NJ-networks, turning greedy into all possibilities

Armin Hoenen, Texttechnology Lab, Goethe University, Frankfurt am Main, Germany

Background

The Neighbor Joining algorithm [1] is well used for the computation of phylogenies. Based on a distance matrix, it proceeds in a stepwise manner. At each step one new minimal value is chosen. With two equally minimal distance values, NJ chooses only one of them. Such an event is fairly rare but could lead to different outputs on the same data set from which various methods can produce single networks [2].

Methods and results

We present an implementation of NJ which at each point of encounter of more than one minimum proceeds following all possible ways of computation assembling all equally likely NJ trees on the same distance matrix. Furthermore, we introduce rounding at the distance matrix update steps as a free parameter in order to force minima to converge, so as to be able to produce NJ-networks from any kind of data set. Through a simulation study randomizing large numbers of input distance matrices, we investigate and report the expectable incidence of multiple minima for NJ and estimate the setting of the free parameter when one aims at producing exactly 2 or more trees from one data set.

Discussion

In some biological cases, (e.g. in bacterial and plant evolution), we deal with certain amounts of lateral gene transfer or hybridization. In these cases sometimes exactly 1, 2, .. n such events or a certain proportion may be expectable. Instead of resorting to other algorithms, NJ-networks and the free parameter allow to use the efficiency and accuracy of NJ and produce 2 or more phylogenies by one algorithm e.g. as input to a minimum hybridization network. The rounding procedure achieving this may theoretically correspond to a degree of uncertainty allowing the evolutionary model some flexibility.

References

[1] Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), 406-425.

[2] Huson, D. H., & Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic biology*, *61*(6), 1061-10.