Today

Optimality Criteria	Search Strategies
Maximum Parsimony	Exhaustive Enumeration
Maximum Likelihood	Star Decomposition
Minimum Evolution	Stepwise Addition
Least Squares	Branch Swapping

The Least Squares (LS) Criterion

Given two trees, the one minimizing the sum of squares is best

The Minimum Evolution (ME) Criterion

Given two trees, the one minimizing the sum of edge lengths is best

In both LS and ME, edge lengths are estimated by least squares

Pairwise Evolutionary Distances

	human	chimp	gorilla	orangutan
chimp	0.09267			
gorilla	0.10928	0.1144		
orangutan	0.17848	0.19413	0.18836	
gibbon	0.2042	0.21591	0.21592	0.21466

Data from: Brown et al. (1982)

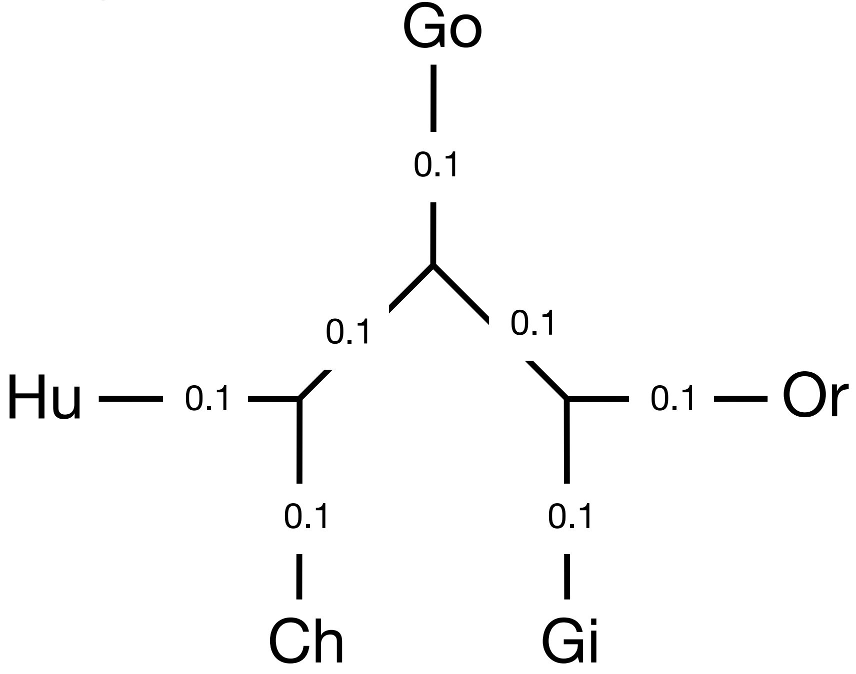
human

gibbon

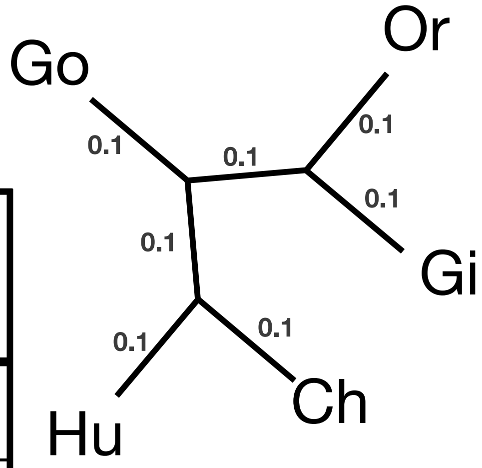
0.2042 substitutions/site

Taxon Pair	distance (data)	distance (tree)
Hu-Ch	0.09267	0.2
Hu-Go	0.10928	0.3
Hu-Or	0.17848	0.4
Hu-Gi	0.2042	0.4
Ch-Go	0.1144	0.3
Ch-Or	0.19413	0.4
Ch-Gi	0.21591	0.4
Go-Or	0.18836	0.3
Go-Gi	0.21592	0.3
Or-Gi	0.21466	0.2

Let's evaluate fit of some totally made-up edge lengths



SS tells us something about goodness of fit



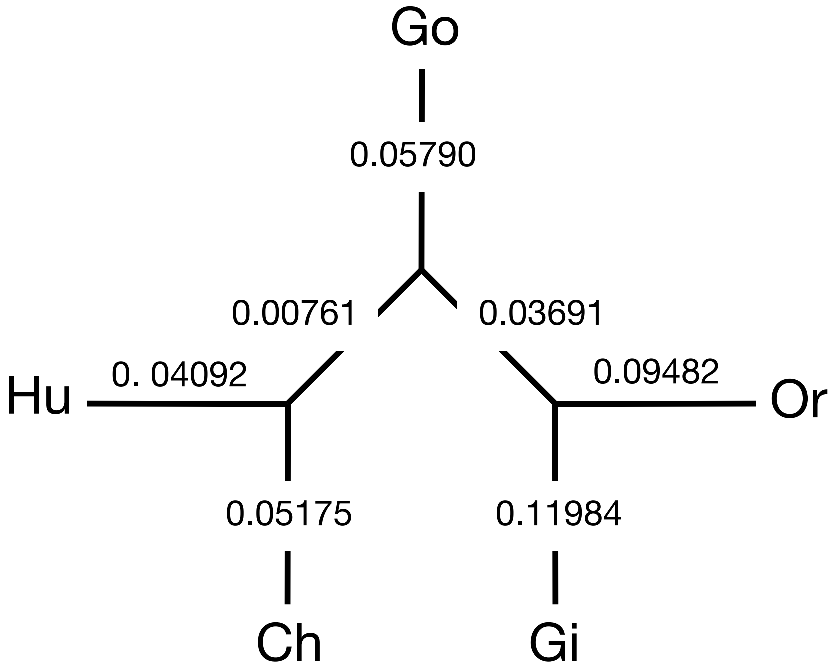
	Taxon Pair	distance (data)	distance (tree)	SS
	Hu-Ch	0.09267	0.2	
	Hu-Go	0.10928	0.3	
	Hu-Or	0.17848	0.4	
	Hu-Gi	0.2042	0.4	
	Ch-Go	0.1144	0.3	
	Ch-Or	0.19413	0.4	
	Ch-Gi	0.21591	0.4	
	Go-Or	0.18836	0.3	
	Go-Gi	0.21592	0.3	
	Or-Gi	0.21466	0.2	

