PHYLOGENY ESTIMATION AND PHYLOGENETIC INFERENCE IN EVOLUTIONARY STUDIES

Summary

Modern phylogenetics, although rooted in Darwin's and Haeckel's ideas on evolutionary relationships among organisms, dates back to the second half of the 20th century and the advance of numerical methods in taxonomy. Its beginnings were marked by a fierce debate between phenetics and cladistics but at present it incorporates a diverse array of methods including those based on distance and clustering algorithms, parsimony, maximum likelihood and Bayesian statistics. The phylogeny of extant organisms is usually inferred using molecular markers, because they are genetic, less arbitrary (do not require arbitrary coding), more additive, less prone to convergence and more universal than traditional morphological markers. Phylogenies inferred using molecular data are usually more stable and have better internal support than those obtained from morphology. However, the informed user of phylogenetics methods must be aware of their assumptions and caveats. The chosen sequences must be orthologous (resulting from a speciation event), as opposed to paralogous (resulting from a duplication event); choosing orthologous sequences does not guarantee that the phylogenetic signal is undisturbed. Reversals, multiple hits and parallel substitutions may result in a higher similarity of sequences than expected from their evolutionary history and therefore affect the phylogenetic reconstructions. Moreover, trees inferred from molecular data are usually gene trees rather than species trees. There are several processes that may result in discordance between a gene tree and an organism tree including interspecific hybridisation, horizontal gene transfer, incomplete lineage sorting and selection for allele polymorphism. The most commonly used phylogenetic methods include those based on parsimony, distance and clustering, maximum likelihood and Bayesian statistics. The last three employ nucleotide substitution models. Each method is based on certain evolutionary assumptions that may not necessarily apply to a given data set. Noteworthy are recent advances in methods of inferring divergence times using relaxed molecular clock. In evolutionary biology, molecular phylogenies are widely used in comparative studies, historical biogeography and for analysing character state evolution.