Sound of distance drums

Vincent Sarich

Molecules and Morphology in Evolution: Conflict or Compromise? Edited by Colin Patterson. Cambridge University Press: 1987. Pp. 229. Hbk £25, \$49.50; pbk £9.95, \$15.95.

"CONFLICT or compromise?" asks the title of this book. The answer, of course, is "neither" — something only Sibley and Ahlquist, among all the contributors, recognize explicitly. They note, first, that "there is only one true phylogeny" (p. 95), and later that molecules and morphology necessarily complement one another. The molecules are able to "reconstruct the phylogeny with a high degree of accuracy" (p. 118); and, once given the phylogeny, the organismal biologist can use it as a framework within which to interpret structure, physiology and behaviour. We thus want to "cease the debate about 'molecules vs morphology' and begin to co-operate on the basis of molecules and morphology" (pp. 118–119). But that is an ideal which the other authors do not even begin to approach. It is instructive to ask why, some 30 years into the renaissance of the molecular approach to systematics, it is still so often *vs* rather than *and*?

A part of the answer is that non-molecular types can often increase their short-term intellectual fitness by keeping it that way; that is, by not having to accept the constraints that molecular phylogenetic frameworks often impose. The most effective way to do this is to reduce the number and resolving power of such frameworks, and this is readily accomplished by restricting the molecular data worthy of consideration to nucleotide and amino-acid sequences — that is, to 'character-state' data which they consider the only type, morphological or molecular, that can be objectively interpreted within a cladistic framework. This means that distance methods (DNA hybridization, immunological comparisons of proteins) are then rejected on 'theoretical grounds', with the 'theory' being little more than the cladist dogma which states that because distance data are not character-state data, they are not phylogenetically relevant.

It is this explicit rejection of distance methods that is the besetting sin of this book, and it is a sin against science because the rejection is imposed by theory, and not by a demonstration that distance data are indeed phylogenetically irrelevant. If they actually were, their irrelevance could readily be shown by demonstrating the total non-congruence of molecular distance trees with those more or less well-known on the basis of other data. This

would be an impossible task, so it has never been undertaken; and the rejection has simply taken the form of ignoring the distance data because they do not conform to 'theory' — that way the inconvenient point that they work in practice could be ignored.

This is most blatant in McKenna's article on high-level mammalian relationships, but is also, and less excusably, omnipresent in the papers by Andrews, and Goodman and his colleagues. I think that a simple challenge is in order. Is there any relationship among mammals that has been resolved using amino-acid sequences that has not already been resolved by using distance data (generally, immunological

Ernst Haeckel's evolutionary tree of life (1910). The picture is reproduced from Rupert Sheldrake's *The Presence of the Past* (a sequel to *A New Science of Life*, reviewed in *Nature* 294, 32; 1981), which is published today by Collins.

comparisons of various serum proteins)? Indeed, do the available sequence data even give as much resolving power? So why do Goodman and his colleagues persist in ignoring those distance data? Similarly, why does McKenna avoid any mention of the results of my own group on inter-ordinal relationships among the mammals he discusses? Those results, for example, clearly associate sirenians, proboscideans, hyraxes and the aardvark; pangolins and carnivores; sloths, anteaters and armadillos to the exclusion of pangolins and the aardvark; cetaceans and artiodactyls; and indicate very strongly that Glires is not a natural unit. I find it difficult to see all this as anything other than institutionalizing ignorance by concentrating on molecular data that cannot have the requisite resolving power, and ignoring those that have been demonstrated to have it.

Given this institutional bias, it is

incumbent upon those doing good work using distance methods to leave as few targets as possible in their presentations. This Sibley and Ahlquist have failed to do. First, they persist in using a distance statistic, $\Delta T_{50} H$, which is quite arbitrary in that it is not inherent in the melting curves themselves, is overly sensitive to variation in the degrees of hybridization achieved, and allows the ascription of greater precision and range than are actually present in the data. All of this could be avoided by the use of the T_{mode} statistic. The mode of a melting curve, provided there is one, is an integral part of it, and if it is missing that tells us that the relationship involved is beyond the useful range of the approach. Second, they persist in drawing trees which claim too much resolving power. Nodes separated by less than 0.5° or so are simply not credible, and the fact that Sibley and Ahlquist are now claiming rate variation, without allowing for it in the construction of their avian trees, only exacerbates the problem. No method, even in theory, could always give us only dichotomous nodes; thus actual practice, which necessarily involves experimental and interpretative uncertainties, must allow for nodes from which three or more lineages stem. Sibley and Ahlquist, sadly, fail to do so.

Those sins, however, are minor and remediable. Those of a number of other contributors are not. Correction of them is going to require the recognition that the cladistic enterprise of recent years has, in effect, reinvented a flawed wheel, and then found that it rolled with no greater success than it had previously. The reason is that a cladistic analysis tends to be after the fact. That is, one has to know one's organisms very well indeed to know which characters are phylogenetically relevant, and which are not. But how do you judge such utility without having already come to some conclusions as to the phylogeny involved? Simply, you cannot, which makes the delusory possibilities here essentially unbounded. If the answer isn't 'obvious', then by definition there will be characters out there to support each of the various phylogenies. Yet because characters can be neither equal, nor objectively chosen, statistical testing of those phylogenies is precluded.

So where does that leave us? If we follow much of the volume under review, the answer is nowhere. If we do not, and still want to make a practical molecular contribution to systematics, it is to save the sequence approach for situations where distance approaches *do not work*. To do so, we of course need to show that they can work. Given some of the closed minds on view in this volume, that will not be a trivial task. □

Vincent Sarich is a Professor in the Department of Anthropology, University of California, Berkeley, California 94720, USA.