Plant Biosystematics

Introduction:

It is now a universally agreed upon fact that different species are not fixed entities but systems of populations which exhibit variation and wherein no two individuals are identical. This concept of variations was first proposed by Lamarck and further developed by Darwin, culminating in his famous book Origin of Species (1859).

Systematics is a unique natural science concerned with the study of individual, population and taxon relationships for purposes of classification. The study of plant systematics is based on the premise that in the tremendous variation in the plant world, there exist conceptual discrete units (usually named as species) that can be recognized, classified, described, and named, on the further premise that logical relationships developed through evolution exist among these units.

The studies on variations, experimental studies and hybridization studies in light of genetic information are commonly covered under the term biosystematics. The term was first proposed by Camp & Gilly (1943 as Biosystematy) to delimit natural biotic units and to apply to these units a system of nomenclature adequate to the task of conveying precise information regarding their defined limits, relationships, variability and dynamic structure.

Clausen *et al.*, (1945) regard genetics, cytology, comparative morphology and ecology as furnishing the critical data which together, when applied to the study of organic evolution, make up biosystematics. These two different approaches aim at the same problem, the study of variations. The study of biosystematics, mainly the experimental systematics and population genetics approach the common aim.

Any studies involving breeding programs among taxa have been regarded as involving biosystematics, even though the data sought were for determining evolutionary relatedness of taxa rather than for discovering "natural units." Böcher remarked that "In my opinion the place of biosystematics is closer to cytogenetics and

ecology than to taxonomy in a narrower sense. Our main interest is not classification but evolution. This, of course, does not mean that we should never deal with classification. But frequently it will be better to leave problems of taxonomic rank and nomenclature to taxonomists, or to cooperate with them. The main goal of a biosystematist is to try to unravel the evolution of a group of taxa, what the evolutionary forces were and how they worked together in each particular case". With the proliferation of other laboratory studies (including DNA) in systematics during the past several decades, however, sometimes biosystematics has been used to refer to any kind of experimental systematic study, i.e., involving any type of data-gathering except traditional morphology and anatomy. Further, because of the very broad definition of systematics, the need for the term biosystematics has surely lessened and, in the minds of some, nearly disappeared. It is still used occasionally, however, and in one instance as a synonym of systematics.

Experimental Taxonomy:

Experimental taxonomy (or experimental systematics) is another term most often used for laboratory-based studies other than (or in addition to) crossing or breeding data. These are not experiments in the strict scientific sense, but rather the gathering and analyzing of different kinds of comparative data that ordinarily are generated in the laboratory (e.g., cytology, phytochemistry, DNA sequences, computer manipulations). The term experimental taxonomy did derive from actual experimental investigations of the nature of plant species by transplants of clones into different environments, which effectively discriminated the genetic vs. environmental components of morphological variation. This was also called "genecology" by Turesson (1923) and others (e.g., Constance 1957).

Although these valuable kinds of studies were continued, experimental taxonomy has come to have a broader usage. Müntzing emphasized the importance of genetics and cytology in experimental taxonomy and commented that "what experimental taxonomy can do is to establish the nature and occurrence of such intra-

and interspecific differentiation that cannot be clarified by morphological, ecological and plant-geographical methods". The most effective experimental approaches in taxonomy have combined work in the herbarium and field with studies in the greenhouse and experimental plot". Some workers (e.g., Stace 1980) have regarded experimental taxonomy and biosystematics as synonyms.

Types of Data:

Over the last few decades, the affinities among plant groups have been redefined as more and more information are accumulated from various sources. Newer approaches in recent years include (a) increasing reliance on phytochemical information (Chemotaxonomy); (b) studies on ultrastructure and micromorphology; (c) statistical analysis of the available data and providing a synthesis of all the available information (Taxometrics); and (d) analysis of phylogenetic data to construct phylogenetic relationship diagrams (Cladistics). The aforesaid disciplines constitute the major modern trends in taxonomy. Data continues to flow from different disciplines, so that the process of analysis and synthesis is an ongoing activity. Taxonomy (Systematics) is as such a field of unending synthesis. The following disciplines have contributed to a greater or lesser extent to a better understanding of taxonomic affinities among plants.

1- MORPHOLOGY:

Morphology has been the major criterion for classification over the last many centuries. The initial classifications were based on gross morphological characters. During the last two centuries, more and more microscopic characters of morphology were incorporated. Although floral morphology has been the major material for classifications, other morphological characters have also contributed in specific groups of plants.

Habit:

Life-forms-though of little significance to taxonomy-allow a means of estimating adaptiveness and ecological adjustment to the habitat. In *Pinus*, bark characters are used for identification of species. Woody and herbaceous characters have been the primary basis of recognition of Lignosae and Herbaceae series within dicots by Hutchinson (1926, 1973).

For several decades it was believed that trees or shrubs with simple leaves represented the most primitive condition within angiosperms. Increased evidence over the last decade, however, is pointing towards the assumption that the perennial herbaceous condition in paleoherbs such as Ceratophyllaceae, Nymphaeaceae and Piperaceae represents the archetype of the most primitive angiosperms.

Underground parts:

Rhizome characteristics are important for identification of various species of the genus *Iris*. Similarly, bulb structure (whether bulbs are clustered on rootstock or not) is an important taxonomic criterion in the genus *Allium*. Davis (1960) has divided Turkish species of the subgenus *Ranunculus* of genus *Ranunculus* based on rootstock and habit.

Leaves:

Leaves are important for identification in palms, *Salix* and *Populus*. The genus *Azadirachta* has been separated from *Melia* among other features by the presence of unipinnate leaves as against bipinnate in the latter. Similarly, the genus *Sorbus* has been separated from *Pyrus*, and genus *Sorbaria* separated from *Spiraea* on the basis of pinnate leaves. Stipules are an important source for identification in *Viola* and *Salix*. Leaf venation is important for the identification of the species in *Ulmus* and *Tilia*.

Flowers:

Floral characters are extensively used in delimitation of taxa. These may include the calyx (Lamiaceae), corolla (Fabaceae), stamens (Lamiaceae, Fabaceae-Mimosoideae), or carpels (Caryophyllaceae). A gynobasic style is characteristic of Lamiaceae. Similarly, the gynostegium characterizes Asclepiadaceae (now recognized as subfamily Asclepiadoideae of family Apocynaceae). Different species of *Euphorbia* have a distinctive cyathium inflorescence with clusters of male flowers each represented by a single stamen.

Fruits:

Fruit characteristics are very widely used in identification. Coode (1967) used only fruit characteristics in delimitation of species of the genus *Valerianella*. Singh *et al*., (1972) used fruit morphology in identification of Indian genera of Compositae (Liguliflorae). In Asteraceae-the cypsela shape (usually called achene), presence or absence of pappus and whether the pappus is represented by hairs, scales or bristles, the presence or absence of beak, and its length, the number of ribs on the cypsela-constitute valuable identifying features. The number of capsule valves is used in segregating genera in family Caryophyllaceae (*Melandrium*, *Silene*, *Cerastium*). Seed characters are valuable identification features in the genus *Veronica*.

2- ANATOMY:

Anatomical features have played an increasingly important role in elucidation of phylogenetic relationships. Anatomical characteristics are investigated with the help of a light microscope; whereas ultrastructure (finer details of contents) and micromorphology (finer details of surface features) are brought out using an electron microscope. Anatomical work of taxonomic significance was largely undertaken by Bailey and his students. Carlquist (1996) has discussed the trends of xylem evolution, especially in the context of primitive angiosperms.

Wood anatomy:

Wood represents secondary xylem constituting the bulk of trees and shrubs, formed through the activity of vascular cambium. It primarily consists of tracheids and vessels. Tracheids are long narrow elements with tapering ends, imperforate at ends, and transfer of water and minerals occurring through pit-pairs (two adjacent pits of two tracheids, separated by primary cell walls). The vessels, on the other hand, are composed of vessel elements, much broader than tracheids and with perforation plates at ends (with opening not having primary walls unlike pit-pair). Vessel elements are joined end to end to form long tubes, the vessels. Perforation plate may be simple with a single opening, or compound with several openings. Latter with elongated openings in a row like a ladder is known as scalariform, a common type in primitive angiosperms.

Vessels are absent in Gymnosperms, but present in Angiosperms. It is commonly believed that there has been a progressive evolution in angiosperms from tracheids to long, narrow vessel elements with slanted, scalariform perforation plates, to short, broad vessel elements with simple perforation plates. Studies on wood anatomy have contributed largely in arriving at the conclusion that Amentiferae constitute a relatively advanced group, and that Gnetales are not ancestral to angiosperms. Bailey (1944) concluded that vessels in angiosperms arose from tracheids with scalariform pitting, whereas in Gnetales they arose from tracheids with circular pitting, thus suggesting an independent origin of vessels in these two groups. Demonstration of vessel-less angiosperms (Winteraceae, Trochodendraceae), also having other primitive features, has led to the conclusion that angiosperm ancestors were vessel-less. The separation of *Paeonia* into a distinct family Paeoniaceae and *Austrobaileya* into a separate family Austrobaileyaceae has been supported by studies of wood anatomy.

Nodal anatomy has considerable significance in angiosperm systematics. The number of vascular traces entering leaf base and associated gaps (lacunae) left in the vascular cylinder of stem at each node are distinctive for several groups. The node may

have single gap (unilacunar) from single leaf trace or three leaf traces (two additional commonly entering stipules) or three gaps (trilacunar) associated with three leaf traces (Figure 1) The genus *Illicium* has been separated from Winteraceae because of unilacunar nodes and the absence of granular material in stomatal depressions.

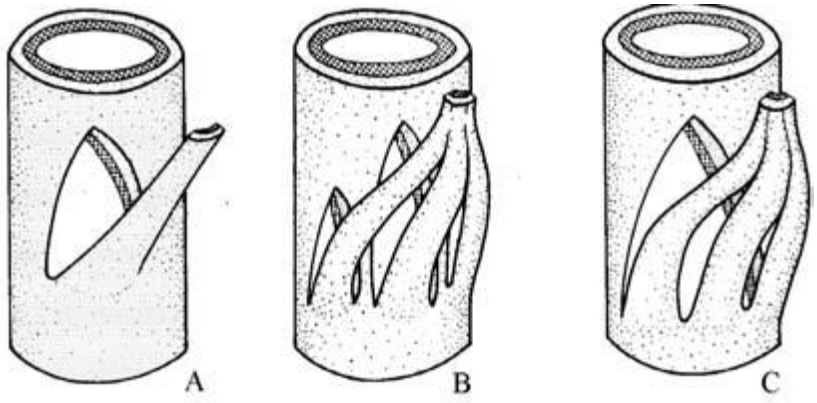


Figure (1): Nodal anatomy. A: Unilacunar node with one leaf trace; B: Trilacunar node with three leaf traces; C: Unilacunar node with three leaf traces.

Trichomes:

Trichomes constitute appendages of epidermis which may be non-glandular or glandular. Non-glandular trichomes may be in the form of simple unicellular or multicellular hairs (common in Brassicaceae, Lauraceae and Moraceae), in the form of vesicles, peltate hairs (*Olea*) or flattened scales.

Branched hairs may be dendroid, stellate (*Styrax*) or candelabrum-like (*Verbascum*). Glandular trichomes may be sessile or stalked and present a variety of forms.

Unicellular glandular hairs of *Atriplex* are bladder-like (Figure 2) with few-celled stalk and basal cell and they secrete salt. Others may secrete nectar (calyx of *Abutilon*), mucilage (leaf base of *Rheum* and *Rumex*). The stinging hairs of *Urtica* are highly specialized with silica tip which readily breaks when hair is touched. The broken tip is sharp like a syringe and easily penetrates the skin injecting irritating cell contents. Trichomes hold considerable promise in systematics of angiosperms.

Trichomes have been of considerable help in Cruciferae (Schulz, 1936), especially in the genera *Arabis* and *Arabidopsis*.

Trichome characters are very useful in the large genus *Astragalus* (with more than 2000 species). The Himalayan species *Hedera nepalensis* is distinguished from its European relative *H. helix* in having scaly trichomes as against stellate in the latter. In family Combretaceae the trichomes are of immense significance in classification of genera, species or even varieties (Stace, 1973). Trichomes are also diagnostic characters for many species of *Vernonia* (Faust and Jones, 1973).

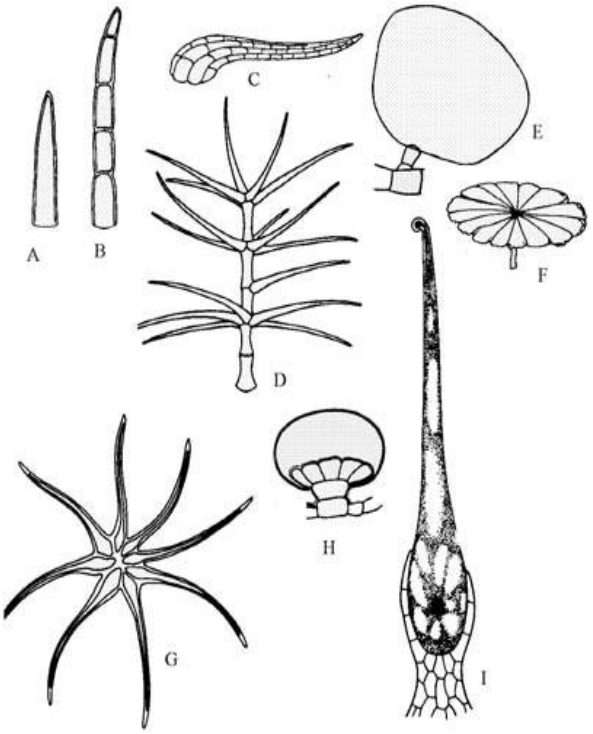


Figure (2): Trichomes. A: Simple unicellular hair; B: Multicellular hair; C: Scale; D: Candelabra trichome of *Verbascum*; E: Vesicular hair of *Atriplex*; F: Peltate hair; G: Stellate hair *Styrax*; H: Secretary gland of *Thymus*; I: Stinging hair of *Urtica*.

Epidermal features:

Epidermal features are also of considerable taxonomic interest. Prat (1960) demonstrated that one can distinguish a Festucoid type (simple silica cells, no bicellular hairs) and Panicoid type (complicated silica cells, bicellular hairs) of epidermis in grasses.

Stomatal types (Figure 3) are distinctive of certain families such as Ranunculaceae (anomocytic), Brassicaceae (anisocytic), Caryophyllaceae (diacytic), Rubiaceae (paracytic), and Poaceae (graminaceous). Anomocytic type has ordinary epidermal surrounding the stomata. In others the epidermal cells surrounding the stomata are differentiated as subsidiary cells. There may be two subsidiary cells at right angles to the guard cells (diacytic), two are more parallel to the guard cells (paracytic), or three subsidiary cells of unequal size (anisocytic). Other types include actinocytic type with stomata surrounded by a ring of radiating cells, cyclocytic with more than one concentric rings of subsidiary cells and tetracytic with four subsidiary cells. The stomatal complex of Poaceae is distinctive in having two dumb-bell shaped guard cells with two small subsidiary cells parallel to the guard cells.

Stace (1989) lists 35 types of stomata in vascular plants. Closely related families Acanthaceae and Scrophulariaceae are distinguished by the presence of diacytic stomata in the former as against anomocytic in the latter. The stomatal features, however, are not always reliable. In *Streptocarpus* (Sahasrabudhe and Stace, 1979), cotyledons have anomocytic while mature organs have anisocytic stomata. In *Phyla nodiflora* (syn = *Lippia nodiflora*) the same leaf may show anomocytic, anisocytic, diacytic and paracytic stomata (Pant and Kidwai, 1964).

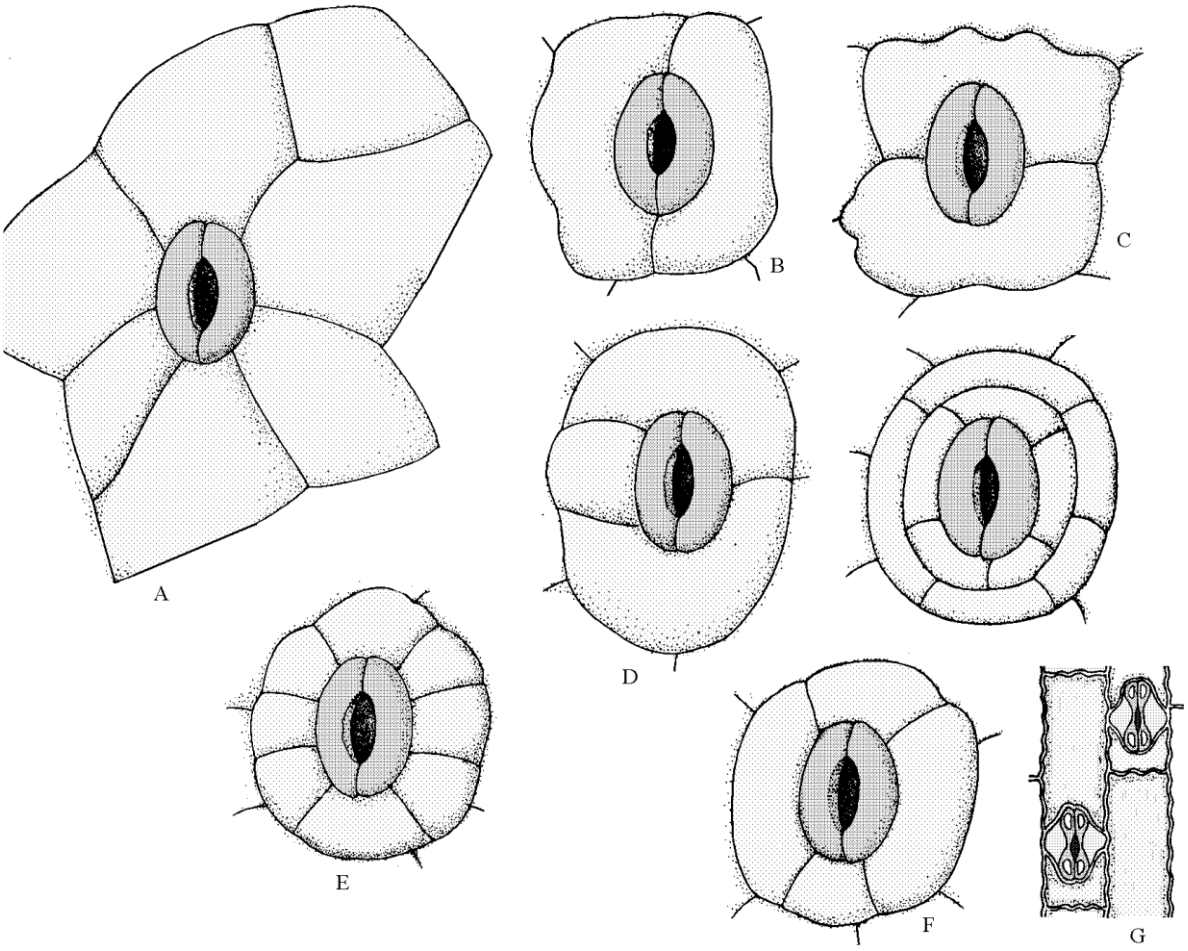


Figure (3): Stomatal apparatus in Angiosperms. A: Anomocytic type with epidermal cells around stomata not differentiated; B: Paracytic type with two or more cells parallel to the guard cells differentiated as subsidiary cells; C: Diacytic type with two subsidiary cells at right angles to the guards cells; D: Anisocytic type with three subsidiary cells of unequal size; E: Actinocytic type with stomata surrounded by a circle of radiating cells; F: Tetracytic type with four subsidiary cells; G: Graminaceous type with dumb-bell shaped guard cells with two small subsidiary cells parallel to the guard cells; H: Cyclocytic type with concentric rings of subsidiary cells.

Leaf anatomy:

The florets of Poaceae are reduced and do not offer much structural variability. Leaf anatomy has been of special taxonomic help in this family. The occurrence of the C-4 pathway and its association with Kranz anatomy (dense thick-walled chlorenchymatous bundle sheath, mesophyll simple), has resulted in revised classification of several genera of grasses. Melville (1962, 1983) developed his gonophyll theory largely on the basis of the study of venation pattern of leaves and floral parts. The rejection of *Sanmiguelia* and *Furcula* as angiosperm fossils from the Triassic has largely been on the basis of detailed study of the venation pattern of leaves (Hickey and Doyle, 1977). The more recent rediscovery of *Sanmiguelia* from the Upper Triassic of Texas (Cornet 1986, 1989) points to presumed angiosperm incorporating features of both monocots and dicots. Discovery of the Late Triassic *Pannaulika* (Cornet) from the Virginia-North Carolina border has reopened the possibilities of Triassic origin of angiosperms.

Floral anatomy:

Floral anatomy has been one of the thoroughly explored areas, with significant contributions to the understanding of the phylogeny of angiosperms. Vascular traces in the carpels of various genera of the family Ranunculaceae have confirmed the origin of achene (*Ranunculus*, *Thalictrum*, etc.) from follicle (*Delphinium*, *Aquilegia*, etc.) through successive reduction in the number of ovules ultimately to one. The additional traces which would have gone to other ovules, now aborted, can be observed in many genera. Thus, there is no justification for Hutchinson's separation of achene bearing genera and follicle-bearing genera into separate families Ranunculaceae and Helleboraceae, respectively.

Melville (1962, 1983) developed his gonophyll theory after studying the vasculature of carpel and other floral parts. He believed the angiosperm carpel to be a modified dichotomous fertile branch adnate to the petiole of a leaf.

The genus *Melandrium* was segregated from *Silene* on the basis of the ovary being unilocular as against partly septate in *Silene*. Detailed floral anatomy revealed that in all the species of both genera, the ovary is multilocular, at least in the early stages of development. The septa break down to various degrees in different species as the ovary develops. Thus structurally, the ovaries are similar. The two genera were consequently merged into the single genus *Silene*.

The inferior ovary in angiosperms has been formed in two ways: appendicular origin (formed by fusion of calyx, corolla and their traces to the ovary wall; in this case, all vascular traces have normal orientation, i.e. phloem towards the outside) or by axial invagination (formed by depression of the thalamus (receptacle); the inner vascular traces have reverse orientation, i.e. phloem towards the inside). Studies on floral anatomy have confirmed that in a large majority of families, the inferior ovary is of appendicular origin. Only in a few cases (*Rosa*, Cactaceae, etc.) is the origin by axial invagination of the thalamus.

Floral anatomy has also supported the inclusion of *Acer negundo* under *Acer*, and does not support its separation into a distinct genus *Negundo*. Although this species is specialized in having a dioecious habit and anemophily, the anatomy of the flower shows unspecialized features of other species. Floral anatomy also supports the separation of *Menyanthes* from Gentianaceae into a distinct family Menyanthaceae. The genus *Centella* is separated from *Hydrocotyle* on the basis of inflorescence being a cyme, and ovules receiving vascular supply from alternate bundles. In *Hydrocotyle*, the inflorescence is an umbel and the ovules receive vascular supply from fusion of two adjacent bundles. *Paeonia* is a classic example of a genus, which was removed from family Ranunculaceae into a distinct family Paeoniaceae. The separation has been supported by evidence from morphology, embryology and chromosomes. Floral anatomy also supports this separation, as both sepals and petals have many traces, carpels have five traces and the stamens are centrifugal. Developmental studies have indicated that some flowers, such as Apiaceae and Ericaceae, that appear to have free

petals, are gamopetalous early in development. They are, therefore, considered to have evolved from gamopetalous ancestors.

