THE MOLECULAR EVOLUTION OF HIGHER PLANT CYTOCHROME *c*

D. BOULTER

Department of Botany, The University of Durham, Durham, UK

ABSTRACT

Higher plant cytochrome *c* phylogenetic trees were constructed using the ancestral amino acid sequence method and the 'flexible numerical' method. These methods have been evaluated and the assumptions used in them stated. The results obtained with 20 higher plants are discussed and compared with existing ideas on the phylogeny of the higher plants.

INTRODUCTION

Phylogenetic relationships can only be established with certainty from an adequate fossil record. As this does not exist for many groups of organisms, attempts have been made to use, instead, comparisons of anatomical and morphological characters of present-day organisms. Recently, in addition, chemical data have been used either by recording the presence or absence of various compounds in different organisms, or by comparing varying structural features of a common constituent.

It is now known, for example, that many different proteins are universally distributed in living organisms. In each case, the equivalent proteins are similar in structure and function, due to a presumed common ancestry of the organisms and of the genetic information involved in the specification of the protein; evidence for this will be given, during the course of this review, for cytochrome *c*. Since there are several thousand different proteins, some or all of which may be useful in establishing phylogenetic relationships, the material potentially available for evolutionary investigations is greatly increased by their consideration.

Figure 1 diagrammatically relates the different kinds of 'characters' involved in evolution. DNA, the genotype, by interaction with environmental factors, gives rise to a population of proteins, the kinds and structures of

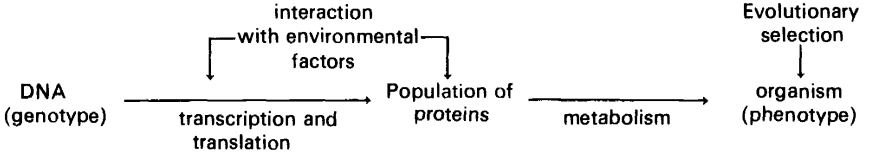

Figure 1. Interrelations between molecular and morphological characters.

which determine the metabolism, and so define the phenotype of the organism. It is the whole organism which is subject to the evolutionary process.

Cytochromes of the *c* type can be separated into at least three groups. Little is known about those which function in anaerobic energy-yielding reactions, and which occur in chemosynthetic bacteria[1], nor of those which function in the photo-reduction processes of photosynthetic organisms. Most of the sequence data have been obtained using cytochromes *c* of the third group, i.e. eukaryotic mitochondrial cytochrome *c*. Members of all three groups contain haem *c* as a prosthetic group, and are, therefore, probably closely related in their origins[2]. However, with the exceptions of *Pseudomonas fluorescens* c_{551}[3], *Desulphovibrio vulgaris* c_3[4], and *Rhodospirillum rubrum* c_2[5], no sequence determinations exist, except for the eukaryotic mitochondrial cytochromes. This last group are chemically and physically very similar, and function as electron carriers in the mitochondrial respiratory chain. This similarity in their properties is reflected in the considerable similarity of their sequence.

These similarities could have arisen during evolution, (1) because of a common ancestry of the specifying genes (homology); (2) by convergence, if similar residues in the sequences of unrelated proteins were necessary for a common function (analogy); or even, (3) by chance.

Whilst it is relatively easy to distinguish whether or not similarities between amino acid sequences are due to chance[6], there is no certain way of distinguishing between analogy and homology. To demonstrate homology, it is necessary to show that two sequences are more similar to each other than is required by a common biological function, and Fitch and co-workers[7,8] have used semi-rigorous methods to show that the cytochromes *c* of animals, plants and fungi, are homologous.

Dickerson[2] has noted the limitations of the above statistical approaches, and has suggested the use of x-ray diffraction data to distinguish between analogy and homology; thus, if the structure determinations show that the three-dimensional structure of two molecules is essentially the same over the entire molecule, this is much more likely to have arisen by homology, since the constraints required by function would not involve the whole molecule. For example, trypsin, chymotrypsin and elastase, all have serine residues at the catalytic site, and have similar tertiary structures[9]. The bacterial subtilisins also have a serine residue at the active centre; they have similar enzymic activity to the above enzymes, but possess little structural relation to them[10,11], thereby indicating convergence.

Dickerson *et al.*[12] have established the three-dimensional structure of horse-heart cytochrome *c* using x-ray diffraction data, and have shown that it is essentially the same as that of *Bonito*, a fish. Boulter and Ramshaw[13], using their projections and the plant sequences, suggest that the structure of the plant cytochrome *c* approximates to that of animals. Although their comparisons do not have the rigour of the full x-ray diffraction method, they suggested that the probable coincidence of the three-dimensional structure between plant and animal cytochromes *c*, is more likely to have arisen because of common ancestry, than by convergence. Thus, even though rigorous proof is lacking, most workers now accept that some of the similarity of the fifty or so cytochromes *c* examined so far, which have identical amino

