Phytochemistry, Vol. 23, No. 5, pp. 945–950, 1984.
Printed in Great Britain.

0031–9422/84 $3.00 + 0.00

REVIEW ARTICLE NUMBER 1

AGE AND DISTRIBUTION OF GALLOYL ESTERS, IRIDOIDS AND CERTAIN OTHER REPELLENTS IN PLANTS

E. C. BATE-SMITH

Agricultural Research Council, Institute of Animal Physiology, Babraham, Cambridge CB2 4AT, U.K.

(*Received* 7 *October* 1983)

Key Word Index—Ellagitannins; benzylisoquinolines; polyacetylenes; glucosinolates; palynology; palaeobotany; repellents.

Abstract—The association of particular constituents with particular groups of plants, and attempts made to fit them into systems of classification are reviewed. Muller's data on the earliest plant records are used to estimate the time of appearance of different repellent constituents, with special reference to galloyl esters (ellagitannins) and iridoids. The evidence is incorporated in a cyclic diagram featuring geological age, chemistry and taxonomy.

INTRODUCTION

In 1964 I was asked by Dr. Max Walters to give a paper to his Biosystematics Seminar on Plant Taxonomy and Chemistry. Having done so, I recorded it in a memorandum for limited circulation [1]. One of the main features was an account of the distribution of ellagic acid, based on the data in my published paper [2], and expressed in the form of a dendrogram in which the primitive dicot stock branches into a Rosalian stock and a Ranalian stock, the former branching into Rosales etc. (families containing ellagic acid) and Rosalian affinities (not Ranalian but not containing ellagic acid). These branches were merely represented by rather crude shapes, no attempt being made to give systematic definition to them. To emphasise the speculative nature of this exercise the dendrogram was christened 'Ellagidendron dicotyledonis', and as such will be frequently referred to later in this paper.

Several other chemical characters were analysed, and the possibility was considered, as an exercise in biosystematics, of a taxonomic system being based on chemical characters paralleling, but not competing with, current systems based on morphological characters.

The discussion that followed elicited the questions that botanists would wish to have answered when confronted with such an objective. Is there an effect of habitat on the chemical constituents? My answer was that there is no evidence (so far as I was aware) on their qualitative distribution, but growing conditions can have a marked effect on their quantitative distribution. An equally important question was that of the variation between different parts of the plant. There are obvious differences between petals, for instance, and leaves. It is the potential of the plant as a whole which is important. However great the differences between different tissues, the constituents are always consistent with the chemical constitution of the plant as a whole, which, by and large, is in turn consistent with its systematic position. Thus, while most of the sub family Papilionoideae of the Leguminosae have no proanthocyanidins in their leaves, they are consistently present in the testas of the seeds. It is obviously desirable to base systematic surveys on like tissues, examined at like stages of development, derived from healthy plants growing under optimal physiological conditions.

Do the constituents we have been considering play a recognisable part in the life of the plant? If they do not—if they are merely waste or excretory products—it might be mere chance if they were distributed in a systematically significant manner. Most people would now agree that products of metabolism are more likely, rather than less likely, to be usefully concerned in the life processes of a plant, but the ones that can be used for comparative purposes must be relatively stable, preferably end products of a metabolic sequence, otherwise they would come and go in an unpredictable manner. Thus, the hydroxycinnamic acids are likely to be used as building stones for lignin in plant cell-wall, but such use is not continuous and at any time the reservoir of unused building material will be recorded as a regular constituent of the tissue of that particular plant. It is then a slow rate of turnover that is responsible for the constituent's presence. In other cases, there may be no destructive mechanism for, or metabolic use made, of the substance and it does, in fact, accumulate as an end-product; but its presence may serve useful ends, such as protection or repellence, and it would be wrong to regard it as no more than a waste or excretory product.

The question was touched upon whether it is the constituents themselves or the processes by which they are formed that are systematically significant. There is really no distinction between these two alternatives because the presence of a substance presupposes a process for its formation.

Finally, there was a good deal of interest expressed in the apparent 'explosion' of so much new chemical virtuosity at the time of the emergence of the angiosperms. It is this question which will occupy most of the remainder of this paper.

