The Least Squares (LS) Criterion

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

The Minimum Evolution (ME) Criterion

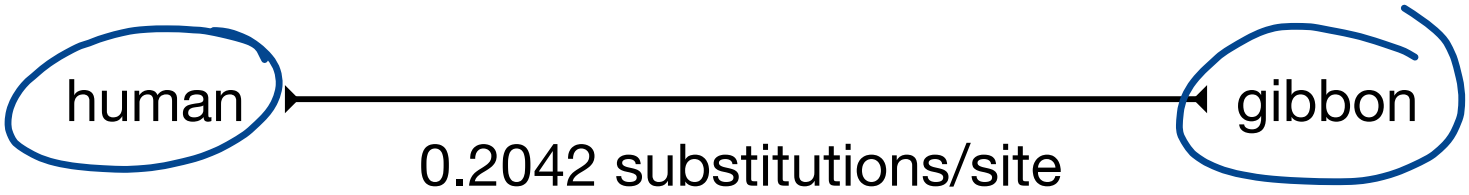
Given two trees, the one minimizing the sum of edge lengths is best

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

Pairwise Evolutionary Distances

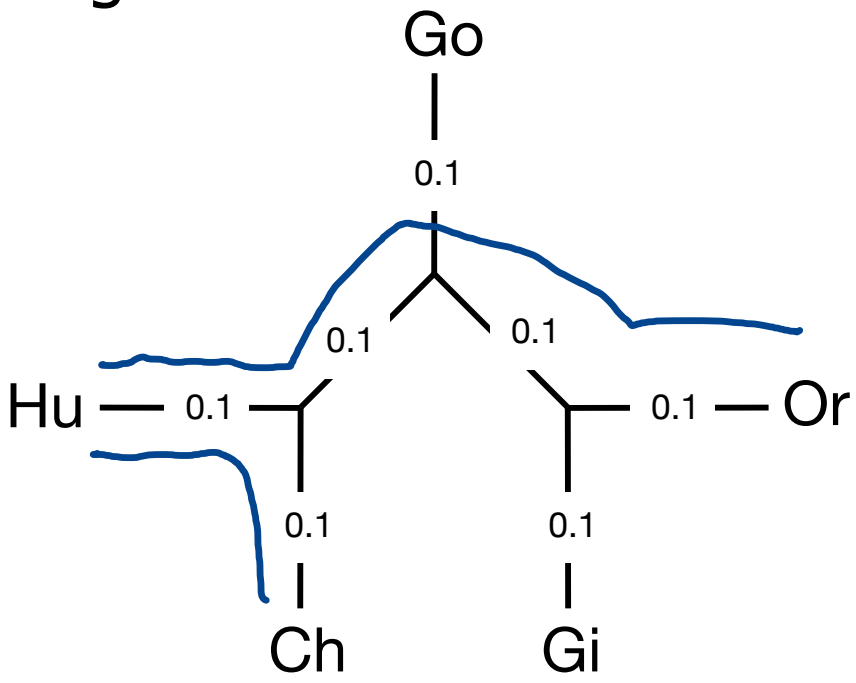
	human	chimp	gorilla	orangutan
chimp	0.09267			
gorilla	0.10928	0.1144		
orangutan	0.17848	0.19413	0.18836	
gibbon	0.2042	0.21591	0.21592	0.21466

Data from: Brown et al. (1982)

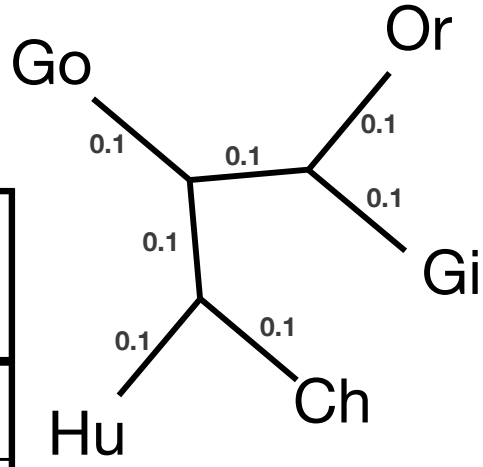


Taxon Pair	distance (data)	distance (tree)
Hu-Ch	0.09267	0.2
Hu-Go	0.10928	0.3
Hu-Or	0.17848	0.4
Hu-Gi	0.2042	0.4
Ch-Go	0.1144	0.3
Ch-Or	0.19413	0.4
Ch-Gi	0.21591	0.4
Go-Or	0.18836	0.3
Go-Gi	0.21592	0.3
Or-Gi	0.21466	0.2

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



SS tells us something about goodness of fit



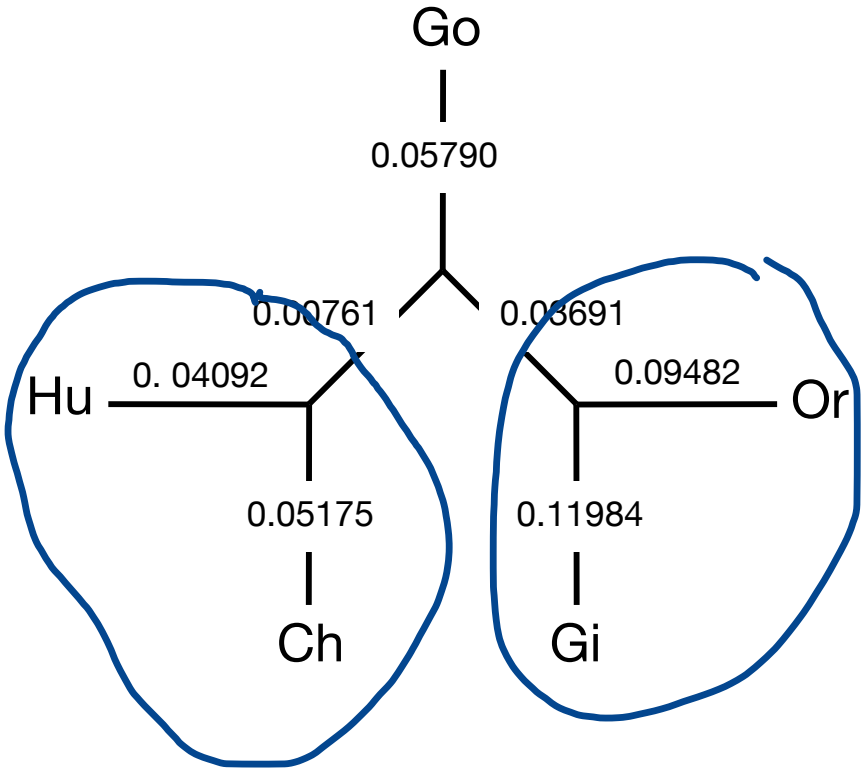
	Taxon Pair	distance (data)	distance (tree)	SS
		x	y	$(x-y)^2$
Amee	Hu-Ch	0.09267	0.2	0.01152
Danielle	Hu-Go	0.10928	0.3	0.03637
Akriti	Hu-Or	0.17848	0.4	0.04907
Savanna	Hu-Gi	0.2042	0.4	0.03905
Mandy	Ch-Go	0.1144	0.3	0.01309
Chelsea	Ch-Or	0.19413	0.4	0.04238
Elena	Ch-Gi	0.21591	0.4	0.03389
Noah	Go-Or	0.18836	0.3	0.01246
Ellie	Go-Gi	0.21592	0.3	0.007069
Jensen	Or-Gi	0.21466	0.2	0.00021
			Brenden crystal	0.2451