3- EMBRYOLOGY

Embryology has made a relatively lesser contribution in understanding taxonomic affinities. This is primarily because of long preparatory work needed for embryological studies. More often, the study of hundreds of preparations may reveal just a single embryological characteristic. It may take many years of laborious and painstaking research to study even a few representatives of a family. The embryological features of major significance include microsporogenesis, development and structure of ovule, embryo sac development, endosperm and embryo development.

Families marked out by distinct embryological features

A number of families of angiosperms are characterized by unique embryological features found in all members. These include:

Podostemaceae

Family Podostemaceae includes perennial aquatic herbs, which have a unique embryological feature in the formation of a **pseudoembryo sac** due to the disintegration of the nucellar tissue. The family is also characterized by the occurrence of pollen grains in pairs, bitegmic tenuinucellate ovules, bisporic embryo sac, solanad type of embryogeny, prominent suspensor haustoria, and absence of triple fusion and, consequently, endosperm.

Cyperaceae

Family Cyperaceae is characterized by the formation of only one microspore per microspore mother cell. Following meiosis, of the four microspore nuclei formed, only one gives rise to pollen grain. Besides Cyperaceae, only Epacridaceae in a few members shows the degeneration of three microspore nuclei. Cyperaceae is distinct from these taxa in pollen shedding at the 3- celled stage, as against the 2-celled stage shedding in Epacridaceae.

Onagraceae

Family Onagraceae is characterized by *Oenothera* type of embryo sac, not found in any other family except as an abnormality. This type of embryo sac is 4-nucleate and is derived from the micropylar megaspore of the tetrad formed.

Specific examples of the role of embryological data

There are a few examples of the embryological data having been very useful in the interpretation of taxonomic affinities:

Trapa

The genus *Trapa* was earlier (Bentham and Hooker, 1883) included under the family Onagraceae. It was subsequently removed to the family Trapaceae (Engler and Diels, 1936; Hutchinson, 1959, 1973) on the basis of distinct aquatic habit, two types of leaves, swollen petiole, semiepigynous disc and spiny fruit. The following embryological features support this separation: (i) pyramidal pollen grains with 3 folded crests (bluntly triangular and basin shaped in Onagraceae); (ii) ovary semi-inferior, bilocular with single ovule in each loculus (not inferior, trilocular, with many ovules); (iii) *Polygonum* type of embryo sac (not *Oenothera* type); (iv) embryo Solanad type (not Onagrad type); (v) one cotyledon extremely reduced (both not equal); and (vi) fruit large one-seeded drupe (not loculicidal capsule).

Paeonia

The genus *Paeonia* was earlier included under the family Ranunculaceae (Bentham and Hooker; Engler and Prantl). Worsdell (1908) suggested its removal to a distinct family, Paeoniaceae. This was supported on the basis of centrifugal stamens (Corner, 1946), floral anatomy (Eames, 1961) and chromosomal information (Gregory, 1941). The genus has been placed in a distinct monogeneric family, Paeoniaceae, in all modern systems of classification. The separation is supported by the following embryological features: (i) centrifugal stamens (not centripetal); (ii) pollen with reticulately-pitted exine with a large generative cell (not granular, papillate and smooth, small generative cell); (iii) seed arillate.

Exocarpos

The genus *Exocarpos* (sometimes misspelled *Exocarpus*) is traditionally placed under the family Santalaceae. Gagnepain and Boureau (1947) suggested its removal to a distinct family Exocarpaceae near Taxaceae under Gymnosperms on the basis of articulate pedicel, 'naked ovule' and presence of a pollen chamber. Ram (1959) studied the embryology of this genus and concluded that the flower shows the usual angiospermous character, the anther has a distinct endothecium and glandular tapetum, pollen grains shed at the 2-celled stage, embryo sac of the *Polygonum* type, and the division of zygote transverse. This confirms that the genus *Exocarpos* is undoubtedly an angiosperm, and a member of the family Santalaceae, with no justification for its removal to a distinct family. The genus placed in Santalaceae in all the major systems of classification.

Loranthaceae

The family Loranthaceae is traditionally divided into two subfamilies: Loranthoideae and Viscoideae, largely on the basis of presence of a calyculus below the perianth in the former and its absence in the latter. Maheshwari (1964) noted that the Loranthoideae has triradiate pollen grains, *Polygonum* type of embryo sac, early embryogeny is biseriate, embryo suspensor present, and viscid layer outside the vascular supply in fruit. As against this, Viscoideae have spherical pollen grains, *Allium* type of embryo sac, early embryogeny many tiered, embryo suspensor absent, and viscid layer inside the vascular supply of fruit. He thus advocated separation of the two as distinct families Loranthaceae and Viscaceae. The separation was accepted by Takhtajan (1980, 1987, 1997), Dahlgren (1980), Cronquist (1981, 1988) and Thorne (1981, 1992).

4- PALYNOLOGY

The pollen wall has been a subject of considerable attention, especially in an attempt to establish the evolutionary history of angiosperms. Some families, such as Asteraceae, show different types of pollen grains (**eurypalynous**), whereas others have a single morphological pollen type (**stenopalynous**). Such stenopalynous groups are

of considerable significance in systematic palynology. Pollen grains present a number of features of taxonomist interest. The number of nuclei present at the time of shedding is also significant. Most primitive angiosperms are shed at 2-nucleate stage, whereas in more advanced groups pollen is shed at 3-nucleate stage.

Angiosperms mostly have pollen grains of radial symmetry; bilateral symmetry being found in several gymnosperms. Most pollen grains are globose in shape, although boat shaped, ellipsoidal and fusiform are also met in different angiosperms. Since most pollen grains at least in early stages form tetrads, the outer end of grain is termed **distal pole**, whereas the inner end where grains meet as **proximal pole**, and the line joining the two poles as **polar axis**. The line running around the pollen at right angles to the polar axis is termed as **equator**.

Pollen aggregation

Microsporogenesis yields four microspores which mature into pollen grains. In large majority of angiosperms, the pollen grains separate prior to release. Such single pollen grains are known as **monads**. In rare cases pollen grains are released fused in pairs, when they are known as **dyads**. In many angiosperms the four microspores do not separate and the pollen grains form a **tetrad**. Five different types of tetrads are differentiated:

1. **Tetrahedral tetrad**- four pollen grains form a tetrahedron: four grains compacted in a sphere. Such pollen grains are found in family Ericaceae.
2. **Linear tetrad**- four pollen grains arranged in a straight line as in genus *Typha*.
3. **Rhomboidal tetrad**- four pollen grains in one plane, with two separated from one another by close contact of the other two.
4. **Tetragonal tetrad**- four grains are in one plane and equally spaced as in *Philydrum*.
5. **Decussate tetrad**- four grains in two pairs, arranged at right angles to one another, as in genus *Lachnanthes*.

In some genera, such as *Calliandra* of Mimosoideae, the pollen grains are connate in a group of more than four. Such pollen grains constitute a **polyad**. A polyad

generally consists of eight pollen grains, and rarely of more than ten. In some members of family Orchidaceae, as for example genus *Piperia*, large number of pollen grains form irregular groups, of which there are more than one groups in a theca. These are known as **massulae**. In subfamily Asclepiadoideae of family Apocynaceae, and several members of Orchidaceae, all pollen grains of a theca are fused into a single mass known as **pollinium**.

Pollen wall

The pollen grain wall is made of two principal layers, outer **exine** and inner **intine**. The exine is hard and impregnated with **sporopollenin**, a substance that makes it resistant to decay, and enables preservation in fossil record. Exine is further differentiated into two layers: outer **ektexine** and inner **endexine**. The ektexine is further distinguished into basal **foot layer,** radially elongate **columella** and roof like **tectum** (Figure 4). In some taxa the columella may be replaced by granular middle layer. Similarly in some primitive angiosperms tectum is lacking (**atectate** pollen grain), and the exine appears granular. Above layers of exine are clearly visible under an electron microscope, but when observed under a light microscope, the inner layer known as **nexine**, includes endexine plus foot layer of ektexine. The upper layers consisting of columella, tectum and the supratectal sculpturing constitute **sexine**.

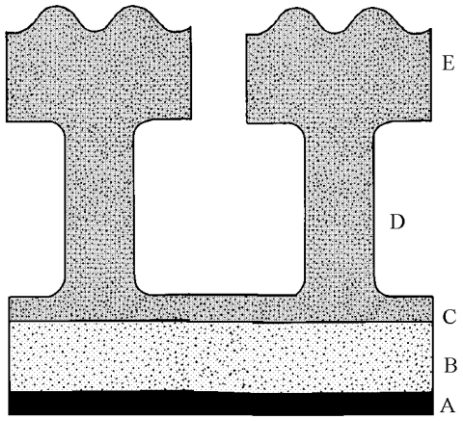


Figure (4): Fine structure of Pollen wall. A, Intine, B, Endexine, C, Foot layer, D, Baculum and E, Tectum. Note the aperture formed due to break in the tectum and baculum layers.

Pollen wall sculpturing

Present on the outer surface of tectum are often certain supratectal projections, which provide a variety of sculpturing to exine wall. In some cases, lacking tectum, the sculpturing is formed by columellae. The common types of sculpturing include: **baculate** (rod-shaped elements, each known as **baculum**), **clavate** (club-shaped elements), **echinate** (spine-like elements longer than 1 micron), **spinulose** (spine-like elements shorter than 1 micron; **scabrate**), **foveolate** (pitted surface with pores), **reticulate** (forming network, each element known as murus and space in between as lumen), **fossulate** (longitudinal grooves), **verrucate** (short wart-like elements), **gemmate** (globose or ellipsoid elements), **psilate** (smooth surface), and **striate** (having thin striations on surface).

Pollen aperture

Pollen aperture is a specialized region of pollen wall through which the pollen tube comes out. The exine may be **inaperturate** (without an aperture) or aperturate. An aperturate pollen may have a single pore (**monoporate**), a single slit running at right angles to the equator (**monocolpate**), three slits (**tricolpate**), three pores (**triporate**) three slits each with a geminate pore in middle (**tricolporate**), with many pores (**multiporate**) accompanied by a variety of surface ornamentations (Figure 5). Pollen with one or more slits located at the polar end is accordingly termed, **monosulcate**, **disulcate** and **trisulcate**, depending on the number of slits. Pollen grain with slits joined at poles is termed syncolpate. Aperture having three branches is termed trichotomosulcate.

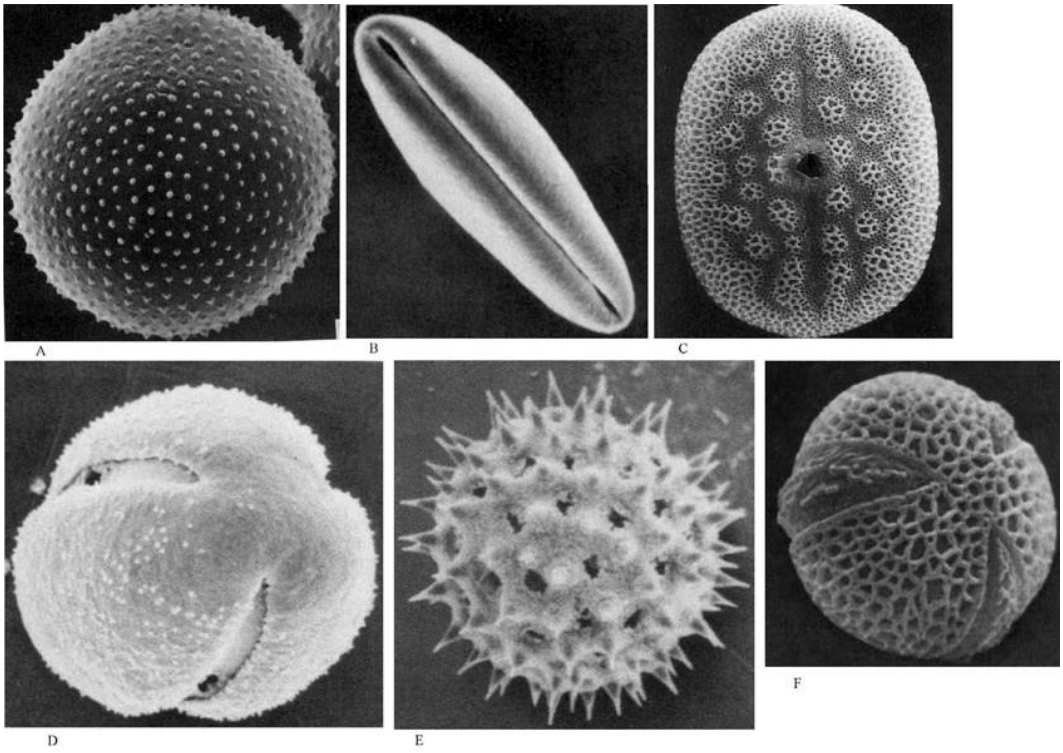


Figure (5): SEM of pollen grains. A: Nonaperturate pollen grain of *Persea americana*; B: Monosulcate pollen grain of *Magnolia grandiflora*; C: Monoporate pollen grain of *Siphonoglossa*; D: Tricolporate pollen grain of *Scaevola glabra*; E: Polyporate spinose pollen grain of *Ipomoaea wolcottiana*; F: Tricolpate pollen grain of *Disanthus cercidifolius*.

Monocolpate condition is widely spread in primitive dicots and a majority of monocots. The pollen of anemophilous plants is usually small, rounded, smooth, rather thin-walled and dry with shallow furrows. Anemophilous pollen is found in *Populus*, Poaceae, Cyperaceae, Betulaceae and several other families. Insect- and bird-pollinated pollen, on the other hand, is large, sculptured and often coated with adhesive waxy or oily substance. The pollen of Asteraceae is generally highly elaborate but simplification towards loss of sculpturing has occurred in several genera with wind pollination. The vestigial scattered patches of adhesive layer on wind pollinated pollen have been considered as evidence of the derivation of anemophily from entomophily.

Fossil studies over the last three decades have confirmed monosulcate pollen of the Early Cretaceous of southern England to be the oldest recorded angiosperm fossil with distinct sculptured exine, resembling the pollen of extant genus *Ascarina*.

Brenner and Bickoff (1992) recorded similar but inaperturate pollen grains of Israel, now considered being the oldest record of angiosperm fossils. This last discovery has led to the belief that the earliest angiosperm pollens were without an opening, the monosulcate types developing later.

Among examples of the role of pollen grains in systematics is *Nelumbo* whose separation from Nymphaeaceae into a distinct family Nelumbonaceae is largely supported by the tricolpate pollen of *Nelumbo* as against the monosulcate condition in Nymphaeaceae.

Brenner (1996) proposed a new model for the evolutionary sequence of angiosperm pollen types. The earliest angiosperm pollen (from the Valanginian or earlier) was small, circular, tectate-columellate and without an aperture. In the Hauterivian, there was possible occurrence of thickening of the intine coupled with thickened endexine and evolution of the sulcus. A considerable diversification of these monosulcate pollens occurred in the Barremian. Tricolpate pollen evolved in northern Gondwana in the lower Aptian. Multicolpate and multiporate pollen arose at a later stage.

5- MICROMORPHOLOGY AND ULTRASTRUCTURE

Although widely used in lower plants, electron microscopy has been a comparatively new approach for flowering plants. The finer details of external features (micromorphology) have been explored in the recent years by Scanning Electron Microscopy (SEM), whereas the minute details of cell contents (ultrastructure) have been discerned through Transmission Electron Microscopy (TEM). On an average basis, the resolution power of SEM is 250A (20 times as good as optical microscope, but 20 times lesser than TEM). Behnke and Barthlott (1983) have made extensive studies of SEM and TEM characters. In most of the examples studied, Electron Microscopy (EM) characters proved to be stable and unaffected by environmental conditions.

Micromorphology

SEM studies have been made primarily on pollen grains, small seeds, trichomes and surface features of various organs. In most of these organs (except pollen grains), the studies involved the epidermis. The value of epidermal studies lies in the fact that an epidermis covers almost all the organs and is always present, even in herbarium specimens. The epidermis is thick and stable in SEM preparations and is little affected by environment. However, it is important to note that only comparable epidermis should be studied (e.g. petals of all plants, leaves of all plants, not petals of some and leaves of others). Most of SEM studies have been concentrated on seed-coats which are usually thick-walled and stable in vacuum, thus facilitating quick preparation for SEM examination without the need for complicated dehydration techniques. The micromorphology of the epidermis includes the following aspects:

Primary Sculpture

This refers to the arrangement and shape of cells. The arrangement of cells is specific for several taxa. In Papaveraceae, seed coat cells by a particular arrangement form a reticulate supercellular pattern (Figure 6-D), which is a family character. The members of Caryophyllaceae, Portulacaceae and Aizoaceae exhibit a specific arrangement and orientation of smaller and larger cells known as “centrospermoid” pattern. There is specific distribution of long and short cells over the veins in the family Poaceae.

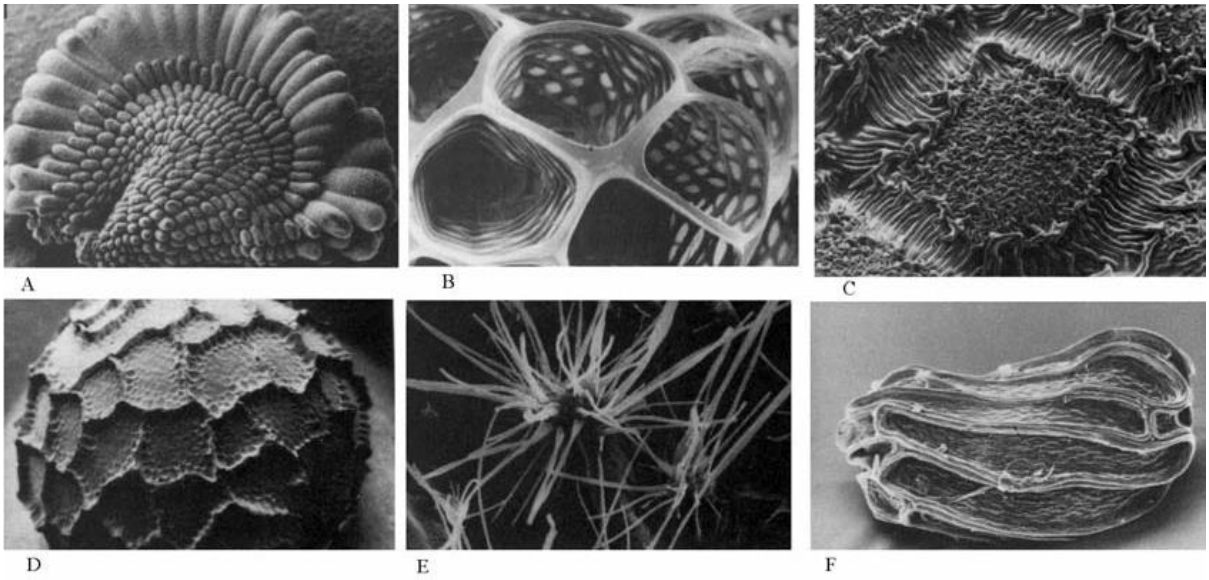


Figure (6): SEM seed characteristics of angiosperms. A: Seed of *Sceletium campactum* (Aizoaceae) showing centrospermoid cell arrangement; B: Seed-coat of *Aeginatia indica* (Orobanchaceae) with isodiametric deeply concave cells and reticulate secondary structure; C: Single isodiametric tetragonal cell of seed-coat of *Matucana weberbaueri* (Cactaceae) with heavy secondary sculpturing; D: Seed of *Eschscholzia californica* (Papaveraceae) with cells arranged to form a supercellular net-like pattern; E: Seed coat of *Jacaranda macarantha* (Bignoniaceae) with stellate epicuticular sculpture; F: Seed of *Dichaea* sp. (Orchidaceae) almost one cell long with heavy marginal thickenings and irregular secondary sculpture.

The shape of cells is mainly determined by the outline of the cells, boundaries of the walls, relief, and cell wall curvature (flat, convex or concave). Outline of cells may be isodiametric (usually tetragonal or hexagonal: Figure 6-B and C), elongated in one direction (Figure 6-F).

Cell boundaries of superficially visible anticlinal walls may be straight (Figure 6-B and C), irregularly curved or undulated and are of high taxonomic significance in family Cactaceae and Orchidaceae. Relief of the anticlinal boundary may be channelled or raised. The curvature of outer periclinal walls may be flat, concave (Figure 6-B) or convex.

Secondary Sculpture

The secondary sculpture (Microrelief) is formed by the deposition of cuticle over the outer wall or due to secondary wall thickenings, often shrinking and collapsing in desiccated cells. It may be smooth, striate (Figure 6-C), reticulate (Figure 6-B) or micro-papillate (verrucose). All members of Urticales have curved trichomes with silicified cuticular striations at the base, and micro-papillations on the trichome body. This single character of trichomes allows for precise circumscription of the order Urticales. Loasiflorae is circumscribed by unicellular irregularly hooked trichomes. Secondary wall thickenings are always of a high taxonomic significance. In Orchidaceae, for example longitudinal striations caused by underlying secondary thickenings are restricted to all members of Catasetinae.

Tertiary Sculpture

Tertiary sculpture is formed by epicuticular secretions such as waxes and other mucilaginous adhesive lipophilic substances and shows a variety of patterns. Secondary and tertiary sculpturing are mutually exclusive as the presence of waxes would invariably mask the cuticle; the cuticle would be visible only if there are no wax deposits. Winteraceae have a particular type and distribution of wax-like secretions (alveolar material not soluble in lipid solvents) on their stomata, similar to gymnosperms, and absent in all other angiosperms. In monocots orientation and pattern of epicuticular waxes seem to provide a new taxonomic character of high systematic significance. Four types of wax patterns and crystalloids have been distinguished (Barthlott and Froelich, 1983):

1. Smooth wax layers in the form of thin films, common in angiosperms.

2. Non-oriented wax crystalloids in the form of rodlets or platelets with no regular pattern. These are common in dicots and Lilianae groups of monocots.

3. *Strelitzia* wax-type with massive compound wax projections composed of rodlet-like subunits that form massive compound plates around the stomata. This wax type is

found in Zingiberanae, Commelinanae, and Arecanae. It is also found in Velloziales, Bromeliales, and Typhales, which further differ from other Lilianae in a starchy endosperm.

4. *Convallaria* wax-type with small wax platelets arranged in parallel rows, which cross the stomata at right angle and form a close circle around each polar end of the stoma, similar to the lines of an electromagnetic field. This type is restricted to Lilianae only.