acids in about one-third of the positions, is due to common ancestry. If this is accepted, since cytochrome *c* is an expression of a small part of the genetic material of the species from which it was obtained, it is possible that the relations between the sequences directly relate to the relationships between the species themselves. This, of course, depends on whether the small part of the genetic material involved is representative, for this purpose, of the whole of the genetic information. Relationships between species, species phylogenies, can only be constructed using sequence data from proteins which are both homologous and equivalent. Thus, originally identical genes may, by duplication followed by the separate evolution of the two or more genes, code for proteins (or polypeptides) with different functions. These proteins (or polypeptides) are homologous but non-equivalent; for example, the α and β chains of haemoglobin. Comparisons between non-equivalent proteins may be used to construct gene phylogenies, but they cannot be used directly to construct species phylogenies, since the time of divergence of duplicated genes is normally not the same as the time of divergence of the two organisms in which they occur.

ACCUMULATION OF SEQUENCE DATA

Although the largest number of sequences for any one protein has been obtained using cytochrome *c*, this protein is not absolutely ideal for these purposes. It is a coloured, stable molecule, and these properties are an aid during purification; also, its small size, as proteins go, makes for faster sequence determination. However, the amount of cytochrome *c* which can be extracted from some organisms, using feasible amounts of starting material, is not very large. Whilst it is possible to get of the order of 100 milligram amounts of pure cytochrome *c* per kilogram of starting material from certain vertebrate sources[14], it is not easy to extract it in sufficient quantities from many invertebrates and from most plants, even if in the latter case, one uses metabolically active, non-green tissues[15]. Looking at the organisms for which sequences have already been determined, one is led to the conclusion that the choice has been dictated by the availability of material as much as by its possible phylogenetic interest (see *Figure 2*). This has been true, to some small extent, in the case of the plant sequences determined in our laboratory. We have extended our range to encompass potentially phylogenetically more interesting species, by using micro-methods[16], but even these require of the order of 1–2 milligrams of pure cytochrome *c*. Fortunately, suitable tropical and sub-tropical sources exist in many important orders, so that it will be possible in the course of the next few years to establish the broad outlines of the higher plant cytochrome *c* phylogenetic tree. The availability of pure proteins from appropriate organisms is the main bottleneck to progress. This will become even more pronounced in the near future, since fast, accurate automatic sequencers have recently become commercially available.

A final drawback to the use of cytochrome *c* is that its rate of evolution is too slow for it to be used to establish the finer delimitations of the phylogenetic tree; compared with other proteins, it has accepted mutations more slowly than most[17].

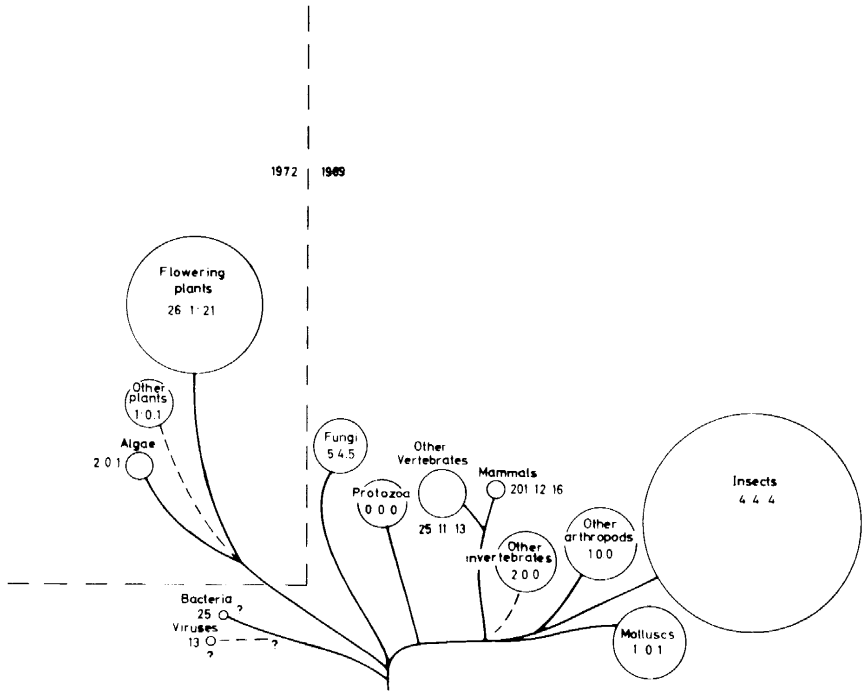

Figure 2. The number of protein amino-acid sequences determined, as modified from Dayhoff[17]. The size of the circles indicates size of the group, the three numbers inside the circles are: first, the total number of sequences on all proteins determined up to 1972 for plants, and to 1969 for other organisms; second, the number of cytochrome *c* sequences determined by 1969 and third the number of cytochrome *c* sequences determined by 1972.

Construction of trees

Computations based on quantitative comparisons of the amino acid sequences of homologous proteins, are used in the construction of phylogenetic trees. For n species there are $n(n-1)/2$ comparisons at each stage, and the number of possible trees which subsequently would have to be evaluated, is so large that the task is not feasible even using a computer. Two main approaches to this problem have been described. They are the numerical matrix method (for example, Fitch and Margoliash[18]), and the ancestral sequence method[19].

