

Genome 559

Instructor: Dr. Mary Kuhner (through 2/5)

Office Hours: By appointment

Phone: (206) 543-8751

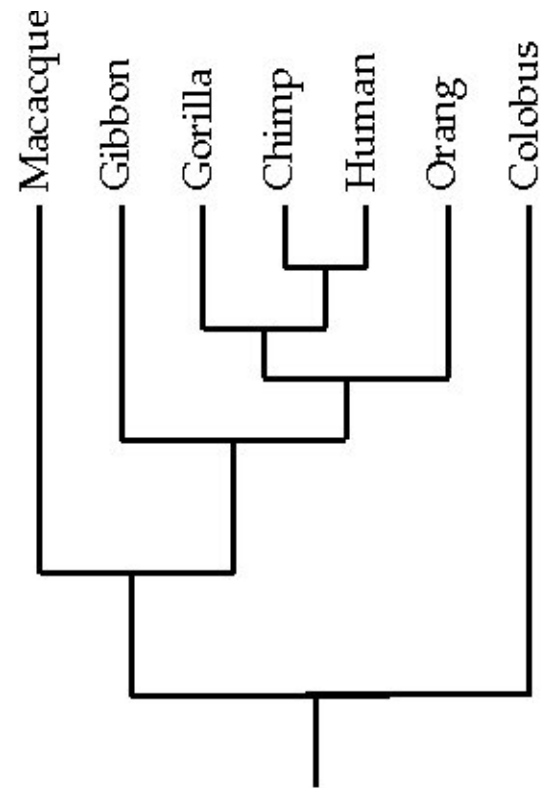
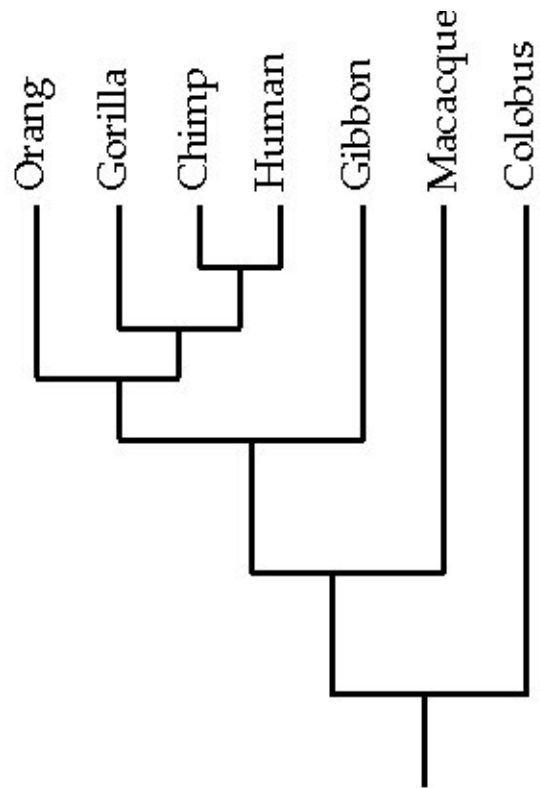
Introduction to Phylogenies: Parsimony

- How to look at phylogenies
- Finding the best phylogeny
- The parsimony principle
- Calculating the parsimony score

