Plan7 pHMMs

Profile hidden Markov models capture the information in a profile, i.e. a multiple sequence alignment. Each alignment column is represented by a state. Passing through the model generates a symbol sequence that represents an *instance* of the modelled domain. We distinguish, in general, three kinds of states

1. match states: You can imagine these states as representing positions in the alignment where the majority of the sequences are represented by an amino acid. Each time you visit a match state, an amino acid is emitted. Different match states can differ in their emission probabilities for the 20 amino acids.
2. insert states: These states represent positions in the alignment where individual sequences have inserted amino acids that are not present in the others. It is not uncommon that all insert states have the same or at least similar emission probabilities for the 20 amino acids.
3. delete states: These states are silent states, as they do not emit any amino acid when visited. Instead, they are used to hop over a match state, by that 'deleting' it.

Plan7 pHMMs have additional emitting states, N, C, and J that model residues inbetween two domain instances (J) as well as sequences flanking the domain (N, J) (Fig. 1).
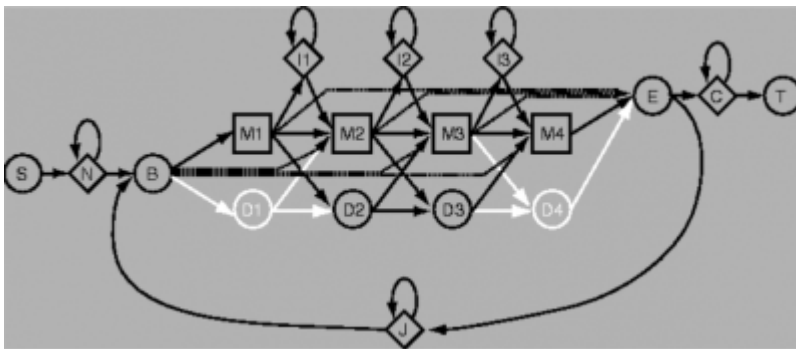

Figure 1: Each triple (Mi, Ii, Di) is called a node. PLAN7 pHMMs have only 7 transitions per node as I→D and D→I transitions are not allowed