AN OVERVIEW OF CHEMOTAXONOMY, AND ITS ROLE IN CREATING A PHYLOGENETIC CLASSIFICATION SYSTEM

COLEEN A MANNHEIMER

National Botanical Research Institute
Ministry of Agriculture, Water and Rural Development,
P/Bag13184, Windhoek.

INTRODUCTION

Chemotaxonomy is the systematic study of chemical variation between plant taxa. Evidence of chemical variation has essentially been used for classification purposes ever since 'folk taxonomies', based on certain obvious plant characteristics were instinctively employed by mankind centuries ago. These categories, such as edibility, taste, colour, smell and medicinal value were founded, however unknowingly, on chemical properties. As long ago as the first century after Christ the aromatic mints had been recognised and grouped together by Dioscorides (Jones & Luchsinger, 1986).

Awareness of the chemical complexity of plants grew from the desires of Europeans for exotic spices and condiments as well as investigations into their medicinal properties. Early knowledge about the subject was summarised in herbals, and concentrated on information about physiologically active secondary metabolites such as alkaloids and saponins (Jones in Street, 1978).

During the eighteenth and nineteenth centuries knowledge in the field increased, and some taxonomists made use of several chemical characteristics in attempts to classify plants and to demonstrate their phylogeny. However, although the chemical characters they used were recognised, they were manifestations of processes or compounds not yet completely identified (Jones & Luchsinger, 1986) and so their use was based on inadequate knowledge and evidence.

Gradually the number of recognised natural plant products increased, extending to include proteins, nucleic acids and the major polysaccharide categories. At the same time research into plant metabolism revealed similarities and uniformities in the chemical functioning of plants, while simultaneously highlighting biochemical peculiarities which might be taxonomically or phylogenetically significant. Successful attempts were made to correlate this variation with known classifications, and many claims were made as to the taxonomic merit of various chemical characters (Smith in Street, 1978). However it is only in recent decades that reasonably rapid surveys of plant extracts have become feasible, due to improved techniques of chemical analysis and the elucidation of the structures of many organic compounds (Radford *et al,* 1974). Technological advances, particularly electrophoresis and chromatography, have simplified and speeded up analyses, and also often made analysis of smaller amounts of material viable. This is particularly valuable when rare herbarium material must be used.

It is now generally accepted that certain compounds and related substances may be characteristic of certain taxonomic groups (Davis & Heywood, 1973), and certainly chemotaxonomic investigations have been employed at all levels of the taxonomic heirarchy, from subvariety to division (Smith in Street, 1978; Stace, 1980). It is thought that when the groups in question differentiated, the ability to form a chemical substance was retained by virtue of metabolic processes retained by the group or its ancestors. By implication we see that if the pathway of chemical evolution were established then this might offer insight to the evolutionary history of the group, as well as to the understanding of the present-day relationships within and between groups (Erdtmann, cited in Davis & Heywood, 1973).

Research has shown that there is *generally* an inverse relationship between the taxonomic distribution of a compound or class of compounds and its biogenetic complexity such that if it is biosynthetically simple and widespread it may be assumed to be primitive, while those more limited in their distribution and more complex to biosynthesise may be assumed more advanced (Gershenzon & Mabry, 1983). However these assumptions prove simplistic when phylogenetic inferences are to be made from chemical data. Davis and Heywood (1973) indicate the following general complications:-

1. Chemical compounds and/or biosynthetic pathways are not always environmentally stable, and so their presence may be affected quantitatively or qualitatively by factors such as temperature or mineral deficiencies.

2. Biosynthetic knowledge is inadequate about a great many of the chemical characters used in taxonomy, so that parallelism or convergence might confuse perception of phylogeny. A compound might arise *via* two or more entirely distinct biosynthetic pathways (convergence), and its presence would thus not necessarily imply a relationship between taxa in which it occurs. (Gershenzon and Mabry (1983) quote the case of napthoquinones, which may arise in higher plants *via* four different pathways). Parallelism may cause similar end products to arise in related groups, particularly when they are derived from ubiquitous metabolic intermediates.

3. In many cases it is not known whether metabolic pathways are reversible or not, and reversal of a pathway confuses the issue of whether a character is primitive or derived in a specific taxon.

4. Large and/or complex molecules are not necessarily the most advanced. Chemical reduction may occur in plants due to metabolic changes.

5. Chemical and morphological features may have evolved at different speeds due to differing selection

pressures, and thus one might find 'primitive' chemical characters together with 'advanced' morphology in a taxon.

6. Chemical variability occurs - for example, over a geographic range, or *via* chemical mutants within species. Chemistry may also vary between ontogenetic stages of organs analysed.