There are several different ways of constructing trees using the numerical matrix method, depending on the choice and calculation of the similarity or dissimilarity measure matrix, and on the choice of the sorting strategy used to generate trees from the matrix. In order to compare two sequences, they must first be aligned and then compared pair-wise over their lengths. The simplest relationship between them is the number of amino acid differences. Alternatively, the minimum mutation distance may be recorded; the latter is defined as a minimum number of nucleotide changes required if the gene for one protein is to code for the other. Since the genetic code is degenerate, there are normally several possible codon pairs for each amino

acid substitution, and it is the minimum number of nucleotide changes necessary which is recorded. Unlike other proposed similarity–dissimilarity measures, for example Gibbs and McIntyre[20] or Sackin[21], the distances along the branches of the phylogenetic trees constructed using these measures, have a numerical meaning related to the sequence differences. In addition to these direct measures relating sequences, it is possible to derive others, for example, by using weighting factors. If the aim were to obtain the best fit between sequence data and an existing established tree, such an approach would be very useful. When sequence data are being used to establish new phylogenetic insights, however, direct measures should be employed.

Similarity or dissimilarity matrices relating species may be used to construct phylogenetic trees in a number of different ways. However, Lance and Williams[22, 23] have shown that agglomerative strategies, the normal method used, can be generalized into a single strategy defined by variable coefficients. Phylogenetic trees are constructed using an hierarchical agglomerative strategy whereby the route by which groups are obtained, is optimized, rather than a 'clustering' strategy, where a property of a group is optimized. The tree is a graphical representation of the order in which groups are fused. Once the sequence is incorporated into a group, only the group is considered in future fusions, and there is, therefore, a great reduction in the number of trees constructed..

With 'space' distorted strategies, groups may move nearer some or all of the remaining elements, so that the chance that a sequence will be added to a pre-existing group is increased, giving a tendency to chain. This is called a space-contracting system. Alternatively, with space-dilating systems, the groups may recede from other elements on their formation[24, 25]. The flexible nature of the Lance and Williams[22] strategy lies in its complete range of space-distorting properties, which may be used with a set of data to detect optimum groups. We have also used the method of Fitch and Margoliash[18], but compared with that of Lance and Williams, this method may be expected to show space-contracting properties, which could lead to chaining; also it is not necessarily monotonic. For these reasons, we have used the flexible strategy of Lance and Williams as the major numerical method for constructing trees from our data.

Numerical methods make the assumption that evolution has occurred with the minimum number of changes. This assumption, that parallel and back mutations have not or have only rarely occurred, is wrong, and the degree to which the assumption affects the final tree will depend upon the method used in its construction and must clearly be considered when evaluating the method.

Further, since comparisons of present-day organisms are used, the resulting groupings generated will only relate to evolutionary groups if the rates of change have been reasonably constant along all the lines of descent (see Jardine *et al.*[26]). It has been suggested that this is probably so for the rate of change of mitochondrial cytochrome *c*[27], if periods in excess of 200 million years are considered. Thus, numerical methods may be of great value when relating distant groups, but results obtained with them have to be considered carefully with recently diverged groups of species, such as those within the higher plant kingdom.

Figure 3 gives a phylogenetic tree obtained using the flexible numerical method. The value chosen for the variable parameter beta was -0.1, since this value gives better results, e.g. *Ginkgo* as a single line of descent. We have, however, constructed a variety of trees using both amino acid differences and minimum mutation differences with the plant data, in which we have varied the beta parameter from -0.4 to $+0.3$. The trees are all essentially the same. Moving from space-conserving to space-dilating values of beta

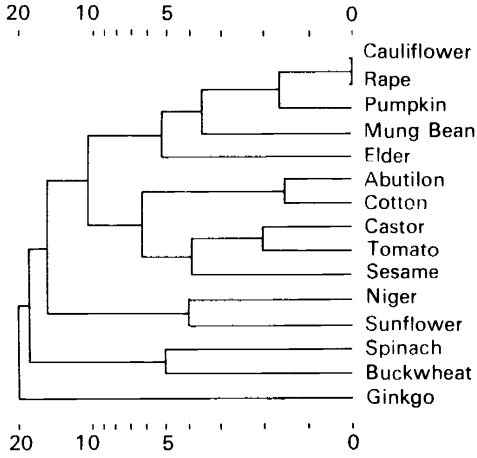

Figure 3. A phylogenetic tree relating the cytochrome *c* sequences of 15 species constructed using the flexible numerical method. $\beta = -0{\cdot}1$. Taken from Boulter *et al.*[28]. Reproduced by kind permission of the Royal Society of London.

causes minor rearrangements in the sesame, castor, tomato group. The results are essentially the same using either amino acid differences or minimum mutation difference, and again the differences lie in minor rearrangements in this group. A comparison was also made between this method and that of Fitch and Margoliash[18]. Once again, the same tree is arrived at, apart from small differences in the positions in the sesame, castor, tomato group.