Tertiary sculpture is generally lacking from seeds. In Orchidaceae, however, certain tribes possess epicuticular waxes on their seed-coats. Seed-coats of *Jacaranda* (Figure 6-E) have stellate epicuticular sculpture known as 'star scales', a feature characteristic of this genus. Many members of Aizoaceae are characterized by seed-coat with epicuticular secretions forming long upright rodlets and small rodlets lying on the cell surface.

Cactaceae is a huge family commonly divided into three subfamilies, of which Cactoideae includes 90 percent species but its classification is difficult because of uniform floral characters, pollen morphology and plasticity. Barthlott and Voit (1979) analyzed 1050 species and 230 genera by SEM for seed coat structure in the family Cactaceae. The simple unspecialized testa of Pereskoideae supports its ancestral position. Opuntioideae has a unique seed with a hard aril, thus confirming its isolated position, also indicated by pollen morphology. Cactoideae shows complex diversity, confirming its advanced position and subtribes have been recognized based on seed-coat structure, each subtribe possessing distinctive features. Thus, the genus *Astrophytum* has been transferred from Notocactinae to Cactinae.

Orchidaceae is another large family with complicated phylogenetic affinities. Minute 'dust seeds' show microstructural diversity of the seed-coat. Studies of over 1000 species (Barthlott, 1981) have helped in better subdivision into subfamilies and tribes. Barthlott also supports the merger of Cypripediaceae with Orchidaceae, a suggestion incorporated in several recent classification (Judd et al., 2002; APG II, 2003; Thorne, 2003; Stevens, 2003).

Ultrastructure

Ultrastructure studies of angiosperms have provided valuable taxonomic information from phloem tissue, mainly sieve tube elements. Besides this, information has also come from studies of seeds.

Sieve-tube plastids

Studies on sieve-tube were first initiated by Behnke (1965) in the family Dioscoreaceae. Since then, nearly all angiosperm families have been investigated for the taxonomic significance of these plastids. All sieve-element plastids contain starch grains differing in number, size and shape. The protein accumulates in specific plastids in the form of crystalloids and filaments.

Thus two types of plastids are distinguished: **P-type** which accumulate proteins and **S-type** which do not accumulate proteins. Starch accumulation is of no primary importance in classification, since it may be present or absent in both types of plastids. P-type plastids are further divided into six subtypes (Behnke and Barthlott, 1983):

(**i**) PI-subtype. The plastids contain single crystalloids of different sizes and shapes and/or irregularly arranged filaments. This subtype is thought to be the most primitive in flowering plants, mainly Magnoliales, Laurales and Aristolochiales.

(**ii**) PII-subtype. This subtype contains several cuneate crystalloids oriented towards the center of the plastid. All investigated monocots contain this subtype. It is significant to note that only members of dicots with this subtype, *Asarum*, and *Saruma* of Aristolochiaceae are widely regarded among the most primitive members of dicots, a possible link between monocots and dicots.

(**iii**) PIII-subtype. This subtype contains a ring-shaped bundle of filaments. PIII subtype is confined to Centrospermae (Caryophyllales) and the removal of Bataceae and Gyrostemonaceae has been supported by the absence of this subtype in these families. Further, forms are recognized based on the presence or absence of crystalloids (Figure 7) into: PIIIa (globular crystalloid), PIIIb (hexagonal crystalloid) and PIIIc

(without crystalloid). Based on the distribution of these forms, Behnke (1976) proposed division of the order into three family-groups which exactly correspond to the three suborders Caryophyllineae, Chenopodineae and Phytolaccineae, earlier established by Friedrich (1956). Whereas Takhtajan had recognized these three suborders in his 1983 revision, in his final revision of his classification, he merged Phytolaccineae with Caryophyllineae, thus recognizing only two suborders Caryophyllineae and Chenopodineae. Of the three orders recognized in Caryophyllidae of Takhtajan, only Caryophyllales contains PIII-subtype plastids while the other two orders, Polygonales and Plumbaginales, contain S-type plastids. Behnke (1977) as such, advocated retention of only Caryophyllales under Caryophyllidae and removal of the other two orders to subclass Rosidae whose members also contain S-type plastids. The suggestion was not accepted by Takhtajan (1987) and Cronquist (1988), who retained all the three orders under Caryophyllidae. Takhtajan, however, placed the three orders under separate superorders (Cronquist does not recognize superorders).

(**iv**) PIV-subtype. The plastid contains a few polygonal crystalloids of variable size. This subtype is restricted to the order Fabales.

(**v**) PV-subtype. The plastid contains many crystalloids of different sizes and shapes. This subtype is found in the order Ericales and family Rhizophoraceae.

(**vi**) PVI-subtype. The plastid contains a single circular crystalloid. This subtype is found in family Buxaceae.

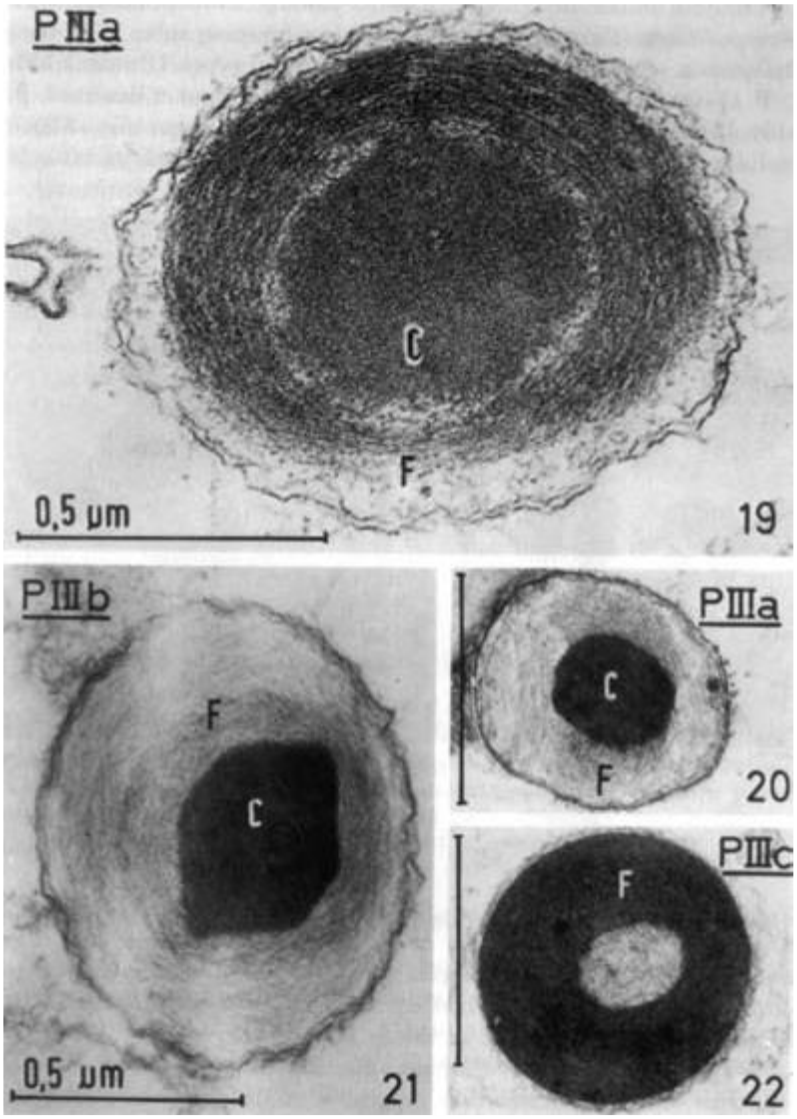


Figure (7): Different forms of PIII–subtype sieve-element plastids all with ring–like bundle of filaments (F). **19** and **20:** PIIIa with globular crystalloid (C); **21:** P IIIb with polygonal crystalloid; **22:** PIIIc without crystalloid.

Dilated Cisternae

Dilated Cisternae (DC) were first described by Bonnett and Newcomb (1965) as dilated sections of endoplasmic reticulum in the root cells of *Raphanus sativus*. Originally found in Brassicaceae and Capparaceae, DC have now been found in several other families of angiosperms but are concentrated in the order Capparales (Brassicaceae and Capparaceae) and form a part of the character syndrome of this order. The DC may be utricular, irregular or vacuole-like in form with filamentous, tubular or granular contents. They have been proposed to be functionally associated with glucosinolates and myrosin cells found in this order.

Phloem (p-) proteins

P-proteins are found only in sieve elements of angiosperms and occur in the form of filaments or tubules. These assemble into large discrete bodies, and are not dissolved during maturation of the sieve-element, unlike single membrane organelles. The composition and three-dimensional arrangement of these proteins exhibit taxonomic specificity. They are dispersed over the entire cell as the cell matures but, in some dicots, a single non-dispersive (crystalline) body of various shapes may be found in addition to dispersive ones. Crystalline bodies are absent in monocots. Their shape is often specific and thus of taxonomic importance. Globular crystalline bodies are found in Malvales and Urticales. Fabanae, which is characterized by PIV-subtype plastids, has spindle-shaped crystalline bodies in the family Fabaceae. The feature has supported the transfer of *Swartzia* to Fabaceae.

Nuclear inclusions

Nuclear inclusions in the form of protein crystals occur in phloem- and Ray parenchyma, primarily in the families of Asteridae. Structural differences are significant for classification within Scrophulariaceae and Lamiaceae. Protein crystals in sieve-tube elements have also been reported in Boraginaceae, and may prove useful.

Non-phloematic TEM Characters

Protein bodies in seeds through TEM, SEM and dispersive X-ray techniques have demonstrated their significance if qualitative and quantitative aspects are both taken into account. Similarly, SEM studies of starch grains are also useful sources of information of taxonomic significance.

6- CHROMOSOMES

Chromosomes are the carriers of genetic information and as such have a considerable significance in evolutionary studies. Increased knowledge about chromosomes and their behaviour has largely been responsible for extensive biosystematic studies and development of the biological species concept. During the first quarter of the twentieth century, chromosomal data were relatively sparse. Such information has markedly increased over the last few decades, however, with ample useful information coming from studies of the banding pattern. Three types of chromosomal information have been of significance in Systematics.

Chromosomal number

Extensive records of chromosome numbers are available in the works of Darlington and Janaki-Amal (1945), Darlington and Wylie (1955), Federov (1969) and Löve et al. (1977). The International Association of Plant Taxonomy (IAPT) has also been publishing an Index to Plant Chromosome Numbers in its series Regnum Vegetabile. Between 1967 and 1977, the series published 9 volumes mostly forming annual lists of chromosome numbers. An updated server of the Missouri Botanical Garden maintains the records of chromosome numbers. The chromosome counts are usually reported as diploid number (2n) from mitosis of sporophytic tissue but when based on mitosis in gametophytic tissue or on meiosis studies, counts are reported as haploid (n). The gametophytic chromosome number of diploid species is designated as base-number (x). In diploid species as such n = x, whereas in polyploid species n is in

multiples of x. A hexaploid species with 2n = 42 will thus have n = 21, n = 3x and 2n = 6x.

The chromosome number in angiosperms exhibits considerable variation. The lowest number (n = 2) is recorded in *Haplopappus gracilis* (Asteraceae) and the highest (n = 132) in *Poa littoroa* (Poaceae). The alga *Spirogyra cylindrica* also contains n = 2, whereas the record of the highest chromosome number (n = 630) is found in *Ophioglossum reticulatum* (Pteridophytes). Such a range of variation, (n = 2 to n = 132), however, within nearly a quarter a million species of angiosperms, may not be very significant in taxonomic delimitation, but there have been instances of the isolated role of studies on chromosomes. Raven (1975) provided a review of chromosome numbers at the family level in angiosperms. He concluded that the original base-number for angiosperms is x = 7 and that comparisons at the family level are valid only when the base-number (and not n or 2n) is used. The family Ranunculaceae is dominated by genera with large chromosomes (and x = 8). The two genera *Thalictrum* and *Aquilegia*—originally placed in two separate subfamilies or tribes (and even two separate families Ranunculaceae and Helleboraceae by Hutchinson, 1959, 1973 along with other achene bearing and follicle bearing genera, respectively) are distinct in having small chromosomes (and x = 7) and as such have been segregated into a distinct tribe. The genus *Paeonia* with very large chromosomes (and x = 5) has been separated into a distantly related family Paeoniaceae, a placement which has been supported by morphological, anatomical and embryological data. Significant records in other families include Rosaceae with x = 17 in subfamily Pomoideae, whereas other subfamilies have x = 7, 8 or 9. In Poaceae similarly subfamily Bambusoideae has x = 12, whereas Pooideae has x = 7.

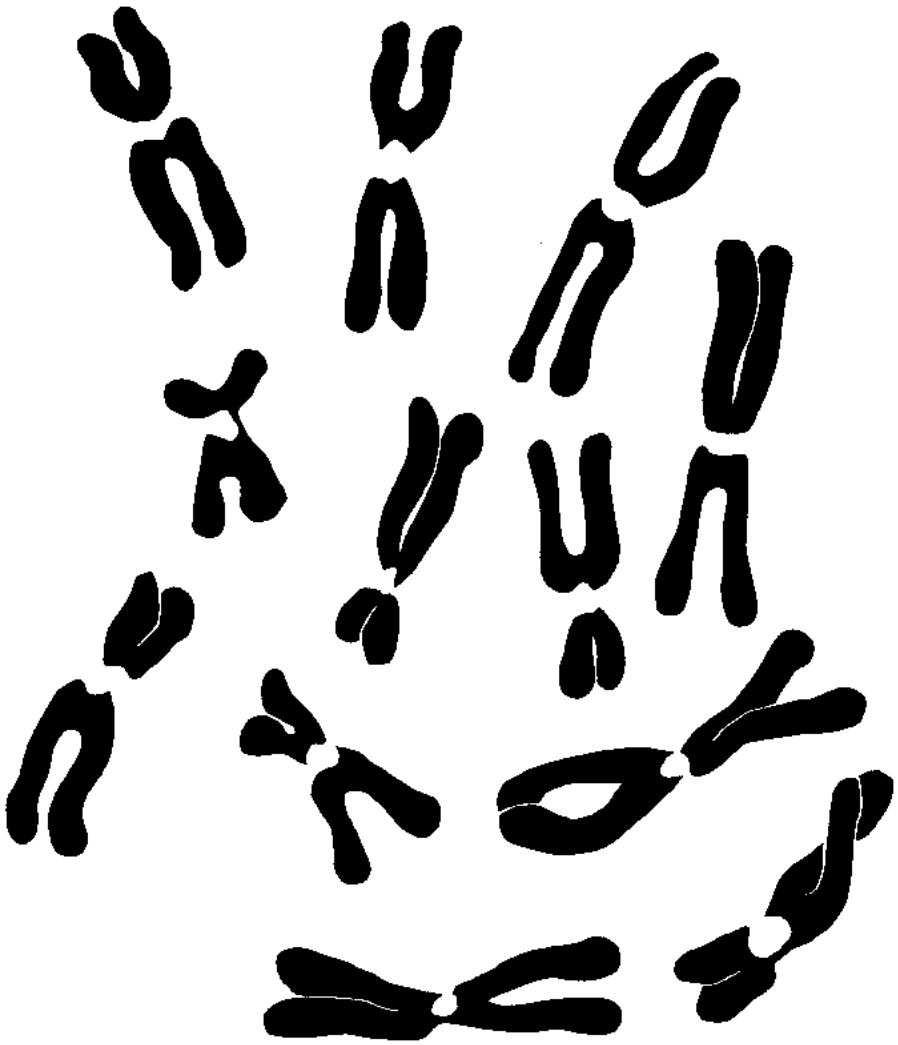


Figure (8): Mitotic chromosomes of *Tradescantia spathacea* (2n=12) with sister chromatids and centromere.

Spartina was for long time placed in the tribe Chlorideae (x = 10) although its chromosomes (x= 7) were at variance. Marchant (1968) showed the genus to have, in fact x = 10, thus securing placement within Chlorideae.

The classical study of the genus *Crepis* (Babcock, 1947) based separation from the closely related genera on chromosomal number and morphology. This led to the separation of the genus *Youngia* and merger of *Pterotheca* with *Crepis*. Similarly, in the genus *Mentha* which has small, structurally uniform chromosomes, the chromosome numbers provide strong support for subdivision into sections Audibertia (x = 9), Pulegium (x = 10), Preslia (x = 18) and Mentha (x = 12).

The duplication of chromosome numbers leading to polyploidy may prove to be of taxonomic significance. The grass genus *Vulpia* contains diploid (2n = 14), tetraploid (2n =28) and hexaploid (2n = 42) species. The genus is divided into five sections, of which three contain only diploids, one diploids and tetraploids and one all three levels of ploidy. It is presumed that tetraploid and hexaploidy species of *Vulpia* arose from diploid progenitors. The duplication of chromosome number of a diploid species may form a tetraploid (autopolyploid). Such a polyploid, however, does not show any or at most may show minor differences from the diploid species, and is rarely recognized as an independent taxonomic entity. The hybrid between two diploid species contains one genome from either parent and thus, generally doesn't survive because of failure of chromosomal pairing during meiosis. Hybridization followed by duplication of chromosomes establishes a tetraploid (allopolyploid; amphiploid) with normal pairing as both genomes are in pairs. Such a tetraploid hybrid with distinct characteristics may be recognized as an independent species.

A triploid hybrid between a diploid species and a tetraploid species may, similarly, not survive as genome from the diploid parent would exhibit the problem of pairing at meiosis but the hybridization followed by duplication leading to hexaploidy can form a perfectly normal independent species. Such facts have led to the detection of hybrids or confirmation of suspected hybrids. *Senecio* (Asteraceae) includes the diploid *S. squalidus* (2n = 20), the tetraploid *S. vulgaris* (2n = 40) and the hexaploid *S. cambrensis* (2n = 60). The last is intermediate in morphology between the first two and is found in the area where this two grow. Additionally, sterile triploid hybrids between two species have been reported. It seems clear that *S. cambrensis* is an allohexaploid between the other two species (Stace, 1989). Similarly, based on chromosome number and karyotype, Owenby (1950) concluded that *Tragopogon mirus* (2n = 24), a tetraploid species arose as an amphiploid between two diploid species, *T. dubius* and *T. porrifolius* (2n = 12).

Whereas a species generally shows a single chromosome number, certain populations or infraspecific taxa (subspecies, variety, forma) may sometimes show a different chromosome number (or even different chromosomal morphology). Such populations or infraspecific taxa constitute cytotypes.

Chromosomal structure

Chromosomes show considerable variation in size, position of centromere (Figure 8) and presence of secondary constriction. The chromosomes are commonly differentiated as metacentric (with centromere in middle), submetacentric (away from middle), acrocentric (near the end) or telocentric (at the end). The chromosomes are also characterized by their size. In addition, the occurrence and position of secondary constriction, which demarcates a satellite is important in chromosomal identification and characterization. The identification of satellites is often difficult, and especially when the secondary constriction is very long, a satellite may be counted as a distinct chromosome. This situation has often led to erroneous chromosome counts. The structure of the chromosome set (genome) in a species is termed karyotype and is commonly diagrammatically represented in the form of an ideogram (Figure 9) or karyogram. An analysis of a large number of studies has led to the conclusion that a symmetrical karyotype (chromosomes essentially similar and mainly metacentric) is primitive and an asymmetric karyotype (different types of chromosomes in a genome) advanced, the latter commonly found in plants with specialized morphological features, such as *Delphinium* and *Aconitum*.

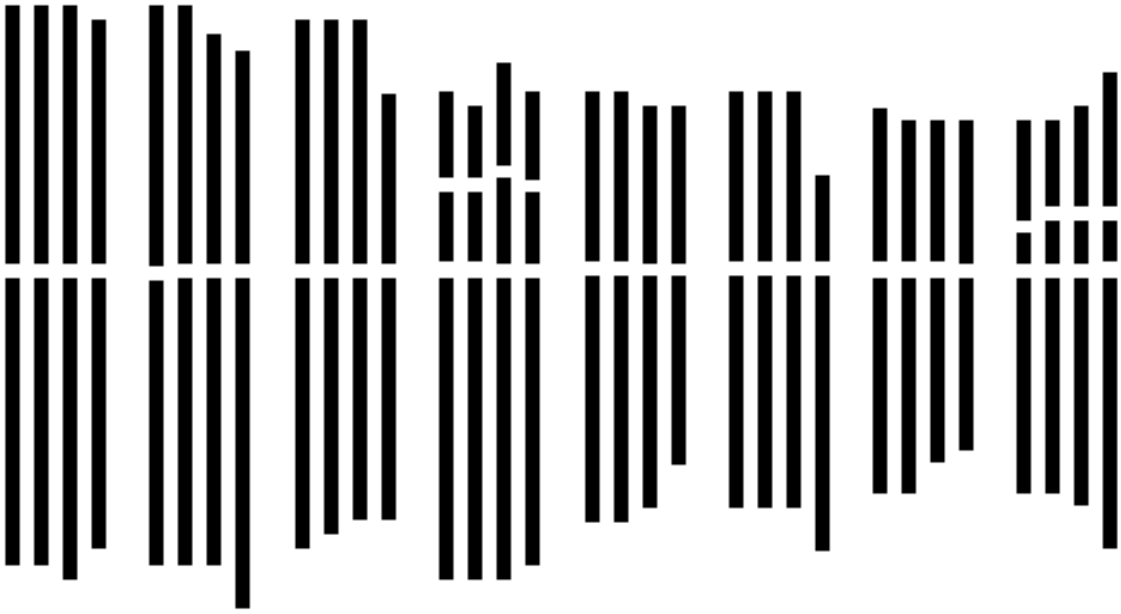


Figure (9): Ideogram of the somatic complement of *Allium ampeloprasm*. Of the 32 somatic chromosomes, 8 show secondary constriction

An interesting example of utilization of chromosomal information is family Agavaceae. The family contains about 16 genera such as *Agave* (and others formerly placed in Amaryllidaceae due to inferior ovary) and *Yucca* (and others formerly placed in Liliaceae due to superior ovary). These genera were shifted and brought into Agavaceae on the basis of great overall similarity. This was supported by the distinctive bimodal karyotype of Agavaceae consisting of 5 large chromosomes and 25 small ones. Rudall et al. (1997) advocated the transfer of *Hosta* (placed in Hostaceae; Hesperocallidaceae by Thorne, 1999), *Camassia* and *Chlorogalum* (both placed under Liliaceae by Hutchinson, 1973; Hyacinthaceae by Thorne, 1999) to family Agavaceae on the basis of possession of bimodal karyotype. Rousi (1973), from his studies on the genus *Leontodon*, showed that data on the basic number, chromosome length, centromeric position and the occurrence of satellites provide evidence for the relegation of the genus *Thrincia* (x = 4) as a section of subgenus Apargia along with section *Asterothrix* (x = 4, 7). The subgenus Leontodon is distinct with x = 6 or 7, and a different chromosome morphology.