5 decimal places

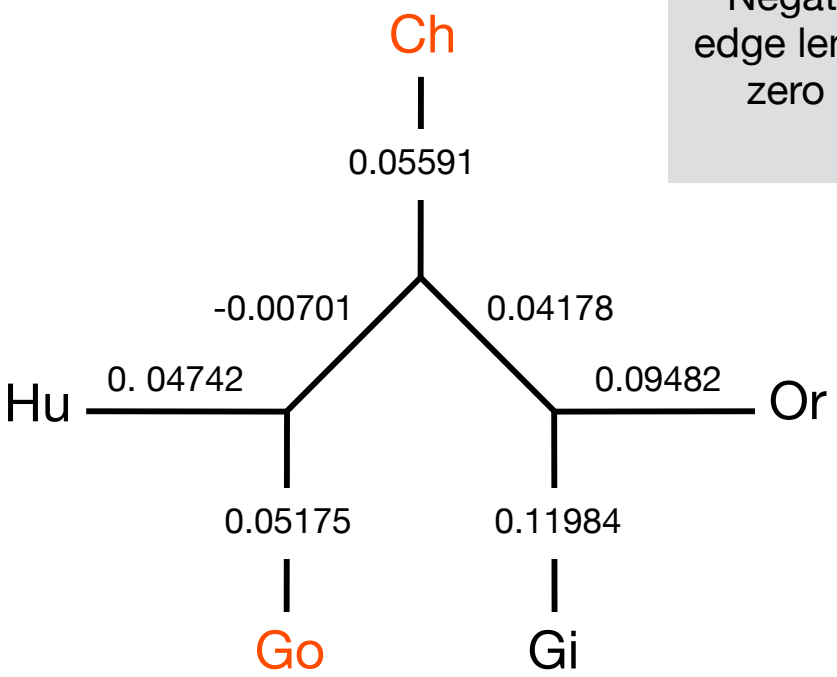
Cavalli-Sforza & Edwards (1967)
Fitch & Margoliash (1967)

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

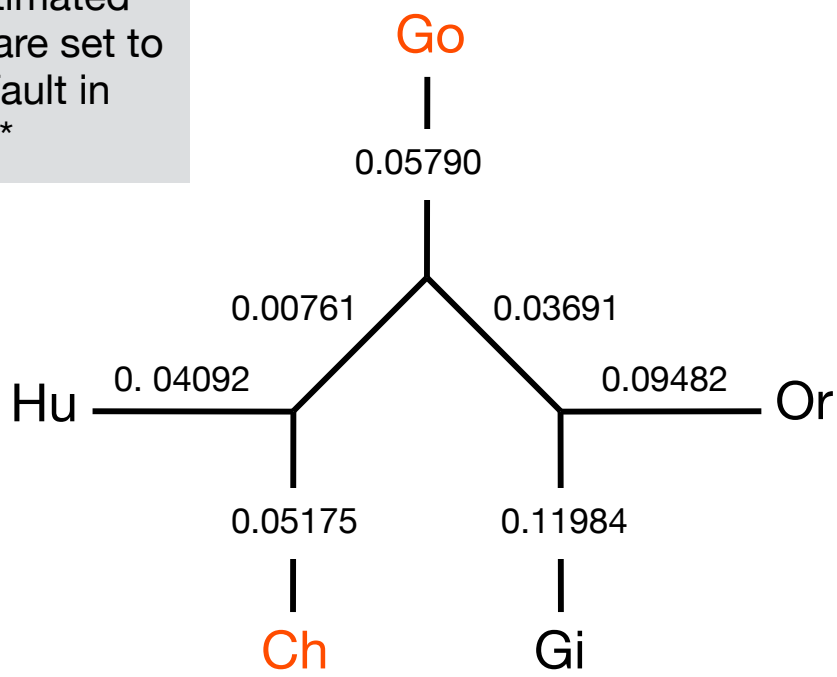
Now choose edge lengths that minimize SS



Least squares criterion ranks tree topologies



Negative estimated edge lengths are set to zero by default in PAUP*



SS = 0.00034

(0.000048541= SS allowing negative edge lengths)