Cavalli-Sforza & Edwards (1967)
Fitch & Margoliash (1967)

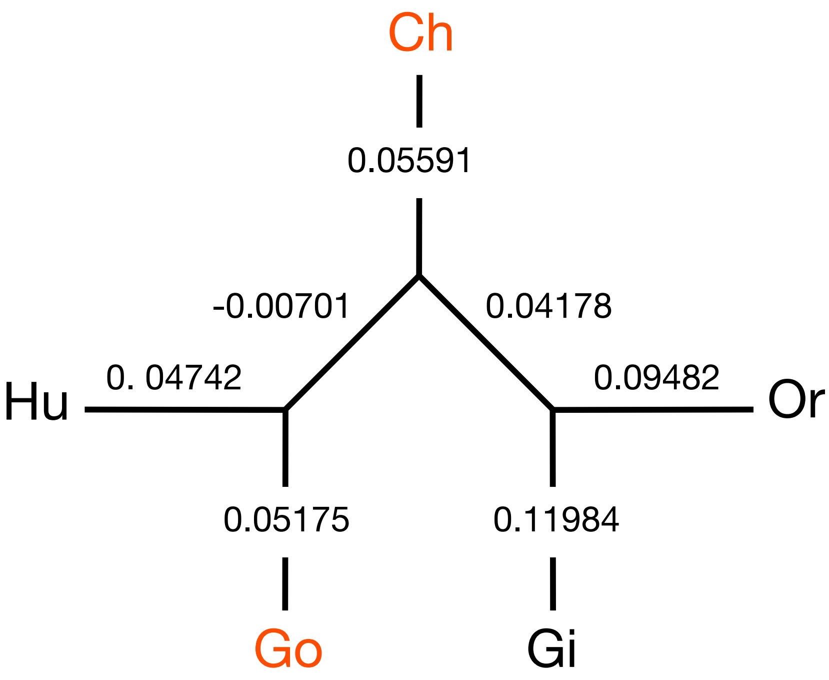
Least squares edge lengths

Taxon Pair	distance (data)	distance (tree)	SS
Hu-Ch	0.09267	0.09267	0
Hu-Go	0.10928	0.10643	0.000008123
Hu-Or	0.17848	0.18026	0.000003168
Hu-Gi	0.2042	0.20528	0.000001166
Ch-Go	0.1144	0.11726	0.00000818
Ch-Or	0.19413	0.19109	0.000009242
Ch-Gi	0.21591	0.21611	0.00000004
Go-Or	0.18836	0.18963	0.000001613
Go-Gi	0.21592	0.21465	0.000001613
Or-Gi	0.21466	0.21466	0
			0.000033144

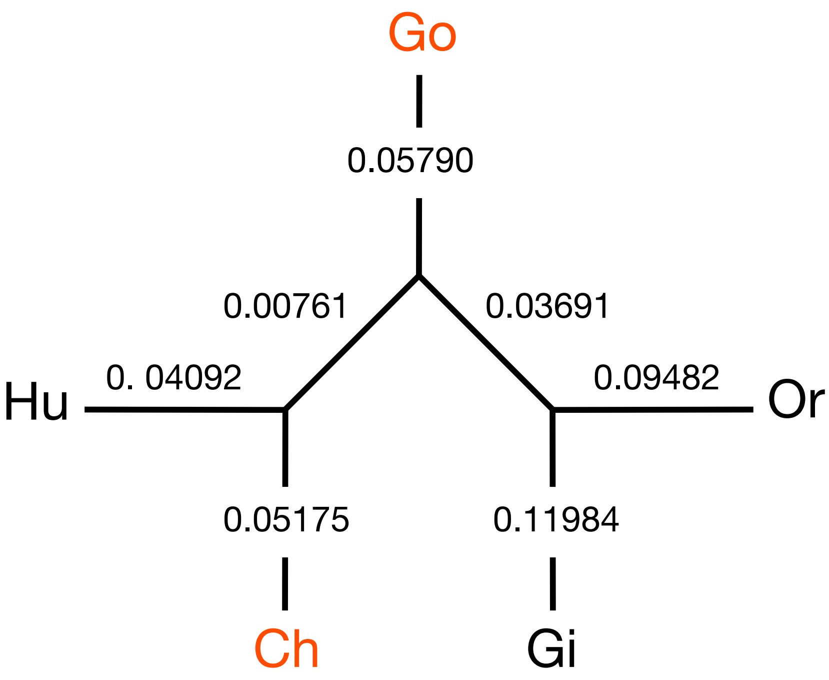
Now choose edge lengths that minimize SS