Cyperaceae and Juncaceae were earlier placed far apart due to distinct floral structure. Both families have small chromosomes without distinctive centromeres, the latter may be diffuse or non-localized. These families as such are now considered to be closely related. Such chromosomes (holocentric chromosomes) do not depend on a discrete centromere for meiosis and mitosis and may undergo fragmentation with no deleterious effect. This may result in variable chromosomal counts. In the *Luzula spicata* group, chromosomal counts are reported to be 2n = 12, 14, and 24. Interestingly, the total chromosomal volume is the same and the higher chromosome number is the result of fragmentation (agmatoploidy) of these holocentric chromosomes. Different chromosome numbers may often occur in different cells of the same root-tip (mixoploidy). The occurrence of accessory chromosomes (known as B-chromosomes) in higher plants generally does not have a significant effect on morphology and, thus, is of little taxonomic importance. B-chromosomes in bryophytes, contrarily, are very small (termed m-chromosomes) and often highly diagnostic. In recent years, considerable breakthrough has been achieved in the study of banding patterns of chromosomes using Giemsa and fluorochrome stains. Already different techniques such as C-banding, G-banding, Q-banding and Hy-banding are in use, and help in clearly distinguishing the heterochromatic and euchromatic regions. C-banding is very useful in indicating the position of centromeres in cases where they cannot be identified by conventional staining.

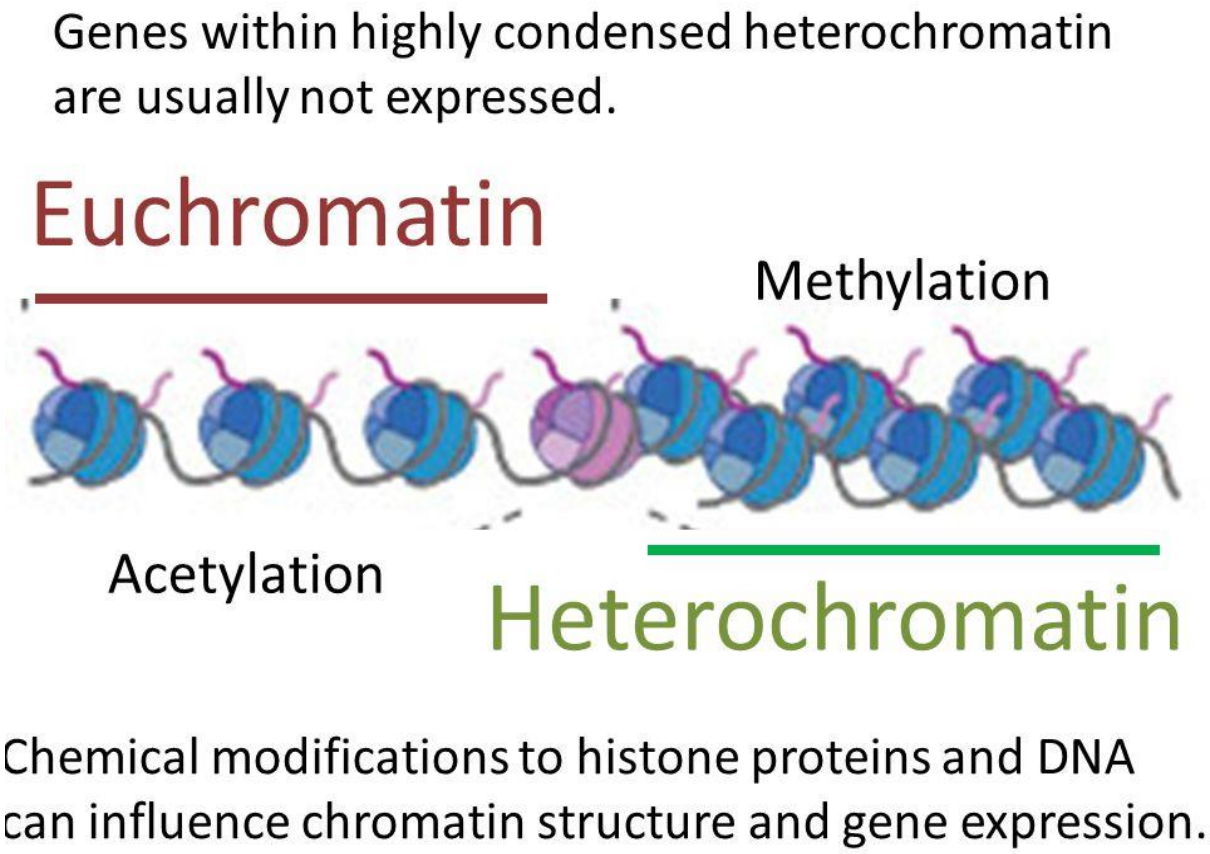


Figure (10): heterochromatic and euchromatic regions.

The technique of silver-staining has been developed to highlight NOR (nucleolar organizing region). An interesting study of the chromosomes of top onion (variously recognized as *Allium cepa* var. *viviparum* or *A. fistulosum* var. *proliferum*) as also those of *A. Cepa* and *A. fistulosum* was done by Schubert, Ohle and Hanelt (1983). By Giemsa banding pattern and silver-staining studies, they concluded that some chromosomes of top onion resemble *A. cepa* and others resemble *A. fistulosum*. Of the two satellites, one resembles either species. Top onion is as such a pseudodiploid with no homologous pair. The study confirmed that top onion is a hybrid between the two aforesaid parents, and thus would be better known as A. x proliferum (Moench) Schrad., and not as a variety of either species. Interestingly, the top onion owes its existence to the bulbils, which are produced in place of an inflorescence and ensure the multiplication of the hybrid, which is otherwise sterile.

Chromosomal behaviour

The fertility of a plant is highly dependent on the ability of meiotic chromosomes to pair (synapsis) and their subsequent separation. The meiotic behaviour of chromosomes enables comparison between genomes to detect the degree of homology, especially when they are a result of hybridization. A greater degree of genomic nonhomology results in either failure of pairing (asynapsis) or a loose pairing of chromosomes without chiasmata so that chromosomes fall apart before metaphase (desynapsis). In extreme cases, the entire genome may fail to pair. The genome analysis of suspected hybrids has helped in establishing the parentage of several polyploid species.

A diploid hybrid between two species generally exhibits failure of meiotic pairing due to non-homology of genomes resulting in hybrid sterility, but when hybridization is followed by duplication of chromosomes to form a tetraploid hybrid, the latter shows normal pairing between the two genomes derived from the same parent and is generally fertile. A triploid hybrid may, similarly, be sterile but a hexaploid one fertile. Genome analysis has confirmed that the hexaploid *Senecio cambrensis* is allohexaploid between tetraploid *S. vulgaris* and diploid *S. squalidus.* Similarly, the tetraploid *Tragopogon mirus* is the result of hybridization between the two diploid species *T. dubius* and *T. porrifolius*. The most significant case, however, is the common bread wheat *Triticum aestivum*, a hexaploid with AABBDD genome. Genome analyses have confirmed that genome A is derived from the diploid *T. monococcum*, B from *Aegilops speltoides*, both genomes being represented in the tetraploid *T. dicoccum*. Genome D is derived from the diploid *Aegilops tauschii*.

7- 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. Chemical information about medicinal plants in herbals published nearly five centuries back was concerned with localization and application of physiologically active secondary metabolites such as saponins and alkaloids. Knowledge about chemistry of plants greatly increased during the eighteenth and nineteenth centuries. The greatest interest has been generated over the last 40 years, however, with the development of improved techniques for studying biological molecules, especially proteins and nucleic acids. In recent years, interest has focused on the study of allelochemy and realization of the concept that the animal kingdom and the plant kingdom have experienced a chemical coevolution. Plants continuously evolve new defensive chemical mechanisms to save themselves from predators, and animals evolve methods to overcome these defenses. In the process, some plant species have developed animal hormones, thus disturbing the hormonal levels of animals if ingested.

A large variety of chemical compounds are found in plants and quite often the biosynthetic pathways producing these compounds differ in various plant groups. In many instances the biosynthetic pathways correspond well with existing schemes of classification based on morphology. In other cases, the results are at variance, thus calling for revision of such schemes. The natural chemical constituents are conveniently divided as under:

- Micromolecules: Compounds with low molecular weight (less than 1000).

Primary metabolites: Compounds involved in vital metabolic pathways: citric acid, aconitic acid, protein amino acids, etc.

Secondary metabolites: Compounds which are the by-products of metabolism and often perform non-vital functions: non-protein amino acids, phenolic compounds, alkaloids, glucosinolates, terpenes, etc.

- Macromolecules: Compounds with high molecular weight (1000 or more).

Non-semantide macromolecules: Compounds not involved in information transfer: starches, celluloses, etc.

Semantides: Information carrying molecules: DNA, RNA and proteins. The utilization of studies on DNA and RNA for understanding of phylogenetic relations has received a great boost over the last decade, meriting the establishment of a new field referred to as Molecular Systematics.

Primary metabolites

Primary metabolites include compounds, which are involved in vital metabolic pathways. Most of them are universal in plants and of little taxonomic importance. Aconitic acid and citric acid, first discovered from *Aconitum* and *Citrus* respectively, participate in Krebs cycle of respiration and are found in all aerobic organisms. The same is true of the 22 or so amino acids forming proteins,

and the sugar molecules, which are involved in the Kalvin cycle of photosynthesis. The quantitative variations of these primary metabolites may, however, be of taxonomic significance sometimes. In *Gilgiochloa indurata* (Poaceae), alanine is the main amino acid in leaf extracts, proline in seed extracts and asparagine in flower extracts. Rosaceae is similarly rich in arginine.

Secondary metabolites

Secondary metabolites perform non-vital functions and are less widespread in plants as compared to primary metabolites. These are generally the by-products of metabolism. They were earlier considered to be waste products, having no important role. Recently, however, it was realized that they are important in chemical defense

against predators, pathogens, allelopathic agents and also help in pollination and dispersal (Swain, 1977). Gershenzon and Mabry (1983) have provided a comprehensive review of the significance of secondary metabolites in higher classification of angiosperms. The following major categories of secondary metabolites are of taxonomic significance:

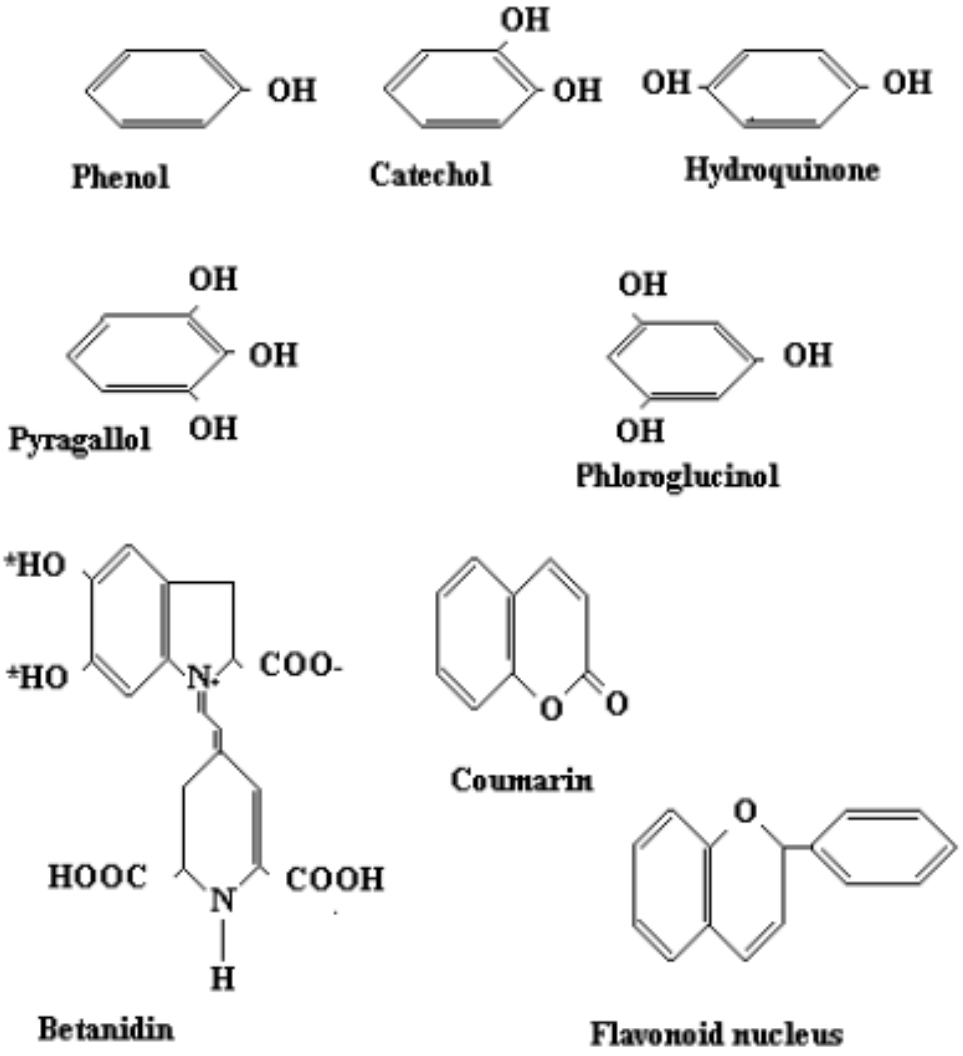


Figure (11): Structure of important phenolic molecules and a betalain (Betanidin). (*indicates the position of sugar.)

Non-protein amino acids

A large number of amino acids not associated with proteins are known (more than 300 or so). Their distribution is not universal but specific to certain groups and, as such, holds promise for taxonomic significance. **Lathyrine** is, thus, known only from *Lathyrus*. **Canavanine** occurs only in Fabaceae and is shown (Bell, 1971) to be a protection against insect larvae. These amino acids are usually concentrated in storage roots and, as such, root extracts are generally used for their study.

Phenolics

Phenolic compounds form a loose class of compounds, based upon a phenol ($C_6 H_5 OH$). **(Simple phenolics** are made of a single ring and differ in position and number of OH groups. These are water soluble and are localized in vacuoles in cell, and are found in combination with sugars as glycoside. (Simple phenolics can be tested by extraction with ethanol. Take 5-6 gm of chopped leaf tissue in a beaker and add 30 ml of 70% ethanol; heat over water bath at 60-70 degree centigrade for 20 minutes; filter and concentrate filtrate over water bath till about 0.5 ml is left; load the sample on Whatman paper (No. 1) using BAW: Butanol, Acetic acid and water in ratio of 4:1:5; run chromatogram, dry and observe under UV light; spray with mixture (1:1) of 1% Ferric chloride and 1% Potassium ferricyanide and calculate Rf value**)**. These are widely distributed in the plant kingdom; common examples being catechol, hydroquinone, phloroglucinol and pyragallol. Coumarins, a group of natural phenolics, have a characteristic smell. The crushed leaves of *Anthoxanthum odoratum* can thus be identified by this characteristic odour. More than 300 coumarins have been reported from nearly 80 families of plants. They are a group of lactones formed by ring closure of hydroxycinnamic acid.

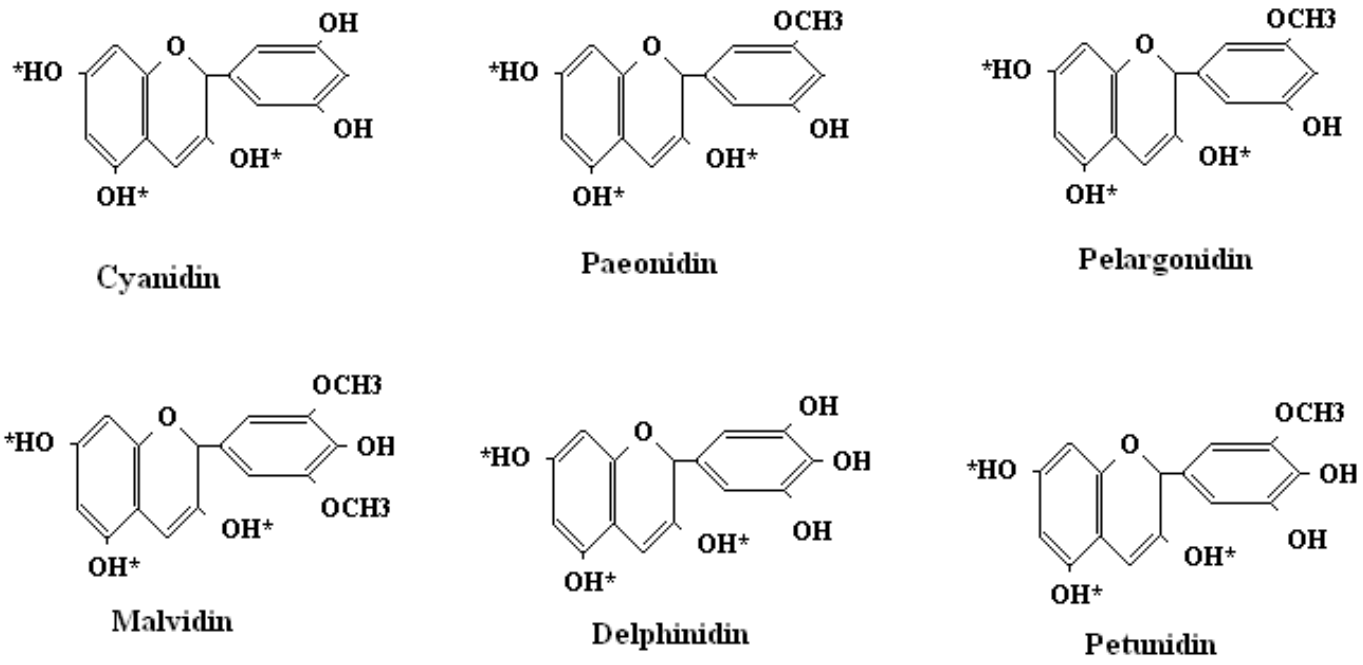


Figure (12): Structure of anthocyanin forming molecules, differing in right three positions, middle position absent in Cyanidin, Paeonidin and Pelargonidin, but having OH in Malvidin, Delphinidin and Petunidin. Cyanidin has OH at both other positions, Paeonidin has one replaced by OCH3, Pelargonin upper is missing, Malvidin has both replaced by OCH3, Delphinidin has OH at both and Petunidin one having OH and other OCH3.

Lignin is a highly branched polymer of three simple phenolic alcohols. Whereas gymnosperm lignin is composed of coniferyl alcohol subunits, the angiosperm lignin is a mixture of coniferyl and sinapyl alcohol subunits. The alcohols are oxidized to free radicals by peroxidase enzyme, and the free radicals react to form lignin.

Flavonoids, the more extensively studied compounds, are based on a flavonoid nucleus consisting of two benzene rings joined by a C3 open or closed structure (Figure 11). **(**Presence of **flavonoids** can be detected as follows: Finely chop 5 gm of flower petals or tepals in beaker; add 20 ml of 2N HCl, cover with aluminium foil, and heat at 80-90 degree centigrade for 30-40 minutes in water bath; filter and extract filtrate with 15-20 ml ethyl acetate in separating funnel, shake and allow solvent to evaporate; two layers are formed, upper organic and lower inorganic aqueous layer (mostly

anthocyanins); collect them in two separate beakers; label beaker with upper organic layer as B; if aqueous layer is coloured heat for 5-10 minutes to expel ethyl acetate; put back in separating funnel, add 2-4 ml of amyl alcohol, shake, transfer upper organic layer to beaker and mark it as A; heat both beakers to dryness on water bath uncovered; to each add 1% methanolic HCl; load each sample on two circular whatman (No. 1) filter paper, load spot in center using Forestall solvent: HCl, Acetic acid and water (3:10:30); mark spots visually; observe one set of chromatograms of A and B under UV; expose second set to ammonia vapours by rotating discs over open mouth of NH3 bottle and observe under UV. **Phenyl propanoids** can be detected similarly except that extraction is done using diethyl ether instead of ethyl estate, only organic layer is retained and the solvent use for chromatography is BAW (Butanol: Acetic acid: Distilled water-4:1:5**)**. Common examples are flavonols (mainly colourless and commonly occurring as co-pigments, yield bright yellow spot in chromatogram after acid hydrolysis), flavones (similar, yield dull brown spots), glycoflavones, biflavonyls (similar, yield dull absorbing spots on BAW), isoflavones (colourless, often found in roots of legumes), flavanones (colourless, occur in leaves, *Citrus* fruits, yield red colour with HCl), chalcones or aurones (usually occur in yellow flowers, yield red colour with NH3), Anthocyanins (red, blue coloured water soluble) and leucoanthocyanins (mainly colourless, mainly in heartwood and leaves of trees, yield anthocyanins).

Anthocyanins and **Anthoxanthins** are important pigments in the cell sap of petals providing red, blue (anthocyanins), and yellow (anthoxanthins) colours in a large number of families of angiosperms. They are formed by anthocyanadins combining with different sugars at different places. Six main categories (Figure 12) of anthocyanin forming molecules are recognized providing different colours: Cyanidin-magenta; Pelargonidin-orange-red; Delphinidin-purple, blue, mauve; Petunidin-purplish; Paeonidin-magenta and Malvidin-purple.

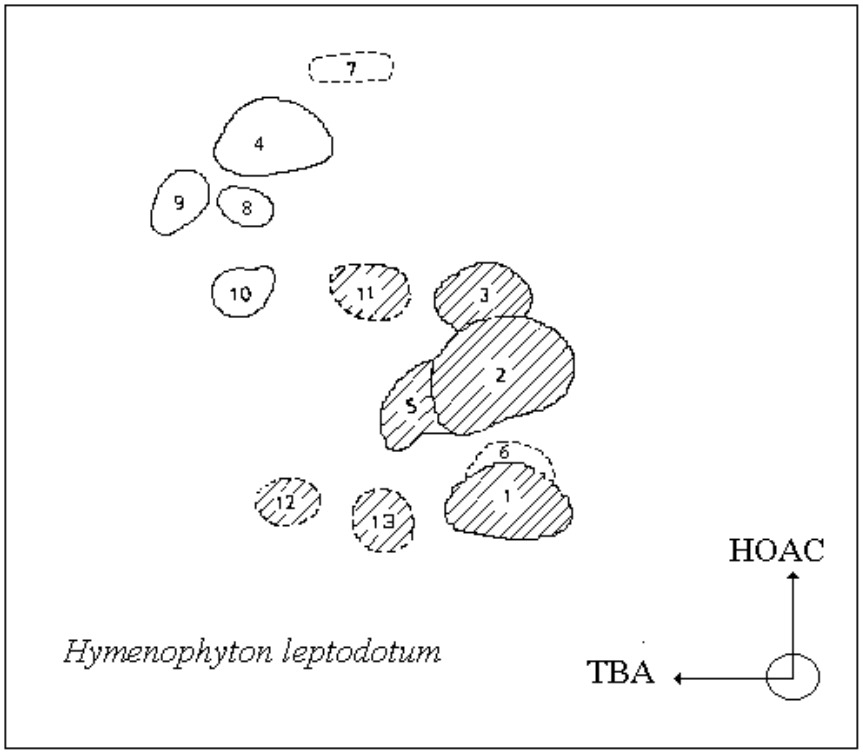
These pigments are absent in some families and replaced by highly different compounds, **betacyanins** and **betaxanthins** (together known as **betalains**), which

consist of heterocyclic nitrogen-containing rings and having quite distinct metabolic pathways of synthesis. However, they carry the same functions as anthocyanins. Betalains are mutually exclusive with anthocyanins, and concentrated in the traditional group Centrospermae of Engler and Prantl, now recognized as order Caryophyllales. Of the nine families which contain betalains, seven were included in Centrospermae, Cactaceae placed in Cactales or Opuntiales and the ninth was placed in Sapindales. Traditional Centrospermae also included Gyrostemonaceae, Caryophyllaceae and Molluginaceae which lack betalains and contain anthocyanins instead. Mabry et al., (1963) on the basis of separate structure and metabolic pathways, suggested the placement of only betalain-containing families in Centrospermae, thus advocating the inclusion of Cactaceae and Didiereaceae and exclusion of Gyrostemonaceae, Caryophyllaceae and Molluginaceae. Whereas the inclusion of Cactaceae and Didiereaceae was readily accepted (thus bringing all betalain containing families in the same order Centrospermae). Behnke and Turner (1971), on the basis of ultrastructure studies, reported P-III plastids in all members of Centrospermae and thus suggested a compromise by including all families within subclass Caryophyllidae with betalain-containing families placed under the order Chenopodiales and the other two (Caryophyllaceae and Molluginaceae) placed under Caryophyllales. Interestingly, Mabry (1976), on the basis of DNA/RNA hybridization studies, found closer affinities among these families and suggested the placement of all these families under Caryophyllales with the betalain-containing families under the suborder Chenopodiineae and the two nonbetalain families under Caryophyllineae. This final compromise has met with mixed response in recent years with the morphological, anatomical and DNA/RNA hybridization evidence.