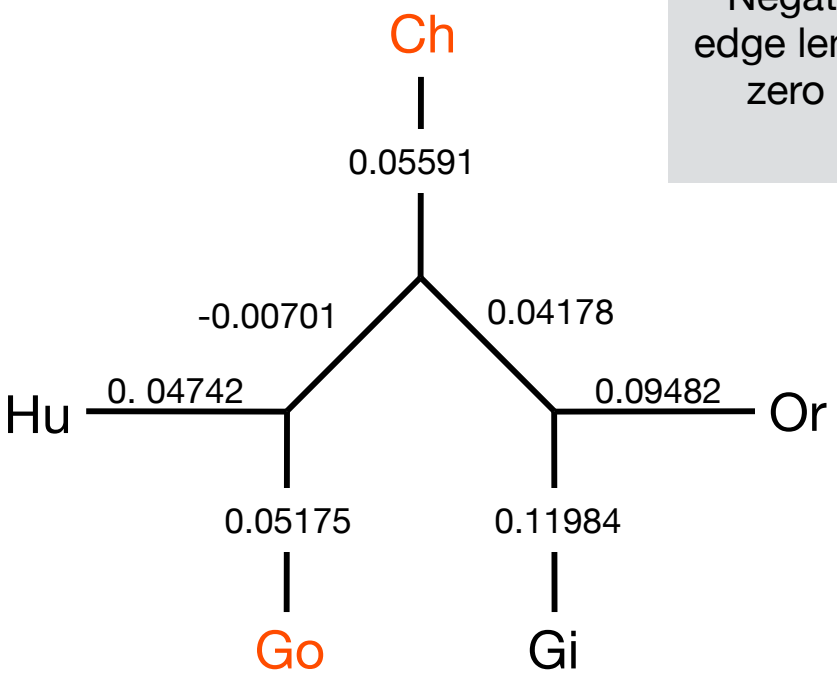
SS = 0.00003314

(best)

Minimum evolution criterion

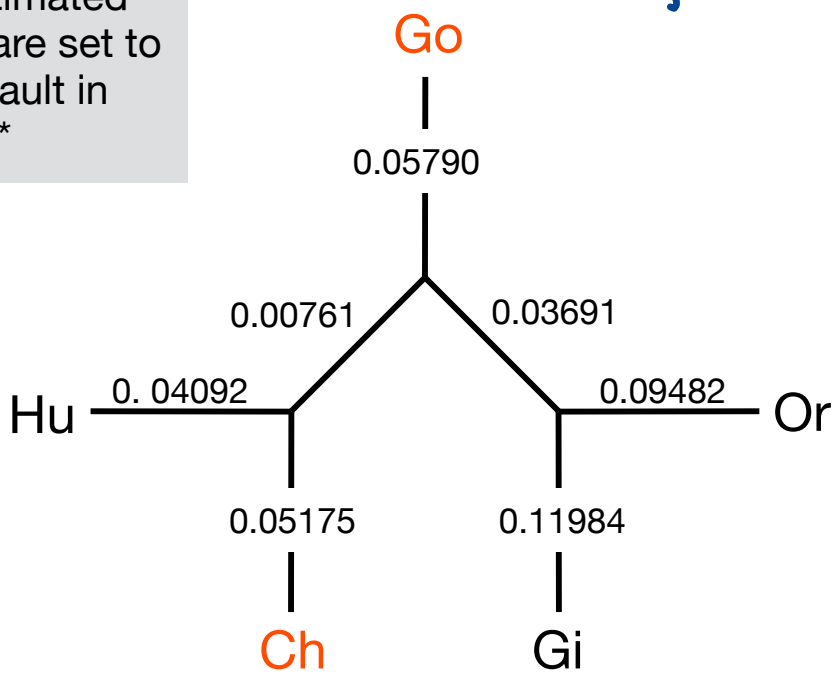
NJ neighbor-joining

Negative estimated edge lengths are set to zero by default in PAUP*



Sum of edge lengths
= 0.41152

(0.40451 = sum allowing negative edge lengths)



Sum of edge lengths
= 0.40975
(best)

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

Sum-of-squares variants

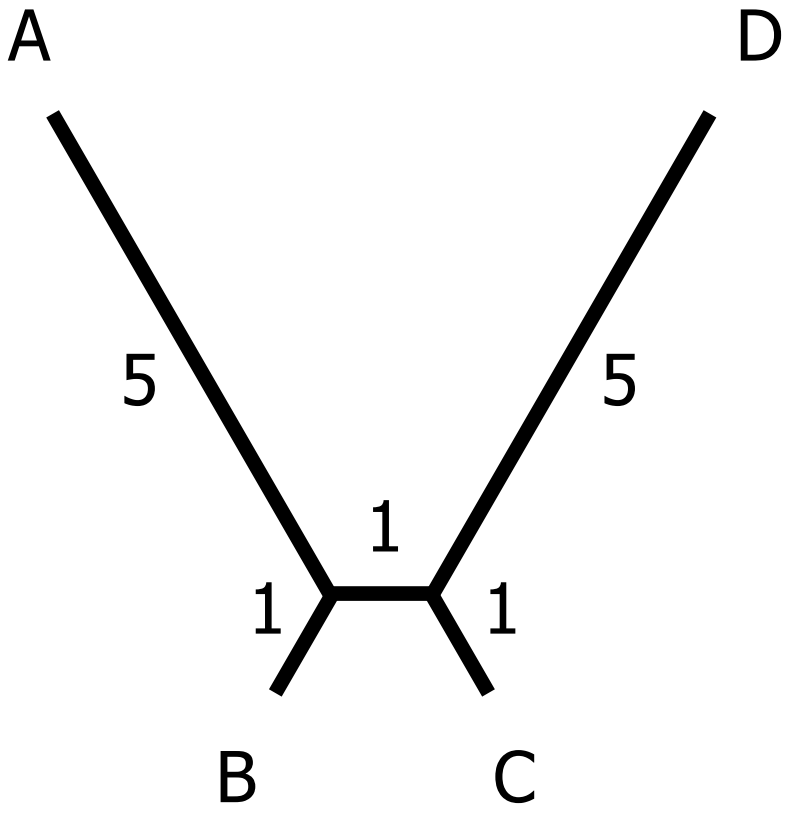
$$SS = \sum_{i<j} \frac{\left(d_{ij} - p_{ij}\right)^2}{d_{ij}^k}$$

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

k = 0 Cavalli-Sforza & Edwards (1967)

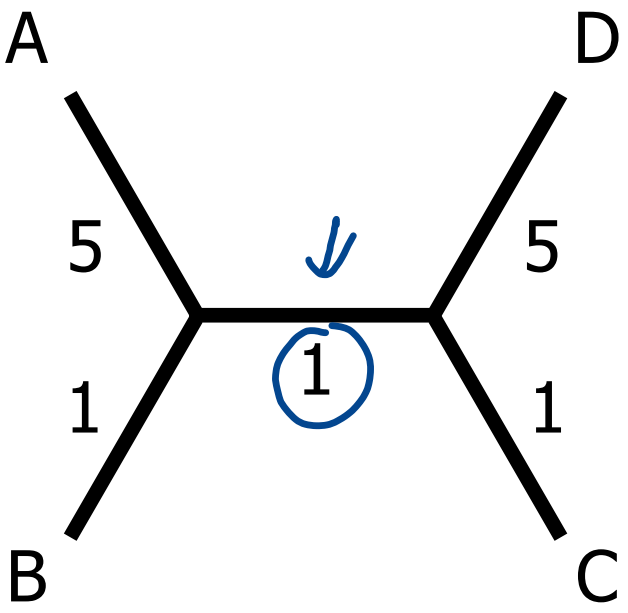
k = 2 Fitch & Margoliash (1967)

