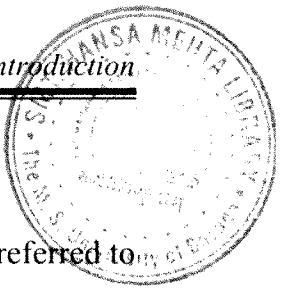


INTRODUCTION:

The variety of life on Earth and its biological diversity is commonly referred to as biodiversity. The number of species of plants, animals and microorganisms, the enormous diversity of genes in these species; the different ecosystems on the planet such as deserts, rainforests and coral reefs are all part of a biologically diverse Earth. Biological diversity means the variability among all living organisms from all sources including inter alias, terrestrial, marine and other aquatic ecosystems and biological diversity within a species and of ecosystems. *Biodiversity is the degree of variety in nature and not nature itself.* This universe is the creation of the supreme power meant for the benefit of all His creations. Individual species must therefore, learn to enjoy its benefits forming a part of the systems in close relation with other species. Let not only one species encroach upon the other's rights (Ishopanishad). The Earth abounds in a range of living organisms, from depths of the sea to the tropical rain forests and from the polar icecaps to the hot springs (Agarwal, 2002), Indian cosmology estimates 84 lakh species of living organisms in the entire universe but the biologists have described only 15 lakh species and estimated to have a total of around 50 lakhs. However, from further research the total number of species is now estimated to be as high as five crores (Gadgil, 1992). Over the years, the amount of work being done in each specialized branch of biological science has increased potentially. Moreover, with the advent of advanced methods, equipments and technologies enormous amount of knowledge with potential applications is being gathered; as a result a vast data is being generated and complied in various forms.

Biodiversity Worldwide:

Out of the 1.4 million known species of living organisms only about 2,50,000 are higher plants and 1.03 million are animals (WWF, 1989). According to another estimate, world wide there are 2,70,000 known species of vascular

plants (Table-1), which include ferns, fern allies, gymnosperms and flowering plants (IUCN, 1997).

Table – 1: Estimated number of known species worldwide

Taxonomic group	Number of Species
Bacteria	3,600
Blue green algae	1,700
Fungi	46,983
Bryophytes	1,700
Gymnosperms	750
Angiosperms	2,50,000

Biodiversity in India:

India is the seventh largest country of the world with an area of about 32,67,500 sq kms (Negi, 1993). India ranks sixth among the 12 mega-biodiversity centers of the world, and is home for an unusually large number of endemic species. It supports 15,000 species of flowering plants 5,000 of them exclusively providing shelter to 317 species of mammals (WCMC; 1992). India is unique, not so much of its numerical species but for the range of biodiversity attributable to a variety of biogeographically and physico-environmental situation. A Characteristic feature of Indian forests is the interspersion of human inhabitation with people dependant on the biological resources for their sustenance and livelihood.

Nature has endowed India with a rich biological diversity, which includes over 40,000 species of plants and 75,000 species of animals. India has about 12% of the global plant wealth amongst which there are nearly 3,000 tree species. However, nearly a third of the total plant species of India are endemic (Negi, 1993).

Floral Diversity in India:

Indian flora is extremely varies in extent, composition and endemism. In India there are over 30,000 species of higher plants belonging to 174 natural orders. There are over 600 species of Pteridophytes including ferns. Of the higher plants, there are 11,124 species of dicots with 1,831 genera. The family Orchidaceae is the largest family of flowering plants, contributing nearly 1,700 plant species.

Table – 2: Estimated number of plant species in India

Taxon	Number of species	Percentage
Bacteria	850	1.87
Fungi	23,000	50.79
Algae	2,500	5.52
Bryophytes	2.843	6.2
Pteridophytes	1,022	2.25
Gymnosperms	0.64	0.14
Angiosperms	15,000	33.1
Total	**45,279**	**100**

(BSI, 1994).

Most of the plant species in India are found in the forest areas, which occupy nearly 20% of the total geographical area of the country. The forests belonging to various categories are found distributed all over India depending upon the geographical conditions.

Salient Features of Indian Flora:

The flora of this region represents the sum total of different types and kinds of plants, which the vegetation represents. Amongst the various families of flowering plants in India the dominant are – Orchidaceae, Leguminaceae, Gramineae, Rubiaceae, Euphorbiaceae, Acanthaceae, Compositae, Cyperaceae,

Labiatae and Urticaceae. Amongst the families, Labiatae and Compositae are more abundant in the temperate regions while the rest are largely tropical in distribution. One interesting feature of the Indian flora is that Compositae, which is the richest family of flowering plants in the world, has a relatively low position in Indian flora. There are hundreds of species of grasses, over 100 species of Bamboos and more than 25 species of conifers in India.

