

Modeling molecular phylogenetic inferences using distance methods

1st *Negar Karimi Ghahfarokhi¹, 2nd Hojjatollah Saeidi²

¹ Department of Plant and Animal Biology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran.

² Department of Plant and Animal Biology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran.

*negar.karimi.gh@gmail.com
damghan@du.ac.ir

Abstract— Background; The principle of distance methods is that when we know the evolutionary distances between each pair of taxa, the average number of substitutions per site in a DNA sequence, then these distances will lead to a phylogenetic tree. To do so we need to calculate the proportion of traits that differ between two taxa and measure their pairwise distance. There are several methods for estimating these numbers. Aim of study; for evolutionary purposes, pairwise distance is usually measured by the number of nucleotide or amino acid substitutions between two taxa. Here we discussed different distance methods to reveal their advantages and disadvantages and to understand their usability. Material and methods; Most of the well-known pairwise distance methods were studied. The methods were divided into three categories of nucleotide substitutions, amino acid substitutions and synonymous and nonsynonymous substitutions. The methods were further discussed using these quantities to compute the proportion of synonymous and nonsynonymous nucleotide differences per synonymous and nonsynonymous DNA and amino acid sites. Results; the methods used to estimate the number of amino acid substitutions are generally similar to those used for estimation of nucleotide substitutions except that there are 20 different states for amino acid substitutions and there are only 4 different states for nucleotide substitutions. All mentioned and discussed distance measures can be computed for both amino acid and nucleotide sequences as well as synonymous and nonsynonymous substitutions. The overall method is based on the calculation of number of substitutions occurred between each pair of comparing sequences. Conclusion; the results suggest that evolutionary distances are useful methods for studying molecular evolution and are useful for phylogenetic reconstruction and estimation of divergence times between different taxa. However; choosing the best distance method depends on the purpose of phylogenetic inference, the number of data used, amount of nucleotide substitutions per site, amount of transition/transversion ratio, frequencies of nucleotides or amino acids. Finally, to avoid large standard errors it is suggested to use genes that evolve more slowly or to use amino acid sequences rather than nucleotide sequences when a coding region of DNA is examined.

Keywords— Pairwise distances, numerical phylogenetic, Tamura distance, Gamma distances, Tamura-Nei distance, Jukes-Cantor distance

REFERENCES

- [1] B. David, A. Stacey, D. Smith, S. Donovan, The tree-thinking challenge. Science: 310.5750, 2005, pp. 980.
- [2] F. Joseph, Distance methods for inferring phylogenies: a justification. Evolution, pp.16-24, 1984.
- [3] Z. Richter, O. Eva, M. Bäcker, S. Vogt, Review of distance education research (2000 to 2008): Analysis of research areas, methods, and authorship patterns. International Review of Research in Open and Distributed Learning, vol. 10.6, pp. 21-50, 2009.