Least squares criterion

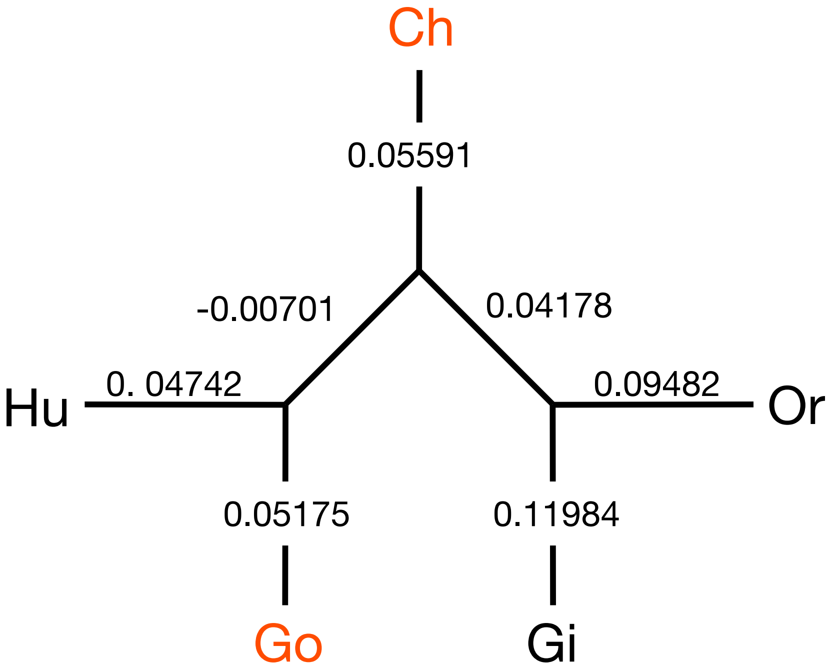


SS = 0.00034

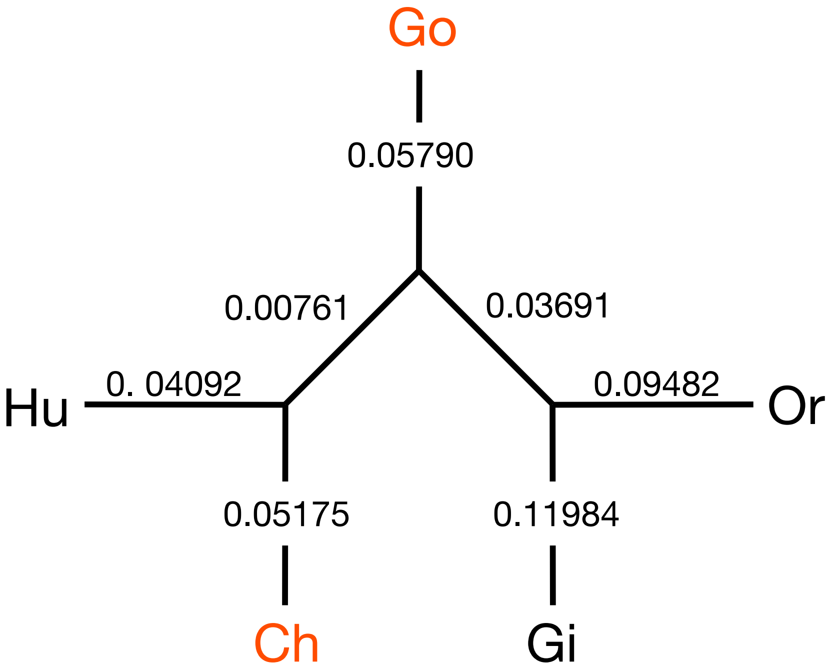


SS = 0.00003314
(best)

Minimum evolution criterion



Sum of edge lengths
= 0.41152



Sum of edge lengths
= 0.40975
(best)

Kidd & Sgaramella-Zonta (1971)
Rzhetsky & Nei (1992)

Sum-of-squares variants

$$SS = \sum_{i<j} \frac{\left(d_{ij} - p_{ij}\right)^2}{d_{ij}^k}$$

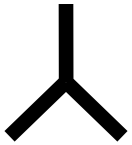
The **power** *k* is most commonly one of these choices:

$k = 0$	Cavalli-Sforza & Edwards (1967)
$k = 2$	Fitch & Margoliash (1967)

Optimality search combinations

Optimality Criteria	Search Strategies
Maximum Parsimony	Exhaustive Enumeration
Maximum Likelihood	Star Decomposition
Minimum Evolution	Stepwise Addition
Least Squares	Branch Swapping