UP-TO-DATE COMMENTS

In the time which has elapsed since 1964, the increase in chemical information has been enormous. This is particularly true of the ellagitannins, now united with gallotan-

Table 1. Distribution of galloyl esters in some Rosalian families and genera*

	Galloyl ester types			
Plants	Depsides	Ellagitannin	Geraniin	Polygalloylglucose
Cercidiphyllaceae	−	−	+	+
Hamamelidaceae	+	+	+	+
Fagaceae	−	+	−	+
Quercus infectoria	−	+	+	+
Paeoniaceae	+	−	−	+
Theaceae	+	+	−	+
Ericaceae	+	−	−	+
Arbutus	−	−	+	+
Rosaceae	−	+	−	+
Saxifragaceae	−	+	−	+
Myrtaceae	−	+	−	+
Punicaeae	−	−	+	+
Combretaceae	−	+	−	+
Terminalia fruit	−	−	+	+
Nyssaceae (*Davidia*)	−	−	+	+
Cornaceae	−	+	−	+
Elaeagnaceae	−	+	−	
Aceraceae (9 species)	−	−	+	+
Aceraceae (3 species)	+	−	−	−
Simaroubaceae	−	−	+	+
Anacardiaceae	+	−	−	+
Geraniaceae	−	−	+	+
Juglandaceae	−	+	−	+

*Condensed from Haslam [3]. Depsides are esters of gallic acid with substances other than glucose; ellagitannins include those with additional galloyl groups and disregard stereoisomerides; geraniins include punicacin, which is an isomer of geraniin; polygalloylglucose includes esters other than those which contain *m*-*O*-galloyl substituents (e.g. tannic acid, which is included as a depside).

nins as a class of galloyl esters (see Table 1) [3], but also true of flavonoids and terpenoids generally. There have also been numerous contributions in plant chemistry and systematics, notably a symposium organized by the Nobel Foundation in 1973 [4] and an outstanding contribution on 'Pflanzenstoffe und Pflanzensystematik' by Hegnauer in 1971 [5]. Okuda [6] has also contributed importantly to the galloyl compounds especially by his discovery in 1980 of geraniin and mallotusinic acid in Geraniales. These are derivatives of ellagitannin by oxidation of one benzene ring in one of the ellagic acid residues, and Haslam [3] has now begun a survey of their occurrence, and that of the different forms of gallo- and ellagitannins in a wide range of plant families (Table 1).

An even more revealing contribution has been made by work on the 'Aucubinartige Glycoside', now known as iridoids. At first, their presence was recognised only by their forming a blue or green colour when heated with acid, resulting in a blackening of the leaves of plants as they died. Chemical examination has now revealed the presence of many more related substances which do not, however, become coloured when heated. Typically they have nine or ten carbon atoms in a bicyclic form, one a heterocyclic oxygen ring which may carry a carboxylic acid group. But many of these derivatives are ring-opened, and are known as seco-iridoids; others with the inclusion of a nitrogen atom, are alkaloidal. As reproduced in Table 2, taken from Dahlgren *et al.* [7], their distribution, individually and collectively, is highly significant systematically, being especially prevalent in many of the sympetalous orders [8]. This has been extended by Kaplan and Gottlieb [9], who have devised a scheme of biosynthesis starting from a simple saturated carbocyclic system developing by single step dehydrogenation and hydroxylation processes to the most oxygenated carbocyclic type and by ring-opening to the seco-iridoid type. Stages in this biosynthetic sequence are related to systematic distribution, the most highly oxidized being regarded as the most advanced. An exception is the Compositae, in which typical, sometimes unique, constituents are sesquiterpene lactones, polyacetylenes, pyrethroids, senecio alkaloids, etc. [10]. Polyacetylenes also occur in the Umbelliferae, and on this and other grounds these two families have been regarded as related [11].