An alternative approach to the numerical matrix methods for constructing phylogenetic trees, is the ancestral amino acid sequence method of Dayhoff and Eck[19]. A description of this method, as we have used it, has been given in Boulter *et al.*[28]. This method gives a closed topology and the earliest point of time must be established by numerical methods or from other biological considerations. Not all possible trees are constructed and evaluated, since, as stated earlier, this is not feasible and the method includes a step designed to limit the number of trees to be considered whilst still setting out to obtain the optimum solution.

The other main assumption of the method is the same as that of the numerical methods, namely, that evolution has taken place by the minimum route. In fact, both back and parallel mutations have occurred and using the higher plant cytochrome *c* data, Boulter *et al.*[28] have estimated their

extent as 27 per cent parallel mutations and 6 per cent back mutations. The question is—what is the extent of back and parallel mutations overall? That is, are these estimates with the existing data representative of the complete data set? Fitch and Margoliash[29] have calculated that 20 per cent parallel and 1 per cent back mutations have occurred in cytochrome *c* during animal evolution, i.e. their findings are of the same order as are ours with plants. We know that in the animal case the molecular tree corresponds, apart from minor details, with the fossil tree showing that this level of parallel and back mutations does not distort the molecular tree significantly, presumably because back and parallel mutations are of a relatively low frequency and are randomly distributed. Furthermore, it is unlikely that parallel, back mutations or both would be responsible for the calculation of a false ancestral sequence, since in each such computation every position in the sequence is considered in turn. Thus, one of the main weaknesses of attempting to establish a phylogeny using morphological characters is minimized with the ancestral sequence method. The inability of morphologists to detect many of the similarities due to convergence, stems from the fact that morphological characters cannot be sharply delineated one from another, and also because very few morphological characters are considered. A further weakness of using morphological characters is that these have evolved in different groups at different rates; this problem is also largely overcome by using the amino acid ancestral sequence method. Even if evolutionary rates are different in different lines of descent, two sequences will still be related to their true common ancestor. When a faster than average rate of evolution has occurred along a branch, there will be more changes between the present-day sequence and the ancestral one with the result that fewer positions can be used to link the sequence to the tree and it will appear more remote, but will still be linked to its own ancestor.

As already mentioned, methods of tree construction have steps to limit the number of trees, but, unfortunately, their effectiveness cannot be evaluated statistically, and the final justification of these methods must rest in the results which they produce. In the case of the phylogenies generated from animal cytochromes *c* data using the methods of Fitch and Margoliash[18] and Dayhoff and Eck[19], the existence of sufficient vertebrate fossils allows one to test the validity of the methods directly. In this instance, the molecular trees agree, apart from certain minor details, with that established from the fossil record, and the importance of this work lies not in the generation of new phylogenetic insights, but in establishing the effectiveness of the molecular method. In the case of higher plants, where an adequate fossil record does not exist, the molecular tree must be evaluated in relationship to all other available biological data. If the molecular tree disagrees with existing ideas on phylogeny based on other biological evidence, the decision will have to be made as to whether the disagreement justifies, (1) the search for additional evidence; (2) a re-interpretation of the existing biological evidence, or (3) disregarding the molecular data in this instance.

If the trees given in the numerical and ancestral methods are compared, only minor differences are found (see *Figures 3* and *4*). In the matrix tree, sunflower (*Helianthus annuus*) and niger (*guizotia abyssinica*) come before the *Abutilon theophrasti* and castor (*Ricinus communis*) group, whereas they come

after this group in the ancestral sequence method; the matrix tree also joins tomato (*Lycopersicum esculentum*) to the castor, sesame (*Sesamum indicum*) group. Since the rules for constructing the two trees are different, support is thereby given to the suggestion that the different assumptions used in their construction are valid. The number of amino acid substitutions is greater for the matrix tree (*Figure 3*) than for the ancestral sequence tree (*Figure 4*),

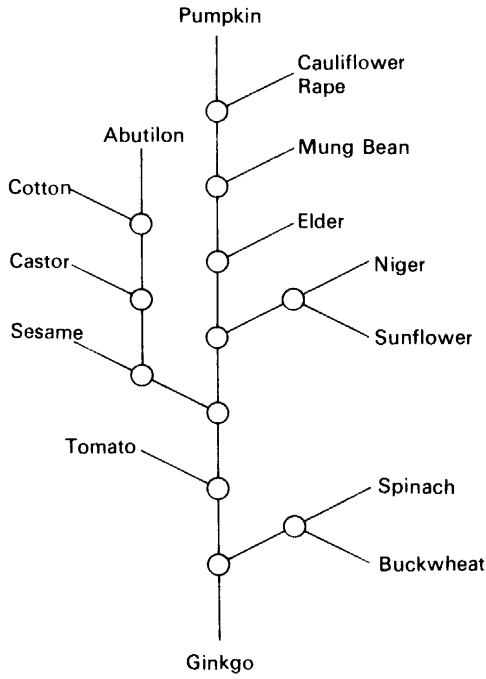

Figure 4. A phylogenetic tree relating the cytochrome *c* sequences of 15 species constructed using the ancestral sequence method. Modified from Boulter *et al*[28]. Reproduced by kind permission of the Royal Society of London.

when the former tree is evaluated using the amino acid sequence method. Thus, of the two methods, the ancestral sequence method is to be preferred; no doubt, this reflects the fact that fewer assumptions are used in its construction. When the species in *Figure 4* are arranged according to Takhtajan's scheme, the number of amino acid substitutions is greater than that required by the present ancestral tree arrangement.

