

Chemotaxonomy

Chemotaxonomy of plants is an expanding field of study and seeks to utilize chemical information to improve upon the classification of plants. Chemical evidence has, in fact, been used ever since man first began to classify plants as edible and inedible, obviously based on their chemical differences.

Stages in Chemotaxonomy:

Hegnauer recognized six chief stages in a phytochemical study, reconciling the need for informed chemical procedure and a project design satisfying formal taxonomic requirements:

- (i) Choice of a group, taxonomic survey and sound sampling.
- (ii) Choice, mastery and modification of suitable chemical techniques in a pilot survey.
- (iii) Full analysis of all material.
- (iv) Interpretation and comparison with data from all other sources.
- (v) Adjustment of classification as necessary, and
- (vi) Treatment of any evolutionary relationship as indicated by the new data.

Significance of Chemotaxonomy:

The occurrence and distribution of the various types of chemical substances present in plants prove to be of taxonomic significance. However, it should be noted that, all kinds of chemical substances present in plants do not reveal information useful to the taxonomist. Phytochemical characters of taxonomic significance have been classified into three types.

These include: classes of Chemical and test character:

a. Primary constituents:

These include the macromolecular compounds directly taking part in metabolism and include proteins, nucleic acids, chlorophyll and polysaccharides. All chemical materials synthesized by an organism reflect the information in DNA, RNA and proteins. These latter molecules have been termed as semantides. Semantides, thus contain useful information of taxonomy and phylogeny.

b. Secondary constituents :

They include compounds lacking nitrogen and not involved directly in plant metabolism i.e., simple phenolic compounds like caffeic, benzoic and nicotinic acids and polyphenolic compounds like flavonoids, terpenes, coumarines, alkaloids and pigments of which flavonoids are most widely studied with respect to plant systematics.

c. Miscellaneous substances:

However, no suitable classification of the chemical characters and their use in taxonomy is developed so far. On the basis of their molecular weight, Jones and Luchsinger (1987) has divided the natural chemical plant products useful in taxonomy, into two major groups.

d. Micro-molecules :

They are low molecular weight compounds with a molecular weight of 1000 or less, e.g. amino acids, alkaloids, fatty acids, terpenoids, flavonoids, etc.

e. Macromolecules :

They include the high molecular weight compounds with a molecular weight of over 1,000, e.g. proteins, DNA, RNA, complex polysaccharides, etc.

Some taxonomically important chemical compounds, along with a few examples of their systematic value is given below:

(a) Proteins:

Among the various semantides, proteins serve as the most important tool in chemotaxonomy.

The importance of proteins in chemotaxonomy has several reasons behind it:

I. Firstly, these are large, complex molecules showing little qualitative variation with changing environmental factors.

II. Secondly, they are universally distributed.

III. Thirdly, they are relatively simple to extract and handle and present in appreciable amount.

IV. Lastly, numerous cheap, simple and rapid methods have developed in the recent past, for protein analysis and comparison by taxonomists, of which electrophoresis is particularly an important method of protein separation.

The variation in protein structure between plants effectively provides an inner view into the cell's genome, as the amino acid sequences of proteins are encoded by the nucleotide sequences of a cell's DNA. Proteins have been used in taxonomic investigations in the following ways:

Comparison of the protein banding patterns :

In recent times protein banding patterns obtained by gel electrophoresis, have been focused on the problem of identification of critical taxa, their relationship and taxonomic status. Taxonomic interpretations should be based on comparison of proteins from homologous organs of the same age in order to avoid any confusion due to variability of proteins.

Considerable variation in protein complements has been recorded at the level of species and genus, and even between the same plants in different populations. These form evidence upon which taxonomic systems may be founded, tested or demolished.

Electrophoresis is one of the most extensively used techniques in protein investigations. Because of the presence of ionizable molecules on the surface of proteins, they will migrate when subjected to an electrical field in a solution of suitable pH. Soluble proteins thus migrate within an electric gradient at a rate that depends on their net electric charge and on their molecular size and shape. The rate of migration of each protein is constant under identical conditions and hence can be used as a reliable character for the detection of homologous proteins. Separation of

proteins can be done either on paper (paper-electrophoresis) or on a gel medium (gel-electrophoresis). The commonly employed gel media include starch, polyacrylamide and cellulose acetate. It has been possible to obtain useful systematic information from electrophoretic analysis of crude biological protein samples.

For example:

a. Johnson and Hall (1965) have demonstrated the phylogenetic affinities in Trichinae by the process of protein electrophoresis.

b. This process also helped in establishing a close relationship between *Vicia* and *Lathyrus*.

c. Interspecific variations among eight species of *Cassia* (Caesalpiniaceae) were evaluated on the basis of seed protein and mitochondrial DNA RFLP by gel electrophoresis as well as pollen protein patterns.

d. Amino acid sequence studies of homologous proteins — Comparison of amino acid sequences of homologous proteins from different taxa has also provided a powerful tool for evolutionary and systematic studies. The degree of similarity in the amino acid sequence is presumably proportional to the degree of genetic relationship.

For example, based on amino acid sequence studies, Martin and his associates have tried to trace the phylogeny of various taxa like Malvaceae, Ranunculaceae, Magnoliaceae, Polygonaceae, Myrtales and some Solanaceae, and have elucidated the phylogeny and taxonomy of angiosperms.

e. Analysis of isoenzymes and alloenzymes — Electrophoresis of enzymes can reveal two distinct types of genetically controlled variation in enzyme phenotype.

f. Alloenzymes — They are different forms of a particular enzyme that are coded by a single gene locus but by more than one different allele, each coding for a slightly different amino acid sequence, which have difference in mobility during electrophoresis.

