

Exercise

part 1: construction of a distance matrix

Below you will find a DNA alignment.

Count the distance between each pair of sequences and enter the numbers in the distance matrix below.

Here, the distance between a pair of sequences is defined as the number of nucleotide positions where the two sequences differ. (In reality you ought to correct these observed distances for multiple substitutions. We will return to that issue later in today's lecture).

Sequence A: TCCCTGACCGTGT
Sequence B: AGCCTGACCGTGT
Sequence C: ACGGATTCGGTGA
Sequence D: ACGGAGAGGCTGA

Alignment

	A	B	C	D
A		2	8	8
B			8	8
C				4
D				

Distance matrix

Exercise

Part 2: tree reconstruction

Use the distance matrix from part 1 as the basis for constructing a phylogenetic tree for the sequences A, B, C, and D.

Specifically, add sequence labels (B, C, D) to the tree outlined below in the correct positions (I have already indicated the position of A). Label all five branches with the proper branch lengths.

Make sure that all pairwise distances measured along branches are identical to the values in your distance matrix!

Note that branch lengths are not drawn to scale in the tree below!

