A critical study of homoplasy in molecular data with the use of a morphological based cladogram, and its consequences for character weighting

H. Philippe, G. Lecointre, H.L. Van Lê, H. Le Guyader. 1996. Mol. Biol. Evol. 13 : 1174-1186.

An improvement in molecular phylogenetic inference is usually expected with the elimination of highly variable positions from nucleotide sequences. In this paper, this point of view is addressed through the evaluation of 28S rNA sequences of vertebrates by the use of the maximum-parsinomy method. A tree based on morphological data, which is considered as true has been used to determine these positions, which are homoplastic. The congruence of the tree constructed after the removal of such positions with the true tree was not better than that of the tree constructed with complete sequences; however, it displayed more multifurcations. The elimination of homoplastic sites appeared to decrease the signal rather than the noise. This is due to the fact that noise is created by multiple changes at a given site only if they occurred within a small time interval; when they occur over a broad time interval they, in fact, provide information. Thus, the global variability of a site is not a reliable indicator of noise. We have shown that, for a given position, changes were often concentrated in a limited area of the tree whatever the total number of substitutions was. This observation was generalized by comparing the number of steps in two sister group for various molecules and taxonomic groups. The evolutionary rate of a given position was thus shown to vary throughout time. Two major conclusions emerged from this study : (1) substitution models should incorporate variation in evolutionary rate at a given site; and (2) the character weighting approach could be significantly improved if the weight given to a position depends on the area of the tree.