g. Isoenzymes — They are different forms of the same enzyme formed from genes at different loci. Nowadays, a large number of such different enzymes are being exploited electrophoretically for taxonomic purposes, particularly to reconstruct phylogenies either within or between species. Most of these include enzymes involved in fundamental cellular processes such as those involved in glycolysis (Hexokinase, phosphofructokinase, aldolase, etc.), Krebs cycle (isocitrate dehydrogenase, malate dehydrogenase, etc.), enzymes involved in protecting cells from free radicals (catalase, superoxide dismutase, etc.), etc. Comparative studies of enzymes may yield data about differences in primary structure, which can be traced to evolutionary adaptations. For example, Natarrella and Sink (1975) studied the peroxidases and proteins of four species of the genus *Petunia* (viz. *P. axillaris*, *P. inflata*, *P. violacea*, and *P. parodii*) and some 11 cultivars of *P. hybrida* Hort., by electrophoresis. Electrophoretic patterns revealed that all the species are closely related and that *P. inflata*, *P. axillaris* and *P. violacea*, are involved, in *P. hybrida*. Similarly based on peroxidase, esterase and acid phosphatase isoenzyme pattern, Symenoidis and Tsekos suggested that the genus *Taeniantherum*, formerly considered as a part of the genus *Horedum*, should be treated as an independent genus.

(b) Nucleic Acids:

The potentiality of the huge amount of phylogenetic information comprising the base sequences of cellular nucleic acids has been recognized relatively recently. Lack of the necessary expertise and equipment, required for the nucleic acid technique, to most of the taxonomists, hindered the progress in this direction. At present this field is expanding at such a prodigious rate that it is now within the reach of most laboratories to obtain informative sequence data from a wide range of plants, and recent advances in DNA technology have created a wealth of new opportunities for taxonomy. It has now become possible to sequence any form of DNA from any plant. However, this is still a new technology, and we are only beginning to understand its problems as well as its potential.

Nuclear DNA and RNA:

The relative homology of DNA or RNA of various plants is useful in a taxonomic study and as a possible screening method for inter-fertility of species. The nuclei of the cells of higher plants contain relatively enormous amount of DNA ranging between 1000-10,000 mega base pairs. However, some plants like *Aesculus hippocastanum*, *Arabidopsis*, etc. have smaller genomes of the size in the region of 100 mega base pairs, while some lilies have DNA with almost 100,000 mega base pairs in each nucleus. Ribosomes, which are essential for protein synthesis, are ubiquitous and virtually abundant in the vast majority of plant cells. Their constituent RNA and proteins are encoded by multiple gene copies, making them particularly amenable for sequencing. Although ribosomes show some variation between organisms, on the whole they are very much similar in both their gross structure and in sequences, even in the most disparate members. All ribosomes are composed of a larger and a smaller subunit, each in turn comprising a RNA molecule (rRNA), the size of which, vary between plants, and a number of protein molecules. It is the rRNA that has made a considerable impact on molecular taxonomy as they have domains with different average rates of nucleotide substitution. Thus, each ribosomal subunit contains some sequence information relevant to divergence in the distant past, as well as other more evolving sequences carrying information relevant to separations almost to the present. For example, Pollard noted the specificity of ribosomal RNA in cabbage, cauliflower, celery, corn and parsnip and showed that 28S and 18S ribosomal RNA from these taxa have distinct base compositions, which are characteristic of the species. Sequencing of rRNA presently is usually restricted to selected segments of one or other subunits. This is based on the fact that some regions evolved at such a low rate that their sequences would be identical or very similar across a group. The use of polymerase chain reaction, along with suitable primers, has greatly enhanced the usefulness of rRNA sequence data at all levels of phylogenetic analysis. Transfer RNA (tRNA), like ribosomes are ubiquitous and highly conserved and forms an integral part of the protein synthesis machinery, coding for the various amino acids. Unfortunately so far there have been few comparative studies of codon usage between higher taxa, but as the pattern of codon usage is a genetically controlled character, it might throw some light on phylogenetic at a range of levels.

Repetitive DNA :

The nuclear genome not only contains the largely single copy DNA that comprises the functional genes and their various regulatory units, etc., but it also typically contains between 25-80% nonsense sequences of varying length which are repeated many times, forming the so called satellite DNA, which are spread widely throughout the genome. It is also possible sometimes to isolate tandemly repeated sequences by restriction endonuclease digestion, for cleaving the DNA at the same relative position in each repeat, yielding a large number of copies of the repeat from each nuclear digest, which can be separated by gel electrophoresis. For example, studies on thirteen species of *Cucurbita* revealed that all the species possess mainband and satellite DNA whose densities are constant throughout the genus.

Mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA) :

Apart from the nuclear DNA, plants also have the circular double-stranded mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA). In higher plants (angiosperms), the total size of mtDNA ranges from 100-240 kilobase pairs and therefore considerably difficult to work with. The cpDNA on the other hand are generally between 1,35,000 and 1,60,000 base pairs long. The complete gene sequences for several cpDNAs are now known and partial sequences are available for many others. As compared to mtDNA, cpDNAs are relatively conservative and therefore of greater taxonomic significance at the level of species, genera and family. Both mtDNA and cpDNA are inherited through the maternal line. Thus, comparison of nuclear genome with cpDNA offers a potentially powerful tool for revealing past hybridization events or introgression. Further, restriction fragment analysis of cpDNA can help distinguish between potentially single or multiple origins of hybrid taxa. For example, presence of two quite different chloroplast genomes in the hybrid species *Aegilops triuncialis*, indicates that it has arisen through two hybridization events, one with *A. caudata* as the female parent and one with *A. umbellulata*.