PHYLOGENETIC IMPLICATIONS

Time of origin of flowering plants

When animal sequences from different classes are compared, it can be seen that whilst intraclass comparisons show a range of variations, interclass comparisons between a member of one class and any member of another are

relatively constant. These relationships have been interpreted to imply that elapsed time is an important parameter in determining the number of effective mutations accumulated by the cytochrome *c* structural gene along any line of evolutionary descent[30–33]. The factors which determine this empirically-derived linearity complex are, at this time, unclear; but it has been suggested[14, 34] that over long periods of evolutionary history, perhaps a minimum of two hundred million years for cytochromes, other factors relating to the rate of fixation of amino acid differences have either cancelled or averaged themselves out, leaving elapsed time as the variable most directly related to primary structure differences. It has been suggested by Kimura and others[35, 36] that this may be interpreted as support for a neutral mutation theory, but this has been contested by others and it is unnecessary for present purposes to assume any mechanism through which the interrelations, discussed above, have been achieved. One can simply accept this relationship which enables a direct correlation to be made between the time and rate of change of cytochrome *c*.

Unit evolutionary periods[37] have been determined using the single amino acid difference and minimum mutation distance measures, and these have been used with the cytochrome *c* data to examine the divergence times of the major taxonomic groupings. The values obtained for the times of divergence of the three major kingdoms are such that no specific order in their descent can be determined[38]. The results with the angiosperms show that by this method of computation they originated at least several geological periods before the Cretaceous, which is the earliest period in the geological record in which authentic angiosperm fossils have been found[38].

Flowering plant phylogenetic tree

Most existing phylogenetic schemes place the Ranales s.l. (Polycarpicae) as the basic group from which other flowering plant groups have arisen. Since the characteristics of the earliest angiosperms are not known, the evidence for this assumption is not strong (see for example, Takhtajan[39] and Cronquist[40]). In Bessey's[41] scheme most other groups are supposed to have descended from the Rosales, and this would accord with much phytochemical evidence, see for example, Meeuse[42]. Following Kubitzki[43], Meeuse[42] associated together a Rosalean–Saxifragalean–Cornalean–Cestralean lineage, with a Dillenialean–Thealean–Cystalean lineage, from which considerable diversification took place leading to the Amentiferae, Hammamelidales, Saxifragaceae, Sympetalae, etc. Bessey[41] derived the Rosales from the Ranales, although there is little evidence in support of this and the weight of phytochemical data suggest that the Ranales (Polycarpicae) are a separate offshoot possibly leading to the Rutaceae, Umbelliferae and Compositae. It is clear that cytochromes should be extracted and sequenced from plants belonging to these key areas, and much of the source material must come from the tropics or semi-tropics. To analyse the contradictions between the various existing schemes[40, 44], even the major ones, would be premature in the absence of this information.

However, it is generally accepted that the polypetalous and polycarpic conditions evolved prior to that of the sympetalous and syncarpic ones. The main arguments for this, as given in Eames[45], rest on the assumption that the

flower is a modified leaf-like organ (sporophyll), although Meeuse[42] and others have contested that this is not the case for all types of flower.

At first sight, therefore, it might appear that certain sections of the tree given in *Figure 5* are the reverse of the conventional view. Thus, the Polypetalae, including here, in terms of affinity, pumpkin (*Cucurbita maxima*), are at the top of the tree; similarly, tomato, sesame, castor, cotton (*Gossypium barbadense*), abutilon, are in reverse order. However, it is more likely that

