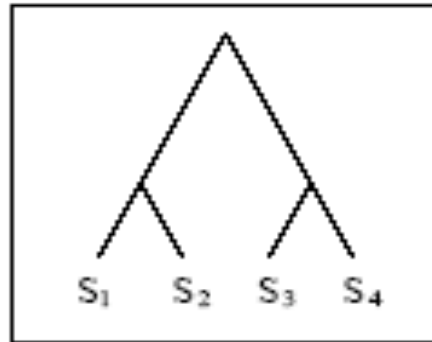# Distance based phylogeny reconstruction

Usman Roshan

Distance-based methods
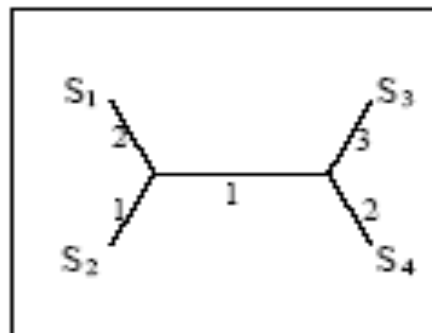

TRUE TREE

S_1 ACAATTAGAAC
S_2 ACCCTTAGAAC
S_3 ACCATTCCAAC
S_4 ACCAGACCAAC

DNA SEQUENCES

STATISTICAL ESTIMATION OF PAIRWISE DISTANCES

METHODS SUCH AS NEIGHBOR JOINING

INFERRED TREE

	S_1	S_2	S_3	S_4
S_1	0	3	6	5
S_2		0	5	4
S_3			0	5
S_4				0

DISTANCE MATRIX

Evolutionary distances

- We will use the Jukes Cantor model

$$d(x,y) = -\frac{3}{4}\ln(1 - \frac{4}{3}p(x,y))$$

- *x* and *y* are DNA/protein/RNA sequences
- *p(x,y)* is the number of mismatches in the alignment divided by the length of the alignment. This is also called the normalized Hamming distance.