The various techniques used for sequencing and characterizing nucleic acids are as follows:

Gel electrophoresis :

It is now possible to study DNA or RNA molecules in a routine manner by gel electrophoresis. This method is based on the principle that dissolved molecules in an electric field move at a speed determined by their charge-mass ratio, which allows the separation of molecules in a mixture according to size. Since the phosphate groups in the nucleic acids are ionized, they have a negative charge in solution and hence migrate towards a positive electrode. However, nucleic acid molecules consisting of long chains have almost identical charge-mass ratios, as each residue contributes about the same charge and mass, whatever may be the length of the chain. Therefore, little or no separation of molecules of varying lengths would occur if the electrophoresis of nucleic acids were simply carried out in solution. Hence, instead of using a liquid solution, nucleic acids are now separated by electrophoresis in an agarose gel, in which the rate at which molecules can move depends on the size of the pores. Nucleic acids with identical charge-mass ratios separate according to length, with the longer ones moving more slowly. Even very long nucleic acids, with chains containing 10,000-20,000 residues that differ by only a few

percentage points in their length can be separated. Nucleic acid mixtures containing chains of 500 nucleotides or less can be separated on polyacrylamide gels, in which each chain length can be resolved making DNA sequencing possible. The location of DNA bands within the gel can be determined directly by staining the gel with the intercalating dye ethidium bromide and very low concentrations of DNA i.e., as little as long can also be detected by direct examination of the gel in ultraviolet light. Thus this technique is simple, fast, and capable of resolving mixtures of DNA fragments that cannot be separated adequately by other sizing procedures.

The Polymerase Chain Reaction (PCR) :

This method was invented by Kary Mullis and it is an in vitro method for the enzymatic synthesis of specific DNA sequences using two oligonucleotide primers that are hybridized to opposite strands flanking the region of interest in the target DNA.

It involves a repetitive series of cycles (Fig. 8.27), each cycle involving the steps of template denaturation, primer annealing, and the extension of the annealed primers by DNA polymerase, which results in the exponential accumulation or multiplication of a specific fragment whose termini are defined by the 5' ends of the primers. The primer, extension products synthesized in one cycle in turn can serve as a template in the next, thus doubling the number of target DNA copies approximately at every cycle. Initially, the PCR technique used the Klenow fragment of *E. coli* DNA polymerase for extension of the annealed primers. But since this enzyme was inactivated by the high temperature required to separate the two DNA strands at the outset of each PCR cycle, the thermo-stable DNA polymerase (Taq polymerase) isolated from *Thermus aquaticus* was introduced, which transformed this technique into a simple and robust reaction which can now be automated by a thermal cycling device. The higher temperature optimum for the Taq polymerase allows the use of higher temperatures for primer annealing and extension, thereby increasing the overall stringency of the reaction. Various reaction parameters such as enzyme, primer, Mg^{2+} concentration, the temperature cycling, etc. are required for this process, optimal set of conditions varying for any given pair of oligonucleotide primers. PCR is thus an important technique when it comes to dealing with small samples. It is also of enormous value as it allows the amplification of particular DNA sequences without the need of cloning.

Random Amplification of Polymorphic DNA (RAPD):

This is a recently developed technique involving PCR, in which random primer sequences are employed to amplify sections of genomic DNA of unknown function. In this method oligonucleotides are used as primers, but unlike PCR which is used to amplify a specific gene sequence, only one primer is added to the reaction mixture and it is a matter of chance as to the number of short sequences, which are bordered at either end by an inverted pair of that particular primer sequence, so that they and the intervening sequence can be amplified. Generally and RAPD amplification with 9 or 10 primers may yield 1-20 or so distinct amplified DNA sequences, which can be separated electrophoretically. RAPD has the advantage over PCR in being simpler and cheaper. This technique allows determining relationships between closely

related species, assessing kinship relationships to the production of species-specific probes, to detect hybridization and gene flow between two samples.

Restriction Fragment Length Polymorphism (RFLP):

This technique involves the cutting of DNA molecules by specific enzymes called restriction endonucleases, which cut the DNA within their own particular recognition sequences, the length of which may vary between 4, 6 and 8 bases. For example the restriction enzyme BamHI cuts between the first two Gs of its recognition sequence GGATCC. The restriction fragments are then separated electrophoretically. The significance of restriction enzymes for phylogenetic research lies in the fact that in many plants the distribution of sites in the genome where a particular enzyme can cut is relatively fixed, generally showing variations between species or genera.

DNA hybridization:

This technique, which was developed by Bolton & McCarthy, is a relatively simple but effective method, for measuring the homology of the DNA and RNA in various groups of organisms and has been shown to be useful in taxonomic investigations. In this method, DNA extracted from one plant is treated to turn it into a single-stranded polynucleotide chain, which is then mixed with DNA from another taxon and the resulting amount of annealing (re- association) between the two DNAs, is an estimate of relationship of the nucleotide sequences (Fig. 8.28). This method has also been used in plant systematics. For example, Bendich & Bolton discussed the relationships among the Leguminosae as measured by the DNA-Agar technique. Based on DNA hybridization studies, they found that of the three cereals, *Secale*, *Hordeum* and *Triticum*, wheat (75%) is more closely related to rye (100%) than is barley (57%).

Southern Blotting Technique
Northern Blotting Technique

DNA-RNA hybridization:

RNA has also been studied by the DNA-RNA hybridization technique in which the quantity of association occurring between the DNA of one organism and a fraction of the RNA (usually ribosomal RNA) of another organism is considered as a basis of similarity (Fig. 8.29). For example, utilizing this technique, Mabry studied plants of the Centrospermae, and concluded that the *Caryophyllaceae* (which contains anthocyanins only) is related to the betalain-producing families but the relationship is not so close as the latter are to each other. Similarly when ribosomal RNAs of different species of *Cucurbita* were hybridized with DNAs, there was a considerable variation (1.7-3.1) occurring among these species. Dissimilar species *C. andreana* and *C. sororia* had the same proportion (1.7) of ribosomal DNA whereas similar species *C. andreana* and *C. maxima* had different proportions with 1.7 and 3.1 respectively.

(c) Amino Acids:

There are two groups of amino acids present in plant cells.

They are:

Protein amino acids :

They are the building blocks of proteins and are released on hydrolysis of proteins. They are universally distributed in plant tissues. Amino acids present in the free form in various plant cells has also been used as a taxonomic factor in some cases. For example, Shellard and Jolliffe studied the free amino acid composition of the pollen of 11 different grass species, and found it to be similar in all the species. Generally total free amino acids are usually higher in pollen than in leaves or other plant tissues. Hence, free amino acid composition of pollen has been used to study the homology between various species, e.g. Asteraceae, Cassia, etc. However, it is very difficult to draw any conclusion on evolution based upon the data on free amino acid composition alone, as amino acid composition greatly varies with climatic and nutritional conditions, as well as with storage and handling patterns.