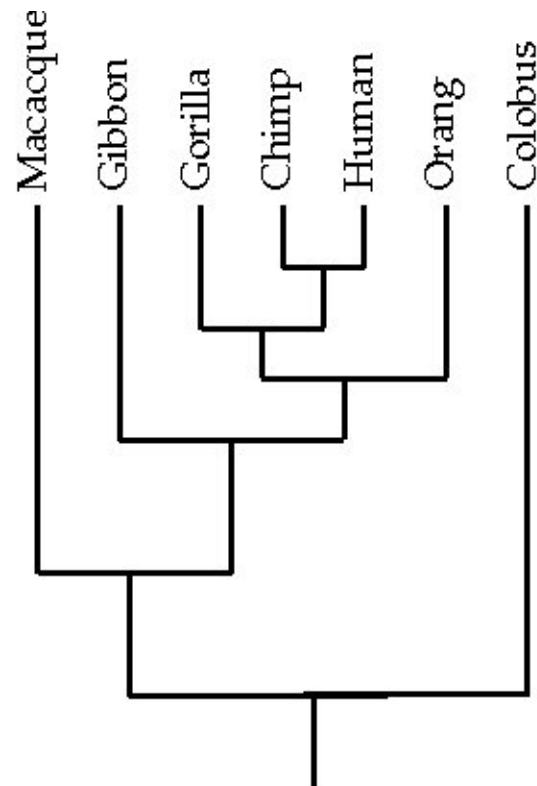
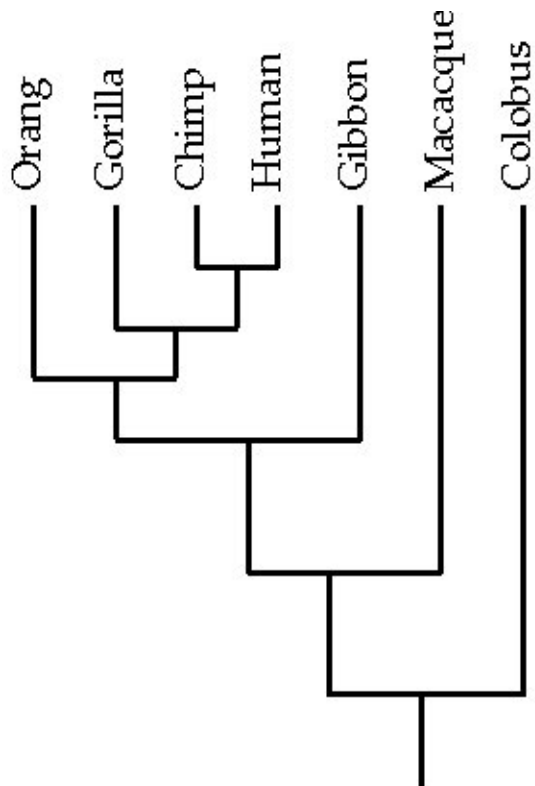
Recommended additional reading

Felsenstein, J (1988) Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22: 521-565.

Looking at a phylogeny

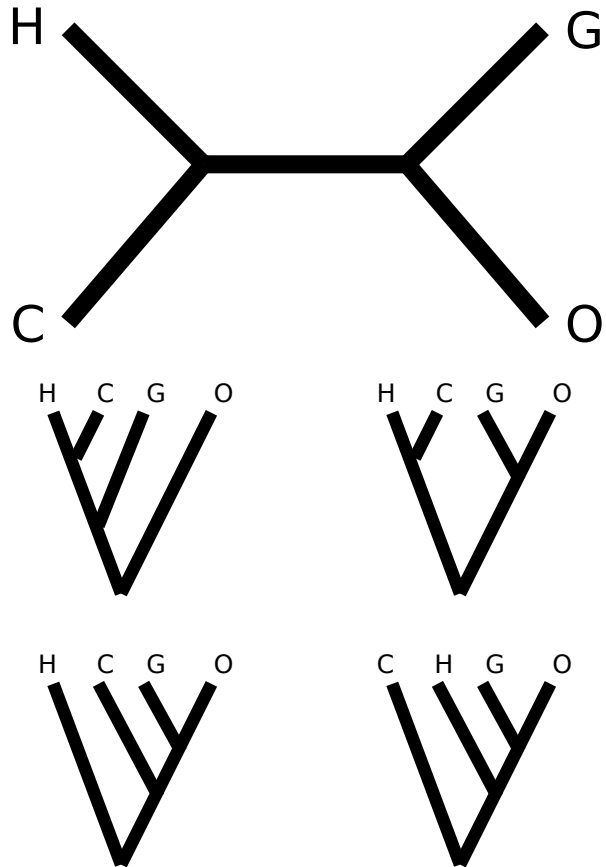


Looking at a phylogeny



These two trees are the same!

