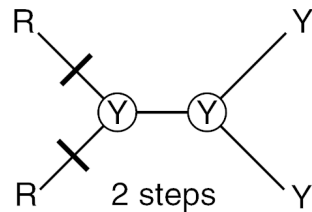
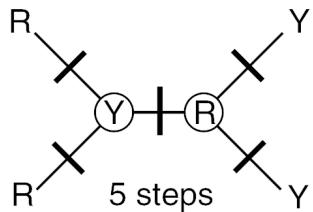
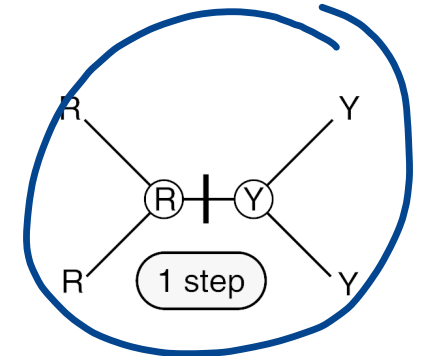
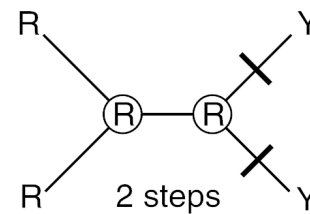
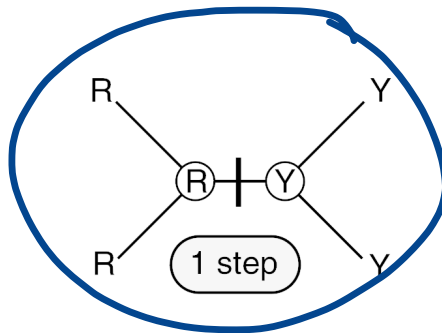
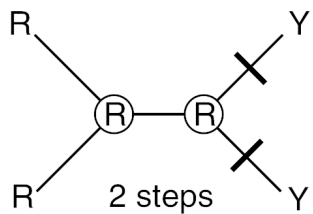
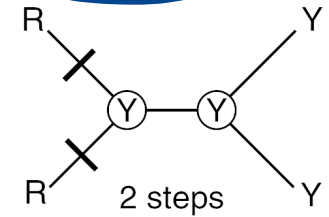
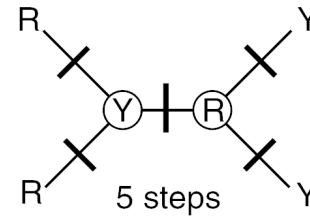
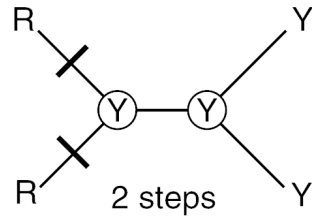
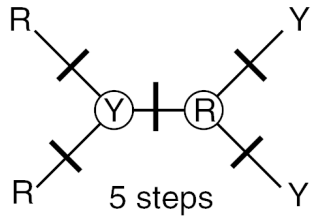
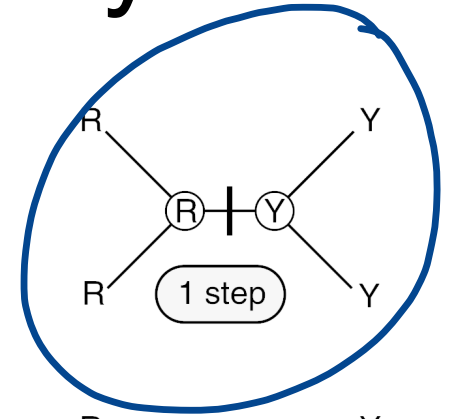
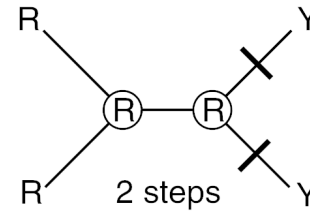
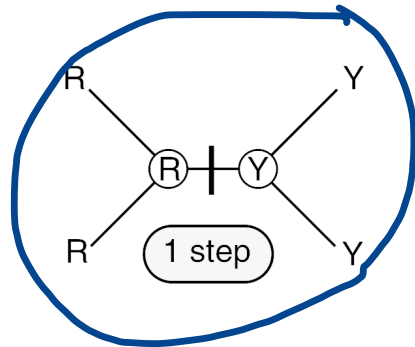
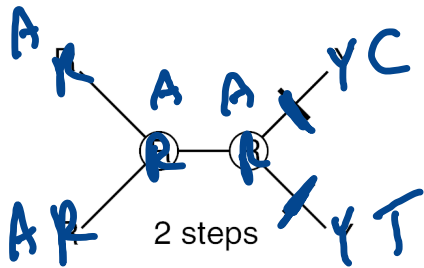
pyrimidine (Y) C, T

transversions more reliable

transversion parsimony

format equate = "A=G C=T"