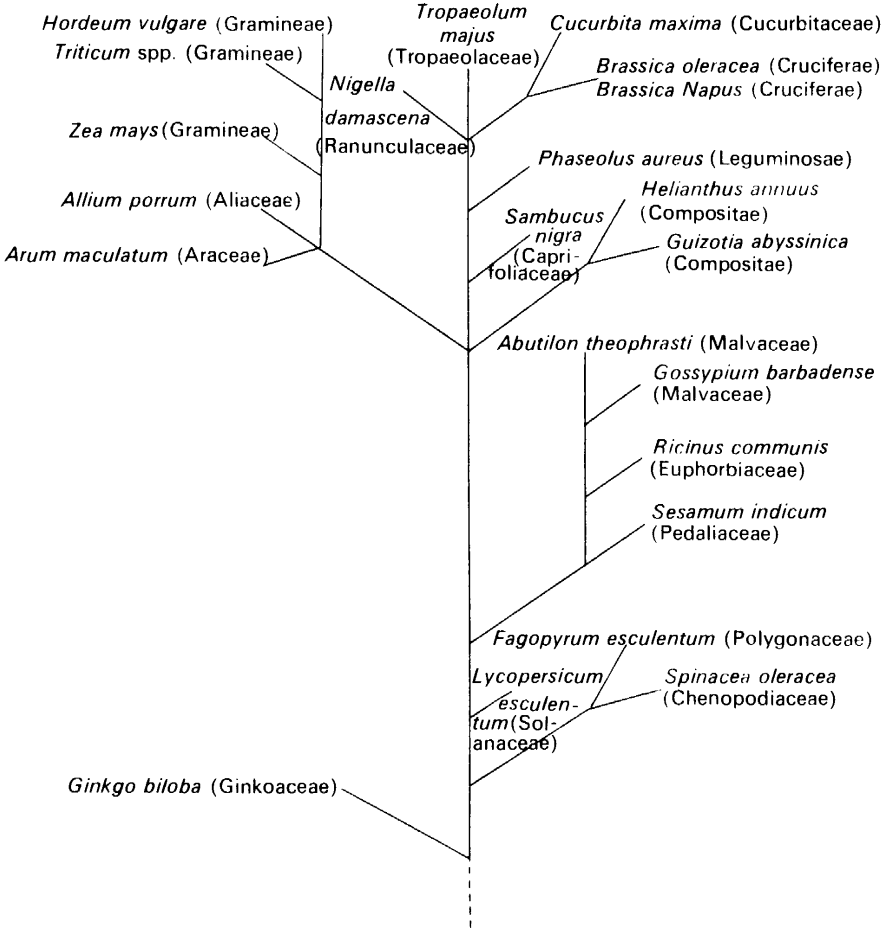

Figure 5. A phylogenetic tree relating the cytochrome *c* sequences of 21 species constructed using the ancestral sequence method. Part modified from Boulter *et al.*[28]. Several of the sequences are the unpublished data of Drs. J. A. M. Ramshaw, R. Scogin, R. H. Brown, and Mr. D. L. Richardson. Wheat data from Stevens *et al.*[52].

Abutilon = Abutilon; *Arum* = arum; *Hordeum* = barley; *Fagopyrum* = buckwheat; *Ricinus* = castor; *Brassica oleracea* = cauliflower; *Gossypium* = cotton; *Sambucus* = elder; *Ginkgo* = ginkgo; *Allium* = leek; *Zea* = maize; *Phaseolus* = mung bean; *Tropaeolum* = nasturtium; *Nigella* = nigella; *Guizotia* = niger; *Cucurbita* = pumpkin; *Brassica Napus* = rape; *Sesamum* = sesame; *Spinacea* = spinach; *Helianthus* = sunflower; *Lycopersicum* = tomato; *Triticum* = wheat.

this apparent contradiction is due to the fact that very few sequences have so far been examined, plus the fact that the trends towards reduction in numbers of parts and fusion of parts, may have happened several times in evolution. Following on from the phytochemical evidence, for example see Meeuse[42], it is possible that basal members of the Rosalean etc., Dillenialean–Thealean–Cystalean linkages, may be scattered along the nodal positions of the main axis (*Figure 5*), and further, that as more cytochrome sequences are determined, polypetalous species may be interpolated near the bases leading to sympetalous ones, e.g. between the node and sesame, and sympetalous species added to the ends of branches which at present contain polypetalous ones. The simplest present interpretation of the results given in *Figure 5*, although tentative, is that there was a basic flowering plant stock, which remained morphologically relatively constant during evolution, and from which various groups, some represented in *Figure 5*, have diverged at different times. Subsequent evolution of morphological characters in these groups has taken place at varying rates along the different lines of descent, so that present-day species have come to acquire, to a greater or lesser extent, those characters which have been used as an index of advancement[46, 47]. The twenty or so trends suggested by Bessey[41], Hutchinson[48], Cronquist[40] etc., upon which the major phylogenetic schemes have been constructed, have probably evolved several times. Also, the use of so few characters has led to the present confusion which exists between the different major phylogenetic schemes.

Examination of *Figure 5* shows that, of the angiosperms investigated, spinach (*Spinacea oleracea*) and buckwheat (*Fagopyrum esculentum*) diverged from the other angiosperms, at the earliest point in time. There is considerable other evidence that the groups of angiosperms to which spinach and buckwheat belong, have a rather isolated position, since they contain a number of so-called 'primitive' characters, e.g. the ultra-microscopic structure of the pollen grain apertures[49]. Furthermore, the Centrospermae, to which spinach belongs, and to which buckwheat (Polygonaceae) is clearly related, is the only group in which betalains are found rather than anthocyanidin pigments. Their seeds contain perisperm, and the protoplasmic inclusions of the sieve-tubes is of a type not found elsewhere in the angiosperms[50].