Another class of alkaloids, the benzylisoquinolines, is typical of many families in the Ranales (sensu Engler) or Magnoliales (sensu sundry modern authors). These orders are also distinguished by the absence of galloyl esters and, except for Illiciaceae, any trihydroxyflavonoids. Benzylisoquinoline alkaloids are also absent from this family, but they are present in several other families in the Choropetalae, and also in the monocot family Liliaceae.

Some kind of form can now be given to the 'Ellagidendron' tree mentioned earlier. Besides the Ranalian branch there will be a large branch corresponding to the iridoids (mainly comprising the Sympetalae) and numerous other branches representing such families as the Malvales (which have cyclopropenoid acids as characteristic constituents) and other families or orders with certain specific products. A representation of this situation which seems to me particularly appropriate is one used by Philipson [12] in his analysis of the systematic distribution of ovular characters in the dicots. He found a remarkable coincidence between the tenuinecellate unitegmic ovule and the sympetalous groups other than the Ebenales and Primulales. As also in the case of iridoids, this feature appears sporadically in the choripetalous families. His Figure 1 might well be used as a model for the elaboration of 'Ellagidendron', but the branches would not be in terms of taxonomic groups, but of chemical constituents.

Table 2. Distribution of iridoids in dicotyledonous families*

	Iridoid type									
Plant family	1	2	3	4	5	6	7	8	9	10
Meliaceae						+				
Daphniphyllaceae	+									
Altingiaceae	+									
Fouquieriaceae	+									+
Sarraceniaceae						+				
Actinidiaceae										+
Pyrolaceae	+									
Ericaceae	+			+						
Monotropaceae	+									
Epacridaceae										+
Roridulaceae										+
Eucommiaceae	+	+								
Symplocaceae			+							
Icacinaceae	+					+			+	+
Montiniaceae										+
Escalloniaceae	+									
Stylidiaceae	+					+				
Alangiaceae									+	
Nyssaceae									+	
Garryaceae	+									
Cornaceae s. str.	+		+							
Davidiaceae						+				+
Mastixiaceae										+
Curtisiaceae						+				
Griselinaceae			+							
Torricelliaceae			+							
Aucubaceae	+									
Aralidiaceae			+							
Hydrangeaceae	+			+	+	+				+
Adoxaceae						+				+
Sambucaceae						+				
Loasaceae				+		+	+			+
Viburnaceae			+		+	+	+			+
Valerianaceae					+	+				+
Caprifoliaceae						+				
Calyceraceae						+				
Dipsacaceae						+	+			+
Goodeniaceae						+				+
Oleaceae				+				+	+	+
Menyanthaceae						+				+
Loganiaceae						+			+	+
Rubiaceae	+	+				+			+	+
Gentianaceae						+	+			+
Apocynaceae	+					+	+		+	+
Hippuridaceae	+									
Verbenaceae	+	+	+	+						+
Callitrichaceae	+									
Lamiaceae	+	+								+
Bignoniaceae	+									+
Myoporaceae	+									
Buddlejaceae	+									
Stilbaceae				+						
Retziaceae				+						+
Selaginaceae		+								+
Globulariaceae	+									
Plantaginaceae	+									
Pedaliaceae	+									
Martyniaceae	+									
Scrophulariaceae	+	+			+					+
Lentibulariaceae	+									
Acanthaceae										+

*Data from Dahlgren *et al.* [7]. Key: carboxylated iridoids—1, 10-hydroxylated compounds; 2, 8-β-oxy-8-methyl compounds; 3, cornin group; 4, 10-carboxylated and 10-decarboxylated iridoids; 5, *Valeriana* compounds. Seco-iridoids—6, simple seco-iridoids; 7, gentiopicroside group; 8, oleuropein group; 9, complex iridoid alkaloids. Primitive or otherwise unclassified alkaloids—10.