It is interesting to note that the betalains have also been reported in Basidiomycetes (Fungi), in some cases the same substance found in both fungi and angiosperms. The above studies on the significance of distribution of betalains in Centrospermae bring home the fact that chemical data are useful in taxonomic realignments when such accord with data from other fields. The significance is reduced

when larger evidence from elsewhere contradicts the chemical evidence. Thus, whereas no questions were ever asked about the removal of Gyrostemonaceae and the inclusion of Cactaceae and Didiereaceae, there has been no agreement about the removal of Caryophyllaceae and Molluginaceae as it goes against the evidence from morphology, anatomy, ultrastructure and DNA/RNA hybridization. This also highlights the danger of relying too much on one type of evidence.

Studies on phenolic compounds have helped in solving some specific problems. Bate-Smith (1958) studied five phenolic characters of different sections in the genus *Iris*. The chemical evidence supported the division into various sections, but *I. flavissima*, originally placed in the section Pogoniris resembled species of the section Regelia on the basis of phenolic characteristics. Chromosomal evidence also supported this transfer. The technique of two-directional paper chromatography, which brings about a more pronounced separation of flavonoids, has proved very useful in taxonomic studies. *Hymenophyton* (Bryophytes) was considered by some researchers to be a monotypic genus, but by others to include two species. Markham et al., (1976) on the basis of rapid flavonoid extraction, two-dimensional chromatographic analysis and identification (Figure 13) concluded that the genus contains two distinct species, *H. leptodotum* and *H. flabellatum*, and that there is no justification for their merger.



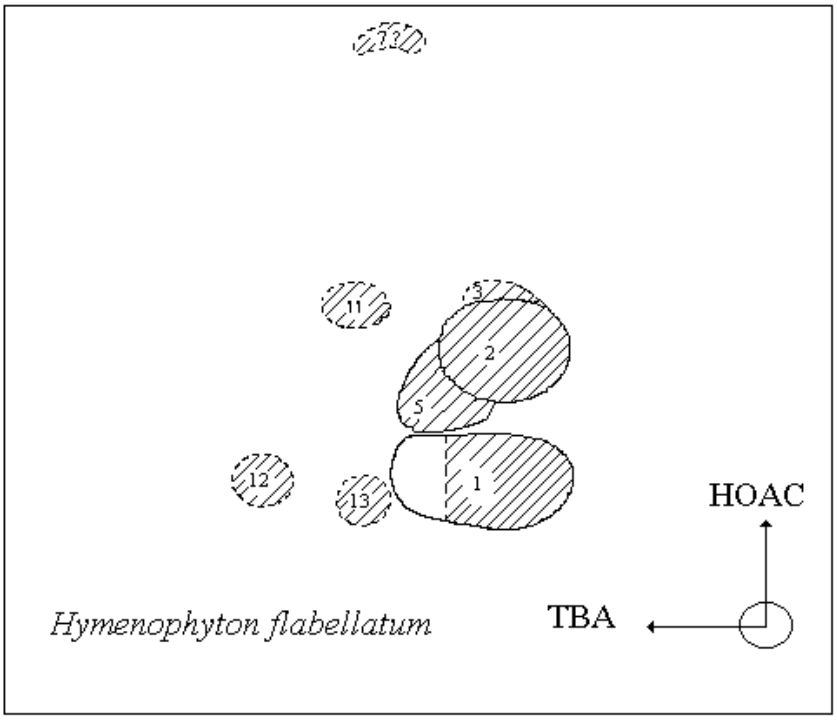


Figure (13): Two-dimensional paper chromatograms of the flavonoids in two species of *Hymenophyton*

Similar studies in the genus *Baptisia* (Fabaceae) by Alston and Turner (1963) have been very useful in the detection of hybridization. Each species of the genus has a distinctive spectrum of flavonoids, and the hybrid can be easily identified by the combination of flavonoid pattern of both parental species in the suspected hybrid.

It is interesting to note that the ten taxa recognized (four parental and six hybrid), could not be differentiated on the basis of morphological or biochemical characters alone, but a combination of both enabled a complete separation. The flavonoids in these studies were extracted from either flower or leaf.

Alkaloids

Alkaloids are organic nitrogen-containing bases, usually with a heterocyclic ring of some kind. They form one of the largest classes of secondary metabolites, with nearly 10,000 different types reported. They are insoluble in water but soluble in organic solvents, but their salts are soluble in water and insoluble in organic solvents. Their distribution is restricted to some 20% of angiosperms. They are mostly present in storage tissues, seeds, fruits and roots. They act as chemical defence of plants against herbivory, and allelopathic reactions among plants.

Alkaloids are generally classified on the basis of predominant ring system present in the molecule. They are synthesized from a few common amino acids like tyrosine, tryptophan, ornithine, argenine and lysine. Tobacco alkaloid Nicotine (*Nicotiana*) is synthesized from nicotinic acid and caffeine (coffee beans and tea leaves) from purine. Isoquinolene alkaloids: morphine, codeine and papaverine are found in opium poppy (*Papaver somniferum*). Their distribution is often specific and thus taxonomically significant (Figure 14). Conalium is the simplest known alkaloid found in *Conium maculatum* (Apiaceae). Alkaloids are present in specialized parts of plant. Higher nicotine content is found in only older leaves. In *Datura*, alkaloids occur only in seeds. Alkaloids are more widely distributed in dicots as compared to monocots. Some of them are of medicinal importance at low concentration, but toxic at high concentration. Some such as lycotonine (*Delphinium*), scopolamine (*Datura*),

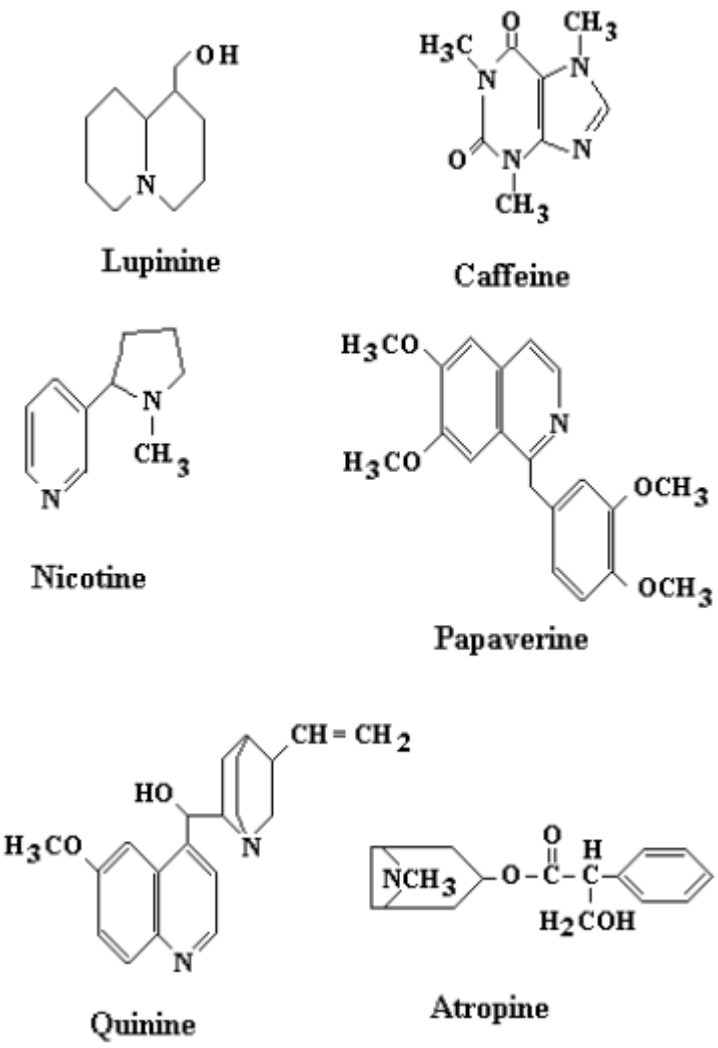


Figure (14): Main Examples of alkaloids found in the plant kingdom. The distribution of some is highly specific

and atropine (*Atropa*) cause poisoning of livestock. Alkaloids are generally extracted from plants in weak acid alcoholic solution and precipitated by NH3. Their presence is tested through chromatic method, and quantified by solvent extraction method.

For chromatographic method, take 10 gm of chopped plant tissue in beaker containing 30 ml of 10% acetic acid in ethanol; heat in water bath at 60 degree centigrade for 40 minutes and filter; concentrate on water bath at 100 degree centigrade to one fourth volume; add NH4OH or NH3 dropwise till pH rises to 9-10 and precipitate alkaloids; centrifuge at 2000 rpm for 5-10 minutes; wash with 1% NH4OH; dissolve in few drops of ethanol; repeat above steps with tea, cigarette and coffee as

control; take circular Whatman paper No. 1, immerse in 5% Sodium dihydrogen citrate for 5 minutes, dry in oven; make pore in center of disc, mark circle of 1 cm, load extract and run chromatogram using solvent (butanol 217 ml, water 32 ml, acetic acid 1.2 ml); and dry; on one chromatogram spray Iodoplatinate reagent (10 ml 5% Platinum chloride, add 240 ml 2% KI, make final volume to 500 ml with water) and to other Dragendroff's reagent (make two solutions, one 0.6 gm Bismith subnitrate + 2 ml HCl+ 1m water; second 6 gm KI dissolved in 10 ml water; mix two solutions and filter, add 7 ml conc. HCl and 15 ml water, dilute to 400 ml with water); dry chromatograms and observe color under UV; calculate Rf value.

For Solvent extraction method take 2 gm of chopped plant material in a beaker containing 20 ml methanol; heat on water bath for 30 minutes at 70 degree centigrade, covering beaker with aluminium foil to prevent evaporation of methanol; filter, suspend residue in 2 ml methanol, add 12 ml HCl to break cell wall and release cell sap, shake and filter; wash residue with 8 ml of 1% HCl, and filter; collect all three filtrates, add NH4OH or NH3 and adjust pH to 10-11; add precipitate formed in separating funnel, and add chloroform, shake and collect lower organic layer; repeat procedure with chloroform, and label all collected solution as A; to aqueous layer in separating funnel add a pinch of sodium sulphate and 20 ml of mixture of ethanol and chloroform (2: 3 ratio); collect and label organic layer as B; put solutions A and B in test tubes heat in water bath till volume is reduced to about 2 ml; add 2 ml of 1% HCl and 2 ml of chloroform in both test tubes, two layers are formed; collect upper layer using dropper; if turbid filter through a piece of cotton film; put a drop of this extracted sample of alkaloids in a groove tile, add 2-3 drops of different reagents (1, Mayer's reagent: dissolve 1.63 gm HgCl2 in 60 ml of distilled water; separately dissolve 5 gm KI in 10 ml of distilled water; mix two and distilled water to make 100 ml. 2, Wagner's reagent: dissolve 1.27 gm iodine and 2 gm KI in 10 ml of distilled water, make volume to 100 ml. 3, Dragendroff's reagent, described above. 4, Scheilden's reagent: dissolve 25 gm sodium tungstate and 17 gm of disodium hydrogen phosphate (Na2HPO4) in

125 ml of distilled water; adjust pH to 4 or 5 by using conc. HNO3 and note the change of color and precipitate formation.

Mears and Mabry (1971), in studies conducted on the family Fabaceae, observed that the alkaloid hystrine occurs only in three genera *Genista*, *Adenocarpus* (both belonging to Genistae) and *Ammodendron* (originally placed in Sophorae). The latter, however, lacks matrine, characteristic of Sophorae. This indicates that the transfer of the last genus also to Genistae is warranted. Families Papaveraceae and Fumariaceae are closely related. This affinity is supported by the occurrence of the alkaloid protopine in both.

Gershenzon and Mabry (1983) reported that tropane alkaloids of Solanaceae and Convolvulaceae are similar, suggesting a close relationship. The families are placed in the same order in recent systems. Papaveraceae, earlier grouped with Cruciferae and Capparaceae, is now removed to nearer Ranunculales on the basis of the absence of glucosinolates and presence of benzylisoquinolene. Nymphaeaceae and Nelumbonaceae differ in the sense that the former lacks benzylisoquinolene alkaloids. Benzylisoquinolene, and the alkaloids that can be derived from it, are characteristic of Magnoliidae, as also family Rutaceae, some Rhamnaceae and the genus *Croton*.

Glucosinolates

Glucosinolates are sulphur containing compounds found in 15 families of angiosperms, mainly concentrated in the order Capparales (Figure 7.15). Mustard oils or isothiocyanates are hydrolytic products of glucosinolates. Originally Cruciferae, Capparaceae, Papaverace- and Fumariaceae were placed in the same order, Rhoeadales. Chemical and other evidence, however, supported the placement of Cruciferae and Capparaceae in the order Capparales (on the basis of the presence of glucosinolates) and Papaveraceae and Fumariaceae in the order Papaverales—or suborder Papaverineae of Ranunculales (Thorne, 2003)—(on the basis of the absence of glucosinolates and the presence of the alkaloid benzylisoquinolene). Bataceae and Gyrostemonaceae were once placed in Centrospermae (Caryophyllales) but subsequently removed due to the absence of betalains. This removal was supported by the presence of glucosinolates, which are absent in Caryophyllales.

Cyanogenic Glycosides Cyanogenic glycosides are phytotoxins which occur in at least 2000 plant species, of filter through a piece of cotton film; put a drop of this extracted sample of alkaloids in a groove tile, add 2-3 drops of different reagents (1, Mayer's reagent: dissolve 1.63 gm HgCl2 in 60 ml of distilled water; separately dissolve 5 gm KI in 10 ml of distilled water; mix two and distilled water to make 100 ml. 2, Wagner's reagent: dissolve 1.27 gm iodine and 2 gm KI in 10 ml of distilled water, make volume to 100 ml. 3, Dragendroff's reagent, described above. 4, Scheilden's reagent: dissolve 25 gm sodium tungstate and 17 gm of disodium hydrogen phosphate (Na2HPO4) in 125 ml of distilled water; adjust pH to 4 or 5 by using conc. HNO3 and note the change of colour and precipitate formation. Mears and Mabry (1971), in studies conducted on the family Fabaceae, observed that the alkaloid hystrine occurs only in three genera Genista, Adenocarpus (both belonging to Genistae) and Ammodendron (originally placed in Sophorae). The latter, however, lacks matrine, characteristic of Sophorae. This indicates that the transfer of the last genus also to Genistae is warranted. Families Papaveraceae and Fumariaceae are closely related. This affinity is supported by the occurrence of the alkaloid protopine in both. Gershenzon and Mabry (1983) reported

that tropane alkaloids of Solanaceae and Convolvulaceae are similar, suggesting a close relationship. The families are placed in the same order in recent systems. Papaveraceae, earlier grouped with Cruciferae and Capparaceae, is now removed to nearer Ranunculales on the basis of the absence of glucosinolates and presence of benzylisoquinolene. Nymphaeaceae and Nelumbonaceae differ in the sense that the former lacks benzylisoquinolene alkaloids. Benzylisoquinolene, and the alkaloids that can be derived from it, are characteristic of Magnoliidae, as also family Rutaceae, some Rhamnaceae and genus Croton. Glucosinolates Glucosinolates are sulphur containing compounds found in 15 families of angiosperms, mainly concentrated in the order Capparales (Figure 7.15). Mustard oils or isothiocyanates are hydrolytic products of glucosinolates. Originally Cruciferae, Capparaceae, Papaverace-which a number of species are used as food in some areas of the world. They are hydrolysed by various enzymes to release hydrogen cyanide, the process known as cyanogenesis, and the plants as cyanogenic plants. Cassava and sorghum are especially important staple foods containing cyanogenic glycosides. There are approximately 25 cyanogenic glycosides known. The major cyanogenic glycosides found in the edible parts of plants used for human or animal consumption include Amygdalin (Figure 7.16; almonds, Prunus dulcis), Dhurrin (Sorghum album, S. bicolor). Linamarin (cassava, Manihot esculenta; lima beans, Phaseolus lunatus), Lotaustralin (cassava, Manihot carthaginensis; lima beans, Phaseolus lunatus). Prunasin (stone fruits, Prunus avium, P. padus, P. persica, P. macrophylla), and Taxiphyllin (bamboo shoots, Bambusa vulgaris). The potential toxicity of a cyanogenic plant depends primarily on the potential that its consumption will produce a concentration of HCN that is toxic to exposed animals or humans. Hydrogen cyanide is released from the cyanogenic glycosides when fresh plant material is macerated as in chewing, which allows enzymes and cyanogenic glycosides to come together, releasing hydrogen cyanide. Cyanides inhibit the oxidative processes of cells causing them to die very quickly. Because the body rapidly detoxifies cyanide, an adult human can withstand 50-60 ppm for an hour without serious consequences. However, exposure to concentrations of

200-500 ppm for 30 minutes is usually fatal. Whereas most of the cyanogenic glycosides are widely spread, others such as cyclopentenoid cyanogenic glycosides are restricted in distribution mainly to Flacourtiaceae, Passifloraceae, Turneraceae, and Malesherbiaceae. Leucine derived cyanogenic glycosides are found in Rosaceae, Fabaceae, and Sapindaceae. Several families belonging to Magnoliales and Laurales contain Cyanogenic glycosides derived from tyrosine. Presence of cyanogenic glycosides can be tested by two methods. Picrate paper test: Cut Whatman no. 1 filter paper into rectangular strips so that they can be fixed in the cap of screw vial; place strips in picric acid (5.7 gm in 500 ml of distilled water, saturated with sodium bicarbonate and filtered) for 5 minutes and dry with drier or in oven, paste strips in the cap of the vial; put chopped leaves or plant material in so as cover the entire base of vial, add 2-3 drops of distilled water and 1-2 drops of toluene, crush material with glass rod; cap the vial; picrate paper in cap should not touch plant tissue or walls of vial; incubate at 600 C for 2 hours; observe the change of paper colour to red. Ferrous hydroxide paper test: Dip rectangular strips of Whatman paper in 10% FeSO4 solution for 5 minutes, remove and dry; next immerse in 20% NaOH solution for 20-30 seconds and dry, and fix in cap of vial; put crushed leaves in vial, add 2-3 drops of distilled water and 2-3 drops of toluene, crush using glass rod; cap the vial and incubate at 600 O for 2 hours; immerse strips in 30% H2SO4; observe colour change to prussian blue due to formation of sodium ferric ferrocyanide.

Terpenes

Terpenes include a large group of compounds derived from the mevalonic acid precursor and are mostly polymerized isoprene derivatives. Common examples are camphor (Cinnamomum), menthol (Mentha), and carotenoids (Figure 7.17). They seem to have a definite role in the allelopathic effects of plants. They are lipid soluble found in single membrane bound liposomes, in glandular cells as essential oils, and can be easily extracted with petroleum ether and chloroform, and can be separated by GLC (Gas Liquid Chromatography), enabling qualitative as well as quantitative

measure of chemical differences. Terpenes are isomeric unsaturated hydrocarbons of the basic 5-carbon isoprene (CH2=C(CH3)-CH=CH2) present in Hamamelis japonica. Terpenoids, the common group of terpenes are distinguished as 10-Carbon Monoterpenoids (Fennel, Menthol-Mentha, Gymnosperms), 15-C Sesquiterpenoids (Asteraceae as sesquiterpenol, ABA, some essential oils), 20-C Diterpenoids (Taxus-Taxol, gibbrellins), 30-C Triterpenoids (sterols, steroids, saponins, betulin in Betula papyrifera), 40-C Tetraterpenoids (Carotenoids) and poly-C Polyterpenoids (Rubber). They have been largely used in distinguishingspecific and subspecific entities, geographic races and detection of hybrids. Studies in Citrus have focused on determination of the origin of certain cultivars. Studies on Juniperus viginiana and J. ashei have refuted previous hypotheses about extensive hybridization and introgression between the two species. Their distribution in Pinus has been used (Mirov, 1961) to understand relationships. P. jeffreyi has been considered a variety of P. ponderosa, but turpentine distribution showed that it strongly resembles the group Macrocarpae and not Australes to which P. ponderosa belongs. A major contribution of terpenoid chemistry has been the use of sesquiterpene lactones in the family Compositae. Many tribes within the family are characterized by distinct types of sesquiterpene lactones they produce. This helped to establish that genus Vernonia has two centres of distribution—one in the Neotropics and the other in Africa. Similarly, studies on Xanthium strumarium (McMillan et al., 1976) have thrown some light on the origin of Old World and New World populations. Old World populations produce xanthinin/ or xanthinosin, whereas the New World populations contain xanthinin or its stereoisomer xanthumin. Plants of the chinense complex (from Louisiana) contain xanthumin and are believed to be the source of introduced chinense populations in India and Australia. Triterpenoids occur in several families. Betulin occurs only in bark of white birch (Betula papyrifera) and its relatives, is waterproof highly flammable, and is taxonomically useful at species level. Triterpene saponins occur in Apiaceae and Pittosporaceae and support their close relationship. Iridoids constitute another important group of terpenes (mostly monoterpene lactones). They are present in over

50 families and their presence is correlated with sympetaly, unitegmic tenuinucellate ovules, cellular endosperm and endosperm haustoria. Assuming that 'independent origin of several groups with this combination of independent attributes is unlikely', Dahlgren brought together all iridoid-producing families. The occurrence of iridoids in several unrelated families, e.g. Hamamelidaceae and Meliaceae, however, suggests that iridoids could have arisen independently several times in the evolution of angiosperms. The occurrence of a distinctive iridoid aucubin in Budleja has been taken to support its transfer from Loganiaceae to Budlejaceae. Aucubin and geniposide, have shown antitumoral activities. Iridioid presence in plant tissues can be tested using Trim-Hill Reagent. Take 5 gm of chopped leaves in a beaker containing 5 ml of 1 % HCl; heat on water bath at 40-500C for 30 minutes; filter using coarse filter paper; take 0.1 ml of filtrate in a test tube; add 1 ml Trim-Hill Reagent (THR: 10 ml acetic acid, 1 ml 0.2 % CuSO4 and 0.5 ml conc. HCl); warm over spirit lamp for few seconds; change of colour commonly to yellow, orange or red confirms presence of iridoids. Quantitative estimation can be done by reading absorption at 609 nm, concentration calculated on the basis of standard curve of aucubin. Cronquist (1977) proposed that chemical repellents had an important role in the evolution of major groups of dicots. The alkaloid Isoquinolene of Magnoliidae gave way to tannins of Hamamelidae, Rosidae and Dilleniidae, which in turn gave way to the most effective iridoids in Asteridae, the family Compositae developing the most effective sesquiterpene lactones.