Table – 3: Endemic Flora in India

Dicots	Total number
Species	11,124
Genera	1,813

Biodiversity in Gujarat State:

Gujarat has a variety of climatic and edaphic conditions. In spite of continuous onslaught of the biological resource, the overall range of biodiversity of Gujarat is still unmatched. Gujarat has many kinds of natural ecosystems due to varied environmental conditions (Chavan and Lal, 1998). The Gujarat flora is representing nearly 13% of the flora of the country. The state of Gujarat covers about 19.6 million km^2 and accounts for about 6.5% of total geographical area of country. Four out of ten biogeographic regions of the country- the highest among all states – are present in the Gujarat (Rodgers and Panwar, 1988) supporting a diversity of ecosystems. This state also has got a large population of tribal who are still dependent on the natural biological resources for their livelihood.

As far as the faunal diversity is concerned, out of 81,000 animal species recorded in India, more than 2,700 animal species is recorded in Gujarat. There are 94 species of mammals which is about 24% of the total mammal species found in the country. The state also has a rich avifauna comprising 498 bird species, which represents about 40% of the avifauna of the country. About 8.6% of the total geographical area of the state is protected under some or other

wildlife Sanctuaries or National parks. The rich biodiversity of the state is indicated by the presence of 4 National Parks and 22 wildlife Sanctuaries. The state has the first Marine National Park of the country at Jamnagar coast in the Gulf of Kuchchh which is considered to be one of the four major coral reef areas in the country having 52 species of corals (Pandey *et al*, 2005).

Table – 4: Recorded number of species in Gujarat

Taxon	Number of Species
Algae	1,933
Fungi	164
Bryophyta	8
Pteridophyta	16
Gymnosperms	1
Angiosperms	2,198
Total	**4,320**

(GEC, 1995)

Plant Wealth of Gujarat:

Floral wealth is one of the vital parts of the biodiversity. Floristic studies in Gujarat received focused attention since mid-nineteenth century. The first botanical work published by Mr. John Graham, Deputy Post master general, Bombay in 1,839 entitled "Catalogue of the plants grown in Bombay and its vicinity". The earliest documentation on plant diversity is of Dalzell and Gibson (1861). Palin (1880) published a list of plants of Kutch in "Bombay Gazetteer". Under the Bombay State, in the Gujarat more than 7,000 plant and animal species were recorded. Cook (1901-08) studied the flora of Bombay Presidency. Thaker (1910) explored the flora of Barda hills of Saurashtra. The first comprehensive account available is Flora of Kutch by Blatter (1908-09) and Thaker (1918) published a list of 614 flowering plants from north Gujarat. This work is supplemented by Saxton (1922). Dixit (1941) also contributed an important work on phytodiversity of Gujarat. After bifurcation from Bombay State in 1960, Santapau (1962) published 1st part of Flora of Saurashtra which was completed with part 2nd and 3rd by Bole and Pathak (1988). The significant contribution to the flora of Gujarat is provided by Shah (1978). The botanical

explorations are never ending and continue to add further knowledge about flora of Gujarat. The important documentation of plant diversity done in any part of state is short listed in table – 5.

Table – 5: Documentation of plant diversity of the State in research works and published floras

Author	Area / Title of work	Year
Hooker	Flora of British India	1872-1897
Cook	Flora of Bombay Presidency	1901-1908
Blatter	Flora of Kutchh	1908
Thaker	Flora of Barda hills	1910-1918
Saxton and Sedgwick	Flora of North Gujarat	1918
Kapadia	Grasses of Junaghadh, Poarbandar and Kutchh	1945-1954
Patel	Plant life of Maha Gujarat	1953
Raizada and Vaid	Grasses of Gujarat state	1953
Santapau	Glimpses of the vegetation of Okhamandal	1953
Santapau and Raizada	Contribution to the flora of Gir forest in Saurashtra	1955
Santapau	The flora of Saurashtra, Part – I, Plants of Saurtashtra A preliminary list, Contribution to the Botany Dangs forests in Gujarat	1962, 1953, 1955
Patel	List of trees, shrubs and woody climbers of Gujarat in Kathiyawad	1957
Chavan and Oza	Flora of Pavagadh	1966
Santapau and Janardhan	Flora of Saurashtra	1967
Sabnis	Flora of Baroda and Environs, Special Account on Cyperaceae	1967
Bedi	Ratanmahal and surrounding hills	1968
Deshpande	Tuwa and Panchmahal district	1968
Suryanarayan	Dangs	1968