The phylogenetic tree form is, however, difficult to maintain. For further taxonomic use to be made of this chemical evidence it has to be grafted on to existing, or newly devised, systems. Dahlgren [13] has done just this, modifying as far as is at present possible the classification of families, orders and superorders to accommodate their chemical characteristics. He is emphatic that his classification is striving towards one which will provide for chemical factors as they become known, and is in no way to be regarded as final. This means accepting families for the most part as at present circumscribed, and the relationships between them as represented by orders (which may differ from those in other authors' systems). He arranges these in a two-dimensional diagram, which is, in fact, essentially centric, representing the plan of a phylogenetic tree as it might be seen from above. The result is successful in concentrating much of the chemical evidence in certain areas (especially that relating to the iridoids) but much is still widely scattered over areas showing distant relationships with each other. This may, of course, mean that the orders are not in fact, related and the chemistry they have in common is polyphyletic. One purpose of the present memorandum is to try and show that, at least in the case of ellagic acid, this is not true.

Help may come from the palaeobotanic history of the plant species which form the different groups. Though much is already known, and more is continually being discovered, especially from the pollen and fossil records, it can only be a matter of opinion as regards the situation before the Tertiary period. As regards pollen, the review by Muller [14] is the most substantial and detailed. He restricts his data to pollen recognisable as attributable to present *extant* genera, placing the first such pollen (that of

Chloranthus) in the early Cretaceous. (The dates of these pollen records are not, of course, necessarily those of the first appearance of the plants concerned.) As regards the fossil record, Doyle [15] and Hickey and Doyle [16] place the origin of the angiosperms "near the Jurassic-Cretaceous boundary", with periods of radiation in the Cretaceous. Typical angiospermous fossil flowers (not, however, recognisable as those of any extant genera) have been found in the Lower Cretaceous [17; Dahlgren, R., personal communication]. Muller (who gives 118 m years as the date of this boundary) has the record for *Chloranthus* pollen just a little later.

The centric diagram had been used earlier by Sporne [18] with orders radiating from the centre as indicated by his advancement index. The data provided by Muller can be used in a similar way, in conjunction with the chemical data. A possible arrangement is illustrated in Fig. 1.

The sectors represent families in which particular constituents (as named) are present. The size of the sector has no quantitative significance, but is determined by convenience in drafting. The order of the sectors is relevant only in so far as it brings the Rosalian and Sympetalous families together, with a small overlapping zone in which both ellagitannins and iridoids are present. The systematic names (when valid) are Englerian. The names 'Rosalian', 'Ranalian' etc. differ deliberately from accepted systematic names (e.g. Ranalean). They have been used repeatedly since 1964 (see for instance ref. [19]) to represent blocks of families of like chemical content. Kubitzki [20] uses the terms Rosiflorean and Magnolialean in similar contexts.

The palaeobotanical data are all taken from ref. [14]. The length of the radial line in each sector represents the earliest recorded identification of a member of an extant family in that sector, as follows:

Leguminosae: Caesalpinioideae
Malvalian: Bombacaceae
Ranalian: Annonaceae; the earliest record of what is now regarded as a member of the Ranales, Chloranthaceae (110 m years) is invalid because this family does not contain benzylisoquinoline alkaloids.
Rosalian: Gunneraceae (see also Table 3)
Sympetalous: Apocynaceae
Compositae and Umbelliferae sic.
Centrospermae: Amaranthaceae/Chenopodiaceae
Capparidalian: Capparaceae, Brassicaceae.

