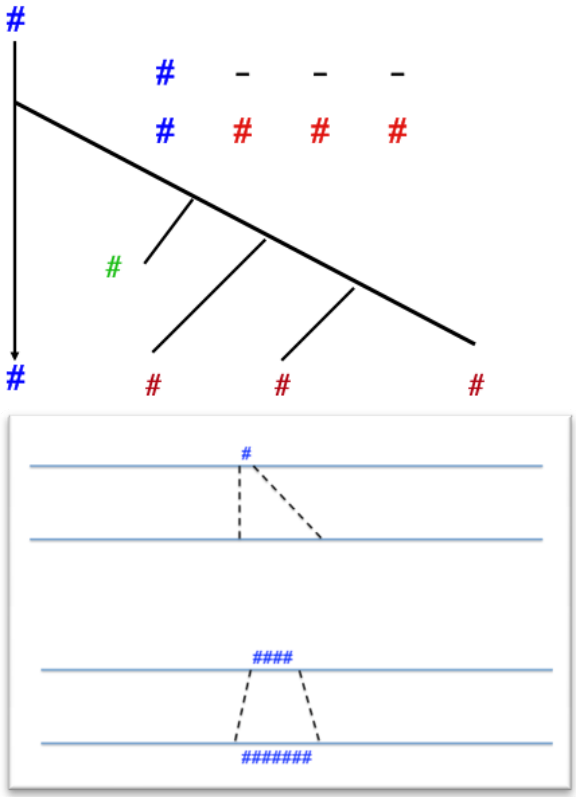
17. Investigating the Miklos-Lunter-Holmes (2004) model of insertion-deletion

Supervisor: Professor Hein

Stochastic Models of sequence evolution that include insertion-deletions (statistical alignment) has risen to prominence in the last decade. They have become important since they allow statistical inference in contrast to earlier methods (optimisation) that was based on mazimizing similarity or minimizing distance. However, the optimisation methods still has the advantage that they allow insertion and deletions of longer segments, while statistical alignment presently only considers insertion-deletions of one nucleotide at a time.

The original model of insertion-deletion (Thorne, Kishino and Felsenstein, 1991 [TKF91]) was immediately followed by a model (Thorne, Kishino and Felsenstein, 1992 [TKF92]) that had longer insertion-deletions, but in a very ad hoc way. Miklos, Lunter and Holmes (2004) [MLH04] considered a proper model with longer indels for two sequences, but it lead to a harder computational problem and was never generalized to more than 2 sequences.



Left: In the simple TKF91 process[the nucleotides are unspecified as #] a nucleotide and die or have children over time creating a descendant string. The overall result of this death-birth process can be found in the insert. Right: if deletions are of length 1, then each nucleotide can still be associated a string which simplifies analysis, but is unrealistic. In presence of longer deletions an segment can evolve into a segment, but the evolution of the complete sequence can be decomposed into segments, when a nucleotides survives into a nucleotide

This project will consider problems of increasing difficulty arising from the MLH04 model:

- Insertion of arbitrary length and deletions of length 1.

- Arbitratrary length for both insertions and deletions.

- Generalisations of multiple sequences

References:

Miklós, I., Lunter, G.A. & Holmes, I. (2004) A 'long indel' model for evolutionary sequence alignment. Mol. Biol. Evol. 21(3), 529–540

Thorne, J. L., Kishino, H. and Felsenstein, J. (1991). An evolutionary model for maximum likelihood alignment of DNA sequences. *Journal of Molecular Evolution*, 33: 114-

Thorne JL, Kishino H, Felsenstein J (1992) Inching toward reality: An improved likelihood model of sequence evolution. J Mol Evol 34:3-16

Prerequisites: good knowledge of probability theory and good programming abilities.