Author	Area / Title of work	Year
Singh	Rajpipla forests (East)	1969
Yogi	North Gujarat	1970
Patel	Forest flora of Gujarat state	1971
Bhatt	Khedbrahma region in North Gujarat	1971
Patel	Rajpipla forests (Gora, Rajpipla and Netrang ranges)	1971
More	Parnera hills, Pardi and Udhwada area in South Gujarat	1972
Padte	Savli Taluka	1973
Vyas	Floristic composition of the river Narmada in Gujarat	1973
Karatela	Chhotaudepur forest division	1974
Parbia	Cyperaceae of Gujarat State	1974
Thaker	Kawant range forest in Cenrtal Gujarat	1974
Bhatt	Panchmahal districts	1975
Desai	Bansda forest	1976
Shah	Flora of Gujarat State	1978
Menon	Some parts of Saurashtra	1979
Joshi	Surat and its environ	1979
Bora	Dharampur, Kaprada and Nana Pondha ranges	1980
Rao	South Eastern Kutchh	1981
Mac	Surat district	1982
Vashi	Umarapada forest in South Gujarat	1985
Contractor	Vapi and Umergaon area in South Gujarat	1986
Thakker	Rajkot	1987
Bhatt	Navsari area	1987
Reddy	Dharampur forest	1987
Jani	Flora around the river Mohr (Kaira district) from origin to Mouth	1988
Oza	Around Bhavanagar	1991
Pradeepkumar	Shoolpaneshwar Wildlife Sanctuary	1993

Author	Area / Title of work	Year
Bhatt	Western Kutchh	1993
Raole	Endangered and endemic desert taxa	1993
Chavan	Gir	1993
Joshi	Cambay taluka	1994
Anonymous (GEC)	Biological diversity of Gujarat.	1996
Nagar	Biodiversity of Barda hills, Saurashtra region	2000
Singh and Parabia	The floral diversity of Gujarat state: A review	2003
Sing *et al*	Gujarat State Biodiversity conservation strategy and action plan	2005
Oza, Sisodia and Jasrai	Counting green heads of Vadodara city	2005
Sisodia, Oza and Jasrai	Tree Diversity in Vadodara City	2005
Oza and Rajput	Biodiversity of Gujarat forest trees	2006
Tyagi, Joshi, Oza, Dalal	Plants of Chhotaudepur	2007
Sisodia	Plant inventory and some ecological studies in Gir National park and sanctuary	2007
Singh and Gavli	Trees of Gujarat	2008
Tyagi, Joshi, Oza	Plants of Baria	2009
Tyagi, Oza, Joshi	Plants of Godhra	2009

A Comprehensive floristic studies are carried out in the state by various workers. Pilo *et al* (GEC, 1996) compiled all the previous publications and provided a complete biodiversity account (Flora and Fauna) of Gujarat state. The scientists of Botanical Survey of India, Jodhpur are working on flora of Gujarat states since 1996. In recent years, Pandey and Padhey (1997, 2005), Singh and Parabia (2003), Singh and Parmar (2004), Jasrai *et al* (2004), Oza and Rajput (2006), Parmar (2006, 2008) etc have expanded current knowledge of flora of the state. Vadodara is one of the major cities of Gujarat state. It is beautiful and well planned city and the main seat of learning and culture in the

state of Mahatma Gandhiji. The general topography is plain. The weather in general remains dry and sunny except during monsoon. The city is situated on the southern bank of the river Vishwamitri, the total city area is 108.22 sq kms The population is estimated to be approximately 1,50,000. The survey has been done which included visits to each street and roads to collect data on number of trees of each species and subsequent analysis of information gathered. All the species were classified according to Bentham and Hooker's system of classification using reference flora. Total 96,475 numbers of trees were noted from Angiosperms and 1,865 from Gymnosperm group. Some of the major tree species with their number of trees recorded in Vadodara city during study were *Polyalthia longifolia* Benth & Hk f (2893) *Azadirachta indica* A Juss (2710), *Peltophorum pterocarpum* (DC) Backeref K Heyne (2271), *Pongamia pinnata* L (1686), *Mangifera indica* L (1426), *Mimusops elengi* L (1351), *Ficus benghalensis* L (686) etc. During an inventory of tree species in the Vadodara city limits 181 tree species were reported from 58 families. The dominant families with major number of tree species were Caesalpiniaceae, Mimosaceae, Bignoniaceae, Moraceae and Arecaceae (Oza, Sisodia and Jasrai, 2005).