Rooted and unrooted trees



Number of different rooted topologies

Tips	Topologies
------	------------

3	3
---	---

4	18
---	----

5	180
---	-----

6	2700
---	------

7	56700
---	-------

8	1587600
---	---------

9	57153600
---	----------

10	2571912000
----	------------

15	6958057668962400000
----	---------------------

20	5644809895887305913369600000000
----	---------------------------------

30	43684666131030695124646801986207638914406400000000000000
----	--

40	302733382994800735654630336455145720004293943205386250170788872192000000000000
----	--

50	3.28632×10^{112}
----	---------------------------

100	1.37416×10^{284}
-----	---------------------------

Principle of Parsimony

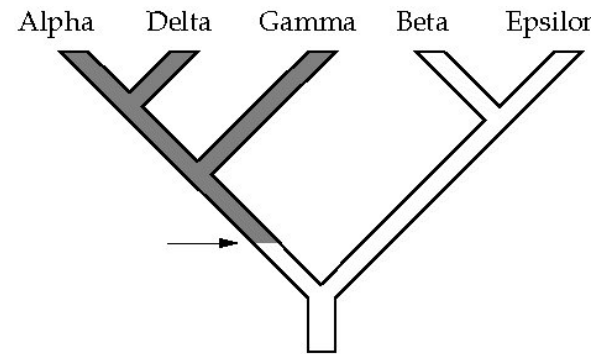
Prefer the hypothesis (tree) that requires the fewest evolutionary events

- Appears to be a “model-free” method
- Implicit model:
 - Evolutionary events are rare
 - They occur independently in different lineages