Non-protein amino acids :

They are not found in combination in proteins and are more numerous, estimated to be about 300 in number. The non-protein amino acids are discontinuous in distribution and less susceptible to rapid change, which increase their taxonomic value. They occur in plants, which are only distantly related. The occurrence of such compounds may be explained by the fact that different plants can probably produce the same amino acid by different biochemical routes. Several workers have studied the taxonomic potential of such amino acids. Following are a few examples of the taxonomic significance of specific non-protein amino acids present in different groups:

I. Acetyl ornithine has been identified as the main non-protein amino acid in Fumarioideae under the Fumariaceae. It has also been found in ferns and grasses.

II. Turner and Harborne, detected canavanine only in 60 percent of the species under Lotoideae of the Leguminosae. The prevalence of canavanine was considered as an advanced character by Birdsong, as it was unknown from the primitive tribes Podalyrieae and Sophoreae. However, this amino acid was subsequently discovered in both the tribes.

III. The amino acid azetidine-2-carboxylic acid is extremely restricted in its distribution, being present in the Agavaceae, Amaryllidaceous and Liliaceae. This amino acid is also reported from the legumes *Delonix* and *Peltophorum*.

IV. *Lathyrus* species could be grouped under seven infra-generic groups on the basis of the association of amino acids within the seeds. Each group is characterized by a different amino acid or group of amino acids.

V. Species of the section Gummiferae of the genus *Acacia* can be recognized easily on the basis of the amino acid contents of their seeds.

VI. Four infra-generic groups have been recognized in the genus *Vicia* on the basis of distribution of amino acids.

VII. Free amino acid, lathyrine is so far known only in the genus *Lathyrus* and this supports the present circumscription of the genus.

(d) Flavonoids:

Of all the various secondary constituents, flavonoids have been the most widely exploited phytochemical constituent in chemotaxonomic studies. They are phenolic glycosides consisting of two benzene rings linked together through a heterocyclic pyrane ring.

Flavonoids are further classified into various types, such as:

I. Flavones — e.g. apigenin, levtolin

II. Flavanones — e.g. naringenin

III. Iso-flavones — e.g. orobol

IV. Iso-flavonoids — e.g. ferreirin

V. Flavanols — e.g. kaempferol, quercetin

VI. Anthocyanidins — e.g. cyanidin, delphinidin

VII. Chalcone — e.g. butein

VIII. Aurone — e.g. sulphuretin

Flavonoids are present in leaves, flowers and fruits. They function not only as insect repellents, but also play a vital role in resisting chronic pests and pathogens and by interaction with plant hormones, they regulate certain growth processes. These chemically complex compounds are easily and quickly identified, have great structural variation, widespread in distribution and physiologically relatively stable, which makes them taxonomically significant. The use of flavonoids in evaluating contemporary classificatory systems has been mainly based on their distribution patterns, i.e. their presence or absence. Phytochemical taxonomists have considered them as evolutionary markers due to the many positive correlations that they display. The primitive and advanced types of flavonoids in angiosperms were, deduced by Harborne. According to Swain, during the course of evolution flavonoids increased in their structural complexity from the primitive glycoflavones to the proanthocyanidins (in ferns, gymnosperms and putative angiosperm progenitors). The content of flavonoids have been found to vary in different plant organs and have also been found to vary in response to light and growth. The flower pigments, which are usually anthocyanins and anthoxanthins, vary greatly and have been shown to be under genetic control. It is also reported that the anthocyanin production increases with the deficiency in nutrients such as nitrogen and phosphorus. However, all variation in pigment production is not genetic in origin. In general, vegetative tissue extracts are more reliable and convenient in chemo taxonomical studies. Flavonoids from vegetative parts

and seeds provide more reliable taxonomic evidence than flower pigments, which are too variable. Many phylogenetic inferences have, been based on these characters in various groups. Following are a few examples:

The flavonoid systematics of the genus *Perideridia* (Umbelliferae), was studied by Giannasi and Chuang, who found that although the 16 species studied by them had the same flavonoids, they could be categorized into three discrete groups.

They are:

I. Those producing only flavonols.

II. Those with flavonols mainly and a few flavones.

III. Those which produce flavones predominantly.

They also tried to correlate between these three different types and their respective geographical distributions in America. The flavonol-producing group was centred in California and the second and third groups in the Pacific North West and eastward to mid-western USA.

An extensive survey of the families Umbelliferae and Araliaceae by Hegnauer revealed that, the two families are closely related. Hegnauer also concluded that they may be regarded as constituting the order Urabellales, which in turn may have evolved from a Rutalean stock. Analysis of flavonoids of leaf of Liliaceae, Juncaceae, Cyperaceae and Poaceae by Williams suggest that all these families have arisen from Liliaceous ancestors.

The flavonoid pattern of 5 species of *Tragopogon* of Asteraceae, and their F1 and F2 hybrids were studied, by Belzer and Ownbey, who noted that all the species and their population showed different chromatographical results. Flavonoids have been used in the estimation of generic affinities in Ulmaceae and of species relationships in *Chenopodium*. The evolutionary pathways in many taxa, e.g. Lemnaceae and *Ruellia* have been traced based on the data of flavonoids. Flavonoid chemistry has been used to distinguish species of *Spirodella*, which cannot be distinguished on the basis of their morphology.

(e) Betalins:

Betalins differ from flavonoids and other phenolic compounds in that they contain nitrogen in them. They are however functionally equivalent to phenolics.