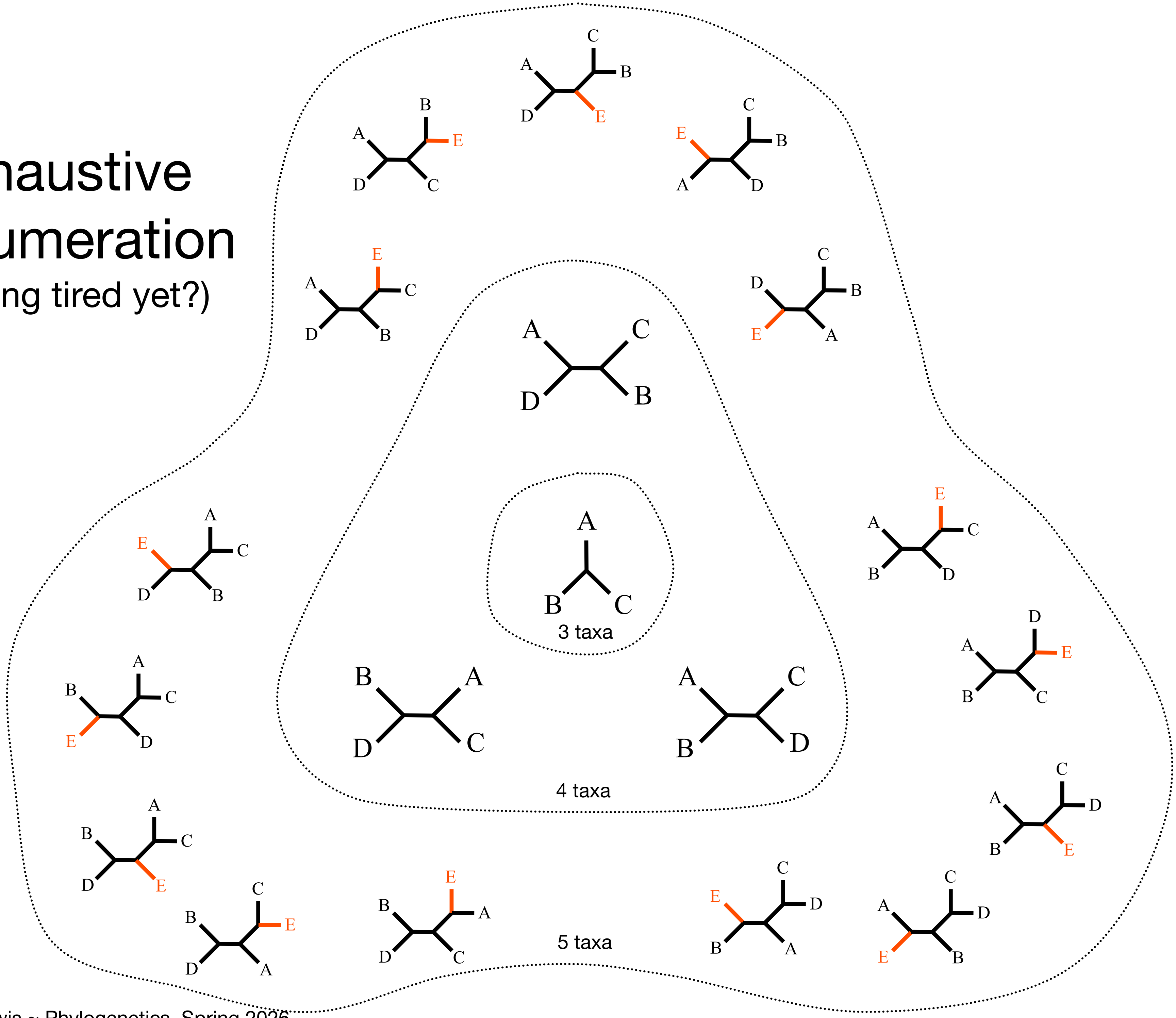
Exhaustive Enumeration



add 2 edges/taxon

Exhaustive Enumeration

(getting tired yet?)



See the pattern?

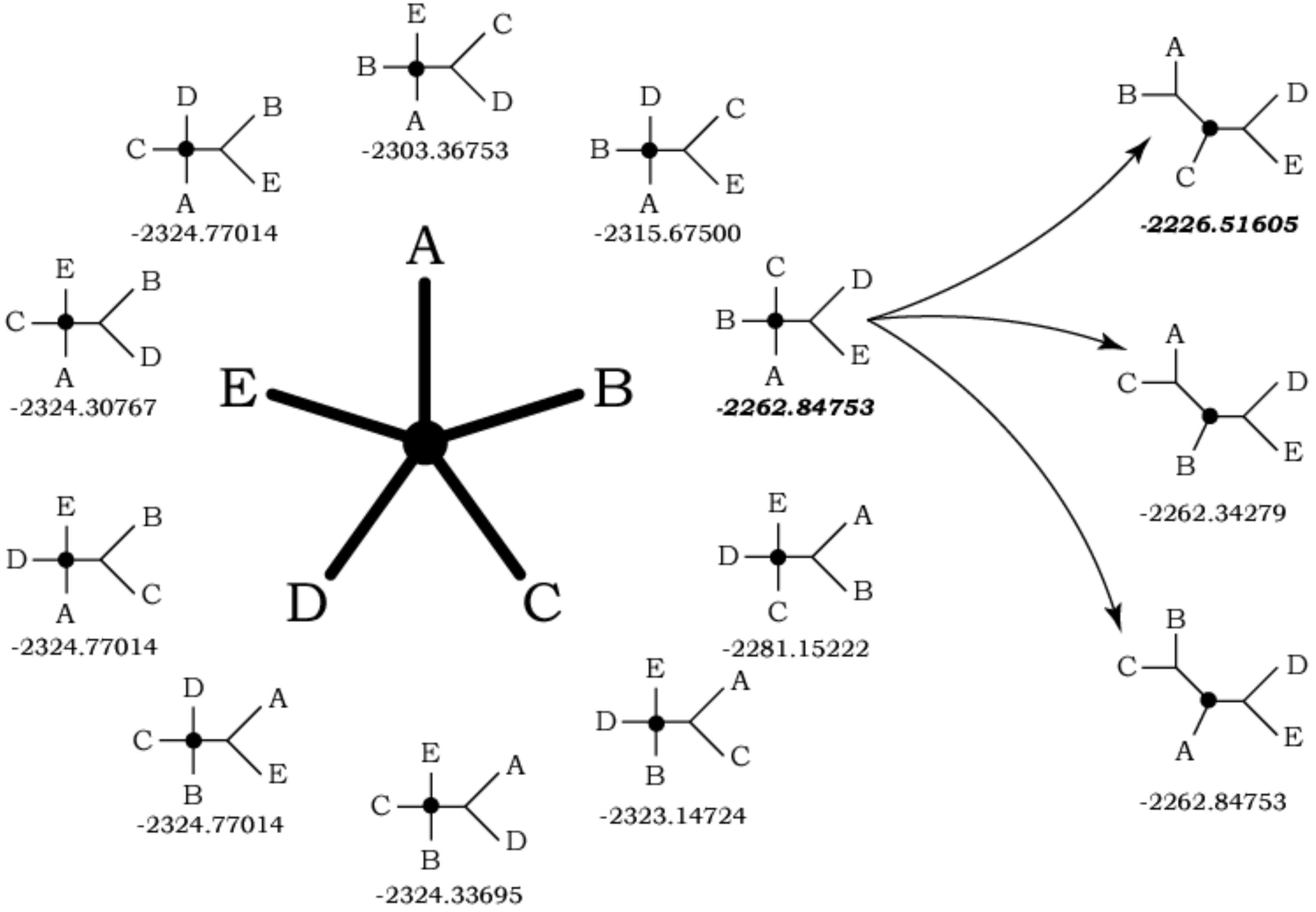
Taxa	Unrooted Topologies	Edges
2	1	1
3	1	3
4	3	5
5	15	7
6	105	9