Transversion Parsimony



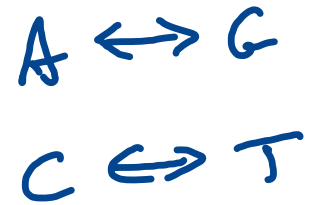
Step Matrices

To

		A	C	G	T
From	A	0	1	1	1
	C	1	0	1	1
	G	1	1	0	1
	T	1	1	1	0

Step matrix for Fitch parsimony

Step Matrices



To

	A	C	G	T
A	0	1	0	1
C	1	0	1	0
G	0	1	0	1
T	1	0	1	0

From

Step matrix for transversion parsimony

Step Matrices

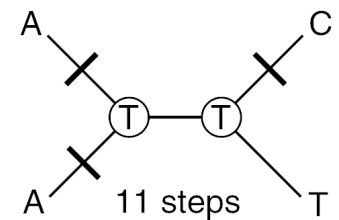
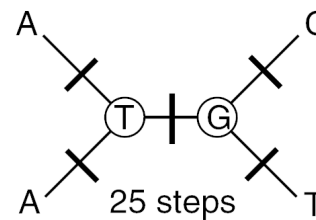
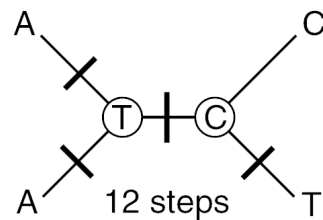
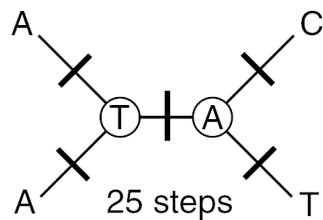
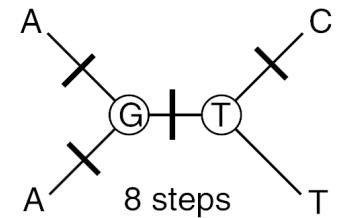
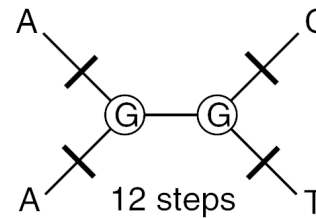
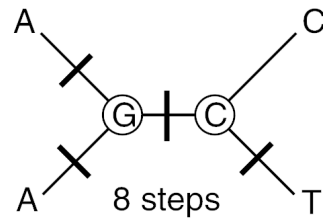
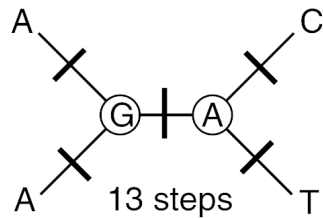
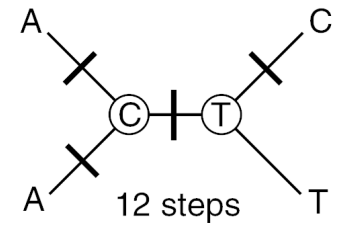
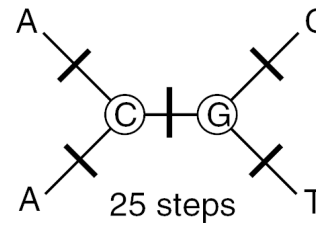
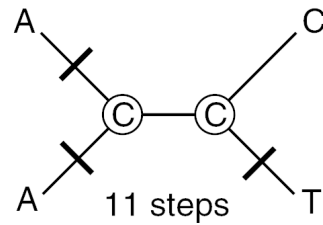
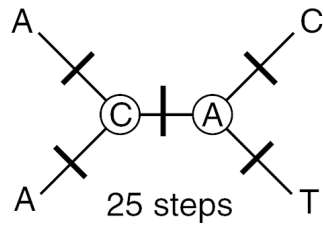
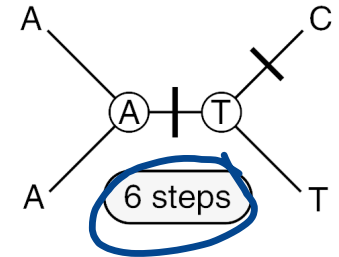
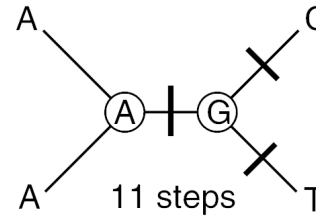
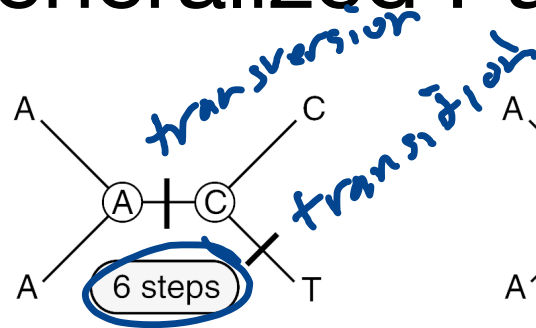
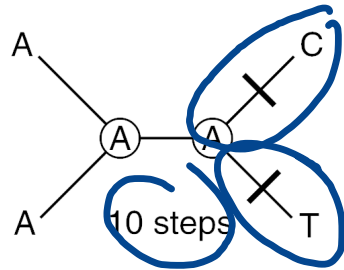
GENERALIZED
PARSIMONY