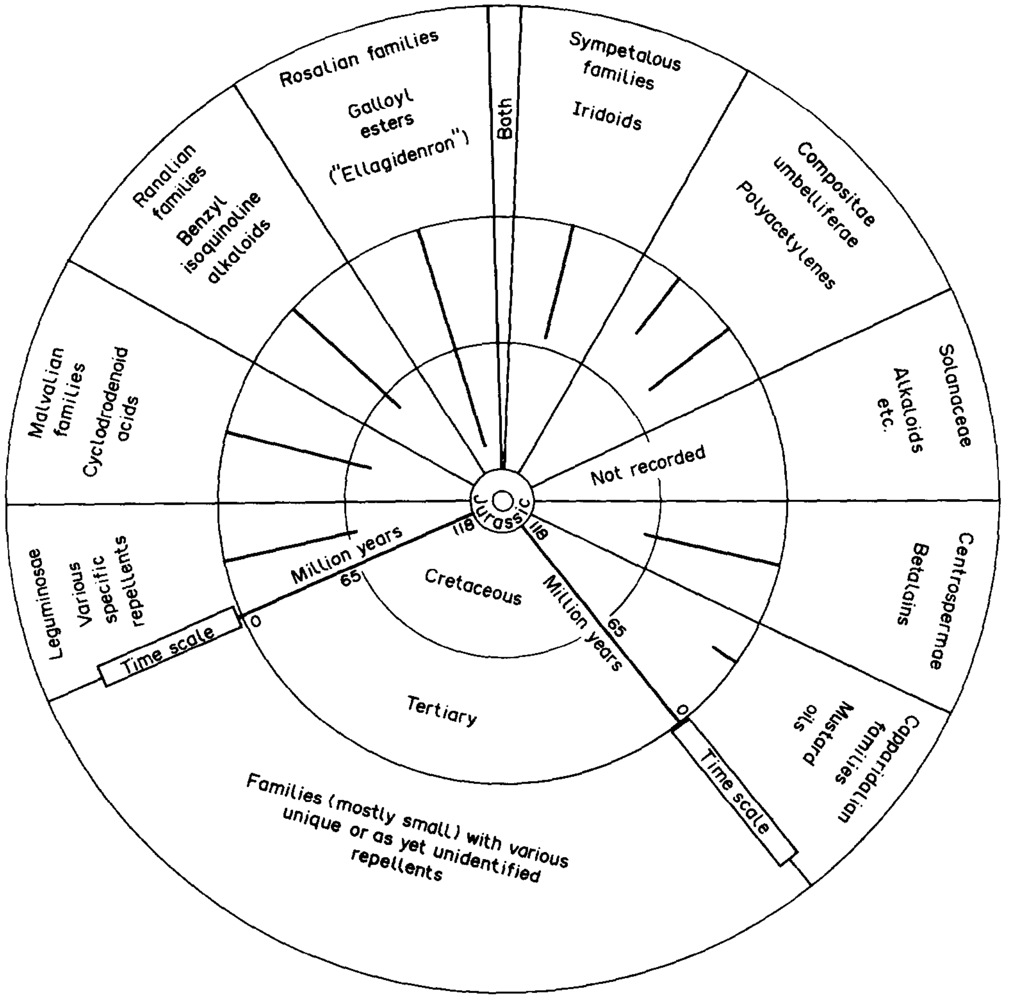

Fig. 1.

Table 3. 'Rosalian' families in Muller's record

Aceraceae
Anacardiaceae
Betulaceae
Caesalpiniaceae
Casuarinaceae
Cercidiphyllaceae
Cistaceae
Clusiaceae
Combretaceae
Coriariaceae
Cornaceae
Cunoniaceae
Dilleniaceae
Dipterocarpaceae
Droseraceae
Ebenaceae
Ericaceae
Euphorbiaceae
Fagaceae
Fouquieriaceae
Geraniaceae
Gunneraceae
Haloragaceae
Hamamelidaceae
Juglandaceae
Lecithydaceae
Loranthaceae
Lythraceae
Melastomataceae
Myricaceae
Myrtaceae
Nymphaeaceae
Nyssaceae
Onagraceae
Plumbaginaceae
Polygonaceae
Rhizophoraceae
Rosaceae
Santalaceae
Simaroubaceae
Theaceae

The Malvalian families are included because they afford such a good example of the restricted presence of a particular kind of constituent. The cyclopropenoid acids are present in all the families of the order malvales. In fact, it has just been reported [21] that they are present in the small Madagascan family Sarcolaenaceae, whose systematic position has been variously ascribed to the Malvales, Theales, Guttiferales, Terebinthales and Parietales. It can now be confidently included in the malvales. Sporadic occurrences are reported in other dicot families and, oddly enough, in *Gnetum gnemon* [22, 23]. It is not known whether these acids are, in fact, repellent.