Non-Semantide

Macromolecules In addition to DNA and RNA, which will be dealt under Molecular systematics, the macromolecules include proteins, and complex polysaccharides such as starches and celluloses. Starches are commonly found in the form of grains which may be concentric (Triticum, Zea) or eccentric (Solanum tuberosum) and present anatomical characteristics which can be seen under a

microscope. Detailed studies of starch grains under SEM also hold promise for taxonomic significance.

Proteins

Proteins, together with nucleic acids, are often called Semantides, which are primaryconstituents of living organisms and are involved in information transfer. Based on their position in the information transfer DNA is a primary semantide, RNA secondary semantide and proteins the tertiary semantides. Semantides are popular sources of taxonomic information, and most of this information has come from proteins. The information about DNA and RNA will be discussed under Molecular Systematics in the next section; only proteins are being discussed here. Proteins are complex macromolecules made up of amino acids linked into a chain by peptide bond, thus forming a polypeptide chain, organized into a three-dimensional structure. Because of their complex structure, special techniques are necessary for the isolation, study and comparison of proteins. These methods include serology, electrophoresis and amino acid sequencing.

Serology

The field of systematic serology or serotaxonomy had its origin towards the turn of the twentieth century with the discovery of serological reactions and development of the discipline of immunology. Precipitin reactions were first reported by Kraus (1897). The technique was originally applied by J. Bordet (1899) in his work on birds, when he reported that immune reactions are relatively specific and the degree of cross reactivity was essentially proportional to the degree of relationship among organisms. The present technique of serology is based on immunological reactions shown by mammals when invaded by foreign proteins. In the study of estimating relationships between plants, the plant extract of species A containing proteins (antigens) is injected into a mammal (usually a rabbit, mouse or goat). The latter will develop antibodies, each specific to an antigen with which it forms a precipitin reaction, coagulating and

thus making it non-functional. These antibodies are extracted from the body of the animal as antiserum. This antiserum is capable of coagulating all proteins in species A, but when mixed with the protein extract of species B, the degree of precipitin reaction would depend on the similarity between the proteins of the two species. The antiserum obtained from the mammal normally contains several immunoglobulins that can bind to the same antigen, is said to be polyclonal. This is because an antigen activates several different lymphocytes within the animal, each producing a different antibody for the same antigen. Techniques have now been developed which can generate monoclonal antibodies. In a method developed by Milstein and Köhler (1975), antibody-producing lymphocyte of mammal (which can not grow and divide in cultures) was fused with malignant myoloma cell (cancer cell which can grow rapidly in cultures) to produce hybrid cells called hybridoma. These hybridomas can grow, proliferate and produce large amount of monoclonal antibody. Antigens are mostly extracted from seeds and pollen. In early works, crude total comparison of precipitin reactions was done but now more refined methods have been developed which can bring about individual antigen-antibody reactions. Major methods include:

Double-diffusion serology

In this technique the antigen mixture and antiserum are allowed to diffuse towards one another in a gel (Figure 7.18). The different proteins travel at different rates and thus the reactions occur at different places on the gel. This method allows comparison of precipitin reactions of several antigen mixtures from different taxa simultaneously on the same gel. In a modification of this method, the antiserum is placed in a circular well surrounded by a ring of several wells containing the samples of antigens.

Immuno-electrophoresis

In this technique the antigens are first separated unidirectionally in a gel by electrophoresis and then allowed to travel towards the antiserum (Figure 7.19). This

method enables a better separation of constituent reactions but has the limitation that only one antigen mixture can be handled on a single gel.

Absorption

Protein mixtures from different species often contain a large number of common proteins, especially those involved in common metabolic processes. The antibodies for these common proteins (antigens) are first removed from the antiserum so that there is a more logical comparison of precipitin reactions.

Radio-immunoassay (RIA)

In this technique the antibodies or antigens are labelled with radioactive molecules enabling their detection even when present in minute quantities.

Enzyme-linked immunosorbent assay (ELISA)

In this technique either the antibodies or antigens are labelled linked with enzymes, thus enabling detection even in very small quantities. It must be noted that there are specific sites on proteins (determinants), which are capable of initiating production of immunoglobulins in specific cells of mammals. Determinants are regions consisting of 10-20 amino acids and one protein may comprise several different determinants and thus several antigens. Extensive studies of the immunoelectrophoretic patterns of the genus Bromus were done by Smith (1972, 1983). Results showed that North American diploids of the genus are reasonably diverse. The study also highlighted that antisera raised from different species could provide different results. On the basis of serological studies, Smith established the distinct identity of B. Pseudosecalinus, previously recognized as a variety of B. secalinus. This separation was supported by cytological evidence also. Serological studies have also supported the removal of Nelumbo from Nymphaeaceace into a separate family Nelumbonaceae, placement of Hydrastis in Ranunculaceae (and not Berberidaceae), and merger of Mahonia with Berberis (Fairbrothers, 1983). Serology

may be done through comparison of protein mixtures or the comparison of single isolated and purified proteins. Schneider and Liedgens (1981) developed a complex but excellent procedure of monoclonal culture of antibodies, but unfortunately used this for construction of a ‘phylogenetic tree’ not parallel with accepted evolutionary schemes. Fairbrothers (1983) cautioned that an evolutionary tree should not be constructed on the reactions of a singleenzyme or a single species. Lee (1981) using purified protein for antigen and using different techniques concluded that Franseria (Asteraceae) should be merged with Ambrosia.

Electrophoresis

The technique of serology serves to compare the degree of similarity between the protein mixtures of different species and does not involve the identification of proteins. The separation and identification of proteins can be done by electrophoresis. Separation is based on the amphoteric properties of proteins whereby they are positively or negatively charged to various extents according to the pH of the medium, and will travel through gel at various speeds across a voltage gradient, usually carried out in a polyacrylamide gel (polyacrylamide gel electrophoresis—PAGE). The procedure involves homogenizing the tissues (containing proteins) in a buffer solution. Sample is loaded into wells in the centre of the gel. The current is run for a specific time, and the proteins run up to different points on the gel. The gel, usually 1 cm thick, is cut into three thin slices, each about 3 mm thick. These slices are subjected to different staining techniques and proteins are identified using various criteria. In commonly used Western blot technique the protein bands are transferred from the gel to nitrocellulose membrane for further processing. In disc-electrophoresis, a gel of larger pores is placed over a gel of smaller pores. The former is used for crude separation and the latter for a complete separation. Method for SDS-PAGE Electrophoresis: Prepare working concentrations of 2, 5, 10 and 25 μg of protein sample by diluting the stock solution; add loading buffer (5 ml 0.5 M Tris, pH 6.8, 8 ml 50% Glycerol, 8 ml 10% SDS, bromophenol blue, 2 ml β− mercaptoethanol added

immediately before loading); boil for 5 min; store in ice; clean, dry and assemble glass plates of casting assembly; prepare 10% resolving gel (3.24 ml Acrylamide-bisacrylamide, 3.5 ml distilled water, 2.5 ml 1.5 M Tris-HCl, 0.1 ml 10% SDS, 0.5 ml 10% APS, 10µl TEMED) and pour into assembly up to its two third position; overlay with water ; allow gel to polymerise for 30 min, remove water; pour 5% Stacking gel (0.82 ml Acrylamide-bisacrylamide, 3.27 ml distilled water, 0.625 ml 0.5 M Tris-HCl, 0.05 ml 10% SDS, 0.25 ml 10% APS, 5µl TEMED) up to one third volume of glass mould, insert comb carefully so that bubbles are not formed, allow to polymerize for 30 min; install gel assembly into electrophoresis apparatus; add 5X SDS Running buffer (6.026 g Tris, 28.8 g Glycerine, 2 g SDS; make to 200 ml with distilled water) in upper and lower chamber, remove the comb from under the buffer; load protein samples of different concentrations into wells by micropipette (also load molecular weight marker proteins in one lane); connect electrodes and run current of 20 mA for 10-15 min, increase current to 40 mA, track the mobility of sample; disconnect power, remove gel carefully; stain gel with Coommassie Blue stain (200 ml methanol, 50 ml glacial acetic acid, 250 ml water, 0.25% Coommasie Blue) for 15-30 min; destain gel till bands are visible. Method for Western blotting: Use gel from electrophoresis without staining; make cut at bottom of gel for orientation; cut nitrocellulose sheet to the size of gel and dip in transfer buffer (14 g glycerine, 3 g Tris base, 0.75 g SDS, 100 ml methanol, make volume to 1 litre with distilled water); soak sponge in transfer buffer and place wet sponge on gel holder; place Whatman paper on sponge; place gel over Whatman paper avoiding air bubbles; keep membrane with shining surface towards gel and roll with glass pipette; place Whatman 3 mm paper over the membrane and a second sponge over the paper; place assembly in transfer tank containing sufficient transfer buffer to completely cover the blot; place assembly in case with gel facing the cathode and membrane the anode; run current for 4 hours at 36 V; lift membrane and stain with Poinceau S.

Staining Solution.

In the technique of isoelectric focusing, a gel of a single pore size, is set up with a pH gradient (usually 3-10), so that proteins come to lie on the gradient corresponding to their iso-electric point. These can be subsequently separated more completely by discelectrophoresis. Isoelectric focusing of Rubisco (Ribulose 1, 5 diphosphate carboxylase) has been very useful in determining relationship between species of Avena, Brassica, Triticum, and several other genera. It is an excellent protein for helping to evaluate hybridization. Electrophoretic studies have supported the origin of hexaploid wheat (Triticum aestivum) from Aegilops tauschii and T. dicoccum. Johnson (1972), working on storage proteins showed that T. aestivum (AABBDD) and T. dicoccum (AABB) possess all proteins of the A genome of the diploid T. monococcum (AA). They also share proteins of the B genome of uncertain origin. The D genome is believed to have come from Aegilops tauschii as evidenced by morphological and cytological data. By mixing proteins of A. tauschii and T. dicoccum it was seen that the electrophoretic properties of the mixture closely resemble those of T. aestivum, thus proving the origin of thelatter from the two previous species. Electrophoretic studies have also helped to assess species relationships in Chenopodium (Crawford and Julian, 1976), by combining data from flavonoids with proteins. A flavonoid survey of seven species showed that in some taxa, the flavonoid data were fully compatible with interspecific protein differences, but in some cases, did not agree. Thus, Chenopodium atrovirens and C. leptophyllum had identical flavonoid patterns but could be distinguished by their different seed protein spectra. C. desiccatum and C. atrovirens, on the other hand, were closely similar in seed proteins but differed in flavonoids. Both flavonoid and protein evidence, however, distinguished C. hians from C. leptophyllum, thus providing support to their recognition as separate species. Vaughan et al., (1966) through the study of serology and electrophoresis have shown that Brassica campestris and B. oleracea are closer to each other than to B. nigra. Electrophoresis has also made possible the separation of allozymes (different forms of the same enzyme with different alleles at one locus) and

isozymes (or isoenzymes with different alleles at more than one locus). Barber (1970) showed that certain polyploids possess isozymes of all their progenitors plus some new ones. Backman (1964) crossed two strains of maize, each with three different isozymes. F1 possessed all six isozymes. The hybrids thus show molecular complementation. Studies of the genus Tragopogon have confirmed that the tetraploid T. mirus is a hybrid between two diploid species, T. dubius and T. porrifolius. Whereas the parental diploids were found to be divergent at close to 40 per cent of the 20 enzyme loci examined, the tetraploid hybrid possessed completely additive enzyme patterns. The evidence thus supported the recognition of a hybrid on the basis of morphological and chromosomal evidence.

Amino acid sequencing

Since only 22 amino acids are known to be the constituents of proteins, the primarydifferences between the proteins result from different sequences of amino acids in the polypeptide chain. It is now possible to break off the amino acids from the polypeptide chain one by one, identify each chromatographically and build up the sequence of amino acids step by step. Cytochrome c is the most commonly used molecule and out of 113 amino acids, 79 vary from species to species, but alteration of even one of the other 34 destroys the functioning of the molecule. Being present in all aerobic organisms, it is ideal for comparative studies. Boulter (1974) constructed a cladogram (Figure 7.20) of 25 species of spermatophytes using the 'ancestral sequence method'. Ginkgo biloba, the only gymnosperm used occupied isolated position in the cladogram. Ginkgo with an isolated phylogenetic position is no new discovery, but rather a long established fact. But the fact that amino acid sequencing also produces a similar cladogram establishes the significance of such studies in understanding phylogeny. Recent data from various fields have pointed to the merger of Aegilops with Triticum. Autran et al., (1979) on the basis of N-terminal amino acid sequencing supported this merger. In general, the number of amino acid differences is roughly parallel to the distance between the organisms in traditional classifications, suggesting

that the method is broadly reliable. There are, however, certain contradictions. The number of differences between the cytochrome c of Zea mays and Triticum aestivum (both members of the same family Poaceae) is greater than between Zea mays and certain dicotyledons. It has been found that cytochrome c and plastocyanin (another protein commonly used in amino acid sequencing studies) can exhibit a large number of parallel substitutions (identical changes from one amino acid to another at the same position in the protein in different organisms), thus rendering them unsuitable for constructing phylogenies. The practical solution is to use evidence from a wide range of proteins, preferably using different techniques.

8- MOLECULAR SYSTEMATICS

The closing years of the past century saw the concentration of macromolecular studies towards DNA and RNA, resulting in the establishment of an emerging field of molecular systematics. Although flavonoids and isozymes also constitute molecular data, molecular systematics commonly deals exclusively with the utilization of nucleic acid data. As molecular data reflects gene-level changes, it was believed to reflect true phylogeny better than morphological data. It has, however, been realized that molecular data may also pose similar problems, although there are more molecular characters available and comparison is generally easier.

Molecular evolution

Traditionally, different taxa, especially the species have been characterized primarily on morphological differences (phenotypes). Additionally, differences in physiology, biochemistry, anatomy, palynology, embryology, gross chromosome structure and behaviour, have been used in refining evolutionary trees. Although, it had been long recognized that evolution is based on genetic changes, only during the last two decades, there have been forceful drive to use genetic material for a better understanding of evolutionary relationships. Those species that are closely related, are expected to have greater similarities in their genetic material than the distantly related

species. During the past decade, molecular genetics has taken a dominant role in enabling us to understand speciation and evolution clearly. Differences in the nucleotide sequences are quantitative and can be analyzed using mathematical principles, utilizing the help of computer programs. Evolutionary changes at the DNA level can be objectively compared among different species to establish evolutionary relationships.

Evolution of Nucleic acids and Proteins

When Earth originated nearly 5000 million years ago the primary atmosphere consisted of only hydrogen and helium, but being too small a planet to hold these light gases, they floated away into space. The earth accumulated its secondary atmosphere because of volcanic activity in early hot earth and the gases consisted of largely steam, variable amounts of CO2, N2, SO2, H2S, HCl, Sulphur and smaller quantities of H2, CH4, SO3, NH3. There was no free oxygen. Our present atmosphere is of biological origin, in which methane and ammonia have largely been consumed, inert components like nitrogen remained unchanged, and oxygen produced by photosynthesis. This happened nearly 2500 m years ago when Cyanobacteria, the first photosynthetic bacteria made their appearance. The ultraviolet radiations from the sun, together with lightening discharges caused the gases to react in the primeval atmosphere forming simple organic compounds such as amino acids, sugars and nucleic acid bases. This mostly happened because of gases dissolving in primeval oceans and continuing to react forming primitive soup, the precursor of life. Further reactions formed polymers, globules and eventually the first primitive cell. The possibility of such reactions in the primitive atmosphere was demonstrated by a Russian biochemist Alexander Oparin in 1920s who proposed that life evolved before there was any free oxygen in the atmosphere. The oxygen if present at that stage would have reacted with precursor organic molecules formed in the atmosphere, oxidizing them back into carbon dioxide and water. These reactions were mimicked by biochemist Stanley Miller in 1950s, who subjected a mixture of methane, ammonia and water vapour to high voltage discharge

or to ultraviolet light, and the products allowed to dissolve and react in water. As long as oxygen was excluded, the results were similar producing several organic molecules such as amino acids, formic acid, glycolic acid, lactic acid, acetic acid, propionic acid, succinic acid, urea, purines, pyrimidines and sugars. These energy sources can also destroy these organic molecules present in the atmosphere. Theoccurrence of primeval oceanic atmosphere helped shielding and preserving these organic molecules and prevented their destruction. Organic acids, particularly amino acids are soluble in water and non-volatile, and little chance of their returning to atmosphere. The polymerization of amino acids and other monomers to form macromolecules requires energy for formation of bonds and removal of water. Such polymers known as proteinoids can be generated by simple heating of amino acids at around 1500 C for a few hours. Such heating could have occurred near volcanoes or when pools left behind by changing coastline evaporated. Inorganic polyphosphates present in the primeval times would have helped in condensation. It is generally thought that RNA probably evolved first through polymerisation of nucleotides present in primeval environment. When RNA template is incubated with mixture of nucleotides and zinc as a catalyst, a complementary piece of RNA is synthesized. The complementary strand in turn will act as template to generate more of original RNA molecule. It is assumed that RNA originated even before proteins. It is also believed that earliest organisms had both genes and enzymes made of RNA and formed RNA world. The examples of enzymatically active RNA are found in Ribozymes and self-replicating introns. Later proteins infiltrated and took over the role of enzymes. This was followed by the evolution of DNA as genetic material, and RNA relegated to the role of intermediate between genes and the enzymes. Changes in DNA sequences (mutations) lead to the changes in the codons, that in turn determine the sequence of various amino acids, deciding the final structure and function of a protein. These changes commonly result from changes of one or more base pairs in a DNA sequence. Two types of nucleotide changes occur in the genome. Some changes give rise nonsynonymous codons, coding for different amino acid, and thus resulting in a corresponding change in the amino

acid sequence of a protein. Other nucleotide changes give rise to synonymous codons, that code for the same amino acid.

Evolutionary Rates within a Gene

It is now well established that different parts of genes evolve at widely different rates, reflecting the extent of natural selection on each part. Some nucleotides code for amino acid sequence of a protein (Coding sequences), whereas others do not code for amino acids in a protein (noncoding sequences). Latter include introns, leader regions, trailer regions (all these are transcribed but not translated), and 5' and 3' flanking sequences that are not transcribed. Pseudogenes, which are nucleotide sequences that no longer produce functional gene products as they have accumulates inactivating mutations, also constitute noncoding sequences. Even within coding regions of functional gene, not all nucleotide substitutions produce a corresponding change in the amino acid sequence of a protein. Many substitutions occurring at the third position of triplet codons have no effect on the amino acid sequence of the protein because such changes often produce synonymous codons. Although synonymous and nonsynonymous nucleotide changes are likely to arise in equal frequency (because enzymes responsible for DNA replication and repair cannot differentiate between the two), yet the rate of synonymous nucleotide changes (conservative substitutions of Kimura) is about five times greater than observed rate of nonsynonymous changes (disruptive substitutions of Kimura). This is because synonymous changes do not alter protein structure and function and are tolerated by natural selection, but the nonsynonymous changes are usually detrimental and are excluded by natural selection. Synonymous substitution rates and not nonsynonymous nucleotide changes are, as such, the fair reflection of actual mutation rate within a genome. Pseudogenes and 3' flanking regions alsoshow high evolutionary rates, comparable to synonymous changes. 5' flanking regions show a little slower rate, whereas leader and trailer regions show very low evolutionary rates, slightly higher than nonsynonymous changes. It is as such obvious that nucleotide changes in noncoding regions or codings that do not

alter amino acid sequences, have high rate of evolution, whereas changes in coding regions, especially those affecting amino acid sequences show very low rate of evolution, as most of them get filtered out by natural selection. It is important note that whereas mutations are changes in nucleotide sequences that occur because of mistakes in DNA replication or repair processes, the substitutions are mutations that have passed through the filter of selection at least at some level.

Location of molecular data

Systematists use molecular data from three different locations within a plant cell: chloroplast, mitochondrion and the nucleus, yielding three different types of genome (DNA). Chloroplast genome is the smallest ranging from 120 to 160 kbp (kilo base pairs) in higher plants (up to 2000 kbp in alga Acetabularia), mitochondrion genome 200 to 2500 kbp, whereas the nuclear genome is much larger often ranging between million to more than billion kbp. Although the former two are inherited from the maternal parent, the latter is biparental. Mitochondrion genome undergoes a lot rearrangements, so that many different forms may be found within the same cell, and hence is of little significance in interpreting phylogenetic relationships, whereas the other two are highly stable not only within the same cell, but also within a species, and present useful taxonomic tools.

Mitochondial DNA

Mitochondrial DNA has been studied from several species of plants. Each mitochondrion contains several copies of mtDNA, andas each cell contains several mitochondria, the number of mtDNA molecules per cell could be very large. Most mtRNA molecules are circular, but linear in Chlamydomonas reinhardtii. In vascular plants, mtDNA is considerably larger, circular, containing many noncoding sequences, including some that are duplicated. The physical mapping of genes in vascular plants has shown that these are located in different positions on mtDNA circles of different

species, even in fairly closely-related species. This renders mtDNA less useful in phylogenetic studies.

Chloroplast DNA

Studies of DNA in plants have largely been undertaken from chloroplast compared to the other two cellular genomes. This is because chloroplast DNA (cpDNA) can be easily isolated and analyzed. It is also not altered by evolutionary processes such as gene duplication and concerted evolution (in rRNA, having thousands of copies of repeated segments so that mutation in one sequence gets corrected to match other copies, this homogenization process is termed as concerted evolution). It also has an added advantage in that it is highly conserved in organisation, size and primary sequence. Chloroplast DNA is closed circular molecule (Figure 7.21) with two regions that encode the same genes but in the opposite direction and known as inverted repeats. Between the inverted repeats are single copy regions. All cpDNA molecules carry basically the same set of genes, arranged differently in different species of plants. These include genes for ribosomal RNA, transfer RNA, ribosomal proteins and about 100 different polypeptides and subunits of enzyme capturing CO2. Most studies of chloroplast DNA have focused on chloroplast gene rbcL, which encodes large subunit of photosynthetic enzyme RuBisCO (ribulose-1,5-biphosphate carboxylase/ oxygenase, carbon acceptor in all photosynthetic eukaryotes and cyanobacteria. The gene occurs in all plants (except parasites), is fairly long (1428 bp), presents no problems of alignment, and has many copies available in the cell. Ready availability of PCR primers has made it possible to generate over 2000 sequences, primarily of seed plants. Other commonly used chloroplast genes include atpB, (beta subunit of ATP synthetase involved in the synthesis ATP), matK (maturase involved in splicing type II introns from RNA transcripts), and ndhF (subunit of chloroplast NADH dehydrogenase, which functions in converting NADH to NAD+H, involved in reactions of respiration. Of these four commonly used genes rbcL, atpB, and matK

belong to large single copy region, where as ndhF is located on small single copy region.