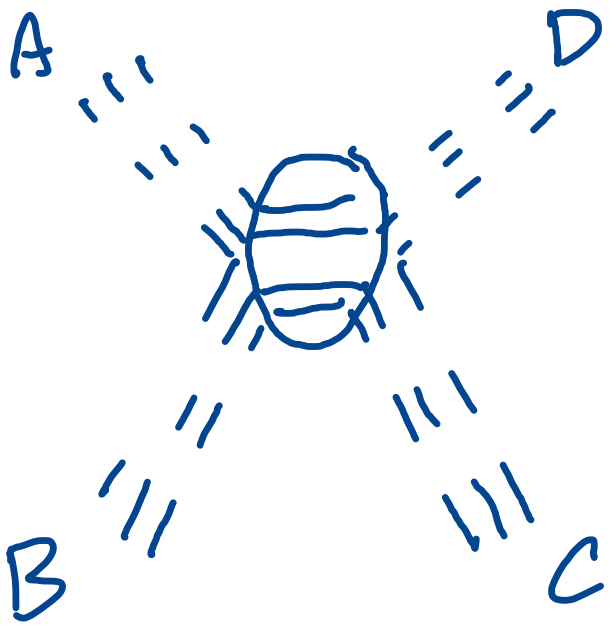
Estimating edge lengths using LS (4-taxon case)



distances from data

	A	B	C
B	6		
C	7	3	
D	11	7	6



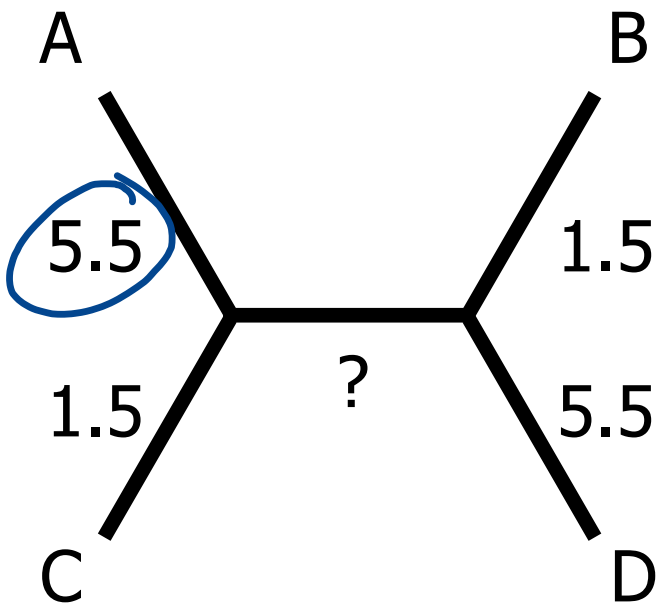


	A	B	C	D
A	0	6	7	11
B	6	0	3	7
C	7	3	0	6
D	11	7	6	0

$$AB + CD + AD + BC + AC$$
$$6 + 6 + 11 + 3 + 7$$

$$+ BD - 3AB - 3CD$$
$$7 - 18 - 18$$

$$= \frac{4}{4} = 1$$

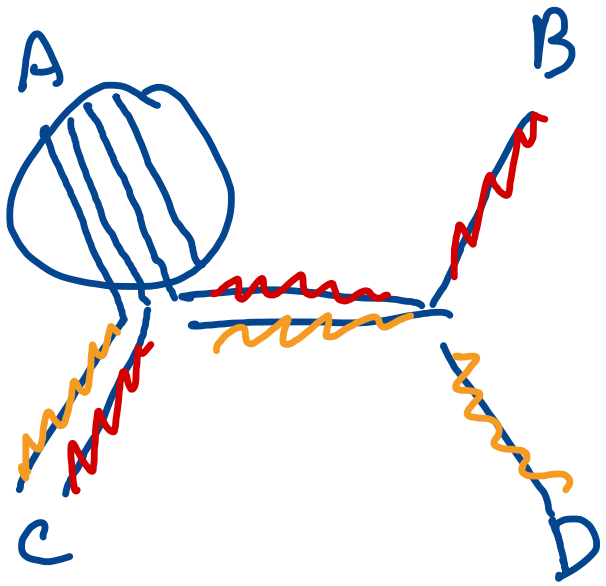


	A	B	C	D
A	0	6	7	11
B	6	0	3	7
C	7	3	0	6
D	11	7	6	0

$AB + {}^2AC + AD - BC - CD$

$6 + 14 + 11 - 3 - 6$

$= \frac{22}{4} = 5.5$

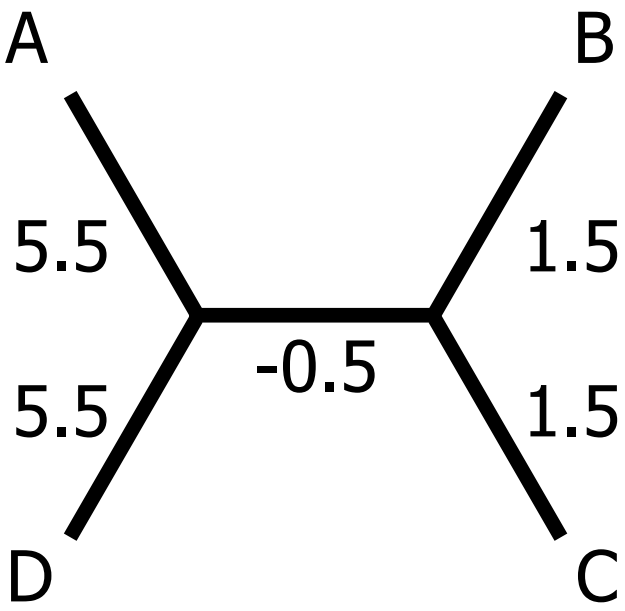


STOPPED HERE 2024-01-25

A C G T - - .
A C C T - - .

