

How many nucleotides are required to resolve a phylogenetic problem ?

The use of a new statistical method applicable to available sequences

G. Lecointre, H. Philippe, H.L. Van Le, H. Le Guyader. 1994. *Molecular Phylogenetics and Evolution*. 3 : 292-309.

The evolution of bootstrap proportions (BP) with sequence length was studied using a 28S ribosomal RNA data set. For different sequence lengths, informative sites were jackknifed several times. Bootstrapping was subsequently performed on each of these subsamples. For each node, Bps so obtained were plotted against sequence length, showing the evolution of the robustness with increasing number of informative sites. For robust nodes (BP of 100%), the pattern of Bps is unvarying and is described by a simple function $BP = 100 (1 - e^{-b(x - x')})$, where x is the number of informative sites and b and x' are two parameters estimated using a nonlinear regression procedure. When a node has a $BP < 100\%$ and the pattern of Bps fits this function, it is possible to estimate the number of informative sites required to obtain a given average BP. The method also identifies nonrobust nodes (nonascending clusters of BP dots), for which it seems to be more cost effective and fruitful to turn to other species and/or genes rather than to continue sequencing longer gene lengths from the same species to reach a BP of 95%.