The Gujarat state forest department has done tree census (2004-05) in the non forest area in the Gujarat state, according to which, the number of trees of Vadodara District are 2,33,211. The survey showed the population of *Acacia nilotica* (L) Del being 14 %, *Azadirachta indica* A Juss is 13%, *Eucalyptus globulus* Labill is 11 % and *Bombax ceiba* L is 9 %. In the district the number of trees per head is only 2 which are lesser than the average. 53 % of these trees are of 10 - 45 cm GBH with the age of 3 to 15 years and 31 % of trees are of 46-90 cm GBH with the age of 16 to 30 years. Among 20 main tree species, *Azadirachta indica* A Juss is found to be the most abundant and distributed all over the Gujarat state. Results also demonstrate that there are 12 trees per hectare in Vadodara District which is much less than the average (Anonymous, 2005).

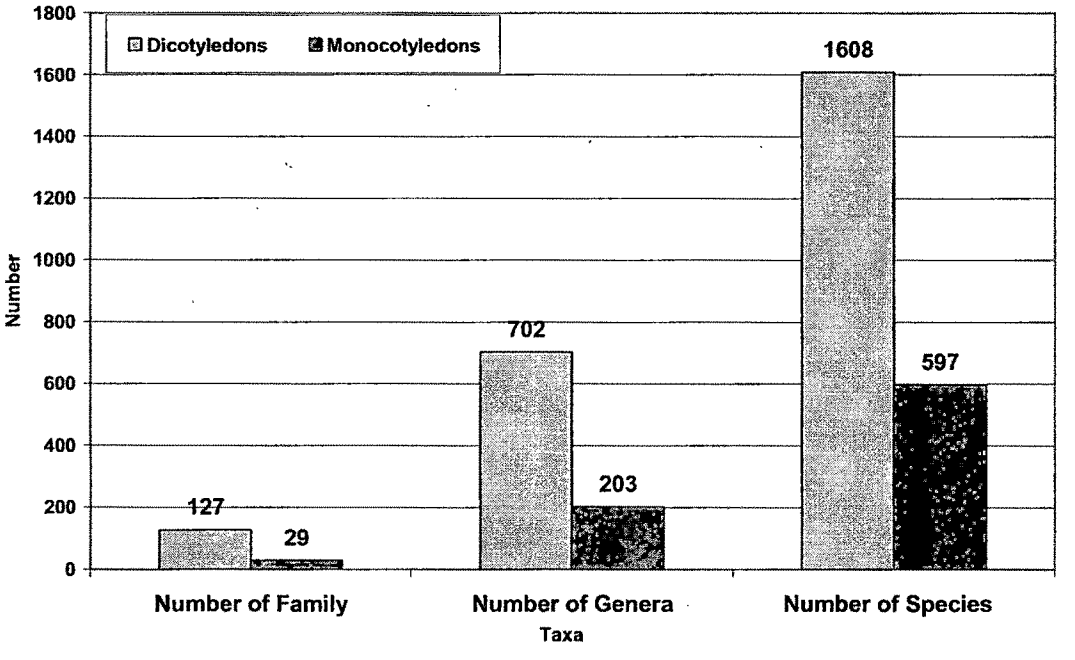
Total number of species recorded in Gujarat is 4,228. These occur on land, freshwater and marine habitats. The species diversity in Gujarat as compared to the country shows that numerically algae are represented the most in the State. More than 1,900 species of algae have been reported from Gujarat of which 210 are marine macro algae. Out of 16,500 species of Fungi reported from the Indian region, only 164 species are documented from Gujarat (GEC 1996). The groups like Bryophytes, Pteridophytes and Gymnosperms have the minimum number of recorded species. The angiosperm flora of the State is varied in extent and composition. According to tree census in the non-forest area done by the Gujarat state forest department, the approximate number of trees in the non forest area of the state is 2,510.33 lac (Anonymous, 2000).

Table – 6: Plant wealth of Gujarat state

Taxon	Number of recorded species		
	Gujarat	**India**	**World**
Algae	1,933	6,500	40,000
Fungi	164	16,500	89,000
Bryophyta	8	2,850	16,000
Pteridophyta	16	1,100	13,000
Gymnosperms	1	64	750
Angiosperms	2,106	17,500	2,50,000
Total	**4,228**	**45,436**	**4,08,750**

(GEC, 1996)

The Flora of Gujarat State was published by G. L. Shah (1978). This was first of its kind in the country. In this flora abut 1,808 angiospermic taxa belonging to 155 families and 827 genera are enlisted. The advent of Gujarat Ecology Commission, Vadodara renewed the interest of all concerned in the floristic exploration. They compiled all the work carried out after G. L. Shah and prepared the base line report. This report was published as "The biological Diversity of Gujarat Report" (1996) and the number of explored plant species reached to 2,198. Kshirsagar *et al* (2002) explored seven more angiospermic plant species from South Gujarat region. Therefore the number has reached to 2,205.