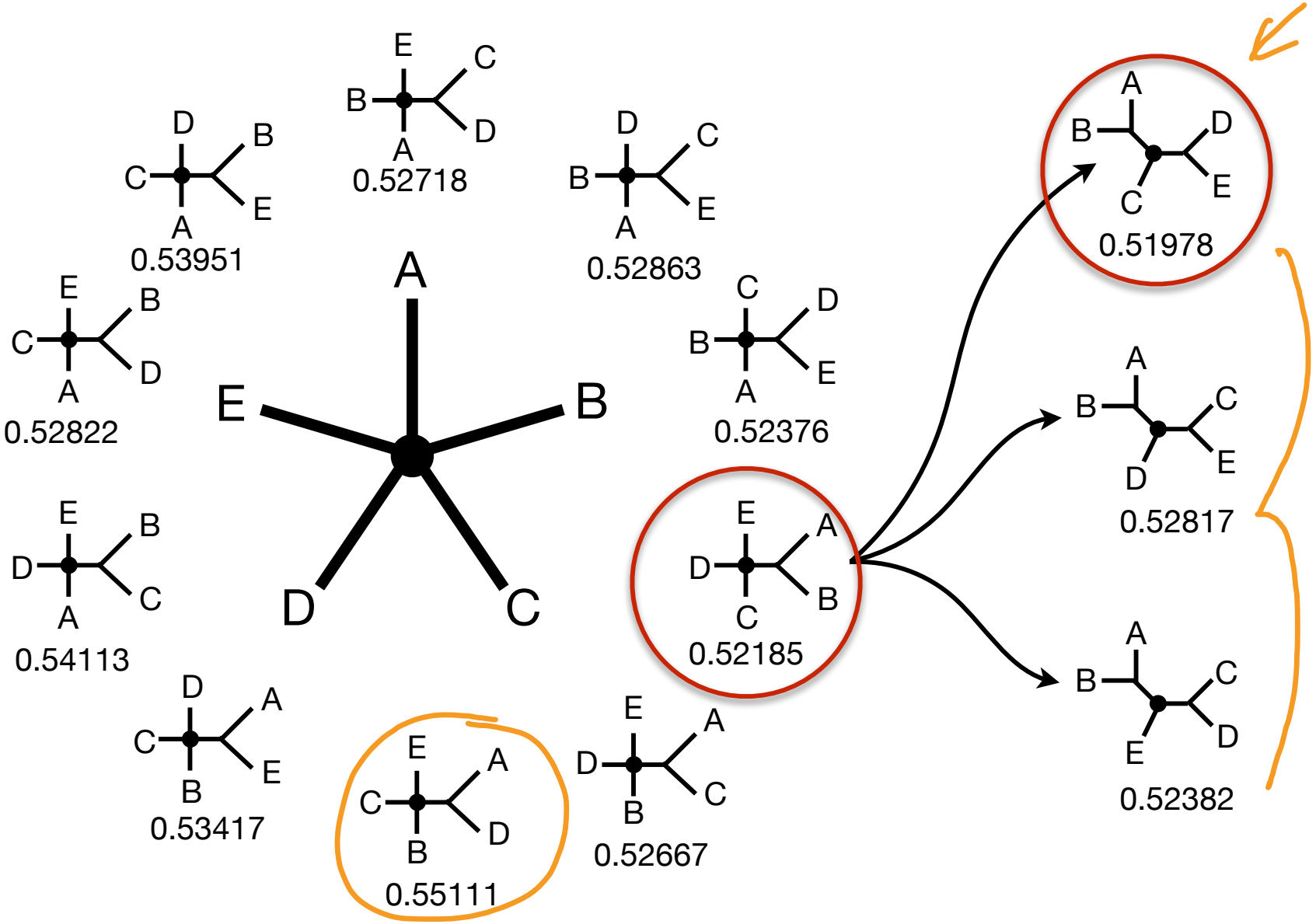
multiple hits

C → G G → C



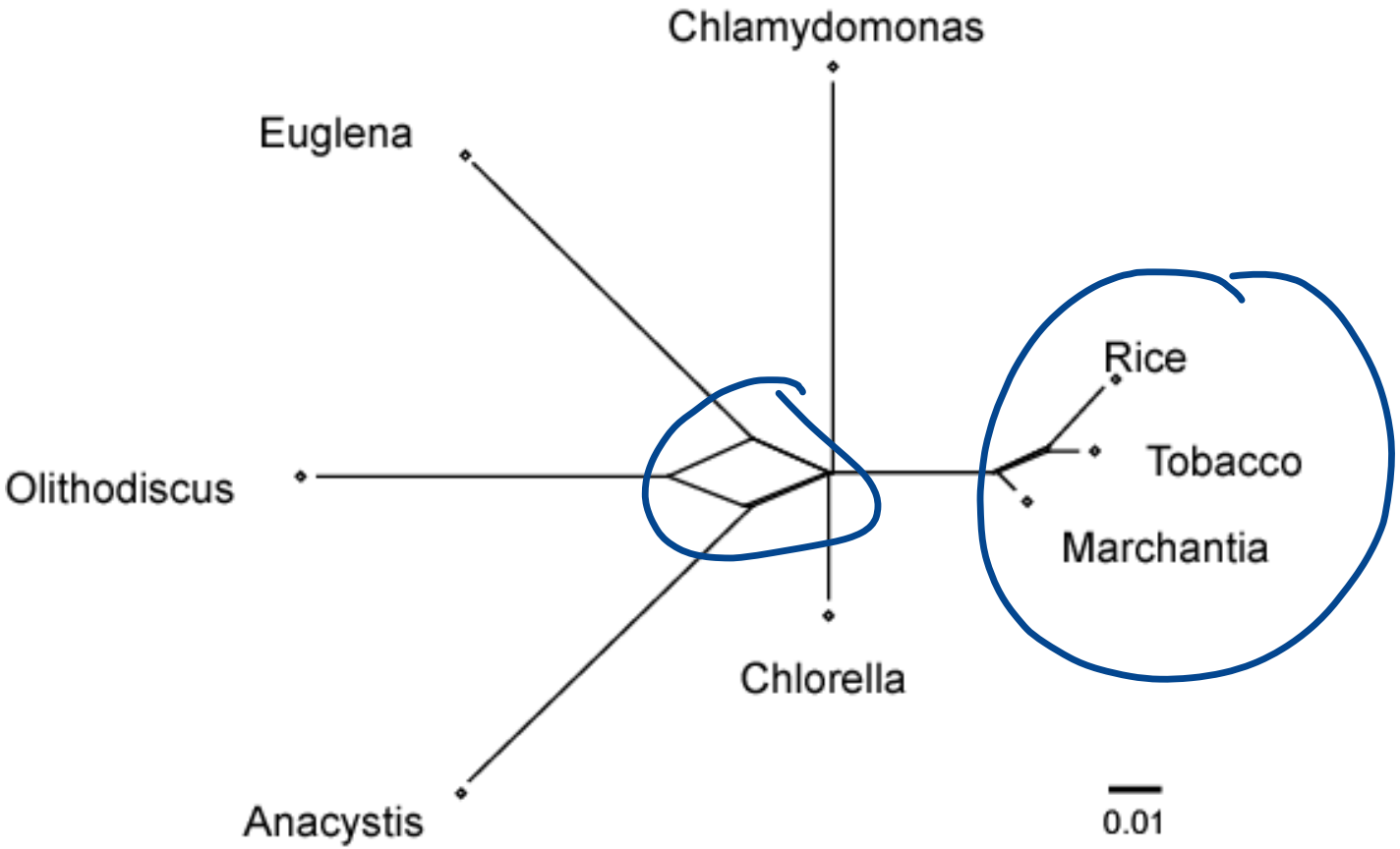
	A	B	C	D
A	0	6	7	11
B	6	0	3	7
C	7	3	0	6
D	11	7	6	0