Nuclear DNA

The nuclear DNA, although more difficult to analyze, and hence used less frequently has two great advantages. Certain nuclear sequences evolve more rapidly than cpDNA sequences, and thus allow finer level of discrimination at population level than cpDNA. Also, whereas the nuclear genome is inherited biparentally, the chloroplast genome is inherited maternally. Thus the hybrid plant will possess the nuclear complement of both parents but only the cpDNA of the maternal plant. The study of nuclear genes has traditionally involved ribosomal RNA. Ribosomal genes are arranged in tandem arrays of up to several thousand copies. Each set of genes has a small subunit (18S) and a large subunit (26S) separated by a smaller (5.8S) gene (Figure 7.22). It must be noted that 5S RNA although also a part of the unit, but of the unknown function is synthesized separately outside nucleolus. The three subunits are separated by internal transcribed spacers (ITS: ITS1 and ITS2). Each set of genes is separated from adjacent one by a larger spacer (variously known as IGS- intergenic spacer, EGS- extragenic spacer or NTSnontranscribed spacer). Sequences of 18S and 26S genes have been used in phylogenetic studies, because they have some highly conservative regions which help in alignment, and other variable regions, which help to distinguish phylogenetic groups. Recently ITS region has been used to determine relationships among species. In general, the ITS region has supported relationships inferred from chloroplast studies and morphology.

Molecular techniques

The techniques of handling molecular data saw great advancements over past few de-cades, starting with comparison of whole DNA molecules. It is now possible to break DNA at specific sites, generate maps of individual genes, determine sequence

of genes, and make multicopies of a DNA through Polymerase chain reaction (PCR) technique. These help in generating enough molecular data for comparison.

Total DNA/DNA hybridization

The early studies on utilization of nucleic acids in systematics involved DNA/DNA hybridization using the whole DNA for study. In a method developed by Bolton and Mecarthy (1962), the extracted DNA is treated to make it single stranded. The DNA of another organism is, similarly, made single stranded. The two are subsequently allowed to hybridize in vitro. The degree of reassociation (annealing) expresses the degree of similarity in sequences of nucleotides of the two organisms. Procedure involves heating DNA so that it becomes denatured into single strands (ssDNA). The temperature is lowered just enough to allow the multiple short sequences of repetitive DNA to rehybridize back into double-stranded DNA (dsDNA). The mixture of ssDNA (representing single genes) and dsDNA (representing repetitive DNA) is passed over a column packed with hydroxyapatite. The dsDNA sticks to the hydroxyapatite; ssDNA does not and flows right through. The purpose of this step is to be able to compare the information-encoding portions of the genome — mostly genes present in a single copy — without having to worry about varying amounts of noninformative repetitive DNA. The ssDNA of species A is made radioactive. The radioactive ssDNA is then allowed to rehybridize with nonradioactive ssDNA of the same species (A) as well as — in a separate tube — the ssDNA of species B. After hybridization is complete, the mixtures (A/A) and (A/B) are individually heated in small (2°–3°C) increments. At each higher temperature, an aliquot is passed over hydroxyapatite. Any radioactive strands (A) that have separated from the DNA duplexes pass through the column, and the amount is measured from their radioactivity. A graph showing the percentage of ssDNA at each temperature is drawn. The temperature at which 50% of the DNA duplexes (dsDNA) have been denatured (T50H) is determined. Bolton (1966) found that only half nucleotide sequences in the DNA of Vicia villosa are similar (homologous) with those of Pisum, while only 1/5th

are homologous between Phaseolus and Pisum. In the technique of DNA/RNA hybridization, the RNA is hybridized with the complementary DNA of related plants. Mabry (1976) used this technique in Centrospermae (Caryophyllales) and concluded that the family Caryophyllaceae (although lacking betalains) is quite close to betalain-containing families, but not as close as the latter are to each other.

Chromosome painting

The technique of chromosome painting provides another way to compare entire genomes. A fluorescent label is attached to theDNA of individual chromosomes of one species. These chromosomes are exposed to the chromosomes of another species. The regions of gene homology will hybridize taking up the fluorescent label and the 'painted' chromosomes can be examined under a microscope. The method is a modification of fluorescence in situ hybridization (FISH). Chromosome painting studies in humans have shown that human chromosome 6 has counterparts in chromosome 5 of chimpanzee, chromosome 7 of pig and chromosome 23 of cow as few examples.

Unravelling DNA Structure

Understanding DNA structure involves complex procedure to unravel the arrangement of genes in DNA, and sequence of arrangement of nucleotides which differentiates different genes and the DNA of different organisms. The procedure involves some distinct steps. DNA Cleaving This technique is a landmark development of 1970s that can be used to generate physical maps of individual genes or the entire genome. The DNA extracted from a species is cut (cleaved) at specific points (recognition- site; restriction site), yielding restriction fragments using restriction endonucleases (REs). The specific enzymes are named using the first letter of the genus and the first two letters of the species of the bacterium from which the enzyme is isolated. Thus, enzyme EcoRI which cleaves DNA at every site where it finds sequence GAATTC (Figure 7.23) is obtained from Escherichia coli. HindIII obtained from

Haemophilus influenzae strain Rd cleaves DNA at AAGCTT, and BamHI from Bacillus amyloliquefaciens cleaves GGATCC. More than 400 restriction enzymes have already been isolated. Their natural function is to inactivate invading viruses by cleaving the viral DNA. Majority of restriction enzymes recognize a 6-nucleotide sequence, but others recognize 4- nucleotide sequence. Thus AluI (from Arthrobacter luteus) recognizes AGCT, TaqI (from Thermus aquaticus) TCGA, and HaeIII (from Haemophilus aegypticus) GGCC. Each restriction enzyme can recognize a sequence four to six nucleotides long, having twofold rotational symmetry, because it can be rotated 1800 without change in the base sequence. Thus, sequence recognized by EcoliRI—if read from '5 to 3' in both strands, of DNA segment—would read GAATTC, but if read from '3 to 5' in both strands it would read CTTAAG. This symmetry is known as palidrome (as, for example, in nonsense phrase: AND MADAM DNA that is read similarly from both ends). This feature combined with the fact that most restriction enzymes give staggered cuts (and not straight cuts) wherein they cut two strands of DNA at different points, produces complementary single-stranded termini that can be rejoined later using enzyme DNA ligase. Such enzymes produce sticky ends or cohesive ends. Others like AluI and HaeIII, however, make simple double stranded cut in the middle of the recognition sequence, resulting in blunt end or flush end. The use of restriction enzymes allows the DNA to be dissected into a precisely-defined set of specific segments. Using different enzymes, sites cleaved by different enzymes can be identified and ordered into a restriction map or physical map. Method of DNA Cleaving: Label three sterile microfuge tubes as E (for EcoRI), H (for HindIII) and l (for control). In tube E add 12µl distilled water, 2µl of 10 X Buffer (prepared by dissolving 108 g Tris, 55 g Borate and 7.4 g EDTA in 700 ml of distilled water, adjust pH to 8.3 and sterilize by autoclaving), 5ml of lamda DNA (1mg) and 1µl of EcorRI restriction enzyme (2 U). In tube H add same chemicals except 1µl of HindIII restriction enzyme (2 U) instead of EcorRI. In l tube add same chemicals but replace distilled water for restriction enzyme. Flick all tubes to mix well and spin for 5 minutes in microfuge. Incubate all tubes at 37oC for 60 minutes in water bath. Stop

reaction by incubating tubes at 65oC for 5 minutes. Subject contents of all three tubes to agarose gel electrophoresis for 1 hour at 100 V. Stain the gel with ethidium bromide. Fragments with be lined in each lane according to size. These can be compared and suitably analysed.

Method of Agarose Gel Electrophoresis:

Prepare 1% agarose in 1X TBE buffer; heat the mixture on a hot plate or microwave until the solution becomes clear; add Ethidium bromide (EtBr) (0.5 μg/ml in 0.5X TBE buffer) to agarose solution when it cools to 45-500C; clean the casting tray, place on table and adjust equilibrium bubble; position the comb 1 mm above the plate; pour solution into gel tray ensuring that there is no bubble between or under the teeth of the comb; allow the gel to set; pour some 1X TBE buffer over the gel; remove the comb; mix 1.5 μl each of DNA sample and bromophenol blue tracking dye (dissolve 70 g sucrose in 50 ml distilled water by heating, add 0.25 g bromophenol blue and 20 ml 0.5M EDTA, raise volume to 100 ml) and slowly load mixture into wells; connect assembly with power supply and run the gel at voltage of 80 V until the dye has travelled 75% of the distance; turn off the equipment, remove the gel and view under UV; alternately stain gel with 0.025% methylene blue for 20-30 minutes, destain with lukewarm water for 30 minutes and observe under white light.

DNA Cloning

A detailed analysis of DNA requires availability in sufficient quantity of DNA or its restriction fragments. DNA cloning is a technique to produce large quantities of a specific DNA segment. The technique of cloning has largely been
made possible through recombinant DNA technology. The DNA molecules from two different sources are treated with restriction enzyme that makes staggered cuts in DNA, leaving single-stranded tails in either of the cleaved DNA. These tails act as sticky ends and complementary ends of two different DNA molecules join to form double stranded recombinant DNA, in the presence of DNA ligase. For successful cloning,

one of the parental DNAs incorporated into recombinant DNA molecule is capable of self-replication, and is known as cloning vector. In practice, the gene or DNA fragment of interest is inserted into a specially-chosen
cloning vector, which is used as a vehicle for carrying foreign DNA into a suitable host cell, such as a bacterium.

Plasmid vector

Plasmids are extra-chromosomal doublestranded circular DNA molecules present in microorganisms, especially bacteria. The plasmid chosen as vector contains a gene for antibiotic resistance. In the most commonly employed technique (Figure 7.24-I), the recombinant plasmids (with foreign DNA inserted into plasmid) are added to an E. coli bacterial culture pretreated with calcium ions. When subjected to brief heat shock, such bacteria are stimulated to take up DNA from their surrounding medium. Once within the bacterial cell, the plasmid replicates autonomously and is passed on to the progeny during cell division. The bacteria containing recombinant plasmid can be separated by treatment with an antibiotic which removes bacterial cells without plasmid. Because a large number of different recombinant plasmids are formed, incorporating different segments, the one of interest can be separated by combined procedure of replica plating and in situ hybridization (Figure 7.24-III). Through replica plating, numerous dishes with representatives of the same bacterial colony are prepared. In one of the replica plates, cells are lysed and DNA fixed on to surface of nylon or nitrocellulose membrane. DNA is next denatured; membrane is incubated with labelled single stranded DNA probe, containing complementary sequence being sought. The unhybridized probe is washed away, and the location of labelled hybrids determined by autoradiography. Inrefined technique of fluorescence in situ hybridization (FISH), probe labelled with fluorescent dyes is used, and labelled hybrids localized with fluorescent microscope. The live representatives of the identified clones can be found on corresponding sites on the original plates, these cells are grown into large colonies, which serve to amplify recombinant DNA plasmid. After

sufficient amplification, the DNA is extracted and recombinant plasmid DNA is separated from bacterial DNA. The recombinant plasmid DNA is again treated with the same restriction enzyme that releases plasmid DNA from the cloned DNA segments. Latter can be separated from plasmid DNA by centrifugation.

Bacteriophage vector

Bacteriophage λ (lambda) is commonly used as a vector. The DNA of the phage is linear 50 kb in length. During treatment with restriction enzyme middle 15 kb segment of phage DNA which contains genes for lysis and can be dispensed with is replaced with foreign DNA. The resultant recombinant DNA is packed into phage heads in vitro (Figure 7.24-II). Phage particles can inject the recombinant DNA molecules into E. coli cells, where they will replicate and produce clones of recombinant DNA molecules. As lambda heads can accommodate molecules of only 45 to 50 kb size, it can accommodate inserts (foreign DNA fragments) of only 10-15 kb.

Cosmid vector

For inserting larger DNA insertion, cosmid vectors are used. A cosmid is a hybrid between plasmid and lambda phage. Cosmids combine plasmid's ability to replicate autonomously with in vitro packaging capacity of lamda phage. A cosmid vector can carry out inserts of 35 to 45 kb. Eukaryotic shuttle vectors Some of the most useful cloning vectors are shuttle vectors that can replicate in both E. coli and another species. Such shuttle vectors are very useful for genetic dissections. A yeast gene can be cloned in shuttle vector, subjected to site-specific mutagenesis in E. coli, and then moved back to the yeast to examine the effects of induced modifications in native host cells. Artificial chromosome vectors Attempts have been made over the recent years to develop vectors which can accommodate DNA sequences larger than 45 kb. One of the most important of these vectors is YAC (yeast artificial chromosome), which can accept DNA fragments as large as 1000 kb. More recently the use of BAC (bacterial

artificial chromosome) has become more common. BACs are specialized bacterial plasmids (F factors) that contain bacterial origin of replication, and can accommodate up to 300 kb of DNA segments.

Amplification through PCR

The earlier procedures for obtaining a large quantity of DNA were very cumbersome, involving the cloning of genes into bacteria, which replicate genes along with their own genome. The development of PCR (polymerase chain reaction) technique has now made it possible to obtain large number of copies of a gene using enzyme in place of bacteria. Small pieces of single-stranded DNA with known sequence are used as primers (Figure 7.25). These primers are built from templates of short regions of DNA that occur at either end (flanking) of DNA segment of interest, do not occur anywhere else in genome (unique), and are invariable (conserved) in all taxa to be investigated. The extracted DNA from a species is mixed with the primer, DNA polymerase (usually taq polymerase, which can tolerate heat), buffers, salts and free nucleotides in a tube. The mixture is alternately heated and cooled. Heating denatures DNA making it singlestranded. The subsequent cooling allows primers to bind to the complementary DNA sequences. Polymers are designed so that they can not bind with each other. The temperature is then raised to make polymerase active, bind to the already formed complex (DNA + polymerase), and begin synthesis of complementary strand (at DNA region not bound by primers) using free nucleotides. The temperature is raised further to denature DNA and the cycle repeated, thus making enough copies of DNA.

Method: Prepare PCR mix E by mixing (all previously placed in ice, taken out and thawed) 35µl autoclaved water (pH 7.0), 5µl 10X PCR buffer, 2µl mixture of deoxyribonucleotides (dNTPs- 10mM of each nucleotide: dATP, dCTP, dGTP and dTTP), 1µl Forward primer (10µm), 1µl Reverse primer (10µm), 5µl Genomic DNA template(25 ng/µl) and 1µl Taq DNA polymerase (5 units/µl); prepare one more PCR

mix without DNA template and label as C; put 20-50µl of each mixture in PCR tubes, briefly centrifuge, insert tubes into thermal cycler; program PCR machine for temperature 94°C (step 1, 5 min: primary denaturation), 94°C (step 2, 1 min: secondary denaturation), Annealing temperature (about 5-10°C less than Tm of the primers, step 3, 1 min), 72°C (step 4 and 5, 1.5 min), hold, 4 min; take out tubes when temperature comes down to 4°C and place in ice; analyze further by gel electrophoresis as described earlier.

DNA libraries

DNA libraries are collections of cloned DNA fragments. Two basic types of DNA libraries can be created. Genomic libraries are produced from the total DNA extracted from the nuclei and contain all of the DNA sequences of the species. cDNA libraries (cDNA—complementary DNA) on the other hand, are derived from DNA copies of usually the messenger RNA, and thus represent DNA sequences which are expressed in the species. This is significant because a large number of DNA sequences do not express themselves, and are of little significance. Sometimes, individual chromosomes of an organism are isolated by a procedure that sorts chromosomes based on size and the DNA content. The DNAs from the isolated chromosomes are then used to construct chromosome- specific DNA libraries, which facilitates the search for a gene that is known to reside on a particular chromosome. This is particularly useful for organisms with large genomes, such as humans. To construct a DNA library, the DNA from a species is randomly cleaved using enzymeswhich recognize short nucleotide sequences, the fragments are incorporated into lambda phage and multiple copies of each recombinant DNA obtained. These are stored and constitute a permanent collection of all DNA sequences present in the genome of a species. To construct a cDNA library using mRNA, a complementary stand of DNA is constructed by reverse transcriptase. RNA-DNA duplexes are converted into double-stranded DNA molecules by combined activity of ribonuclease H, DNA polymerase I, and DNA

ligase. The double-stranded DNA is incorporated into lambda phage and further processed as detailed above.

Gene Mapping

Above techniques contribute in developing the physical maps of gene. Whereas restriction enzymes enable cleavage at specific sites, the cloning and amplification techniques help in obtaining a large number of copies of fragments. Identification of the location of genes and DNA sequences on restriction fragments separated by gel electrophoresis constitutes an important step of genome mapping. The process of gene mapping has been simplified with the availability of cloned organelle genomes which are used as probes. In the commonly used Southern blot hybridization method (named after E. M. Southern, who published it in 1975), a cloned piece of chloroplast DNA (to be used as probe) is labelled with radioactive phosphorus and denatured to produce single-stranded DNA. The cleaved DNA from the specific species, after electrophoretic separation of fragments, is placed on a nylon or nitrocellulose membrane, and denatured by using alkaline solution and finally immobilized by drying or UV irradiation. It is renatured and allowed to bind to the radioactive probe on a nylon membrane. Only matching sequences will bind, and carry the radioactive tag. When transferred to an X-ray film the bound bands will appear as dark bands, which will show the positions of DNA sequences that have hybridized with the probe. The segments of different sizes can be ordered to generate physical maps.

Method: Perform gel electrophoresis of DNA sample; look for fluorescent bands; treat gel with 200 ml of 0.25 M HCl for 15 minutes; rinse with distilled water and treat twice with 200 ml denaturing solution (DS: 1 M NaCl, 0.5 M NaOH) for 15 minutes each; neutalize the gel by soaking in 200 ml of Neatralising solution (2.5 M NaCl, 0.5 M Tris-HCl (60.5 g/l), adjust pH to 7.4 with conc. HCl) for 15 minutes; Lay transfer buffer TB (20 X SSC, 3 M NaCl (175 g/l), 0.3 M Tri Sodium Citrate (88 g/l)), prewetted double layer of Whatman paper onto transfer tray, so as to reach both ends

of reservoir; put gel upside down on Whatman paper, roll gently with pipette to remove bubbles; place TB-prewetted nylon membrane on the gel; put 3 layers of TBprewetted Whatman paper on top of it; place 1 dry Whatman paper on top; put 3 layers of blotting papers on top of Whatman papers; pour 200-400 ml 20X SSC buffer into tray; add 10-20 cm layers or 2/3 of a pack of paper towels; place 500g weight on top and allow the transfer for 12-16 hours; take out membrane from assembly; crosslink DNA on nylon membrane by exposing it to UV light for 3-5 minutes; for detection either expose membrane to X-Ray film or else stain the membrane with 0.025% methylene blue for 20 minutes and then destain in water. The technique of Northern blot hybridization, (so named as it is opposite of Southern blot technique), is used to hybridize RNA molecules separated by electrophoresis. Denaturing is affected by formaldehyde, and after transfer to the membrane, the RNA blot is hybridized either with RNA probe or DNA probe. The procedure of gene mapping is sufficiently complex. It involves crossing two plants, selfing F1 and producing a large number of F2 plants. Genotypes of parents and offsprings are determined using various markers. Although physical maps can be constructed by identifying and aligning overlapping DNA fragments, more elaborate genetic maps are constructed using genetic markers. Genetic map can be unified with physical map using molecular markers. The physical map thus obtained will afford single framework for organizing and integrating diverse types of genetic information, including the position of chromosome bands, chromosome breakpoints, mutant genes, transcribed regions, and DNA sequences. Gene Sequencing Sequencing determines the exact order of the bases (adenine, cytosine, guanine and thymine) constituting nucleotides in a portion of a DNA and thus building an ultimatefine structure map of a gene or chromosome (Figure 7.26). Today, sequencing is a routine laboratory procedure. A complete sequence of human genome has been developed, as also the small annual weed Arabidopsis thaliana, developing into a strong genetic tool. Two main procedures of DNA sequencing are commonly used. In the first procedure developed by Allan Maxam and Walter Gilbert, the DNA chain is cleaved using four different chemical reactions, each targeting A, G, C or C+T. In the second procedure

developed by Fred Sanger (chain termination method) and colleagues, there is in vitro synthesis of DNA in presence of radioactive nucleotides and specific chain terminators to generate four populations of radioactively-labelled fragments that end with As, Gs, Cs and Ts, respectively. The procedure begins with obtaining identical DNA fragments up to about 500 bp using a restriction enzyme. The preparation is divided into four samples. Each sample is denatured into single strands, incubated with a short radioactively-labelled oligonucleotide complementary to 3' end of single strands. To each sample is also added DNA polymerase and all the four deoxyribonuclease triphosphate precursors (dNTPs). To one sample is now added chain terminator ddATP (2', 3' -dideoxyadenosine triphosphate), to the second ddGTP (2', 3' - dideoxyguanosine triphosphate), to the third ddCTP (2', 3' -dideoxycytidine triphosphate), and to the fourth ddTTP (2', 3' - dideoxythymidine triphosphate). The first sample after reaction will have all the segments terminated at As, the second at Gs, the third at Cs and the fourth at Ts. The fragments are separated on gel electrophoresis, and their positions determined by autoradiography. Different bands, representing different segments will be arranged like a ladder. By reading the ladder, a complete nucleotide sequence of DNA chain can be determined. In conventional slab-gel procedure, four different samples are loaded in four different wells on a gel. Nowadays, automated DNA sequencing machines are used which make use fluorescent dyes instead of radioactive nucleotide. The products of all four samples are run through single well, and photocells are used to detect the fluorescence as they pass through the well (tube or gel). The output is directly analyzed by a computer, which analyses, records and prints out the results. The PCR product can be sequenced directly using restriction enzymes. Since restriction sites are spread at several places on the DNA, the results are less sensitive to local vagaries of selection or differences in mutation rate. Sequencing of both the strands often minimizes errors.

Analysis of Sequence data

For the analysis of changes at the level of nucleotides and the amino acids, the alignment of DNA sequences derived from different taxa constitutes an important step. Alignment helps in detection of insertion, deletion or substitutions of one or more base pairs at different sites within a DNA. When comparing two sequences with L positions (nucleotides), of which D positions are different, the evolutionary distance counted A number of different models have been proposed to explain evolutionary distance between two sequences on account of nucleotide
changes.