formula

Taxa	Number of unrooted trees
4	3
5	15
6	105
7	945
8	10,395
9	135,135
10	2,027,025
11	34,459,425
12	654,729,075
13	13,749,310,575
14	316,234,143,225
15	7,905,853,580,625
16	213,458,046,676,875
17	6,190,283,353,629,375
18	191,898,783,962,510,625
19	6,332,659,870,762,850,625
20	221,643,095,476,699,771,875
21	8,200,794,532,637,891,559,375
22	319,830,986,772,877,770,815,625

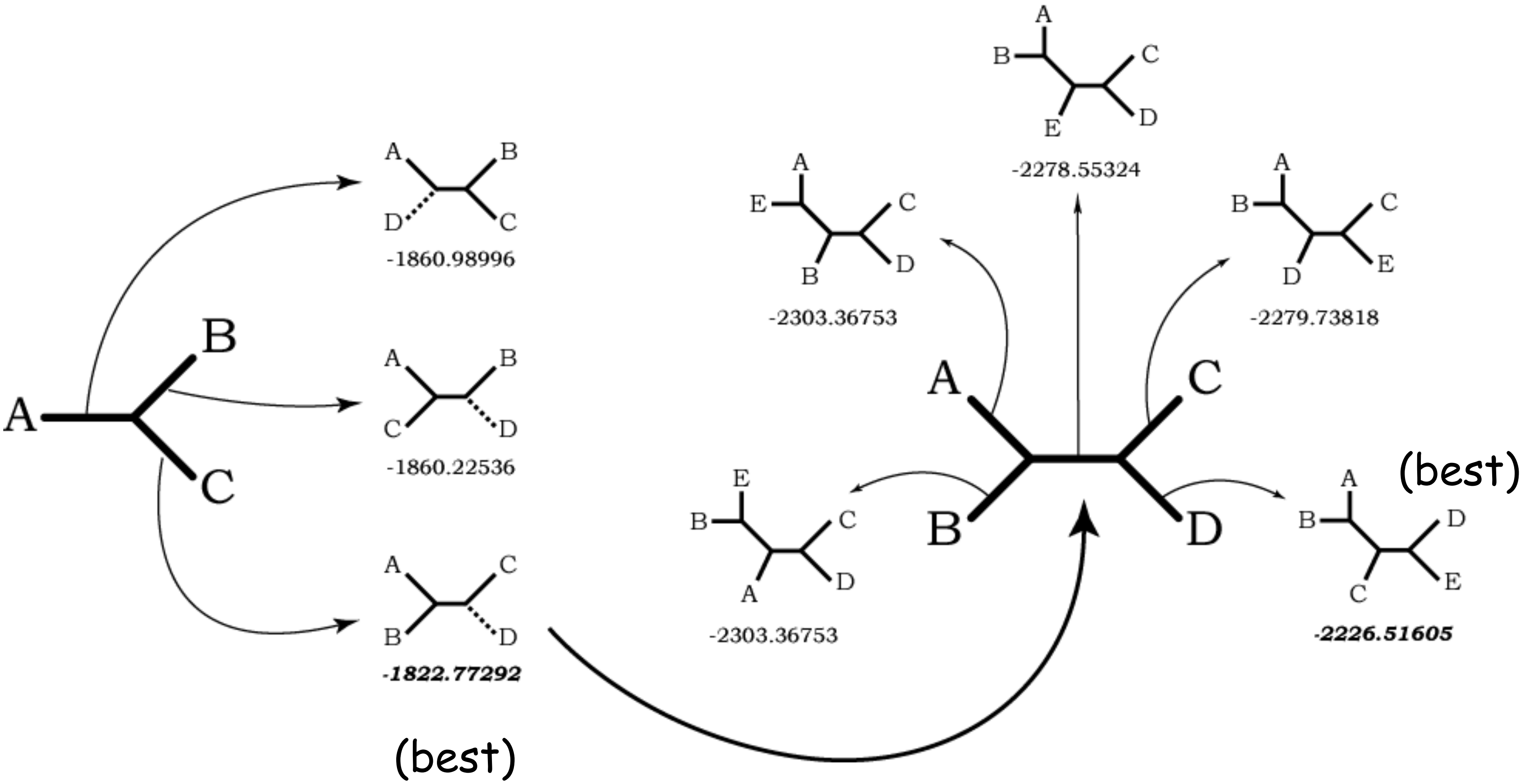
Why not enumerate?