Neighbor Joining



Saitou & Nei (1987)

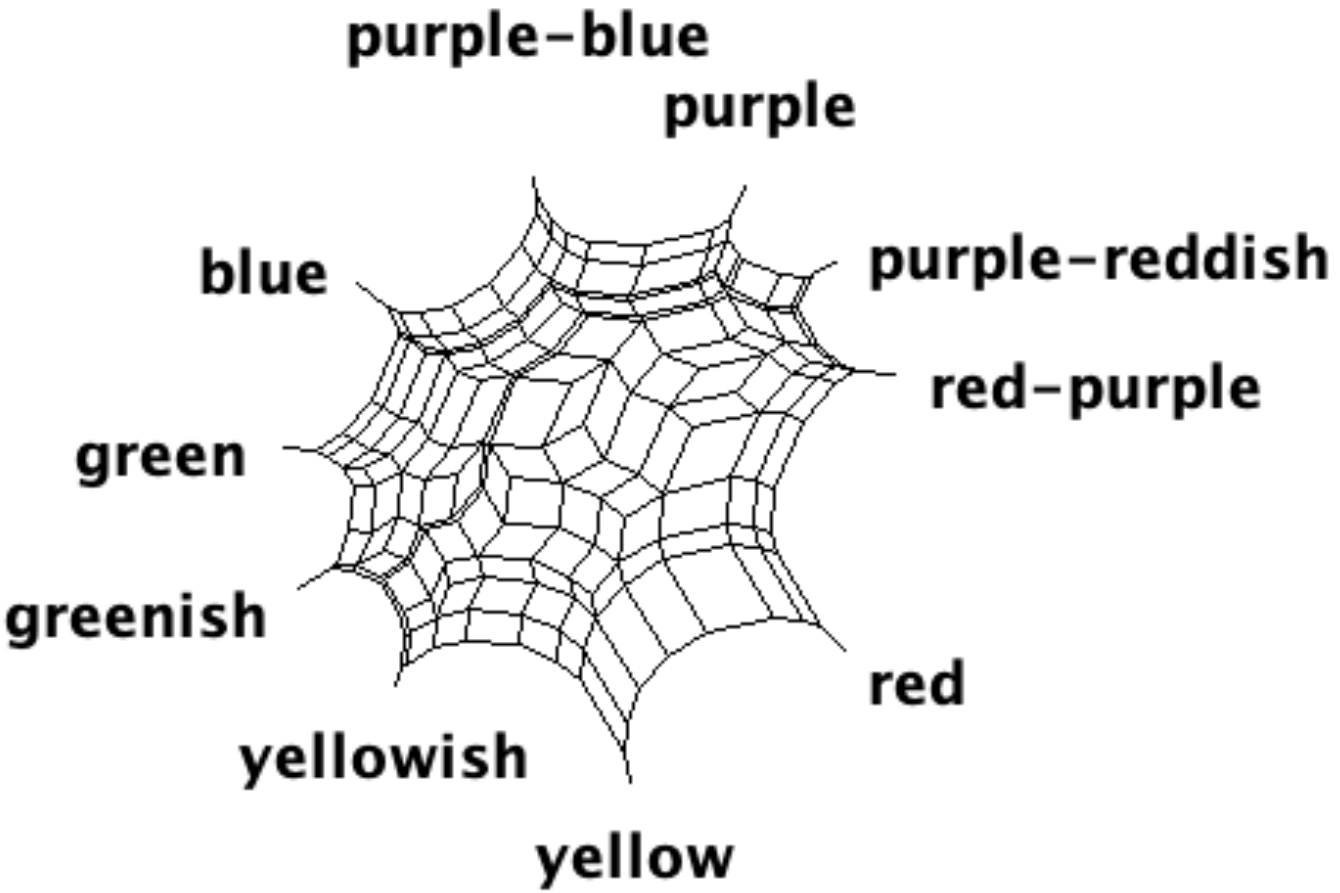
Split Decomposition



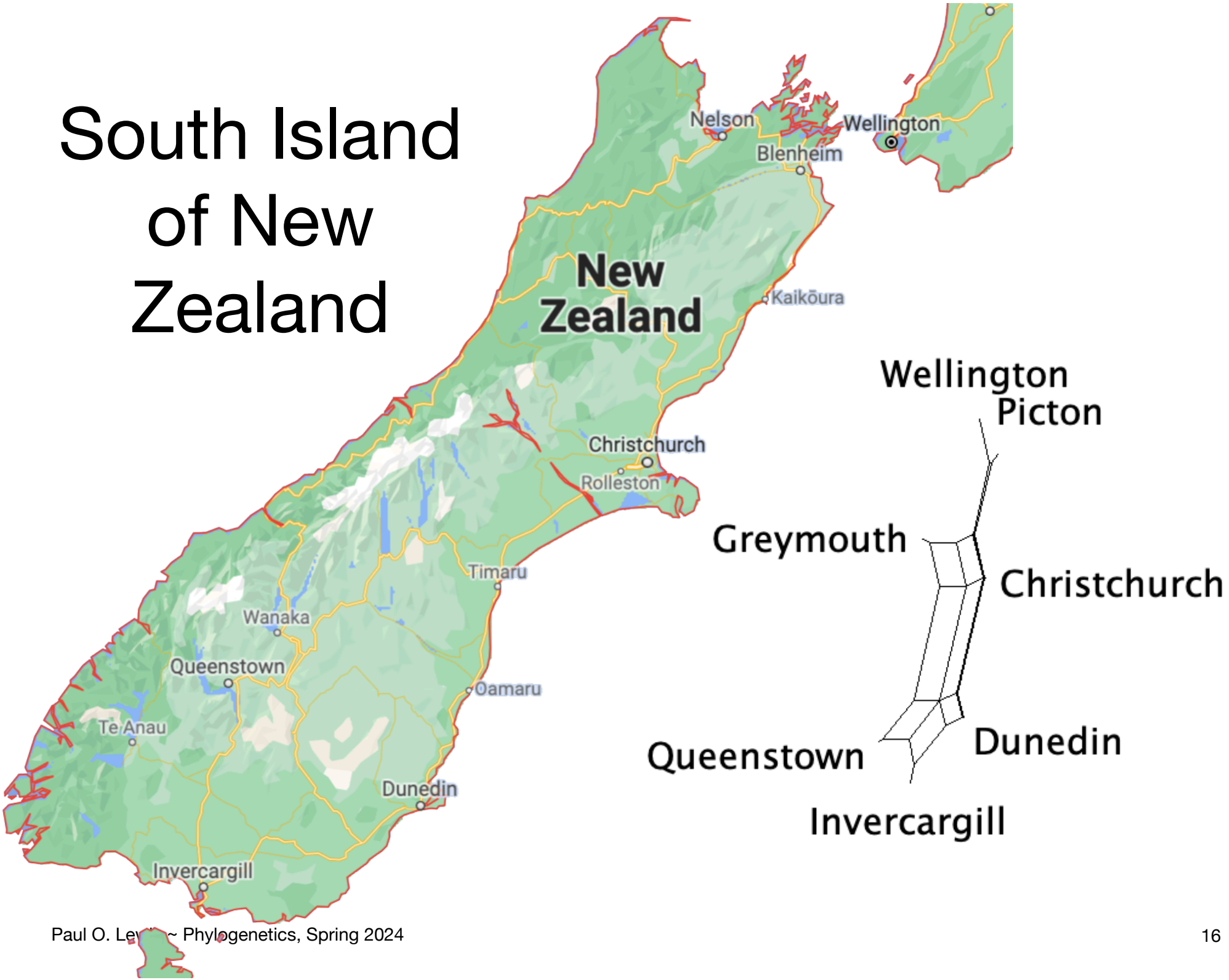
Produced by SplitsTree 4: https://tinyurl.com/4mwhcb49

Huson & Bryant (2006)

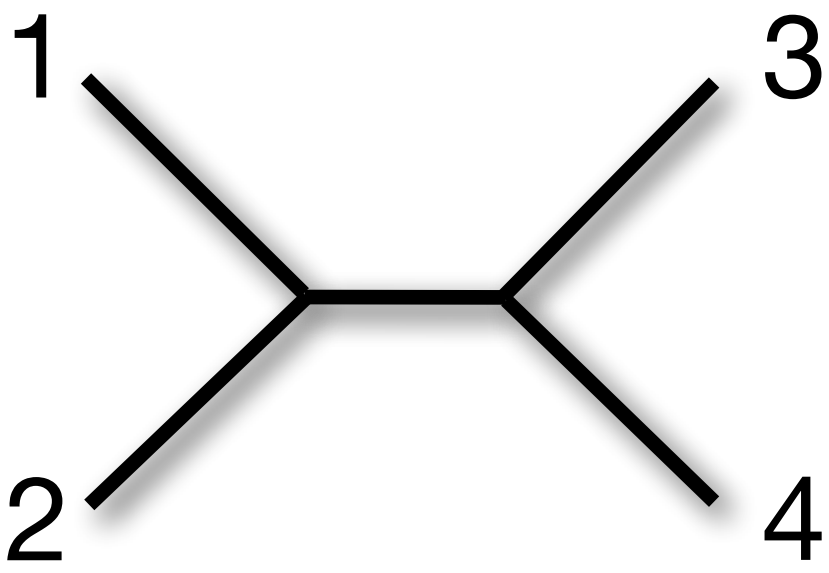
What if data are not phylogenetic?



South Island of New Zealand