Graph – 1: Current status of angiospermic plant species from Gujarat

A comparison of dicot and monocot in Gujarat indicates that state represents 70% of the monocot families in the country. This is because of the general semi-arid to arid conditions in large part of the state, which favor grass growth.

Table – 7: Comparison of Dicots and Monocots species in Gujarat State

	Group	India	Gujarat	% In Gujarat over India
Families	Dicot	210	126	60.00
	Monocot	44	31	70.00
Genera	Dicot	2,281	682	29.80
	Monocot	705	204	28.90
Species	Dicot	12,543	1,525	12.16
	Monocot	4,236	577	13.62

Taxonomy – The Documented science:

Originally the term taxonomy referred to the science of classifying living organisms (known as alpha taxonomy); however, the term is now applied in a wider, more general sense and now may refer to a *classification* of things, as well as to the *principles* underlying such a classification. The world is full of individual objects, living and non-living. They all vary in all ways and to understand and discuss them individually is an impossible task. But if we divide them in to groups, they can be amenable. Any act of grouping objects into categories is an act of classification. The science that deals with classification is 'Taxonomy'. The term is first used by the famous French botanist, A. P. de Candolle in 1,813. The taxonomy is the study of the principles and practice of classification (Sivarajan, 1999).

Over the years plant taxonomy has become a monotonous and repugnant subject. Thus often it gives impression that it does not have a sound theoretical foundation and that it means nothing more than telling the names of plant species. Actually taxonomy is fundamentally a highly documented science. No other scientific subject relies as heavily on accumulation of literature as taxonomy. Main taxonomic literature are monographs and revisions, floras, manuals, general taxonomic indexes, bibliographies, catalogues, review serials, periodicals, glossaries, dictionaries, research reports, synopsis, maps and castography, supporting literatures and references. Plant identification is one of the important functions of taxonomy. It is done by comparison with the predetermined material. Plant identification is followed by naming of plants correctly. Plant identification and nomenclature is carried out with the help of herbaria, botanic gardens, taxonomic literature etc. plant taxonomy has shown a considerable stride from herbarium record to databanks, recording information from all possible corners of plants world. But this advancement is at various levels of its development in different parts of the work because of extreme climate diversity, floristic variability, degree of accessibility and

economic disparity in the different regions. Tropical Asia and Tropical Africa have richest floristic diversity in the world but are among the poorest in their economic resources to pursue floristic study. European countries, although rich in economic resource and have ill defined floristic variability, have been explored tremendously. India, with great climatic, geographical and floristic variability and insufficient resources, has a vast scope of floristic study as evidenced in the past few decades. Identification, classification and nomenclature are three fundamental components of plant taxonomy. A plant taxonomist is aimed at identification, nomenclature and classification of plants in order to study them in orderly and scientific manner. This facilitates a botanist to know what plant he is working on, before he passed on information about it to others. For proper identification, well identified and arranged herbarium specimens, dichotomous keys, polyclaves and computer aided identification should be adopted as an aid.

Identification is to recognize an unknown specimen and is the first and foremost aspect of the study. The identification of unknown plant is based on similarities and differences with an already identified plant and is generally made by comparing the identified pant directly with that of another plant which has been authoritatively predetermined or indirectly with the help of the literature on keys, floras, monographs, manuals, description and illustrations which are important source of identification. The literature regarding any particular specimen or taxa can also be obtained from the Black and Atwood's Guide to the flora of the world. The most recent and widely used method is of taking photographs of plants and plant parts and uploading these pictures on the websites and making an enquiry from concerned websites to look at the photograph and send their comment to enquirer (Pandey and Misra, 2008).

Nomenclature deals with the determination of a correct name to be applied to a known taxon or to a known plant and a system of assigning names to plant on scientific basis is called as plant nomenclature. Botany can not make

satisfactory progress with out a precise system of nomenclature which is used by majority of botanists. In fact, nomenclature is the basic ground work on which the edifice not only of taxonomy but the study of entire botany is built. With the accumulation of knowledge in the fields of ecology, anatomy, embryology, genetics, physiology etc the classical concept of taxonomy has undergone a radical change. There is resurgence on interest in the understanding of plant nomenclature. But, despite availability of considerable literature on plant nomenclature there has been no effort in the direction of consolidation and regimenting of this vast literature into an organized manner so that a student of nomenclature would find it easy to comprehend the basic knowledge about nomenclature (Naqshi, 1993).