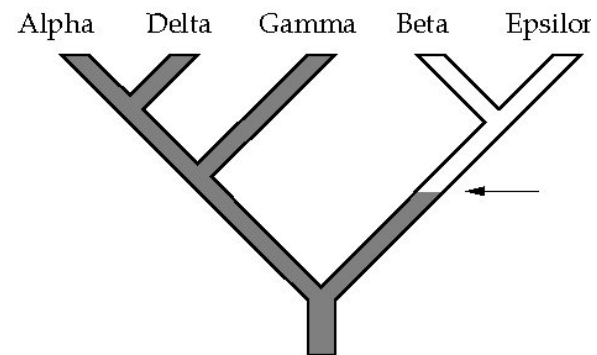
Example data matrix

	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Epsilon	0	0	1	1	1	0

Site 1

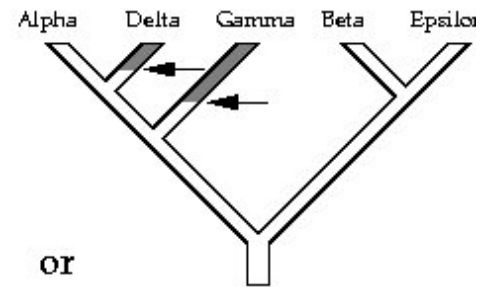


or

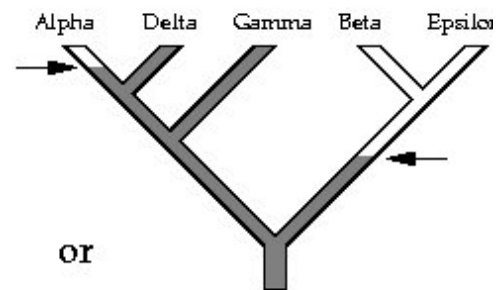


	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Epsilon	0	0	1	1	1	0

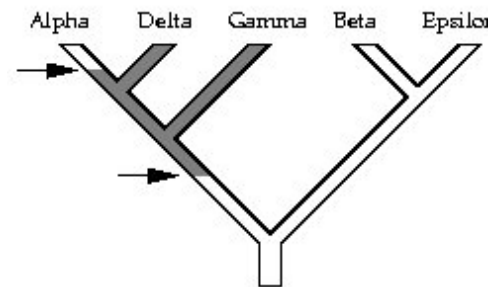
Site 2



OR

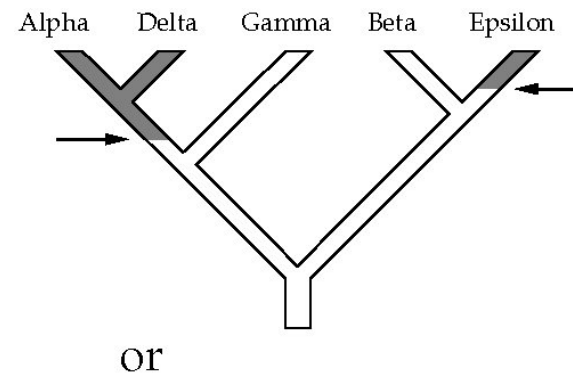


OR

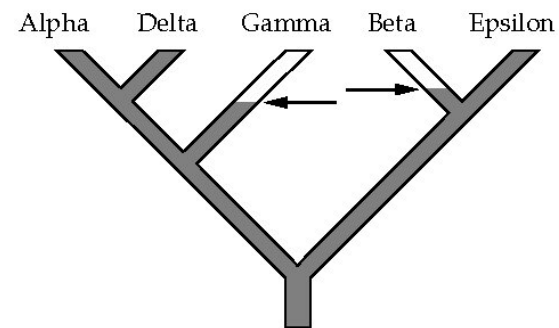


	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Epsilon	0	0	1	1	1	0

Site 4

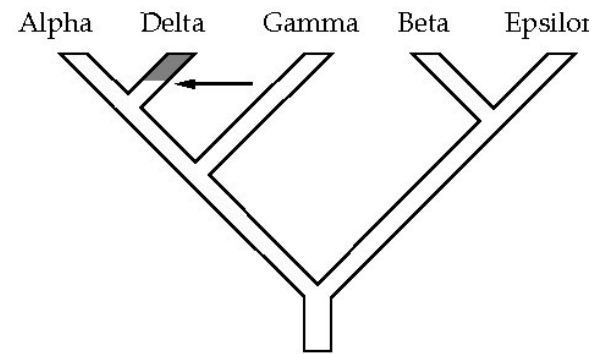


	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Epsilon	0	0	1	1	1	0

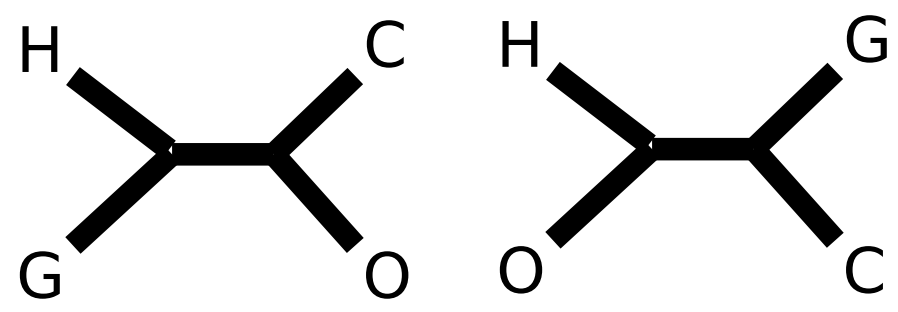
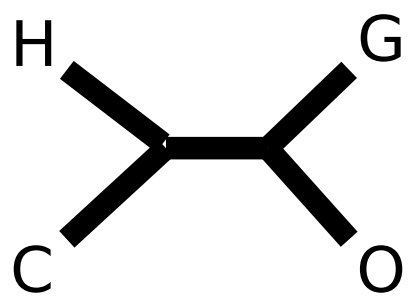


Site 6

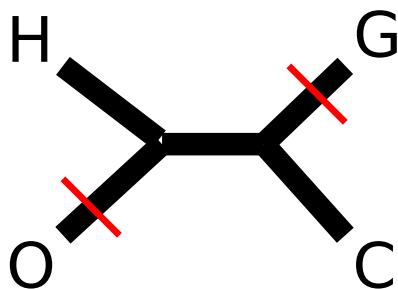
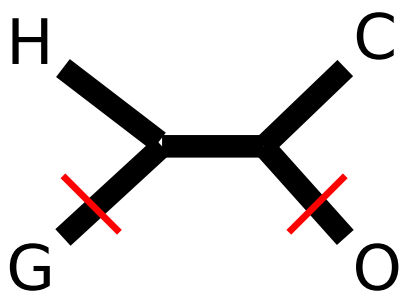
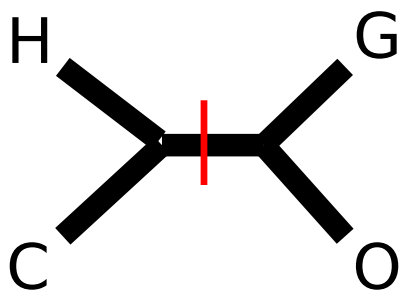
	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Epsilon	0	0	1	1	1	0