Additivity and the 4-point condition



Question: On the tree at left, which of these three sums should be the smallest?

1) $d_{12} + d_{34}$
2) $d_{13} + d_{24}$
3) $d_{14} + d_{23}$

Buneman (1971)

Split Decomposition

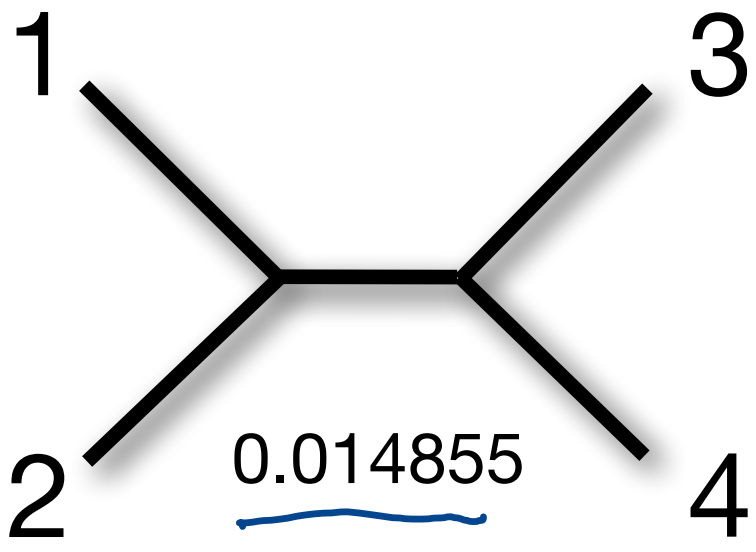
Pairwise distances estimated from real data