Star decomposition



(best)

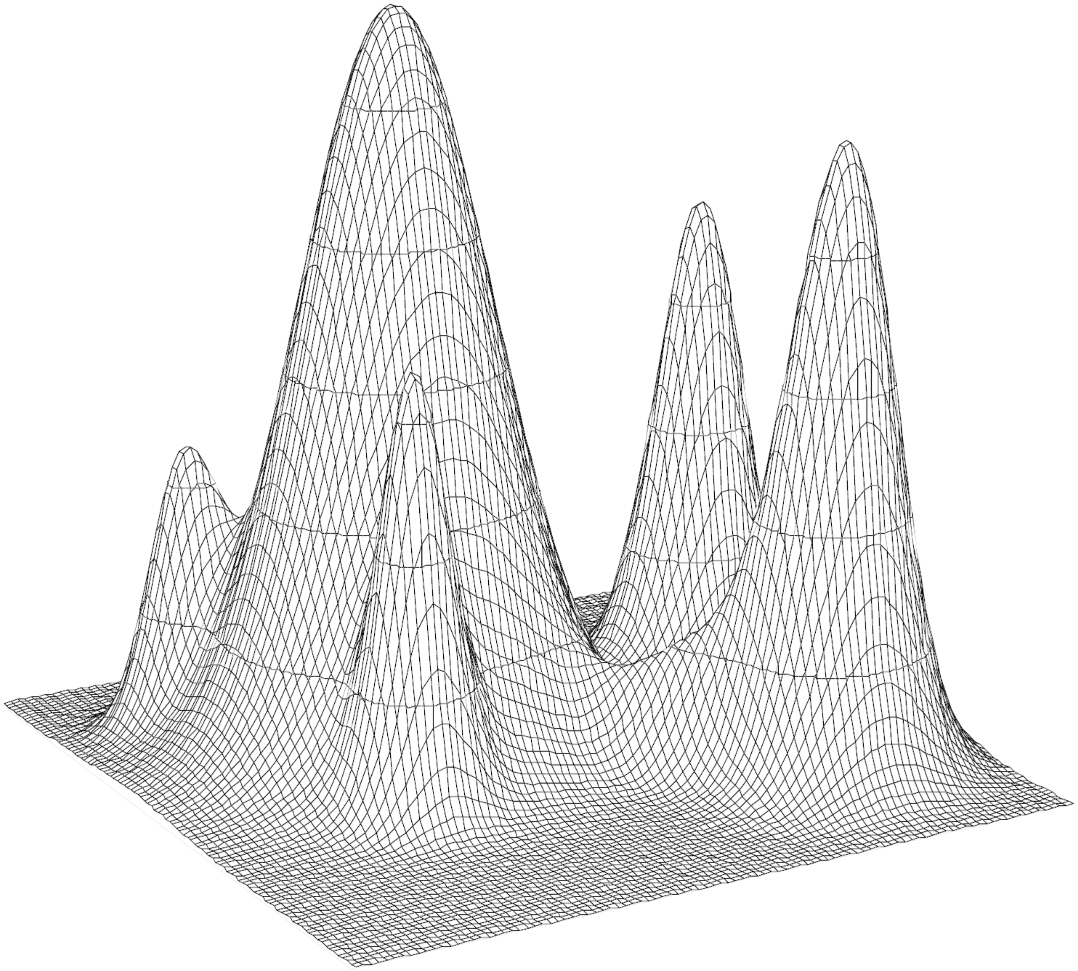
Stepwise Addition



(best)

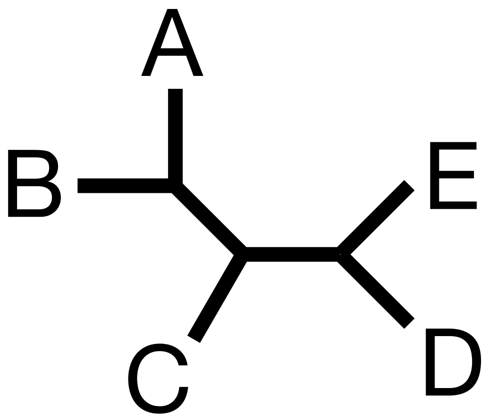
(best)

