A Systematist Looks at Cytochrome c

R. A. Crowson
Zoology Dept., University of Glasgow
Received September 21, 1972

Summary. The available data (as of June 1972) on amino-acid sequences of cytochrome c are reviewed from the point of view of a traditional phylogenetic systematist. The Darwinian assumption, that phylogenetic changes in the sequence have been controlled by natural selection, is made, and some tentative phylogenetic and systematic conclusions are drawn. Attention is drawn to apparent correlations between substitutions at different points in the molecule. Suggestions for further investigations are made.

Key words: Cytochrome c — Systematics — Phylogeny — Ungulata — Vertebrata.

If the biochemists who have been concerned with comparisons of amino-acid sequences in various proteins in different species have failed to appreciate the relevance of classical systematic procedures to their problems, the generality of systematists concerned with the same species have been as much to blame for ignoring the evidence available from protein studies, except perhaps for the serological data. The aim of this article, however, is not to apportion blame but to help bridge a gap.

The complexity of the biochemical procedures involved in the extraction and analysis of specific proteins is such that it would need exceptional courage, and exceptional financial support, for a systematist to engage in them; until now, these procedures have been the preserve of specialist biochemists. From the publication of Anfinsen's (1959) notable book onwards, the evolutionary and systematic interpretation of macromolecular data have also been left in the hands of these same biochemists. The one among them who has shown the best understanding of classical systematics, Marcel Florkin, indeed wrote "La seule route féconde que puisse suivre la biochimie comparée dans son domaine taxonomique, est celle qui trouve à chaque pas ses fondations dans l'acquis actuel de la systématique animale", but in his book (1966) the macromolecular data are only briefly considered.

Dayhoff (1968, 1970), Dickerson (1972), Boulter (1972) and others have adapted the methods of the "numerical taxonomists" of the Sokal-Sneath school to deal with comparisons of protein sequences. No doubt there has been a feeling that only the newest technological procedures would be fitted
to deal with information of such modern and sophisticated kinds—furthermore, the application of these methods demands no prior systematic knowledge of the organisms concerned. I have published elsewhere (Crowson, 1970) a critique of the theoretical claims of the Sokal-Sneath school, as applied to the ordinary comparative data of systematics, and this critique should have almost equal force in relation to the ways in which biochemists have tried to apply such methods to protein sequence data. I see no fundamental difference between the problems of interpreting comparisons of residues at homologous positions in the cytochrome c molecule and those of comparing appendages on homologous segments in Arthropoda, except perhaps in so far as the Dollo principle is more easily applicable in the latter case than the former.

The recent appearance of Dickerson’s useful review of the then (1971) available complete sequences of cytochrome c, together with Boulter and Ramshaw’s (1972) valuable account of the same protein in plants, provide the opportunity for a traditional systematist to review the information on this protein from his own point of view. It is my hope and belief that such an approach will in the end prove both more phylogenetically rewarding and more biochemically suggestive than the computer methods used by Dayhoff (1970). One incidental advantage of classical methods is that they, unlike the numerical procedures, can make use of information from incomplete or partially uncertain sequence determinations.

A Darwinian phylogenetic systematist would be rather shocked to read in the first issue of the “Journal of Molecular Evolution” that “it was mutation that composed the genetic message, while selection merely edited it”, or words to that effect, and to find this idea permeating most papers so far published in the journal. The journalistic metaphor is a highly misleading one; the content of the genetic message no more resides in the individual codons composing it than that of this article lies in the letters of the alphabet—the essential message lies in their arrangement and coordination. The arrangement and the coordination are to be attributed to selection rather than to mutation. In relation to codon substitutions, Clarke (1970) provided statistical evidence to support the conclusion that “When all such factors are taken into account, it seems likely that our estimate of the proportion of ‘neutral’ mutations will be nearer 0.1% than 10%”.

A traditional systematist would certainly begin by trying to characterize taxa on the basis of available cytochrome c sequences. Naturally, with only some 43 species represented by complete sequences, plus a few more incomplete or uncertain ones, to cover the entire range of the Eukaryota, he would not expect to be able to do this in any but a very limited and tentative way. The best represented taxa in the data are Mammalia and Angiospermae, each with some 13 complete sequences recorded, the sample in each case being a reasonably representative one. The Mammalia, however, can be com-