OTHER FOSSIL EVIDENCE

Only very recently a fossil flower has been discovered with an estimated date of 78 million years, in Upper Cretaceous chalk-pits [24] which is closely similar in all morphological detail to an existing saxifragacaeous species, *Vahlia capensis*. Saxifragaceae is one of the families containing ellagic acid. The importance of this observation is that it relates to an identifiable flower and not merely to a type of pollen characteristic of a particular family. However, the date provided by Muller to the earliest occurrence of pollen of a recognisable extant genus is very much the same. Although neither Chloranthaceae nor the next recorded species, an *Ulmus* type, have ellagic acid (at the present time), extant species of *Gunnera*, Sapindaceae, Fagaceae and Betulaceae, estimated at 80–90 million years all contain it. The first of these is exceptionally confidently identified. It does appear that these families are very close to the origin of the Angiosperms, if Doyle's estimate is accepted. It is unfortunate that the systematic position of Chloranthaceae is at present uncertain, but that of Ulmaceae, whether correctly assigned to Urticales or not, is clearly of Rosalian affinity.

The evidence so far suggests, therefore, that the origin of ellagic acid was early in the evolution of the Angiosperms, but preceded by that of the monocots and possibly also by that of the Ranales, the stock being that which I have called Rosalian. This was in the early Cretaceous, and must have been a period of intensive radiation and diversification. Especially involved in this was Myrtales, in so many of the families of which ellagic acid is so richly present as to be almost characteristic of the order.

A similar approach could be made to the dating of the appearance of the other chemical groups, but at present these have nothing like the extensive coverage of ellagic acid. The most important of these are the iridoids, since they occupy such a large area in Dahlgren's classification. The data given in his paper with Jensen and Nielsen [7] can be applied to Muller's review, but only 23 of the 61 families examined by them appear in Muller. The earliest recorded occurrence is that of the cornin group in Symplocaceae in the Maestrichtian (latest Cretaceous, 65–69 million years), i.e. very much later than the earliest ellagic acid. Icacinaceae and Apocynaceae are recorded from the Palaeocene, but all the rest are more recent, most from the Miocene and Eocene. This would agree with the association of the iridoids, and especially of the secoiridoids and iridoid alkaloids with the Sympetalous families.

There is a mysterious feature in the co-occurrence of iridoids and ellagic acid in four different families—in Cornaceae, Symplocaceae, Ericaceae and Fouquieriaceae. It would be unlikely considering the difference in time of the origin of these constituents, that their formation in these taxa was simultaneous. It is notable that they are all woody families, Ericaceae and Symplocaceae sympetalous but now removed from the Sympetalae. *Fouquieria* (the only genus in Fouqueriaceae) has at different times been variously placed in Parietales, Tubiflorae, Ebenales and Tamaricales. Dahlgren now makes a separate order for it in the superorder Corniflorae, in which he also places Ericales and Cornales. Symplocaceae and of course Cornaceae are also placed in the latter. This brings all four of these anomalous families into the same superorder.

The only explanation seems to be that the iridoids in these families have a separate origin from those found in the Sympetalae as a whole, i.e. the iridoids are to this extent polyphyletic. It is unnecessary to assume that ellagic acid also is polyphyletic, but neither can it be assumed that it is monophyletic; this requires further exploration of its early history.

Another group of constituents to which the same analysis can be applied is that of the benzylisoquinoline alkaloids. A list of these is given in Hegnauer [5]. While most occurrences are in the Ranalian families, there are sporadic occurrences in Euphorbiaceae, Rutaceae, Rhamnaceae and Symplocaceae, and also in the monocots. The first named family is Annonaceae, in the Maestrichtian, but the Magnoliopsid pollen is of such common occurrence and coeval with that of Chloranthaceae that it is justifiable to conclude that their origin preceded that of the ellagitannin-containing families if only by a short head.