Treespace



tree islands hill climber

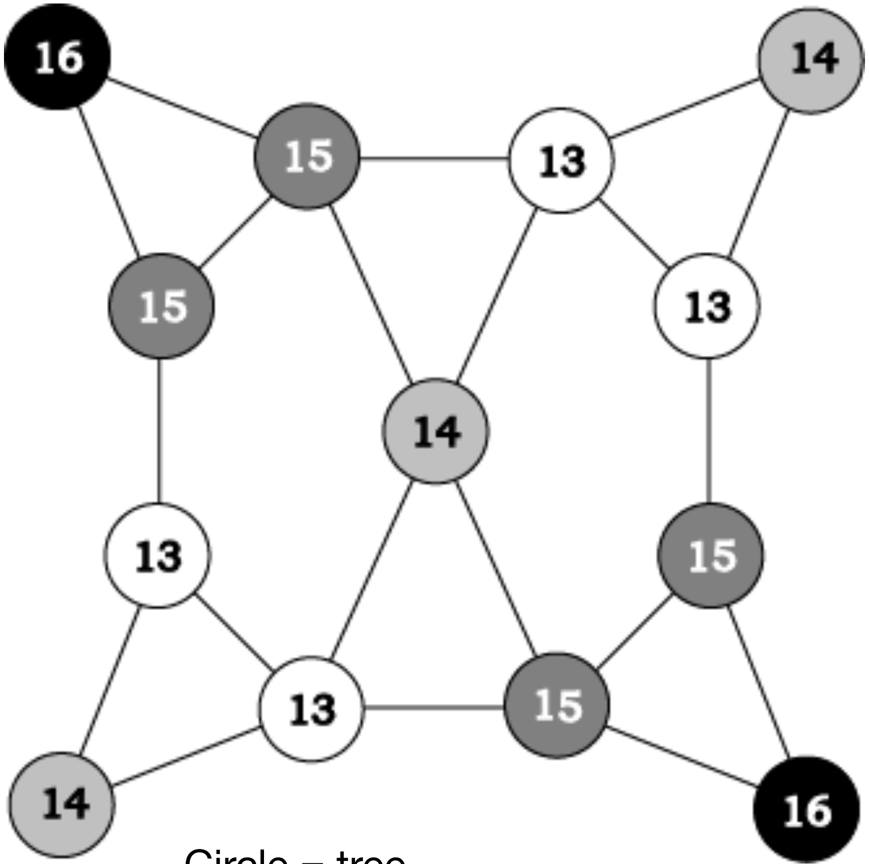
NNI branch swapping



NNI = Nearest Neighbor Interchange

NNI islands

A line in this graph means that an NNI move between trees is possible



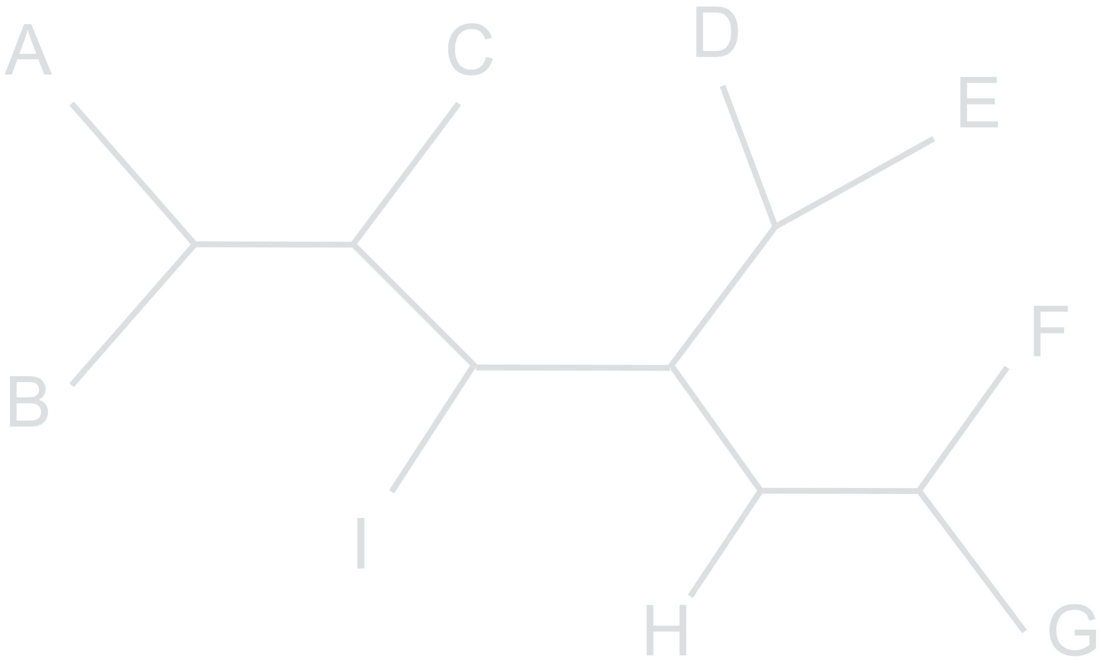
Circle = tree
Number = parsimony tree length
Note: 2 trees omitted for clarity

Sites

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

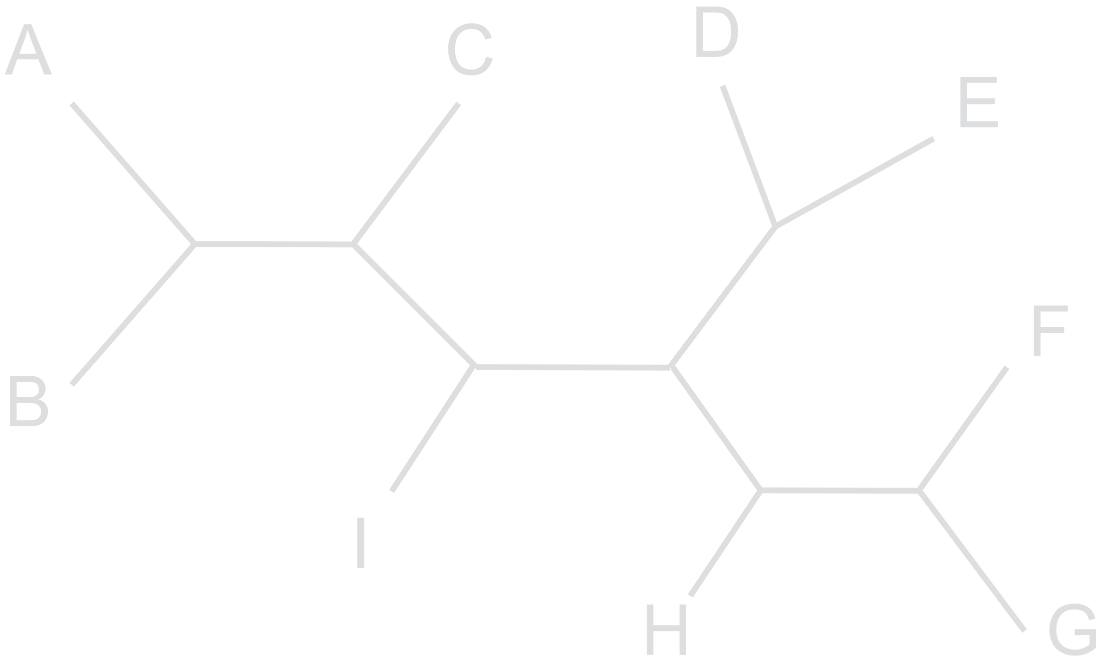
The (contrived) data above produces two NNI islands under parsimony.