Another interesting feature of the results presented in *Figure 5*, is the possible early divergence and isolation of the line leading to the Compositae, e.g. niger and sunflower. These appear to have diverged earlier than elder (*Sambucus nigra*), and not to be derived from the Rosidian complex containing mung bean (*Phaseolus aureus*), as suggested by Cronquist[40], who also noted, however, that the morphological evidence for the derivation of the Asteridae from the Rosidae is no stronger than that suggesting a Dilleniidean origin. In our view, these groups, Asteridae, Rosidae etc., are too large and probably heterogeneous for phylogenetic statements such as the above to be particularly meaningful. The present results with tomato, sesame, castor, cotton, abutilon, support the suggestion by Thorne[51] of a Solanaceae, Euphorbiaceae, Malvaceae link. *Figure 5* also gives the relationship of the five monocotyledonous sequences determined to date. This topology has been grafted on to the dicotyledon tree using the ancestral sequence method. With so little sequence data available, these results must be considered very tentative.

They do, however, agree with existing ideas that the Monocotyledons are a natural group which arose from unspecified dicotyledon ancestors.

In the absence of suitable fossil evidence, and in view of the conflicting evidence from comparative morphology, confirmation of these preliminary cytochrome *c* results is best achieved by determining the sequences of another protein, and plastocyanin has been chosen for this purpose. Plastocyanin is a photo-synthetic protein of similar size to cytochrome *c*, and can be readily purified from the leaves of a wide range of plant species. So far, insufficient data have been established to tell whether the plastocyanin tree confirms the cytochrome *c* one. Plastocyanin is antigenic, and in view of the correlations established between serological crossreactivity and the sequences of albumin and lysozyme by Prager and Wilson[53], we are also collecting serological data as a possible quicker method for filling in areas of the tree, the broad outlines of which have been obtained from the sequence data.

CONCLUSIONS

Cain [54] has cautioned against the facile acceptance of the view that molecules are not under strong selective pressure, and the corollary that because of this, similarities in their structures directly reveal their evolutionary histories. It is certainly true, as pointed out by Boulter[55], that there is no direct evidence within natural populations that macromolecules are not under selective pressures. On the other hand, if we can accept the preliminary evidence quoted previously, that the three-dimensional structure of cytochrome *c* is preserved over the animal and plant kingdoms, in spite of considerable amino acid substitutions, it is likely that this variation is permissible without prejudice to function. Such evolutionary changes are called 'fortuitous' by Heslop-Harrison[56] and could give evidence of phylogenetic relationships.

This latter view does not exclude the possibility (see Woolhouse[57]), that some of the variation encountered may be environmentally conditioned, and, if so, this would tend to distort the establishment of correct phylogenetic relationships. It is important to stress that we are not really in a position to differentiate properly between these views at this stage, and the trees which have been presented here are phylogenies relating cytochrome *c* amino acid sequences and, at the moment, cannot be equated exactly with phylogenies relating the species from which the molecules were extracted.

ACKNOWLEDGEMENTS

I would like to thank Drs. J. A. M. Ramshaw, M. E. Boulter, A. D. J. Meeuse and Mr. B. L. Burtt, for helpful suggestions.

REFERENCES

[1] R. M. Klein and A. Cronquist, *Quart. Rev. Biol.* **42**, 105 (1967).
[2] R. E. Dickerson, *J. Molec. Biol.* **57**, 1 (1971).
[3] R. P. Ambler, *Biochem. J.* **89**, 341 (1963).
[4] R. P. Ambler, *Biochem. J.* **109**, 47P (1968).