All these problems make it difficult to assess the relative degree of advancement of different compounds or classes of compounds in any given taxon. According to Davis and Heywood (1973) chemosystematics is of far greater use in establishing consanguinity (kinship) than in elucidation of phylogeny proper, which involves tracing past evolutionary history.

Despite drawbacks to potential phylogenetic value, there are many instances today in which phytochemical data have contributed to substantial taxonomic improvements (eg. documentation of hybridisation), although they are still underutilised to a large extent (Stace, 1980). In general, restricted or unique occurrence of compounds is of greater taxonomic value than widespread occurrence (Davis & Heywood, 1973). Broadly speaking, four main categories of characters are used in chemosystematics:-

DIRECTLY VISIBLE PARTICLES

These are particles such as starch grains and raphides, which are often also regarded as morphological characters. An example of their use is in the utilisation of starch grains to classify the Poaceae (Polhill and Raven, eds, 1981).

PRIMARY METABOLITES

Although some primary metabolites are utilised chemosystematically, on the whole their usefulness is limited, as they are intermediates in or products of essential metabolic pathways, and as such are of ubiquitous distribution (eg. sugars that participate in the Calvin cycle). Occasionally they accumulate in unusually high concentrations or occur as unusual storage products, and it is largely in such instances that they are used taxonomically (Stace, 1980).

SECONDARY METABOLITES

These are very numerous and of more restricted occurrence than primary metabolites, which makes them more valuable as a source of taxonomic evidence (Stace, 1980), and a large amount of information on their distribution in the angiosperms is available. However, the data demonstrate that some secondary compounds are more useful than others, particularly on a broad scale (Gershenzon & Mabry, 1983). The following are those groups of compounds that have thus far proven to be most useful generally:-

Flavonoids

Jones and Luchsinger (1987) indicate that the flavonoids are one of the most useful taxonomic markers for a variety of reasons:-

- they demonstrate a wide range of chemical structures, which have a demonstrable genetic basis for their variation.
- they are chemically stable, so that analysis of material can be done years after the material is collected.
- they are easily isolated and identified even from small amounts of plant material. Large numbers of plants can easily and rapidly be surveyed for flavonoids using paper, thin-layer, or one or two-dimensional chromatography, and in recent years many useful results have been achieved (Vickery & Vickery, 1981).
- they occur variously but ubiquitously in almost all plants.
- they can be used at all taxonomic levels in most groups of plants.

Although flavonoids are present in all higher plants (Vickery & Vickery, 1981), they are absent from bacteria and the majority of algae. The most primitive group that exhibits them is the Charophyceae (stoneworts), a group of green algae considered to be advanced for a number of reasons. Simple flavonoids have been found in primitive Bryophyta, while far more complex ones have been isolated from the most advanced angiosperms (eg.Orchidaceae).

Flavonoids have been used to revise the families included in the group Centrospermae. The red anthocyanin and betacyanin pigments are mutually exclusive, and so too are the yellow betaxanthins and the carotenoids. A survey of the Centrospermae showed that while ten of the twelve families included in the group contained betacyanins and/or betaxanthins, the Caryophyllaceae and Molluginaceae contained anthocyanins, in common with non-Centrospermae plants. This argued for their removal to a separate order from a chemosystematic viewpoint (Vickery & Vickery, 1981). However, this removal is still debated, as their exclusion is in direct contradiction to the anatomical evidence. Nevertheless, their anomaly was highlighted by their flavonoid pattern.

Flavonoids can be used as a 'fingerprint' in some genera where it has been found that each species has a distinct flavonoid pattern. The *Baptisia* genus (Fabaceae), for example, contains 62 flavonoids, and each of its 17 species has a characteristic flavonoid pattern. However, closely related genera exhibit several species with almost identical patterns (Vickery & Vickery, 1981).

Often the geographic centre of origin of a plant can be deduced from its chemical characteristics. *Geranium* species (Geraniaceae) exhibit a wide range of flavonoids, with primitive compounds predominating in Central Eurasian species, and diminishing in species to the east and west (Vickery & Vickery, 1981).

Parentage of natural hybrids may be deduced if flavonoid biosynthesis is assumed to be additive (Vickery & Vickery, 1981). By recombination, hybrids arising from species with differing flavonoids can synthesise compounds found in neither parent. Such chemical recombination may result in the hybrid exhibiting intermediate exomorphic characters (Davis & Heywood, 1973).

Terpenoids

Despite their comparatively wide occurrence and variety, terpenoids have been used less extensively than flavonoids in taxonomy, possibly due to difficulty of analysis. However they have been used extensively in the chemosystematics of some groups in which they occur (eg. mints, umbellifers). Comparison of terpenoid content between plants has been facilitated by gas chromatography, and they have been used to clarify specific and subspecific taxa as well as geographic races and hybrids (Jones & Luchsinger, 1987).