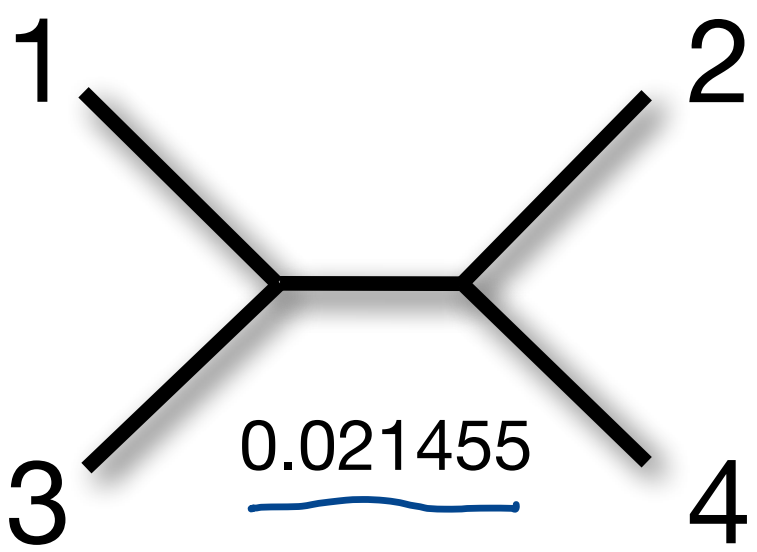
	1	2	3
2	0.13531		
3	0.11771	0.15842	
4	0.13751	0.13531	0.13091

$$d_{12} + d_{34} = 0.26622$$
$$d_{13} + d_{24} = 0.25302$$
$$d_{14} + d_{23} = 0.29593 \text{ (reject)}$$

Split Decomposition

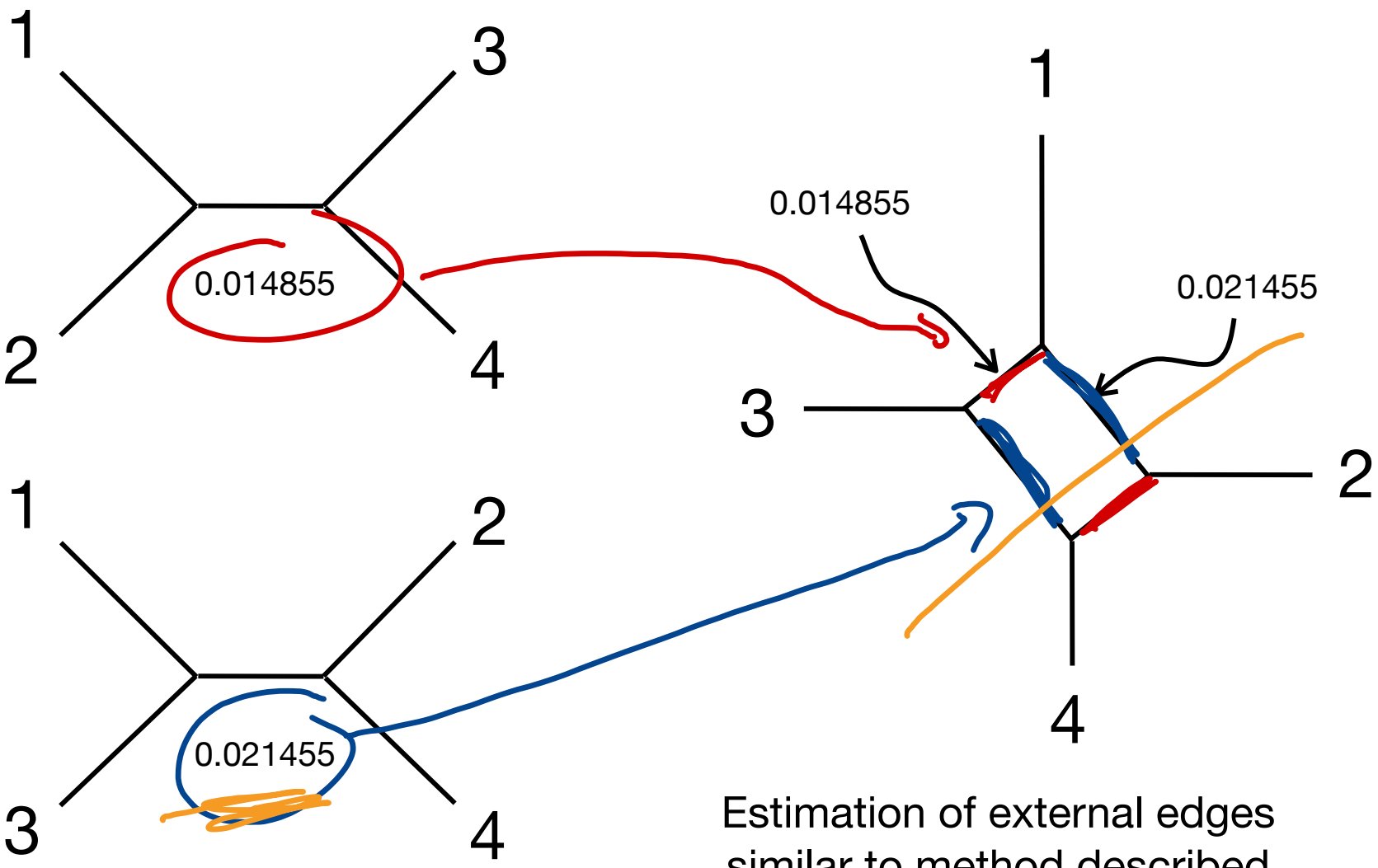
Estimate central edge for both remaining splits





Can make one "SplitsTree" that shows both splits simultaneously

Split Decomposition



Estimation of external edges similar to method described for central edge.