The betalins are popularly known as nitrogenous anthocyanins and comprise of:

I. Red to violet betacyanins.

II. Yellow betaxanthins.

Plants containing betalin do not contain anthocyanins, which are the normal pigments found in other angiosperm families. The two groups are unrelated both chemically and

biosynthetically. The distribution of betalain has proved to be notable chemotaxonomic criteria. Some examples of the taxonomic value of betalins are mentioned below:

The betalins are confined to ten families of angiosperms (which contain only betalain but no anthocyanin) i.e., Chenopodiaceae, Portulacaceae, Aizoaceae, Cactaceae, Nyctaginaceae, Phytolaccaceae, Stegnospermaceae, Basellaceae, Amaranthaceae and Didieraceae, which have been placed under a single, order Centrospermae. Centrospermae also includes two anthocyanin-containing families, Molluginaceae and Caryophyllaceae. Mabry had suggested the transfer of the anthocyanin-containing families, Molluginaceae and Caryophyllaceae, to a distinct order Caryophyllales, restricting the Centrospermae to betalain-containing groups alone. However, other workers did not support this suggestion. Mabry again re-examined his earlier view in 1977, and concluded that the Centrospermae including both betalain and anthocyanin-containing families/is a monophyletic group, and suggested the recognition of two Centrospermous suborders, based on the presence of the respective pigments. The systematic position of the family Cactaceae under Centrospermae, which contain betalain, has been a matter of dispute in the past. It was often placed in the order of its own i.e., the Cactales or Opuntiales. However the presence of betalain in the Cactaceae establishes its position in the Centrospermae. The genus *Gisekia*, based on morphological similarities was traditionally placed under Molluginaceae. However, it has been found to be anomalous in that family because it contains betalains instead of anthocyanins. Takhtaja has recently transferred this genus to Phytolaccaceae, where it seems to belong more naturally than in any other family of Centrospermae. Mabry and Behnke found that the genus *Dysphania*, assigned to various families (Chenopodiaceae, Caryophyllaceae and Illecebraceae) by various authors, possessed betalain pigments and sieve tube plastids, characteristic of the Centrospermae. They have strongly supported its allocation to the Chenopodiaceae.

(f) Alkaloids:

Alkaloids are a heterogeneous group of organic nitrogen containing bases, often with a heterocyclic ring. The true alkaloids have a nitrogen-containing heterocyclic nucleus derived from a biogenetic amine and they can be related structurally to parent bases such as isoquinoline, pyridine, piperidine and tropane.

They are by-products of plant metabolism and are distinguished:

I. True alkaloids — They contain a nitrogen-containing heterocyclic nucleus derived from a biogenetic amine.

II. Protoalkaloids — They are derived from amino acids but lack any heterocyclic ring.

III. Pseudo alkaloids – They are biogenetically unrelated to amino acids, and are derived from terpenes, sterols, aliphatic acids, nicotinic acids or purines. They are distributed throughout the plant tissues, and are present in the vacuoles in the form of salts. However they are not essential for the growth of plant, but have striking properties of affecting the nervous system of animals. They are well known for their medicinal, poisonous and systematic viewpoints. Alkaloids are not

universally accumulated by plants, but are unexpectedly widely distributed in flowering plants and ferns. They are somewhat characteristic of particular families. According to Manske, there are about 38-39 families that can be regarded as alkaloid-containing families on the other hand Hegnauer has suggested that none of the large families are free from alkaloids. Over 5000 alkaloids have been reported from angiosperms mostly from the Dicotyledons, as for example:

I. Families like Berberidaceae, Fabaceae, Ranunculaceae and Solanaceae are especially rich in alkaloidal species. Other families include Annonaceae, Fumariaceae, Hydrastidaceae, Menispermaceae, etc.

II. The members of the Papaveraceae synthesize isoquinoline alkaloids always including protopine, those of the Fabaceae lupin alkaloids, while Solanaceae have tropane alkaloids, Asteraceae and Poaceae; on the other hand, produce many different types of alkaloids.

III. Sometimes alkaloids have a very narrow distribution. For example morphine is restricted to *Papaver somniferum* confined to a few Papaveraceae and strychnine to a few species of Strychnos.

Alkaloid content can be considered as a source of taxonomic evidence as alkaloids characterizing species of a particular taxon are frequently of the same chemical or biogenetic group. This suggests that related plants share the same pathways of alkaloid synthesis.

Therefore alkaloid content can be considered as a source of taxonomic evidence. However, mere accumulation of alkaloid does not imply taxonomic relationship at family level.

Some of the taxonomic aspects of alkaloids are mentioned below:

I. Distribution of alkaloid has proved useful in the taxonomy of the Fabaceae. Of the three genera viz. Genista, Ammodendron and Adenocarpus under this family, Genista and Adenocarpus were included in the tribe Genisteae, whereas Ammodendron was placed in the tribe Sophoreae, which is characterized by the presence of matrine alkaloids.

However, phytochemical studies have shown that all the three genera contain ammodendrone-hystrine alkaloids, which suggests that Ammodendron should be transferred to the tribe Genisteae.

II. Lupin alkaloids have been found in three tribes under the subfamily Lotoideae of the Fabaceae viz. Sophoreae, Genisteae and Podalyrieae, which suggest that these tribes may have originated from a common ancestral stock.

III. According to Jones and Luchsinger, alkaloids are useful in taxonomic studies in Papaver and Argemone (Papaveraceae), Veratrum (Liliaceae), as well as in Lycopodium, Lupinus and Caryophyllales.

IV. The presence of isoquinoline alkaloids in the families Fumariaceae and Papaveraceae indicates very close relationship between the two families.

V. On the basis of the distribution of benzyloisoquinoline alkaloids, several modern taxonomists, including Takhtajan, Cronquist, Thorne and Dahlgren, (rearranged the families with apocarpous gynoecium in Magnoliidae).