	123456
human	agtctc
chimp	agagtc
gorilla	cggcag
orangutan	cgggac

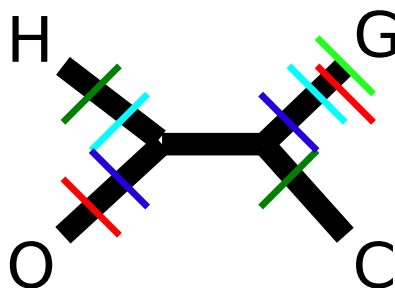
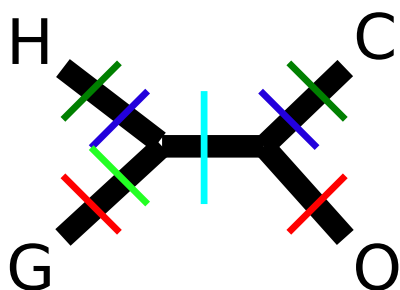
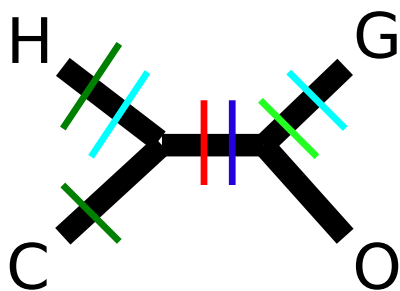


	1	23456
human	a	gtctc
chimp	a	gagtc
gorilla	c	ggcag
orangutan	c	gggac



Fill in the rest yourself on your worksheet.

	1	2	3	4	5	6
human	a	g	t	c	t	c
chimp	a	g	a	g	t	c
gorilla	c	g	g	c	a	g
orangutan	c	g	g	g	a	c