Jukes-Cantor Model

T. Jukes and C. Cantor (1969) realized, even before the DNA sequences were available for analysis, that alignments between sequences with many differences might cause a significant underestimation of the actual number of substitutions that occurred since sequences last shared a common ancestor. They assumed that each nucleotide was as likely to change into any of the other three nucleotides. A can thus equally well change into T, C or G. Based on this assumption they created a mathematical model in which rate of change to any one of the three alternative nucleotides was assumed to be a, and the overall rate of substitution for any given nucleotide was 3a. According to this model, if a site within a gene was occupied by a C (t = 0), then the probability (P) that this site would still be same nucleotide at time1 (t = 1) would be PC(1) = 1 - 3a. On the other hand if C changed to some other nucleotide, the probability that after time t, the site would contain C can be calculated as: PC(t) = (3/4)e-4at The probability rate matrix for the changes in four nucleotides can be represented as under: It was , however, subsequently realized that transitions (change from purine to purine; pyrimidine to pyrimidine) proceed at much faster rate than transversions (purine to pyrimidine or vice versa), but the Jukes-Cantor model can still be taken into account for calculating the number of substitutions (the distance between two sequences) per site (K) when multiple substitutions were possible: K or djc= –3/4 1n (1 - (4/3)p)

Where p is the fraction of the nucleotide that a simple count reveals to be different between two sequences. It follows from the equation that if two sequences have fewer mismatches, p is small and the chance of multiple substitutions is also small. On the other hand if number of mismatches are large, the actual number of multiple substitutions per site will be considerably larger than what is counted. Once number of substitutions per site (K) is calculated, knowing the time taken for divergence (T), the rate of substitution (r) can be calculated as: r = K/(2T) For calculating substitution rates, data from at least two species should be available. If evolutionary rates between species are similar, substitution rates can help in calculating the dates of evolutionary events. Kimura two-parameter (K2P) Model The model was proposed by M. Kimura (1980) and accounts for different rates of nucleotide changes involving transitions and transversions. Supposing we assign value a for transitions and b for transversions, the probability rate matrix would be represented as: where P and Q are observed fractions of aligned sites whose two bases are related by a transition or a transversion, respectively. Once the sequences are generated, they must be aligned. First the sequences of a given length are aligned by arranging homologous nucleotides in corresponding columns. Alignment is simpler for conserved genes, where all taxa will have same number of nucleotides per gene. Some other genes which have some deletions, additions, inversions or translocations in some taxa, are difficult to align. Similarly DNA with multiple copies of a gene makes it difficult to assess homology. Several computer programs are available to produce alignment, but the assumptions used in each program should be carefully examined before the program canbe used for a particular set of taxa. In phylogenetic analysis each nucleotide position is considered as one character, and each of the four nucleotides as one character state. A large number of nucleotide positions, however don’t show variation among taxa, and of others that are variable are often uninformative because of being autapomorphic for a given taxon. This leaves only a small proportion of nucleotide positions that can be used for phylogenetic analysis. Chromosomal mutations such as additions, deletions and translocations are identified as evolutionary novelty, and are generally given more

weightage than individual nucleotides (Figure 7.27). Such chromosomal changes representing apomorphy are important and often used in establishing a lineage. Thus all members of subfamily Faboideae lack one of the inverted repeats found in the chloroplast DNA of most angiosperms. In our example illustrated in Figure 7.27, the four nucleotides are given coding from 0 to 3 for different nucleotides. Other strategies could also be used. Transitions (changefrom A to G or vice versa; or from C to T and vice versa) are more common than transversions (A to T, A to C, G to C, G to T; C to A, C to G, T to A, T to G). The latter are often given more weight depending upon the frequency of distribution in the taxa, more frequently the transitions are distributed, greater weight is consequently given to transversions. Thus if transitions occur 4 times more than transversions, a transition may be given weight of 1 and transversion a weight of 4. Computer programs such as DNAPARS, DNADIST, etc. of PHYLIP are available, which can read and analyse the DNA sequence data directly. Details are described under chapter on Developing Classifications. Whereas alignment of simple chloroplast genes such as rbCL is easier, others such as genes encoding RNAs, secondary structure (folding) of the molecule is also accounted for. The nucleotide differences that result in major changes in the structure of a product, such as ribosomal RNA or a protein, and may have greater effect in the plant function, often receive greater weight thanthose changes that do not affect the function. Several computer algorithms are available to evaluate and handle such analysis.

DNA Polymorphism

Utilization of sequence data in phylogenetic analysis involves the identification of unique sequences which show certain differences in different organisms or populations. These sequences, which could be used as genetic markers in identification of character-state differences between the target taxa, and ultimate construction of phylogenetic trees. The phenomenon is also known as DNA Fingerprinting or DNA polymorphism. The technique is now widely used in forensic investigations. A variety of methods have been developed to detect this polymorphism. Each method has its

own advantages and limitations, and suitable for a particular situation. New methods are being continu-ously developed. Some of the commonly used procedures are discussed below.

Single-Nucleotide Polymorphisms (SNPs)

DNA differences in a population may often be the result of differences in single nucleotide pair at a particular locus, say from C-G to TA. This may result in three genotypes in a population: homozygous with C-G at corresponding sites on both homologous chromosomes, homozygous with T-A at corresponding sites on both homologous chromosomes, and heterozygous with C-G in one chromosome and T-A in homologous chromosome. However, all SNPs are not located on coding sequences or genes. In human genome, for example, any two randomly chosen DNA molecules differ at one SNP site about every 1000- 3000 bp in protein coding DNA, but 500-1000 bp in noncoding DNA segments. SNPs aremost common types of genetic differences among human populations, and are uniformly distributed over the chromosomes. The SNPs can be easily detected if they are located in a cleavage site (Figure 7.28). Thus a sequence GAATTC can be cleaved by EcoRI, but a corresponding GAACTC sequence can't be cleaved as T has been replaced with C (and on the complementary segment A replaced by G). This will result in larger DNA fragment in the latter case.

Restriction Fragment Length Polymorphisms (RFLPs)

RFLP results from the fact that a mutation that causes changes in base sequence may result in loss or gain of a cleavage site, thus alleles differing in the presence or absence of a cleavage site. This may also result from SNPs located at cleavage sites as indicated earlier. As a result fragments of different lengths are yielded. The method is widely used for identification of individuals, species or populations. The DNA from a species is cleaved using a restriction enzyme (say EcoRI) yielding a certain number of fragments (Figure 7.29). These fragments can be separated using Southern blotting procedure, and a map of these constructed. These fragments are further fragmented

using another enzyme (say HindIII), and the data incorporated into original map. Restriction site fragments obtained are coded as characters and character-states for phylogenetic analysis. The absence or presence of a restriction site in closely related species and the presumed hybrids can also be detected by Southern blotting procedure. Species A, for example lacks restriction site at 3000 bp position (allele a, genotype aa), where this site is present in species B (allele A, genotype AA). Southern blotting technique will yield longer first restriction fragment of 5000 bp length for aa genotype, whereas it will yield fragment of 3000 bp length in AA genotypes. The heterozygous Aa genotype, presuming the alleles are codominant will yieldtwo fragments from homologous chromosomes, one of 3000 bp length and another of 5000 bp length. RFLP analysis, however, contains much lesser data than complete DNA sequencing, accounting only for presence or absence of sites 6-8 base pairs long, but the method affords advantage of surveying larger segments of DNA. The use of this method has, however, declined with the development of improved and less expensive sequencing techniques in the recent years.

Random Amplified

Polymorphisms (RAPDs)

RAPD method is commonly used for population studies and involves short (10bp) random PCR primers that will bind to the matching sequences on genome. The approach is useful for species where cloned DNA probes are not available (essential for Southern blotting method), or where DNA sequences are not known (necessary for PCR amplification where oligonucleotide primers have to be constructed). The method uses PCR primers of 8-10 nucleotides with random sequence. These primers are tried singly or in pairs in PCR reactions to amplify segments of DNA from a species. These short primers anneal at multiple sites on DNA, and those that anneal at suitable distance are able to amplify unknown region between them. The presence or absence of such amplified regions in different individuals can be suitably coded for analysis. The procedure helps in identifying different genotypes in the population. The morphologic

characters of interest are mapped according to their linkage to markers. The results of one of several primers used are shown in Figure 7.30. Gel electrophoresis yields 13 bands, of which four show polymorphism, the rest nine are monomorphic. Each of the polymorphic allele can be represented similarly as + for the presence of band, - for its absence, and if + is dominant, both genotypes +/+ and +/– will show this band, whereas it will be lacking in –/– genotype. Thus for the last band in the gel species B and C have –/– genotype, where as A and D are either +/+ or +/–.

Amplified Fragment Length Polymorphisms (AFLPs)

AFLP (amplified fragment length polymorphism) technology is used for nucleic acid fingerprinting, exploiting molecular genetic variations existing between closely related genomes in the form of restriction fragment length polymorphisms. AFLP procedure involves four basic steps (Figure 7.31). In first step DNAs from different sources are isolated and digested with appropriate restriction endonucleases (REs). For most plant DNAs, two REs are used: one a rare cutter having 6-bp recognition site, and the other a frequent cutter with 4-bp recognition site. In the second step, specific double-stranded oligodeoxynucleotide adapters (primer adapters) are ligated to the ends of the digested DNAs to generate chimeric molecules. These primers are so designed that they bind at both cut ends of fragments. In the third step the chimeric fragments are subjected to PCR amplification to provide sufficient template DNA for fingerprinting PCRs. During the fourth step, PCR products are resolved on through electrophoresis using polyacrylamide sequencing gel, which separates the amplified DNA fragments that exhibit length polymorphisms, enabling the recognition of numerous genetic markers. One of the earliest significant results of this method were obtained by Jansen and Palmer (1987), who found a unique order of genes in the large single-copy region of the chloroplast genome in Asteraceae. This unique order could be explained by single inversion of the DNA, a feature lacking in all other angiosperms, strongly confirming that the Asteraceae family is monophyletic. The family Poaceae, similarly, has three inversions in the chloroplast genome. Out of these three inversions,

one is unique to the family and confirms its monophyletic status. Of the other two, one is shared with Joinvilleaceae and one with both families Joinvilleaceae and Restionaceae, suggesting that these two are the sister groups of Poaceae.

Simple Tandem Repeat Polymorphisms (STRPs)

STRP results from the fact DNA molecules may differ in the number of copies of a sequence of few nucleotides repeated in tandem at a particular locus. In TGTGTG sequence, for example, two base pairs are repeated. Such repeated nucleotides are known as tandem repeats. STRPs present at different loci may differ in sequence andlength of repeating unit, and in minimum and maximum number of tandem copies occurring in the DNA of a population. A repeating sequence of 2-9 bp is often known as microsatellite or SSLP (Simple sequence length polymorphism), whereas one of 10-60 bp as minisatellite. If these repeated sequences show variation within a population or a species, they are known as variable number tandem repeats (VNTRs). At a given locus in different individuals, the length of tandem repeats may vary, because of irregularities of crossing over and replication, and as such can be used as genetic marker. Identification of microsatellites involves constructing primers that flank tandem repeats, and then using PCR technology to generate multiple copies of tandem repeat DNA, whose length can be determined by gel electrophoresis (Figure 7.32). VNTR technology generates data quickly and efficiently and is often used for population studies, for examining relationships within a species, or between closely related species. STRP is very useful in mapping, as a large number of alleles present in the population often have high proportion of genotypes that are heterozygous for different alleles. STRPs are widely used in DNA typing (DNA fingerprinting) involving identification of human individuals in criminal investigation.

Allozymes

Different forms of an enzyme differing in different alleles at the same locus constitute allozymes, as distinct from isozymes showing differences at different loci.

Allozymes are separated and detected using starch electrophoresis as against gel electrophoresis for DNA sequencing. Allozymes differing slightly in amino acid composition will take different charges and migrate differently, and can be identified using specific stains. Allozymes have traditionally been used to assess genetic variation within a population or a species, but they can also be used for phylogenetic analysis of closely related species. Allozyme data can be coded in a variety of ways. Each allele may be coded as a character and its presence or absence as character states. Alternately a locus may be treated as character, and unique allele combinations as character states (Figure 7.33). A comparison electrophoresis bands of four species for enzyme I with two allozymes can be coded as 0 for allozyme separated at position 18 and 1 for allozyme separated at position 21. Similarly, Enzyme II with three allozymes can be coded as 0 for 27, 1 for 31 and 2 for 35. Enzyme I, as such would be coded as 1 for species A-C with band at 21 and 0 for D for band at 18. For enzyme II, similarly species A has band at 31 coded as 1, B at 35 coded as 2 and C as 3 having bands at 31 and 35. Allozyme data can also be coded as loss of each allele as one state and gain as another state. Allozyme data can also be coded on the basis of allele frequency. A species with two alleles in the frequency of 90/10% would be coded differently from another species with same alleles but with frequency of 40/60%.

Examples of Molecular studies

Whereas considerable progress has been made in the mapping of chloroplast genome, similar success in nuclear genome is at its infancy. Questions of speciation are being addressed through genome mapping in Helianthus. Some progress has also been made in grasses and the family Solanaceae. In Helianthus, Riesberg and his co-workers (1996) reported that H. annuus and H. petiolaris differed by at least seven translocations and three inversions, which affected recombination and possibilities of introgression. The genome of hybrid derivative H. anomalus, was rearranged relative to both parents, and the species was partially reproductively isolated from both parents. They also created new hybrids between the two parental species and found that

chromosomal rearrangements were similar to the naturally occurring hybrid species, H. anomalus. Belford and Thomson (1979), using side-copy sequence hybridization in Atriplex concluded that division into two subgenera in this genus is not correct. Bayer et al., (1999) on the basis of sequence analyses of the plastid atpB and rbcLDNA, found a support for an expanded order Malvales, including most of the genera previously included in Sterculiaceae, Tiliaceae, Bombacaceae and Malvaceae. They propose to merge Sterculiaceae, Tiliaceae and Bombacaceae with Malvaceae and subdivide this enlarged family Malvaceae into nine subfamilies based on molecular, morphological and biogeographical data.

Grass Genome

Genome analysis of cereal grasses has provided useful information. Of the common cereal grasses, rice has the smallest genome (400 mb). Maize genome is 2500 mb, whereas the largest genome is found in wheat (17,000 mb). In spite of large variations in chromosome number and genome size, there are a number of genetic and physical linkages between single-copy genes that are remarkably conserved amid a background of very rapidly evolving repetitive DNA sequences. By comparison of rice chromosomes numbered R1 to R12 (Figure 7.34-I), with conserved regions marked in lower case (R1a, R1b, etc.), it is found that conserved regions homologous to rice are found in other cereals. The wheat monoploid chromosome set is designated W1 through W7. One region of W1 contains single-copy sequences that are homologous to those in rice segment R5a, another contains singlecopy sequences that are homologous to those in rice segment R10, and still another contains single-copy sequences homologous to those in rice segment R5b. Each of such conserved physical and genetic linkages is called a synteny group. It is notable that maize genome has repetition of segments, confirming that maize is a complete, very ancient tetraploid with two duplicated genome blocks rearranged relative to each other. Simultaneous comparison of above cereal grass genomes is better represented with the help of a circular diagram (Figure 7.34-II). The segments are arranged into a circle in the same

order in which they were aligned in the hypothetical ancestral chromosome. Because of the synteny groups in the genomes, homologous genes can often be identified by location alone. It must, however, remembered that the circular diagram is only for convenient representation; there is not indication that the ancestral grass chromosome was actually circular. It was a normal linear chromosome. A large number of workers have targeted the family Poaceae using different criteria and techniques. All molecular phylogenies point to the Stipeae to be an early-diverging lineage. The morphological characters of the Stipeae are thus a mixture of synapomorphies linking them with pooids and symplesiomorphies, which they share with many other grasses. The studies based on chloroplast gene: cpRFLP (Davis and Soreng, 1993), ndhF sequences (Catalan et al., 1997), and nuclear genes: through ITS (Hsiao et al., 1994), phytochrome b (Mathews and Sharrock, 1996), and granule bound starch synthase I (Mason-Gamer et al., 1998) all supported the same placement of Stipeae. Similar studies of comparison of results from chloroplast DNA and nuclear DNA in Triticeae, however, produced different results, although two chloroplast phylogenies constructed from RFLP (Mason-Gamer and Kellog, 1996) and rpoA sequences (Petersen and Seberg, 1997) produced similar results. New World Tetraploid Cottons Genomic studies in genus Gossypium (Wendel et al., 1995) using isozymes, nuclear ITS sequences, and chloroplast restriction site analysis, indicated that New world diploids are monophyletic, as are the Old World diploids. The New World tetraploid cottons, including G. hirsutum were formed by allopolyploidy of genomes A (from the Old World) and D (from the New World). It was found that H. hirsutum has a chloroplast derived from one of the African species, and it must have acquired it only about 1-2 million years ago, well after the formation of the Atlantic Ocean. Arabidopsis Genome Insignificant small crucifer, Arabidopsis thaliana (Figure 7.35), often ignored in the field, holds great promise for opening new frontiers of phylogenetic analysis. With its small genome size of 114.5 mbp (as compared to 165 mbp in Drosophila melanogaster and 3000 mbp in humans), the species is the most completely known genetically among all flowering plants. During the last 8 to 10 years, Arabidopsis thaliana has

become universally recognized as a model plant for such studies. Although it is a noncommercial member of the mustard family, it is favored among basic scientists because it develops, reproduces, and responds to stress and disease in much the same way as many crop plants. The choice of Arabidopsis as a genetic tool has been forced by the following attributes:

1. Small genome (114.5 Mb/125 Mb total).
2. Extensive genetic and physical maps of all 5 chromosomes.
3. A rapid life cycle (about 6 weeks from germination to mature seed).
4. Prolific seed production and easy cultivation in restricted space.
5. Efficient transformation methods utilizing Agrobacterium tumefaciens.
6. A large number of mutant lines and genomic resources.
7. Multinational research community of academic, government and industry laboratories.
8. Easy and inexpensive to grow.
9. Compared to other plants, it lacks the repeated, less-informative DNA sequences that complicate genome analysis.

The Ara bidopsis Genome Initiative (AGI) is an international collaboration to sequence the genome of the model plant Arabidopsis thaliana. Begun in 1996 with the goal of completing the genome sequence by 2004, the genome sequencing was completed at the end of 2000. Comprehensive information on Arabidopsis genome is available on the internet via The Arabidopsis Information Resource (TAIR), which provides a comprehensive resource for the scientific community working with Arabidopsis thaliana. TAIR is a collaboration between the Carnegie Institution of Washington Department of Plant Biology, Stanford, California, and the National Center for Genome Resources (NCGR), Santa Fe, New Mexico. Funding is provided by the National Science Foundation. Important studies on Arabidopsis thaliana have been devoted to genetic control of development. Transgenic plants of this species have been created that either overexpress or underexpress cyclin B. Overexpression of cyclin B results in accelerated rate of cell division; underexpression results in

decelerated rate. Plants with faster rate of cell division contain more cells and are somewhat larger than their wild type counterparts, but otherwise they look completely normal. Likewise, plants with the decreased rate of cell division have less than half the normal number of cells, but they grow at almost the same rate and reach almost the Figure 7.35 Arabidopsis thaliana, small annual herb from family Brassicaceae, whose genome is most completely known among the angiosperms, is aptly known as the guinea-pig of plant kingdom. same size as wild-type plants, because as the number of cells decrease, the individual cells get larger. The plants thus have abilityto adjust to abnormal growth conditions, asopposed to animals which frequently develop proliferative cancer cellsThe studies on genetic control of flower development in Arabidopsis have revealed interesting results. During floral development (as in other tetracyclic plants), each whorl of the floral parts (sepals, petals, stamens and carpels) arises from a separate whorl of initials. Three types of mutations result in three different phenotypes, one lacking sepals and petals, the second lacking petals and stamens and the third lacking stamens and carpels. Crosses between homozygous organisms have resulted in identification of four genetic groups (Table 7.2). Mutations in the gene ap2 (apetala-2) result in phenotype without sepals and petals. The phenotype lacking petals and stamens is caused by mutation in either of two genes, ap3 (apetala-3) or pi (pistillata). The genotype lacking stamens and carpels is caused by mutations in the gene ag (agamous). Each of these genes has been cloned and sequenced. They are all transcription factors, members of MAD box family of transcription factors, each containing a sequence of 58 amino acids. An interesting finding from this study is that mutation in any of the genes eliminates two floral organs belonging to adjacent whorls. The pattern suggests that ap2 is necessary for sepals and petals, ap3 and pi are both necessary for stamens and ag necessary for stamens and carpels. As mutant phenotypes are caused by loss-of-function in alleles, it may be inferred that ap2 is expressed in whorls 1 and 2, ap3 and pi expressed in whorls 2 and 3, and ag is expressed in whorls 3 and 4. The floral development in this plant is thus controlled by combinational effect of these four genes. Sepals develop from tissue in

which ap2 is active; petals by combination of ap2, ap3 and pi, stamens by combination of ap3, pi and ag; and carpels where only gene ag is expressed. This is graphically represented in figure 7.36. It is pertinent to remember that ap2 expression and ag expression are mutually exclusive. In presence of ap2 transcription factor ag is repressed, and in the presence of ag transcription factor, ap2 is repressed. Accordingly, in ap2 mutants, ag expression spreads to whorls 1 and 2, and, in ag mutants, ap2 expression spreads to whorls 3 and 4. This assumption enables us to explain the phenotypes of single and even double mutants. This pattern of gene expression has been assayed by in situ hybridization of RNA in floral cells with labelled probes for each of the genes. The results confirm the above assumption of repressive action of concerned genes. It is significant that triple mutation involves all the genes. The phenotype of ap2 pi ag triple mutant does not have any normal floral organs. There are concentric whorls of leaves instead.

Gene trees

Molecular systematics presents powerful tools for constructing phylogenetic trees. Commonly used methods over the recent years include studies on chloroplast DNA using restriction site polymorphism (cpRFLP), analysis of chloroplast gene for subunit F of NADP dehydrogenase (ndhF, in the small copy region), for ‘a’ and ‘b’ subunits of RNA polymerase II (rpoA and rpoC2, in a large single copy region), for ‘b’ subunit of ATP synthase (AtpB), ITS region of ribosome, phytochrome B, and granule bound starch synthaseI. An encouraging congruence of results of these diverse studies was met in tribe Stipeae of grasses. In other cases, results from chloroplast phylogeny and nuclear phylogeny did not agree, suggesting caution in relying on any attribute singly for constructing molecular phylogenies. The gene trees constructed from rbcL have great utility in angiosperms. Chase et al., (1993) attempted to yield the phylogeny of all seed plants using 499 rbcL sequences. The analysis proved a few sequences to be pseudogenes, and entire families were represented by single sequences. The data set have been reanalyzed by other authors to yield parsimonious

trees (Rice et al., 1997). RbcL data has supported that Caryophyllidae is monophyletic. It has also supported the union of family pairs Asclepiadaceae-Apocynaceae, Araliaceae-Apiaceae, and Brassicaceae-Capparaceae. The data also supported the polyphyletic nature of Saxifragaceae and Caprifoliaceae.