Heuristic search: SPR branch swapping



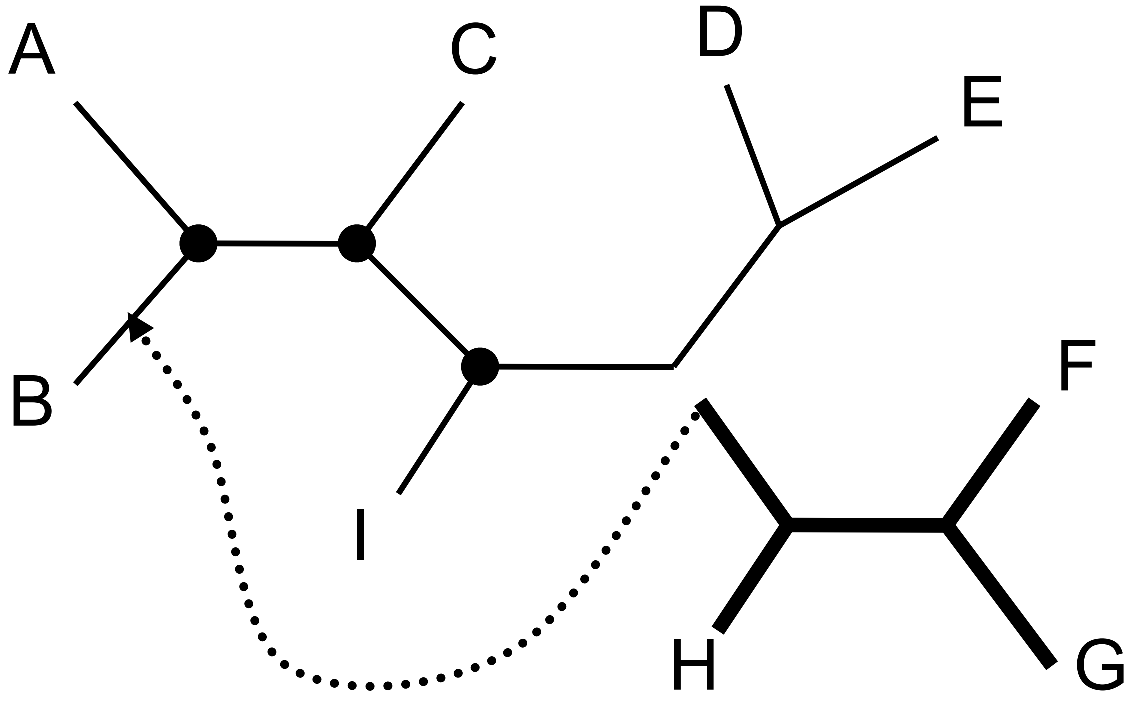
SPR = Subtree Pruning/Reconnection

TBR branch swapping

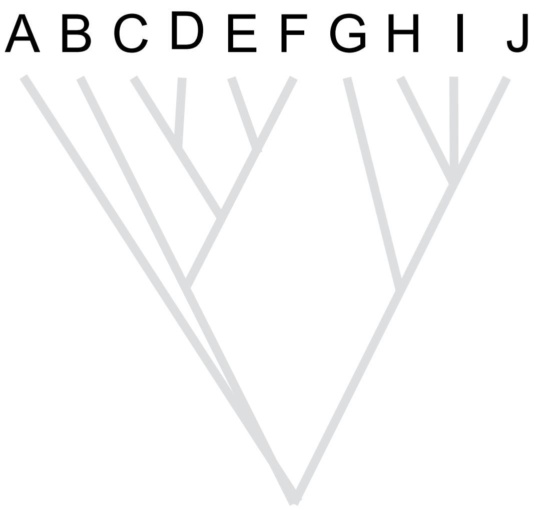


TBR = Tree Bisection/Reconnection

Reconnection Limit



Consensus Trees



Maximum Agreement Subtree

PAUP* commands

If you want this...	then this is the command in PAUP*
Exhaustive enumeration	alltrees
Stepwise addition	hsearch swap=none start=stepwise addseq=random
Star decomposition	stardecomp
Neighbor-joining	nj
NNI heuristic search	hsearch swap=nni
SPR heuristic search	hsearch swap=spr
TBR heuristic search	hsearch swap=tbr reconlimit=infinity
Semistrict consensus tree	contree all / strict=no semistrict
Majority-rule consensus tree	contree all / nostrict majrule percent=50
Maximum agreement subtree	agree all