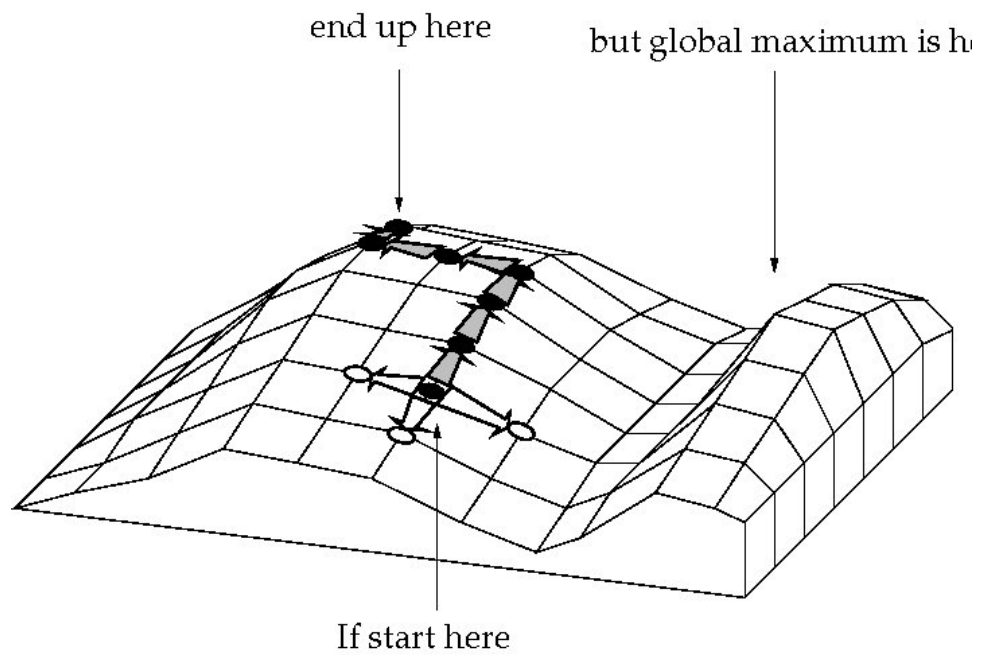
Tree search

- Exhaustive search
- Branch-and-bound
- Heuristic search

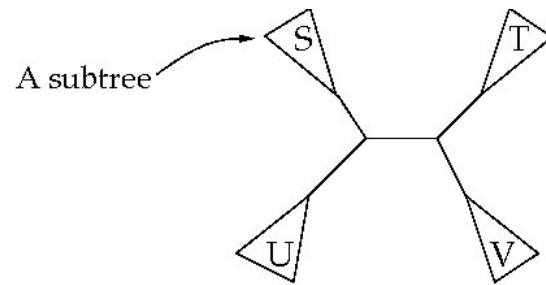
Tree search

- Exhaustive search: up to 8-10 tips, guaranteed results
- Branch-and-bound: up to 15-20 tips, guaranteed results
- Heuristic search: 100+ tips, but may not find correct solution

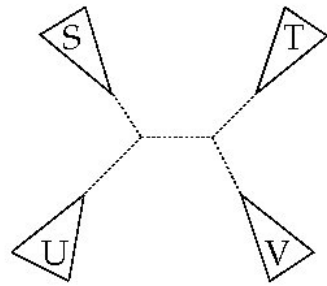
Hill-climbing



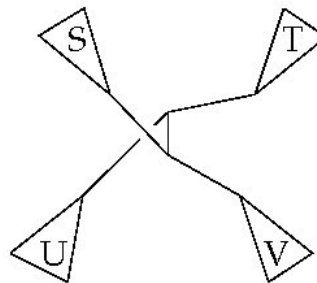
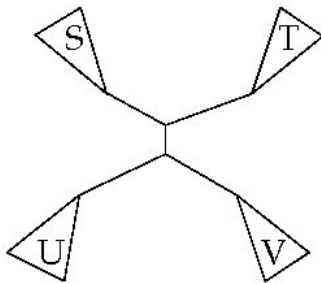
Nearest neighbor interchange



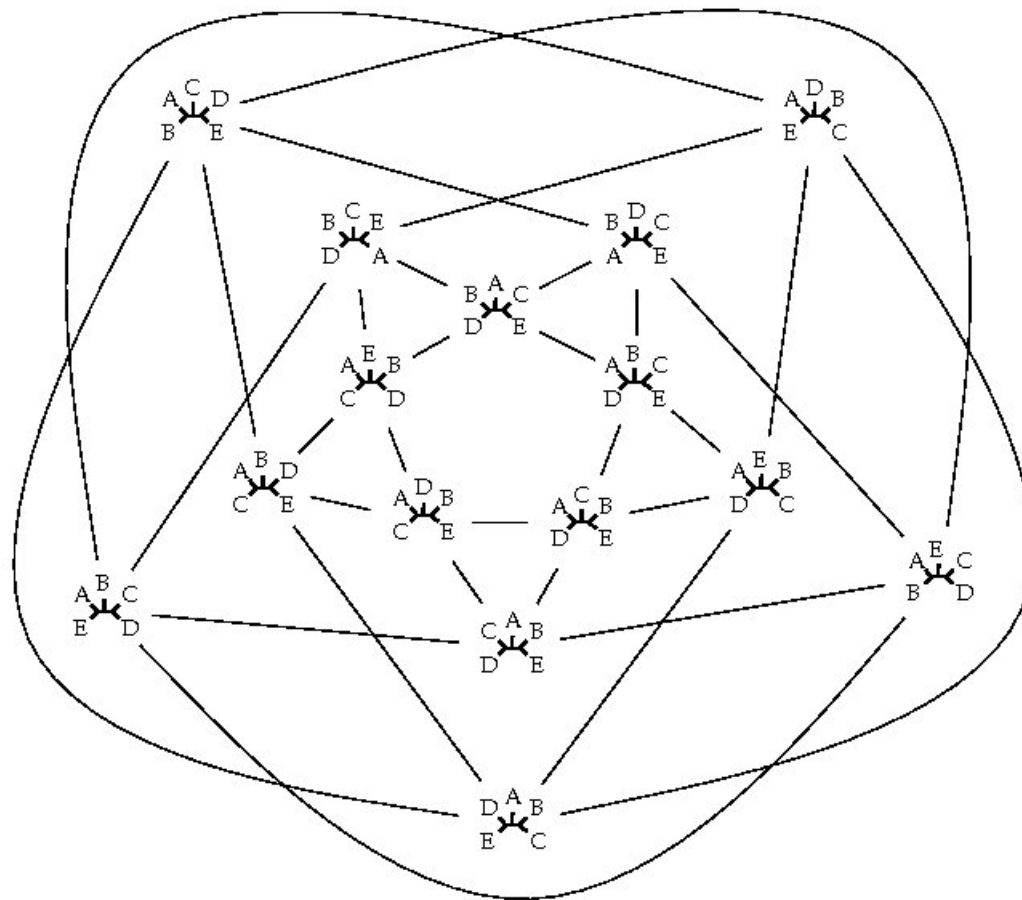
is rearranged by dissolving the connections to an interior branch



