

## Message from Prof Westra

It took some time to investigate this matter, and understand what was happening. By the way, I assume that you used the special matlab M file and data that can be downloaded from the case study page left-above – anyway, I included them with this mail.

Well, I must say, you (and the students) are right! Indeed, there seems to be no discernable difference between the trees obtained with the available methods to compute the distances between every pair of sequences and to correct them for multiple substitutions. The choices in the matlab-function `seqpdist` are: 'p-distance': proportion of sites at which the two sequences, 'Jukes-Cantor' (JC): maximum likelihood estimate of the number of substitutions between two sequences, and the 'alignment-score'.

After some careful investigations, my personal thought on this matter is that evolution did not get hold on this situation - yet: there are hardly any multiple substitutions. And if there are no multiple substitutions, the Jukes-Cantor correction makes no sense as it does not add new information.

To be honest, I composed the assignment in the 'intuition' that there *should* be a difference. In real cases where aeons have accumulated numerous mutations on the genetic material, the real phylogeny is obscured, and corrections like JC are obligatory. Here, however, this does not hold because of the short time spans involved.

Anyway, I hope it was a positive 'learning experience'!