(g) Terpenoids:

These are a biogenic group of volatile compounds, which are mostly polymerized isoprene residues (isoprene unit -2-methyl 1, 3 butadiene). According to the number of isoprene units present in a terpenoid molecule, they may be of following categories:

I. Hemiterpenes (single isoprene unit) — e.g. tiglic acid

II. Monoterpenes (two isoprene units) — e.g. menthol

III. Sesquiterpenes (three isoprene units) — e.g. farnesol

IV. Diterpenes (four isoprene units) — e.g. phytol

V. Triterpenes (six isoprene units) — e.g. squalene

VI. Tetraterpenes (eight isoprene units) — e.g. carotenoid

VII. Polyterpenes – e.g. rubber

Terpenoids are almost universal, though not as widespread as flavonoids and have been used extensively in the chemotaxonomy of mints, umbellifers, Citrus plants, and gymnosperms. Unlike flavonoids, which are more difficult to work with at the technical level, monoterpenes can be rapidly identified by combined gas-liquid chromatography (GLC) and mass spectrometry. Some specific terpenoids are found in certain families, e.g. sesquiterpene lactones in Asteraceae, cucurbitacins in Cucurbitaceae and asperuloside in Rubiaceae. Some of the chemotaxonomic applications of terpenes are mentioned below:

Sesquiterpene lactones are common in the family Asteraceae, in which the oxidation level of these sesquiterpenes is sometimes specific to a tribe, subtribe or even a genus. The absence of sesquiterpenes in the Astereae may be rare. It has been found that if *Ambrosia* and its allies viz /va, *Franseria* and *Xanthium* are removed from the *Heliantheae*, then the residual elements are similar to the *Helenieae* in terms of lactone distribution. Thus, there is a need to construct a separate tribe or a distinct family to accommodate allied forms of *Ambrosia*, which exhibit diversity in lactone production. Geographical races of *Pseudotsuga menziesii* were distinguished based on the terpenoids of their conical oleoresins. The classification of *Eucalyptus* has been very difficult on the basis of gross morphology. Baker and Smith divided this genus into larger groups containing different oil- combination, which they correlated with leaf venation and bark structure:

a. Butterfly venation, shows high yield, with phellandrene and piperitone as main components.

b. Obtuse feather venation shows a low yield, with pinene as the dominant compound. More acute venation, having marginal veins, shows a slightly higher oil yield with cineole and pinene as chief constituents.

c. Origin of certain Citrus cultivars was determined by studying their rind and leaf terpenoid pattern.

d. A survey of monoterpenes in 19 species of Salvia by Emboden and Lewis, indicates that terpene composition is a valid morphological trait in the analysis of introgression within a group.

e. Gum terpenes, have been used by Mirov in the taxonomy of Pinus.

f. Irwin found that there are three sympatric taxa of Hedeoma (Labiatae) coexisting without any sign of natural hybridization, and used their terpenoids in identifying them.

(h) Iridoid Compounds:

They are monoterpenoid cyclopentanoid lactones, which represent a separate class of taxonomically significant compounds. The distribution of iridoid compounds has attracted the attention of plant taxonomists as a character of systematic importance in recent times and is considered to be an evolutionary marker. The significance of the monoterpenoid cyclopentanoid lactones called iridoids has been reviewed by Bate-Smith and Swain. Among these compounds, asperuloside is particularly common in the Rubiaceae. Aucubin is frequently noted in the Cornaceae, Scrophulariaceae, Orobanchaceae and some closely related families.

Some of the aspects of their taxonomic value are mentioned below:

a. On the basis of iridoid distribution, some taxonomic changes have been suggested in the 12th edition of Engler's syllabus. The aucubin-containing genus Buddleia must be removed from the Loganiaceae to the Buddleiaceae, with a position near the Scrophulariaceae.

b. Removal of Garryaceae from apetalous orders to the vicinity of the Cornaceae is suggested in Engler's syllabus on the basis of morphological evidence and this treatment receives support from chemical evidence as both Garryaceae and Cornaceae produce aucubin.

c. An extensive survey of the iridoids among angiosperms, was made by Jensen, who discussed their systematic importance. From their studies, they concluded that the sympetalous subclass Asteridae, is not natural as defined by Takhtajan, but polyphyletic. The orders Dipsacales, Gentianales and Scrophulariales (excluding Solanales) are characterized by the presence of iridoids while the others lack it.

They have proposed the splitting of Scrophulariales and also the separation of Goodeniales from Campanulales and the Loasales from Violales. All these studies thus, suggest a double or even multiple ancestry for the Sympetalae, which has to be rejected as an unnatural group.

d. Kubitzki connected the Rosalean and Guttiferalian complexes on the basis of the presence of iridoid compounds. According to Cronquist both these groups came from Magnoliidae, which, however, is completely devoid of iridoids.

e. Meeuse pointed to the presence of iridoids in some taxa of the traditional Rosiflorae and their absence from others to support his concept of the polyphyletic origin of angiosperms.

f. The restriction of iridoid compounds mainly to groups with unitegmic ovules, suggests that they developed along an evolutionary line where the ovules were just about to evolve from bitegmic to unitegmic condition. According to Jensen, 'Altingiaceae and Daphniphyllaceae (Hamamelidales) are perhaps relicts of primitive iridoid-bearing groups with bitegmic ovules.

g. Studies on iridoids in correlation with other traditional taxonomic characters of the monogeneric Fouquieriaceae and related orders Tamaricales (in which the family is usually included), Ericales, Cornales and Solanales by Dahlgren, suggests that the Fouquieriaceae is quite distinct and that it should be placed in an order of its own.

h. A minor modification in Takhtajan's classification can lump all iridoid-producing orders into a single group, indicating a possible common origin.

(i) Oils, Fats and Waxes:

Along with the proteins and carbohydrates, lipids or fats form the bulk of the organic matter of plant tissue and are therefore a potential source of taxonomic evidence.

Lipids are the esters of fatty acids with glycerol, and are mostly made up of carbon, hydrogen and oxygen (simple lipids). The fatty acids present may be saturated or unsaturated. The greater the proportion of saturated fatty acids in a lipid, the higher is the melting point.

Saturation is usually measured by the iodine number, i.e. the number of grams of iodine absorbed by a hundred grams of fat, oil or fatty acid.