Botanical keys are very important for identification of taxa. The essential feature of published taxonomic work on plants is keys. A key is a device for easily identifying an unknown plant by a sequence of choices between two or more statements. It is far more efficient to identify a specimen by the use of keys compared to shuffling through a stack of previously named herbarium specimens. A key is fundamentally similar to an outline where the topics are arranged in order of descending importance. The use of key provides the correct identity of a specimen by a process of elimination. Keys are constructed using contrasting characteristics to divide the possible names in the key into smaller and smaller groups. Statements in the keys are based on the characters of the plants. The first step in identification of an unknown plant involves the use of keys to determine the family. Next is the key to genera which will provide the generic name. After the genus is determined, the keying process is repeated within the genus for a determination of the species.

There are mainly two types of keys *viz* Diagnostic keys and artificial keys. Monographs and floras usually incorporate diagnostic keys by use of which the reader may identify an unknown plant. Key do not offer descriptions of the plants concerned, but state only the essential diagnostic characters by means of

which the taxa can be identified. Ideally they use the most conspicuous and clear-cut characters, without special regard to those considered taxonomically the most important. Artificially keys are of two main types (i) single-access or sequential keys and (ii) multi access keys. Sequential keys are nearly always in the form of the dichotomous keys, apparently first used by R. Morison in his *Plantarum Umbelliferarum Distributio Nova* (1672). The dichotomous keys consist of a series of couplets or mutually exclusive pairs of statements, each statement of a pair leading on to a further couplet. At each couplet a decision to follow one lead or the other has to be taken, so that the number of possible taxa with which the unknown specimen can be identified is successively reduced until there is only one possibility. These keys can be written in one of the two ways: bracketed or indented. Multi-access keys are usually produced not on pages in a book but on separate punched cards. They were first successfully operated by A. T. J. Bianchi in 1931 and by S. H. Clarke in 1936. Various more simple or visually attractive methods based upon the principle of multi-access key have been used in more recent years, such as Tabular key to Ranunculus subgenus Batrachium and the Lateral key to the common British Poaceae. The existence of so many sorts of key in addition to the traditional dichotomous keys indicates disenchantment with the latter. The dichotomous trees are inflexible and fallible; if the character is missing in the set of sequence then progress is halted. One quite minor error by the user can lead to a totally wrong answer, and often the mistakes can be realized only at the end of the procedure. The use of keys is often one of the main difficulties experienced by the beginners in taxonomy, and new and better methods of identification are evidently desirable.

Computers in identification:

In recent practices computers have been increasingly used in data collection, processing and integration. Computers can also be used in plant identification even if the user is not well trained botanist. Appropriate programs may be developed, which can construct a taxonomic key based on the taxonomic

information about the taxa, in the same way and based on the same logic which is used by man to construct keys manually. Such keys permanently stored in a computer can be handled as above for the step wise process of identification. Simultaneous character-set identification methods can be used in which whole set of characters of the unknown plant may be fed into the computer simultaneously and a computer programme used to compare the description matches. If the complete information is not available, the computer program may be able to suggest possible alternate identification. Even though such computerized methods hold more hope for the future than the present day tool, but will never have much direct effect on field botanists. The construction of progarmmes for computerized database and identification system may be the solution to overcome the limitations of current methods.

Furthermore, the taxonomic literature has increased quantitatively to such an extent that some people called it an embarrassment. It is disturbing that one who wants to collect comparative data on any given group may find even elementary comparisons unavailable. The situation is so exasperating that Watson (1971) commented that it is easier and probably more reliable to start at ground level, making one's own observations, than to search the literature of data on salient aspects of general morphology. This, in fact, has persuaded several biologists to discuss alternative devices for more efficient storage and retrieval of information. The enormous number of organisms, and the potential number of characteristics that could be recorded of reach, call for the use of modern Electronic Data Processing Equipment to store and retrieve information in meaning full configuration. Thus the concept of use of Bioinformatics in taxonomy came in to existence.

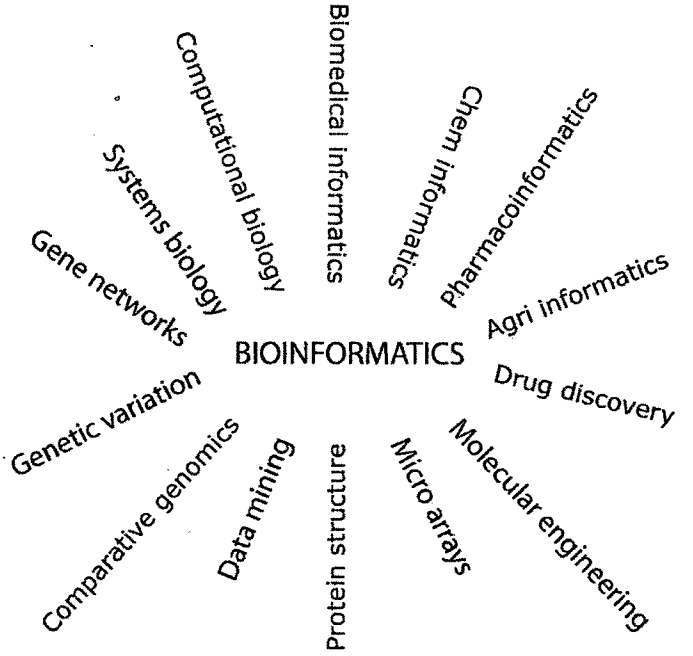
Bioinformatics – A revolutionary:

Bioinformatics, in simple terms can be considered as an application of computer technology to the management of biological information. Bioinformatics is the use of IT in biotechnology for the data storage, data

warehousing and analyzing. It is the fusion of biology and computer science. The science of informatics is concerned with the representation, organization, manipulation, distribution, maintenance and use of information particularly in digital form, the fictional aspect of bioinformatics is the representation, storage and distribution of data. In Bioinformatics knowledge of many branches are required like biology, mathematics, computer science, laws of physics & chemistry, and of course sound knowledge of IT to analyze biotech data. Bioinformatics is not limited to the computing data, but in reality it can be used to solve many biological problems and find out how living things works. According to Attwood and Parry smith (2001), the term bioinformatics is used to encompass almost all computer applications in biological sciences, it was originally coined in the mid-1980's for the analysis of biological sequence data. If we look at the history, bioinformatics can possibly be traced back to the 1920s. During that period, A. J. Lotka and V. Voltera introduced mathematical models representing host/prey interaction. This was perhaps the first attempt to mathematically represent a model to achieve a cyclic balance in mean population density, i.e. to attain a dynamic equilibrium. It was however only recently that the development of powerful computers and the availability of experimental data that can be readily treated by computation launched bioinformatics as an independent field. It is now a rapidly evolving field with multi disciplinary applications (Rastogi, Mendiratta & Rastogi, 2004). Till date bioinformatics is mainly used in the field of medicine and genomics – to reveal the information about the genomes of various organisms, but the rate at which the bioinformatics is growing and spreading is enormous and has empowered the research and development in the biological field right from the level of single nucleotide to the vast classifications.

Bioinformatics is being used in following fields: Molecular medicine, Personalized medicine, Preventative medicine, Gene therapy, Drug development, Microbial genome applications, Waste cleanup, Climate change Studies, Alternative energy sources, Biotechnology, Antibiotic resistance,

Forensic analysis of microbes, Bio-weapon creation, Evolutionary studies, Crop improvement, Insect resistance, Improve nutritional quality, Development of Drought resistance varieties, Vetinary Science etc.



Application in plant science - In Taxonomy and phytoresources:

The very important and one of the main branches of plant biology – Taxonomy can also assisted by bioinformatics, it can be useful as the medium to compile the data of each species, or the data of specific area. Lots of work is being done in this direction. Online floras provide description about plants and simply by inserting the characters user can get the genus to which the plant belongs. The concept of e – herbarium has been already introduced (Jasrai *et al*, 2000) and has generated over whelming response. In addition to commonly adopted taxonomic practices, adapting information technology and preparing computer herbarium sheets of plant taxa, can further strengthen the science of taxonomy. Such a modified technique will help to check some of the existing gaps in plant

identification procedures. Preparation of computer herbarium sheets involves photographing, scanning and systematic arrangement of the collected data in the form of a herbarium sheet.

Components of bioinformatics:

There are mainly three components of bioinformatics.

- The creation of database allowing the storage and management of large of biological data sets.
- The developments of algorithms and statistics to determine relationships among members of large data sets.
- The use of these tools for the analysis and interpretation of various types of biological data.

The computer hardware and its peripherals without an operating system are like a car without fuel.

Operating system:

An operating system may be defined as a set of system programs that control and coordinate the operation of computer system.

Language: These are the interpreters between the user and the computer; they convert the human readable source codes into the binary codes of computer and vice versa. There are several languages like C, C^{++}, JAVA, FORTRAN, Visual Basics etc. that can be used for the purpose.

