Introduction to Bioinformatics for Computer Scientists

Lecture 8a

Outline – next Lectures

- Lecture 8
  - Lecture **8a**: a few more words about search algorithms
  - Lecture **8b**: Introduction to Markov Chains
- Lecture 9
  - Lecture **9a**: Maximum Likelihood – motivation
  - Lecture **9b**: Computing the Likelihood on a tree
  - Lecture **9c**: Computing the Maximum Likelihood on a tree
- Lecture 10
  - Likelihood Models – Implementation, Optimization, Parallelization
- Lecture 11
  - Discrete Operations on trees

Question

- How could one design a search algorithm using, for instance, the least squares criterion given a function *f()* and a distance matrix *D* to compute the least squares score on a given tree?

Answer

- First, we need a starting tree
  - Random starting tree
  - Neighbor Joining Tree
  - Stepwise randomized addition order tree using the least squares criterion → our function *f()*
- Once we have a comprehensive (containing all taxa from the input dataset) tree → apply topological moves to it to try and improve (minimize!) the least squares score
- Numerous options to design such an algorithm
  - Just randomly apply NNI, SPR, or TBR moves for a long time
  - Design a genetic search algorithm
  - Use simulated annealing
  - Use some sort of hill climbing, for instance, continue applying all possible NNI moves on all inner branches of the tree until no NNI moves yields a better tree (i.e., we have reached a local minimum)

An example - The Parsimonator Algorithm

- Build a randomized stepwise addition order parsimony tree
- Apply SPR moves to all subtrees of the current (comprehensive) tree with a rearrangement radius of *20*
- If the rearrangement of a subtree yields an improved parsimony score, keep it immediately

  → this is somewhat greedy as opposed to a steepest ascent hill climbing algorithm
- Continue applying SPR moves with a radius of *20* to all subtrees until no tree with a better parsimony score can be found
- There are much more sophisticated algorithms available

  → TNT tool by Pablo Goloboff
- Keep in mind that parsimony returns discrete scores, that is, there may be many equally parsimonious trees among which we can not distinguish!