Alkaloids

A great deal of information on alkaloid occurrence is available, and their distribution has contributed to taxonomic studies in various groups. However, they are less chemically stable than flavonoids, as well as being structurally and biosynthetically complicated (Jones & Luchsinger, 1987). Due to the lack of knowledge as to the relative advancement of the different biosynthetic pathways leading to the various skeletal types, little attempt has been made to compare different alkaloid types phylogenetically (Gershenzon & Mabry, 1983). Their contribution to such classifications has thus been restricted. An exception among the alkaloids are the bensylisoquinoline alkaloids, which are important taxonomic markers in the angiosperms. Gershenzon and Mabry (1983) attribute this to their biosynthetic uniformity and 'coherent distribution'.

Glucosinolates (Mustard oil glucosides)

These, together with the alkaloids, have been used to divide the four families comprising the old order Rhoeadales into two new orders:
Capparales, containing the Capparaceae and Cruciferae, which produce glucosinolates.

Papaverales, containing the Papaveraceae and Fumariaceae, which produce alkaloids.

Research has shown that glucosinolate patterns may also be used to document hybridisation as well as provide infrageneric characters (Jones & Luchsinger, 1987).

Iridoids

This group of secondary metabolites is of increasing taxonomic importance (Gershenzon & Mabry, 1983; Jones & Luchsinger, 1987), and they show promise in the clarification of relationships at various levels. They have also contributed to solving the debate over the ancestral progenitor of the Asteraceae, which do not produce iridoids. Several putative progenitors' cases were weakened when they were shown to produce iridoids, leaving only the Campanulaceae, Araliaceae and Umbelliferaceae, which do not produce iridoids, in the running (Jones & Luchsinger, 1987).

In order to use secondary metabolites more extensively in a systematically meaningful way a great deal more knowledge about their biosynthetic pathways, including the metabolic intermediates and enzymes involved, will have to be established.

SEMANTIDES

These comprise DNA, RNA and proteins. Because each of these is so intimately connected with genetic characteristics, many researchers consider them to be of immense taxonomic value, as they potentially are. However, the time, equipment and know-how necessary for their effective analysis often limits their usefulness.

Stace (1980) mentions three main methods used in protein taxonomy - electrophoresis, amino-acid sequencing and systematic serology.

Electrophoretic techniques enable proteins to be 'fingerprinted' by establishing their relative size, charge and isoelectric point by separating them in variable gel mixtures across a voltage gradient (Stace, 1980). Protein profiles produced *via* electrophoretic separation and subsequent staining have been used in various systematic studies investigating polyploid taxa, as well as at interspecific, intraspecific and population levels. Particular care and expertise are required in the use and interpretation of protein profiles (Jones & Luchsinger, 1987).

Amino-acid sequencing attempts to establish the variation in the precise sequence of amino acids in a single homologous protein throughout a range of organisms. This analysis relies on the fact that a particular protein may vary to a certain extent without altering its essential function. One molecule used extensively for this purpose is cytochrome c, in which 79 out of the approximately 113 amino acids vary interspecifically, but alteration of even a single one of the other 34 destroys the functioning of the molecule. Generally the number of differences parallels the relational distance between the organisms in traditional classifications, but anomalies do present themselves, inferring that a measure of protein structure is not an infallible guide to degree of kinship.

Certain assumptions are made when analysing the results of this technique:-

It is assumed that:
- the molecule has evolved *via* the minimum number of mutations
- no convergent evolution or back-mutation has occurred
- different positions on the molecule are equally susceptible to substitution.

These assumptions weaken the evidence when the sequencing of a number of homologous proteins yields conflicting evidence (Stace, 1980). Some researchers have pointed out that to apply cytochrome patterns in chemosystematics requires that one should take into account the quantitative and qualitative effects of growth conditions on the cytochrome content (Kandler & Schleifer in Ellenberg *et al*, 1980). Stace (1980) suggests that in order to prevent interpretative mistakes, results from a wide range of proteins, preferably studied by a number of different techniques, should be pooled rather than placing total reliance on the screening of a single protein.

Synthetic serology is an immunological technique relying on the relative specificity of the immune reaction, and the

fact that the degree of cross-reactivity is proportional to the degree of relationship between the organisms.

In plant serology, antisera to antigens from various taxa are raised in animals, using various plant extracts (Jones & Luchsinger, 1987), and then the antisera can be used as a standard test against other plant extracts. The degree of coagulation that the other extracts cause in them is used as a measure of their similarity to the original antigen. Refinements in the technique have made this method more specific than it was previously, and serology has been extensively used throughout the taxonomic levels from above family to below species, yielding many valuable data (Stace, 1980).