As regards other systematically restricted constituents, the restriction of cyclopropenoid fatty acids to the Malvales has already been mentioned (first record by Muller, in Bombacaceae, in the Maestrichtian); the association of betalains with Centrospermae is almost the prototype of chemical systematics (first record Amaranthaceae/Chenopodiaceae also Maestrichtian); and glucosinolates (precursors: of mustard oils) in various families, but mainly in Capparales (first record in the Miocene). The Compositae have several partially restricted constituents, some of which are shared with other families—polyacetylenes and sesquiterpene lactones with the Umbelliferae to such an extent as to suggest a close relationship of the two families. Both these constituents are confined to particular tribes in the Compositae, the polyacetylenes being absent from the Senecioneae, in many of which alkaloids are present [cf. 10]. First fossil record for Compositae is in the Oligocene, for the Umbelliferae in the Eocene.

In conclusion, it is not intended to attempt to relate the results to any particular system of classification, but they would be usable, in fact should be used, in any system that is erected. Thorne [25], who does use chemical characters to some extent in his new realignments in the angiosperms says that "the best classification we can construct with present information is a poor semblance of what it should be if the fossil record were more complete, or what it will be in botanical generations yet to come". Dahlgren [7, 8] develops his system with chemical characters as a virtual starting-point for modifications of more orthodox systems.

Acknowledgements—I am much indebted to Professor Dahlgren for information prior to publication, and for reprints of his own and other authors' papers. I owe the same obligation to Dr. K. R. Sporne, especially for fossil and pollen papers, which would otherwise not have been available to me; and to Professors R. Hegnauer and E. J. H. Corner, with whom I have been in close touch for many years.

REFERENCES

1. Bate-Smith, E. C. (1964) Low Temperature Research Station Record Memorandum, 460.
2. Bate-Smith, E. C. (1962) *J. Linn. Soc.* (*Bot*) **58**, 95.
3. Haslam, E. (1982) *Phytochemistry* **21**, 1049.
4. Bendz, G. and Santesson, J. (eds) (1973) *Chemistry in Botanical Classification.* The Nobel Foundation, Stockholm; Academic Press, London.
5. Hegnauer, R. (1971) *Naturwissenschaften* **58**, 685.
6. Okuda, T., Mori, K. and Hatano, T. (1980) *Phytochemistry* **19**, 547.
7. Dahlgren, R., Jensen, S. R. and Nielsen, R. J. (1981) in *Phytochemistry and Angiosperm Phylogeny* (Young, D. A. and Siegler, S., eds) p. 149. Praeger, New York.
8. Dahlgren, R. (1983) *Nordic J. Botan.* **3**, 119.
9. Kaplan, M. A. C. and Gottlieb, O. R. (1982) *Biochem. Syst. Ecol.* **10**, 329.
10. Heywood, V. H., Harborne, J. B. and Turner, B. L. (eds.) (1977) *Biology and Chemistry of the Compositae.* Academic Press, London.
11. Hegnauer, R. (1974) *Chemotaxonomie der Pflanzen* **3**, 544.
12. Philipson, W. R. (1974) *Bot. J. Linn. Soc.* **68**, 89.
13. Dahlgren, R. (1975) *Botan. Notiser* **128**, 119.
14. Muller, J. (1981) *Bot. Rev.* **47**, 1.
15. Doyle, J. (1978) *Ann Rev. Ecol. Syst.* **9**, 365.
16. Hickey, L. J. and Doyle, J. A. (1977) *Bot. Rev.* **43**, 3.
17. Dilcher, D. L. (1979) *Rev. Palaeobot. Palynol.* **27**, 291.
18. Sporne, K. R. (1974) *The Morphology of Angiosperms.* Hutchinson, London.
19. Bate-Smith, E. C. (1965) *Bull. Soc. Bot. France, Memoires*, 16.
20. Kubitzki, K. (1983) in press.
21. Gaydon, E. M. and Ramanoelina, A. R. P. (1983) *Phytochemistry* **22**, 1725.
22. Berry, S. K. (1980) *J. Sci. Food Agric.* **31**, 657.
23. Vickery, J. R. (1980) *J. Am. Oil Chem. Soc.* **59**, 87.
24. Friss, E. M. and Skarby, A. (1982) *Ann. Bot.* **50**, 569.
25. Thorne, R. (1981) in *Phytochemistry and Angiosperm Phylogeny* (Young, D. A. and Seigler, D. S., eds.). Praeger, New York.