[5] K. Dus, K. Sletten and M. D. Kamen, *J. Biol. Chem.* **243**, 5507 (1968).
[6] W. M. Fitch, *J. Molec. Biol.* **49**, 1 (1970).
[7] W. M. Fitch and E. Margoliash, *Biochem. Genet.* **1**, 65 (1967).
[8] W. M. Fitch and E. Markowitz, *Biochem. Genet.* **4**, 579 (1970).
[9] D. M. Shotton and B. S. Hartley, *Nature* **225**, 802 (1970).
[10] E. L. Smith, F. S. Markland, C. B. Kasper, R. J. Delange, M. Landon and W. H. Evans, *J. Biol. Chem.* **241**, 5974 (1966).
[11] C. S. Wright, R. A. Alden and J. Kraut, *Nature* **221**, 235 (1969).
[12] R. E. Dickerson, T. Takano, D. Eisenberg, O. B. Kallai, L. Samson, A. Cooper and E. Margoliash, *J. Biol. Chem.* **246**, 1511 (1971).
[13] D. Boulter and J. A. M. Ramshaw, *Phytochem.* **11**, 553 (1972).
[14] E. Margoliash and A. Schejter, *Adv. Prot. Chem.* **21**, 113 (1966).
[15] M. Richardson, D. L. Richardson, J. A. M. Ramshaw, E. W. Thompson and D. Boulter, *J. Biochem. (Tokyo)* **69**, 811 (1971).
[16] C. J. Bruton and B. S. Hartley, *J. Molec. Biol.* **52**, 165 (1970).
[17] M. O. Dayhoff, *Atlas of Protein Sequence and Structure* **4**. National Biomedical Research Foundation, Silver Springs, Md. (1969).
[18] W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).
[19] M. O. Dayhoff and R. V. Eck, *Atlas of Protein Sequence and Structure* **2**. National Biomedical Research Foundation, Silver Springs, Md. (1966).
[20] A. J. Gibbs and G. A. McIntyre, *Eur. J. Biochem.* **16**, 1 (1970).
[21] M. J. Sackin, *Biochem. Genet.* **5**, 287 (1971).
[22] G. N. Lance and W. T. Williams, *Nature* **212**, 218 (1966).
[23] G. N. Lance and W. T. Williams, *Comp. J.* **9**, 373 (1967).
[24] W. T. Williams, J. M. Lambert and G. N. Lance, *J. Ecol.* **54**, 427 (1966).
[25] L. Watson, W. T. Williams and G. N. Lance, *J. Linn. Soc.* **59**, 491 (1966).
[26] N. Jardine, C. J. van Rijsbergen and C. J. Jardine, *Nature* **224**, 185 (1969).
[27] C. Nolan and E. Margoliash, *Ann. Rev. Biochem.* **37**, 727 (1968).
[28] D. Boulter, J. A. M. Ramshaw, E. W. Thompson, M. Richardson and R. H. Brown, *Proc. Roy. Soc.* B, **181**, 441 (1972).
[29] W. M. Fitch and E. Margoliash, *Brookhaven Symp. Biol.* **21**, 217 (1969).
[30] W. Feller, *Probability Theory and its Applications*, Vol. 1, p. 71. Wiley, New York (1950).
[31] G. G. Simpson. Personal communication cited by Smith and Margoliash (1964) (1964).
[32] E. Zuckerkandl and L. Pauling, *Horizons in Biochemistry* (M. Kasha and B. Pullman, eds.), p. 189. Academic Press, New York (1962).
[33] E. Margoliash, *Proc. U.S. Natl. Acad. Sci.* **50**, 672 (1963).
[34] E. Margoliash and E. L. Smith, *Evolving Genes and Proteins* (V. Bryson and H. J. Vogel, eds), p. 221. Academic Press, New York (1965).
[35] M. Kimura, *Nature* **217**, 624 (1968).
[36] J. L. King and T. H. Jukes, *Science* **164**, 788 (1969).
[37] E. Margoliash and W. M. Fitch, *Ann. N.Y. Acad. Sci.* **151**, 359 (1968).
[38] J. A. M. Ramshaw, D. L. Richardson, B. T. Meatyard, R. H. Brown, M. Richardson, E. W. Thompson and D. Boulter, *New Phytol.* **71**, 773 (1972).
[39] A. Takhtajan, *Flowering Plants—origin and dispersal*. Oliver and Boyd, Edinburgh (1969).
[40] A. Cronquist, *The Evolution and Classification of Flowering Plants*. Nelson, London and Edinburgh (1968).
[41] C. A. Bessey, *Ann. Mo. bot. Gdn.* **2**, 109 (1915).
[42] A. D. J. Meeuse, *Acta Bot. Neerl.* **19**, 61 and 133 (1970).
[43] K. Kubitzki, *Taxon* **18**, 360 (1969).
[44] G. H. M. Lawrence, *Taxonomy of Vascular Plants*. MacMillan, Oxford (1951).
[45] A. J. Eames, *Morphology of the Angiosperms*. New York, Toronto, London (1961).
[46] K. R. Sporne, *New Phytol.* **68**, 555 (1969).
[47] K. R. Sporne, *New Phytol.* **69**, 1161 (1970).
[48] J. Hutchinson, *The Families of Flowering Plants*, 2nd ed. Oxford University Press, Oxford (1959).
[49] F. Roland, *Rev. gen. Bot.* **78**, 329 (1971).
[50] H. D. Behnke and B. L. Turner, *Taxon* **20**, 731 (1971).
[51] R. F. Thorne, *Aliso* **6**, 57 (1968).
[52] F. C. Stevens, A. N. Glazer and E. L. Smith, *J. Biol. Chem.* **242**, 2764 (1967).

[53] E. M. Prager and A. C. Wilson, *J. Biol. Chem.* **246**, 5978 (1971).
[54] A. J. Cain, *Chemotaxonomy and Serotaxonomy* (J. G. Hawkes, ed.), p. 229. Academic Press, London and New York (1968).
[55] D. Boulter, *Progress in Phytochemistry*, (L. Reinhold and Y. Liwschitz, eds.) p. 199. Interscience, London (1972).
[56] J. Heslop-Harrison, *Chemotaxonomy and Serotaxonomy* (J. G. Hawkes, ed.), p. 279. Academic Press, London and New York (1968).
[57] H. W. Woolhouse, *Phytochemical Phylogeny* (J. B. Harborne, ed), p. 207. Academic Press, London and New York (1970).