Nucleic acids have not yet been used very extensively in plant systematics due to the complexity of their analysis. Most techniques are of relatively recent origin, and so the data accumulated thus far are limited (Jones & Luchsinger, 1987). Theoretically these characters should be able to solve many phylogenetic problems, firstly because each organism has DNA with a unique base sequence, and secondly because the theory of evolution is based on the premise that related organisms should show similarities in their DNA which are not shown by unrelated species (Vickery & Vickery, 1981).

The most useful technique in this regard at present is DNA hybridisation, in which DNA double helices are induced to unwind, and then allowed to recombine with each other as well as similarly treated DNA from other species. This results in some hybrid double helices being formed, the number and fidelity of recombinations theoretically depending on the compatibility of the two DNA base sequences. Some useful results have been obtained which show the potential value of this method, but techniques have not yet been perfected. Jones and Luchsinger (1987) point out that variable results have been obtained depending on experimental conditions. Some evidence has suggested that *in vitro* replication of the DNA template is affected by factors such as temperature, and the absence of regulatory phenomena or specific factors that are present *in vivo* (Wackernagel in Ellenberg *et al*, 1980), and it seems well possible that this might just as well apply to recombination also.

Other techniques have been used to investigate DNA and RNA, but results, according to Stace (1980), are of limited application. He suggests that advances in gene cloning and genetic engineering may lead to more extensive use of nucleic acid characters in taxonomy, but a potential drawback to their extensive use in phylogeny is that live material is often a prerequisite.

Chemical variation is of considerable taxonomic value in several ways:-

1. Confirmation or support of putative classifications derived from other sources of taxonomic characters, such as morphology.

2. Resolution of problems where relationships based on other evidence are ambiguous or conflicting.

3. Providing evidence to suggest more natural positioning of anomalous taxa, as well as to separate taxa. Often the presence of anomalous taxa in a group is accentuated by their chemical peculiarities.

4. Detection of confirmation of hybridisation.

5. Providing additional on/off characters for numerical taxonomy by their presence or absence in taxa.

However, as with all other taxonomic characters, chemical variation must constantly be subject to critical appraisal of techniques and interpretations. Two major problems that appear to need addressing are the lack of standardisation of methodology and the inadequate sampling of groups.

Chemotaxonomy has undoubtedly made a big contribution to taxonomic work in the past and will most certainly continue to do so in future. However, given the lack of fossil evidence and the need for live material in some analyses it seems that its contribution to a phylogenetic classification must perforce remain limited. The valuable information it offers is best used in conjunction with other sources of taxonomic evidence and thus a multidisciplinary approach is required in order to establish a system of classification which reflects natural relationships as accurately as possible.

REFERENCES

DAVIS, P.H. AND HEYWOOD, V.H., 1973. Principles of Angiosperm Taxonomy. Robert Kreigler Publ. Co. Inc., N.Y.

GERSHENZON, J. AND MABRY, T.J., 1983. Secondary metabolites and the higher classification of Angiosperms. *Nord. J. Bot.* 3: 5-34.

JONES, B.J.JNR. AND LUCHSINGER, A.E., 1987. Plant Systematics. Mc-Graw Hill Book Co. New York.

KANDLER, O. AND SCHLEIFER, K.H. Systematics of Bacteria. IN: ELLENBERG, H., ESSER, K., KUBITZKI, K., SCHNEPF, E. AND ZIEGLER, H. (eds), 1980. Progress in Botany. Springer- Verlag, Berlin.

POLHILL, R.M. AND RAVEN, P.H. (eds), 1981. Advances in Legume Systematics, Part 2. Royal Botanic Gardens, Kew.

RADFORD, A.E., DICKISON, W.C., MASSEY, J.R. AND BELL, C.R., 1974. Vascular Plant Systematics. Harper and Row, N.Y.

SMITH, P.M. Chemical Evidence in Plant Taxonomy. IN: Street, H.E., 1978. Essays in Plant Taxonomy. Academic Press, London.

STACE, C.A., 1980. Plant Taxonomy and Biosystematics. Edward Arnold (Publishers) Ltd., London.

VICKERY, M.L. AND VICKERY, B., 1981. Secondary Plant Metabolism. The MacMillan Press Ltd., London and Basingstoke.

WACKERNAGEL, W. REPLICATION. IN: ELLENBERG, H., ESSER, K., KUBITZKI, K., SCHNEPF, E. AND ZIEGLER, H. (eds), 1980. Progress in Botany. Springer-Verlag, Berlin.