Some lipids may contain nitrogen, a carbohydrate group, phosphorus or some other group, in addition to carbon, hydrogen and oxygen. They are called conjugated or complex lipids. The waxes on the other hand are esters of long-chain alcohols with long-chain fatty acids and may contain free alcohols, free fatty acids, aldehydes, ketones or hydrocarbons.

Lipids and waxes are somewhat heterogeneous group and are completely or partially soluble in organic solvents such as ethanol, ether or chloroform. Fats and oils can be distinguished by their physical state at normal temperatures i.e., fats are solids while oils are liquids.

Lipids are found in all parts of plants but are dominant in the storage organs, seeds and fruits, forming droplets suspended in the cytoplasm. Waxes occur in the cuticular layers of plants.

The taxonomic significance of lipids was assessed for the first time, by McNair. According to him, the iodine number of lipids was higher (i.e. lipids are more saturated) in more advanced plant groups. However, since the iodine number is a rather unstable character, it must be used carefully.

Fatty acids specific to particular plant groups may be of taxonomic significance. For example, ximenyric acid is found in the Olacaceae and Santalaceae, petroselinic acid is almost completely restricted to the Umbelliferae and erucic acid to Cruciferae. Certain species of the Flacourtiaceae are characterized by the presence of chaulmoogric acid and some related acids.

Bacterial lipids contain fatty acids, which are not found in other plants. Similarly, the highly evolved Asteraceae do not contain lipid with unsaturated fatty acids, while the algae, possess very unsaturated lipids. Thus, keeping in mind the fact that lipids are not constant in absolute composition, characters relating to their intact structure would not be of any value in taxonomy.

Some of the taxonomic applications of lipids and waxes are under mentioned:

a. The fatty acid composition of lipids has been used by Shorland as a possible source of taxonomic importance, and on the basis of the major fatty acids released by hydrolysis of their lipids, he recognized groups of families containing:

I. Linolenic acid – rich seed fats.

II. Linoleic acid – rich / oleic acid – rich seed fats.

III. Linoleic acid and oleic acid rich seed fats rich, with linolenic acid or a conjugated polyethanoic acid as principal component.

IV. Seed fats rich in palmitic, oleic and linoleic acids.

V. Seed fats with other characteristic acids in addition to oleic, linoleic and palmitic acids.

However, he found that, almost all fatty acids are present in all groups at low concentrations, but oleic acid is the main fatty acid in most groups. Simultaneously, distantly related families sometimes occur under the same group and a family may appear in more than one group. However, the family Palmae is an exception, where the proportions of fatty acids in lipids are constant in both species and genera.

The presence of unusual fatty acids in groups has also been of some taxonomic significance.

For example:

i. Linolenic acid and the unusual fatty acid, octadecatetraenoic acid, are present in the lipids of both leaf and seeds of eight members of the Boraginaceae. At the same time, the same unusual fatty acids are present in ten species of the unrelated Caryophyllaceae.

ii. Capric acid, which is present in *Ulmus*, is also dominant in seed lipids of *Zeikova* (Ulmaceae) as well as members of the Lauraceae and Lythraceae.

iii. Malvalic acid is present in members of the Malvaceae.

Proportions of fatty acids in lipids are constant in both genera and species in Palmae.

The Eucalyptus species of the Australian arid zone form a homogeneous group on the basis of wax characters. Wax alkanes also appear to be attractive taxonomic characters. However according to Martin-Smith, a systematic investigation into the possible influence of season, climate, geographical distribution and the kind and age of organs on the composition of plant surface waxes is essential before the method can be accepted.

- i. Alkane variation in cuticular waxes serves as a useful source of taxonomic information in the Poaceae.
- ii. There is a general uniformity in proportions of alkane hydrocarbon constituents within the genera of the Crassulaceae.
- iii. Wax alkane variation can be related to morphology, cytogenetics and protein analysis in 22 tuber-bearing species of Solanum.

(j) Steroids:

Steroids may be considered as derivatives of a fused and fully saturated ring system called cyclopentanoperhydrophenanthrene or sterane, which is formed by the fusion of 3 cyclohexane rings in non-linear or phenanthrene manner and a terminal cyclopentane ring. True steroids possess two methyl groups and are mostly alcohols or esters.

In plants they serve the role of water-proof, being located in the plant cutins. Steroids have also proved to be of some taxonomic significance in some taxa. As for example, their distribution has proved helpful in the taxonomy of the genera of tribe Veratreae of family Liliaceae. According to Kupchan, these genera contain the steroid veratrum.

(k) Polysaccharides:

Polysaccharides perhaps offer the greatest hope for taxonomic evidence because of their complexity and diversity. But so far they have been examined from a systematic viewpoint only, particularly due to the difficulties in the procedures for isolation and fractionation of these compounds. Following are some of the examples of the use of polysaccharides in taxonomic evaluation:

MacLeod and McCorquodale studied the water-soluble polysaccharides in the seeds of 22 grass species, and observed considerable taxonomic variations. Lesser or greater amounts of D-glucosane was present in all the Festuceae, but Festuca and Lolium are distinctive in having an unusual trisaccharide. On the other hand, due to the presence of fructosans and absence of raffinose, the Bromelaeae forms a very natural tribe. The sugar alcohols have been studied by Plouvier, from a taxonomic standpoint.