and reforming them in one of the two possible alternative ways



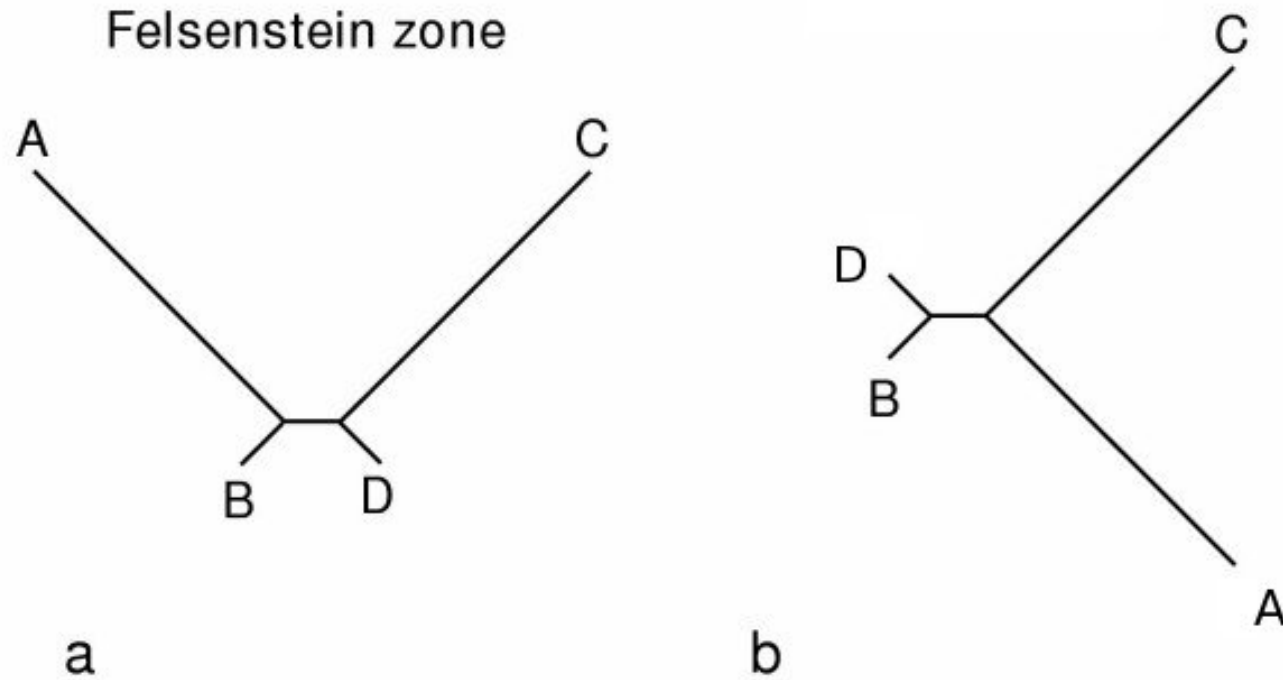
Tree space for unrooted trees of 5 tips



Parsimony assumptions

- Billed as an “assumption-free” method, but....
- In practice it assumes:
 - Changes are rare
 - Changes are independent among sites
 - Ideally, the same site would not change multiple times; the more often this happens, the more trouble it causes

A case in which parsimony fails badly



If the data come from the tree on the left, they will prefer the tree on the right. AKA “long branch attraction”.