stimulating as its predecessors, reflecting the enormous effort devoted by Kissmeyer-Nielsen and his associates to its organisation and showing once again the value of such world-wide collaborative studies. The next workshop will be organised in Oxford in about 2 years' time and will have as its main theme the definition of human Ia specificities, their relation to MLC determinants and their significance for disease association and clinical transplantation. □

Reliability of molecular phylogenetic trees

from David R. Thatcher

IN principle a comparison of the primary structure of the same protein from a number of related organisms should provide evolutionary information of high precision. The amino acid sequence of a protein is a translated copy of the sequence of nucleotide bases in the structural gene and therefore is the one phenotypic character that is a direct reflection of the structure of the genome. The comparative morphology of amino acid sequences should therefore permit the deduction of some of the discrete mutational steps which have occurred during the evolution of an organism. Providing no genetic transfer has occurred and the gene has retained the same function throughout its history, a phylogenetic topology derived from amino acid sequence comparisons should represent the true ancestral relationships of the organisms expressing that particular gene. Molecular phylogenetic methods have the potential advantage of being able to distinguish evolutionary relationships where the classical methods of evolutionary biology do not produce a convincing answer, for example where the fossil evidence is controversial or absent or where there is little morphological diversity to compare.

In practice little new phylogenetic information has been derived from the construction of trees based on molecular evolutionary data and the method has received a somewhat hostile reception from other evolutionary biologists (these points are fully discussed in the excellent review of J. Williams in *The Chemistry of Macromolecules: Proteins*; Medical and Technical Publishers, 1974). The most damaging criticism proposed is that data based on a single structural gene cannot possibly reflect the evolution of the whole genome. Nevertheless, trees constructed by comparing sequences of vertebrate cytochromes *c*, haemoglobins and fibrinopeptides are each more or less in agreement with the accepted phylogeny for this group, showing that the method at least works even if its precision is not as great as at first hoped.

The resolution of the sequence method is unquestionably restricted and if future projects are going to be attempted with an eye to solving particular phylogenetic problems, these limitations must be fully defined. These limitations are due to two main factors. First, to what extent is the assumption of parsimony justified at the molecular level? There is no absolute way of knowing the degree to which parallel, back or convergent mutations affect observed amino acid similarities. Second, if one accepts the most parsimonious solution as being the actual evolutionary history of a gene, the numerical problems involved in calculating this solution for even a small number of alternative topologies is formidable.

Peacocke and Boulter (*J. molec. Biol.*, **95**, 513; 1975) have begun to assess the extent to which these technical limitations have affected their conclusions on higher plant evolution. The evolution of a protein was simulated in a computer and the reconstruction of this hypothetical phylogeny was attempted using two different methods. The accuracy of each procedure was then estimated by determining the number of times a branch had to be moved in order to restore the original topology. The two methods employed were the ancestral sequence method of Dayhof (*Atlas of Protein Sequence and Structure*, 1972) and the matrix method of Moore, Goodman and Barnabas (*J. theor. Biol.*, **38**, 423–457; 1973). The ancestral sequence method infers the structure of nodal ancestral sequences as it generates a tree by adding one sequence at a time to the topology. New branches are added with respect to the most parsimonious sequences (real and ancestral) already in the tree. In the matrix method a distance matrix is calculated and an approximate tree is produced by successive pairwise clustering. This tree is then adjusted until hypothetical mutational distances (computed from the tree by comparing the mutational distances of two neighbours with the average mutational distance of all other members of the tree) approach the actual values of the original distance matrix. Applied to the model data the accuracy of the ancestral sequence method became progressively worse as the number of observed substitutions in the data set was increased. In contrast the matrix method, although less accurate than the ancestral sequence method at lower degrees of substitution, improved as the variation between the sequences increased. Phylogenetic trees were then constructed from model data which varied to the same extent as the real data Boulter's group has accummulated on higher plant cytochrome *c*. The error observed in tree building with the model data and the ancestral sequence method ranged between 2% and 8% whereas the matrix method operated with an error of 9% to 15%. For a rapidly evolving protein like plastocyanin methods were equally successful in predicting the correct topology (error 8% to 12%).

Before molecular phylogenies are considered seriously by systematicists, some attempt must be made to quantify the errors inherent in each method; from the extent to which tree topologies may be influenced by sequencing mistakes to the probability of a given tree being the most parsimonious solution possible. Peacocke and Boulter have shown that an optimal amount of amino acid variation is required for the most accurate topological assignments by the ancestral sequence method. As structural genes have characteristic rates of evolution, which vary greatly from gene to gene, only a limited range of accurate phylogenetic information can be obtained from the study of a single gene.

Convincing evidence is obtainable but only by investigating the best gene for a particular phylogenetic problem and applying the numerical method most suited to the data. □

Consequences of hydrothermal circulation

from Peter J. Smith

LAST year, Lister (*Eos*, **55**, 740; 1974) made a strong plea for the testing of the hydrothermal circulation hypothesis using deep sea boreholes near oceanic ridges. His argument was that there is growing evidence that certain observations, notably in the fields of oceanic heat flow and ocean floor mineralisation, can only be explained in terms of the penetration of water to substantial depths (possibly several kilometres) in the new crust forming at ridge crests.

Presumably it is still too soon to expect any results to have emerged from such a proposal. In the meantime, however, indirect support for hydrothermal circulation continues to accumulate. Lee and Von Herzen (*Geophys. Res. Lett.*, **2**, 201; 1975), for example, have now reported 15 new heat flow measurements from the South Atlantic triple junction near 55 °S, 0 °E —the zone in which the boundaries of the African, Antarctic and South