He found that:

- I. Quercitol is exceptionally common in the Menispermaceae, although it is found elsewhere. Though it is present in all 35 species of Quercus, it is absent from Castanea and Fagus of the Fagaceae.
- II. Sorbitol distribution is in accordance with the taxonomy of the Rosaceae, supporting the transfer of Ulmaria to the Rosoideae (usually without sorbitol) from the Spiraeoideae (which contains genera with sorbitol).
- III. Similarly, Pinitol is widespread but is particularly common in the Caryophyllaceae. It is found only in the genera Magnolia under the Magnoliaceae, but absent in closely related families.

iv .Among storage polysaccharides, the distribution of seed amyloids, which consist of a principal chain of glucose with side-chains of galactose and xylose, has also proved to be of some taxonomic significance.

v. Amyloid was detected in 16 families of the Dicotyledons and none of the Monocotyledons by Kooiman in a test involving over 2,500 species. Structures of these compounds have been fully worked for *Annona muricata* and *Tamarindus indicus*, being basically similar. Although Leguminosae and Acanthaceae are not closely related, yet both of them contain amyloid and the amyloid distribution follows taxonomic lines within these two families.

vi. *Paewia* is the only amyloid-producing genus among the 30 tested genus from Ranunculaceae, although some taxonomists would prefer to transfer it to a monotypic family.

vii. Amyloid is present in a few species of the Lotoideae.

viii. Amyloid is found in many genera of the Caesalpinioideae. However, all compounds are restricted to the Amherstieae, Cynometreae and Sclerobiae. Linear and branched hemicellulose fractions of species of the Leguminosae and Poaceae were compared by Gaillard. He found prominent differences between the two families. While branched polymers of Leguminosae contain a high proportion of arabinose, galactose and uronic acid, those of Poaceae have comparatively more of xylose. Uronic acid is attached to arabinose in the legume polymer and to xylose in the grass polymer.

(l) Ellagitannins:

The systematic importance of ellagitannins have been focussed by Bate-Smith, Harborne and Spome. Extensive surveys of the flowering plants have revealed that these substances are exclusive to the Dicotyledons, and absent from the Monocotyledons. They have indicated a fundamental cleavage between the Magnoliidae, which lack ellagitannins and the Hamamelidae-Dilleniidae, which possess them, which in turn probably points to the origin of the Monocotyledons from a Magnolian stock.

(m) Cyanogenic Compounds:

Cyanogenic compounds are poisonous compounds produced by plants after injury to their cells, such as hydrocyanic acid, amygdalin, etc. The ability of these plants to release such poisonous compounds is called cyanogenesis and plants containing these compounds are called cyanophoric plants. Hegnauer defined the term cyanogenesis as the ability of certain plants to release hydrocyanic acid (Prussic acid) after injury of cells. Cyanophoric plants usually contain one or several cyanogenic glycosides. Most cyanogenic activity is located in leaves and seeds of plants. Cyanogenic glucosides are known to be transported from one tissue in a plant to another e.g. in cassava from the young leaves to the tubers. In the cell, cyanogenic glucosides are thought to be stored in the vacuole. The same plants also contain degradative enzymes, which upon cellular disruption of the plant tissue get in contact with the cyanogenic glucosides thereby causing the rapid release of hydrogen cyanide. This binary system – two sets of components, which are inert individually, comprises the “cyanide bomb” and plays a role in the chemical warfare of plants against herbivores, pests and pathogens. Cyanide, or prussic acid, is a naturally

occurring glycoside in certain plants. Corn, Sudan grass, sorghums, cherry, apple, and peach may accumulate large quantities of cyanide. This accumulation occurs when plants are injured or drought stressed, which can result in release of large amounts of cyanide. It was Bohm, who first detected HCN and amygdalin (the first plant glycoside) in seeds of *Prunus amygdalus* (Rosaceae) in the early 1800s. Till date about 2,056 species of vascular plants are known to be cyanophoric, including ferns, monocots and dicots. The taxa belonging to Araceae, Poaceae, Juncaceae, Juncaginaceae and Scheuchzeriaceae are common. Several dicotyledonous families belonging to Asteridae, Rosidae and Dilleniidae are also cyanogenic. In angiosperms these compounds have been found to occur erratically. The most frequently occurring cyanogenic glycosides fall into five groups:

- I. The prunasin group
- II. The taxiphyllin group
- III. The linamarin group
- IV. The gynocardin group
- V. The cyanolipid group

The type of biogenesis of cyanophoric compounds has been found to differ in the major groups of tracheophytes viz. Pteridophytes, Gymnosperms and Angiosperms:

- a. Phenylalanine acts as the precursor of cyanophoric compounds in the pteridophytes.
- b. In Gymnosperms, tyrosine is the only precursor.
- c. Angiosperms have valine, leucine and cyclopentenyl glycine, as well as phenylalanine and tyrosine as the precursors of cyanophoric compounds. According to Hegnauer, cyanogenesis, may prove to be of considerable value in the classification of plants, if adequately and carefully used as a character in plant systematics, even at the higher systematic levels. More than one biosynthetic group of cyanophoric compounds occurs only in very large genera or families, and in such cases cyanogenic compounds are of systematic value at subfamily levels. Following are a few examples of the role of cyanogenic compounds in plant systematics: The theory that the Liliopsida (Monocotyledons) evolved from ancestors resembling present day Magnoliidae is well supported by the study of the distribution of individual cyanogenic compounds in the tracheophytes. In both these groups, cyanogenesis proceeds in exactly the same way, i.e., with only tyrosine as the precursor. The tribe Calendulae of Asteraceae and Trifolieae and Phaseoleae of Fabaceae, are characterized by the presence of the cyanogenic compound linamarin, while amygdalin or prunasin are characteristic of the many cyanophoric taxa of Rosaceae having a basic chromosome number 9. The taxonomic evaluation of cyanogenesis in vascular plants based on chemical characterization of the cyanogenic constituents, attributable to distinct biosynthetic pathways, is much more meaningful for plant systematics than cyanogenesis itself. For example, the family Poaceae is biochemically homogeneous with regard to cyanogenesis, as although the subfamilies of Poaceae are characterized by three different cyanogenic compounds, namely dhurrin (Andropogonoideae), triglochin (Festucoideae and Eragrostioideae) and taxiphyllin (Bambusoideae), all are derived from tyrosine, and these compounds only differ in the position

of attachment of glucose and not in biogenesis. In certain cases, cyanogenesis is however quite variable. For example: The genus *Glyceria* comprises species, which may be cyanogenic, facultatively cyanogenic or non-cyanogenic. Extreme polymorphism has been revealed in the cyanogenesis of 12 species of *Lotus* (Leguminosae) from Israel. Some species are cyanogenic and others acyanogenic, while still others contained both cyanogenic and acyanogenic populations.