To

		A	C	G	T
From	A	0	5	1	5
	C	5	0	5	1
	G	1	5	0	5
	T	5	1	5	0

Step matrix for analysis in which all changes are allowed, but transversions are weighted 5 times more than transitions

Generalized Parsimony



Sankoff 1975

Terminology of cladistics

apomorphy vs. plesiomorphy

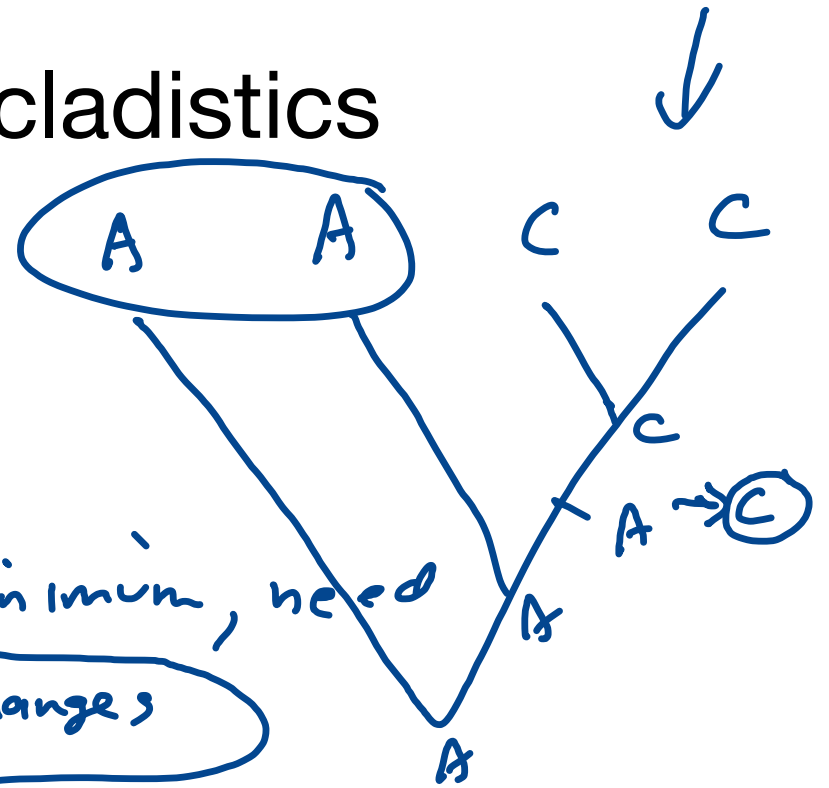
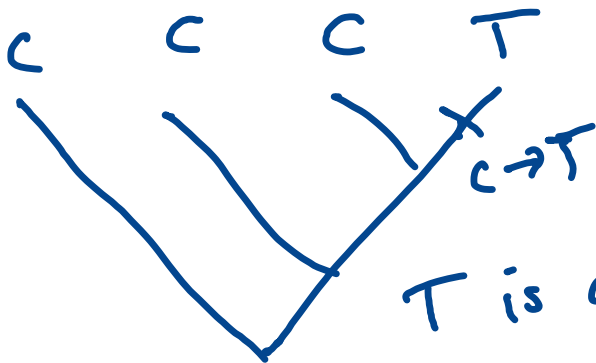
synapomorphy vs. symplesiomorphy

synapomorphy vs. autapomorphy

homoplasy

k states, $k-1$ changes

constant vs. variable vs. parsimony-informative



C is an apomorphy
synapomorphy

A is a plesiomorphy
symplesiomorphy