Database:

Assemblage of the analyzed biological information into central, shareable resources is a database. It deals with the pool of information in a concerted way so as to establish, maintain and disseminate the data, providing easy – to – use software so as to access the information it contains, and designing state of the art analysis tools to visualize and interpret the structural and functional clues latent in the data. Databases prove to be effectively electronic filling cabinets, a convenient, efficient and highly specified method for storing of vast amounts of information. There are various types of database that vary from one another

based on the requirements depending upon the query the user can decide the database to be used. A database can be defined as a structured collection of records stored in a computer so that a program can consult it to answer queries. The records retrieved in answer to queries become information that can be used to make decisions. The computer program used to manage and query a database is known as a database management system (DBMS). The properties and design of database systems are included in the study of information science. Database management systems are usually categorized according to the data model that they support: relational, object-relational, network, and so on. The data model will tend to determine the query languages that are available to access the database.

The first database management systems were developed in the 1960s. A pioneer in the field was Charles Bachman. Bachman's early papers show that his aim was to make more effective use of the new direct access storage devices becoming available: until then, data processing had been based on punched cards and magnetic tape, so that serial processing was the dominant activity. Two key data models arose at this time: CODASYL developed the network model based on Bachman's ideas, and (apparently independently) the hierarchical model was used in a system developed by North American Rockwell, later adopted by IBM as the cornerstone of their IMS product. While IMS along with the CODASYL IDMS were the big, high visibility databases developed in the 1960's, several others were also born in that decade, some of which have a significant installed base today. In the 2000s, the fashionable area for innovation is the XML database. As with object databases, this has spawned a new collection of startup companies, but at the same time the key ideas are being integrated into the established relational products. XML databases aim to remove the traditional divide between documents and data, allowing all the information resources of an organization to be held in one place, whether they are highly structured or not. Databases are used in many applications, spanning virtually the entire range of computer software. Databases are the preferred

method of storage for large multiuser applications, where coordination between many users is needed. Even individual users find them convenient, and many electronic mail programs and personal organizers are based on standard database technology. A database is also a place where you can store data and then arrange that data easily and efficiently.

More recently, SOAP-based interfaces have been developed for a wide variety of bioinformatics applications allowing an application running on one computer in one part of the world to use algorithms, data and computing resources on servers in other parts of the world. A large availability of these SOAP-based bioinformatics web services, along with the open source bioinformatics collections, lead to the next generation of bioinformatics tools: the integrated bioinformatics platform. These tools range from a collection of standalone tools with a common data format under a single, slick standalone or web-based interface, to integrative and extensible bioinformatics workflow development environments.

Advantages India has:

India is well placed to take the global leadership in genome analysis, as is in a unique position in terms of genetic resources. India has several ethnic populations that are valuable in providing information about disease predisposition and susceptibility, which in turn will help in drug discovery. The genomic data provides information about the sequence, but it doesn't give information about the function. It is still not possible to predict the actual 3-D structure of proteins. This is a key area of work as tools to predict correct folding patterns of proteins will help drug design research substantially. India has the potential to lead if it invests in this area. Looking at this biotech and pharma companies need tremendous software support. Software expertise is required to write algorithms, develop software for existing algorithms, manage databases, and in final process of drug discovery.

Some major opportunity areas for IT companies include:

- Improving content and utility of databases
- Developing better tools for data generation, capture and annotation
- Developing and improving tools and databases for comprehensive functional studies
- Developing and improving tools for representing and analyzing sequence similarity and variation
- Creating mechanisms to support effective approaches for producing robust software that can be widely shared.

As major pharmaceutical and genome-based biotech companies invest heavily in software, Indian IT companies have a great business opportunity to offer complete database solutions to major pharmaceutical and genome-based biotech companies in the world. With an objective to help and rise bioinformatics sector to the world map, the Bioinformatics Society of India (Inbios) has been working since August 2001. The Inbios already has over 270 members in a short span of one and half years. It has become a common informal platform for the younger generation to learn and contribute to this sun rising field in India.

Data collection:

To plan any type of database first we need to decide the proper method of data collection. There are two types of data i.e. primary and secondary. The primary data are those which are collected fresh and for the first time, and thus happen to be the original in character. The secondary data, on the other hand, are those which have already been collected by someone else and which have already been passed through the statistical process. The primary data is been collected during experiments in an experimental research but in case we do research of the descriptive type and perform surveys, whether sample surveys or census surveys, then we can obtain primary data either through observation or through personal interviews. There are several methods of collecting primary data, particularly in surveys and descriptive researches. Important ones are:

(i) Observation method – To study behavioral sciences and group studies. Observation methods includes structured observation which is appropriate in descriptive studies, whereas unstructured observation is an exploratory study during observational procedure. If the observer observes by making himself, more or less, a member of the group he is observing so that he can experience what the members of the group experience, the observation is called as the participant observation. When the observer observes as a detached emissary without any attempt on his part to experience through participation what others feel, the observation of this type is termed as non-participant observation. If the observation takes place in the natural setting, it may be termed as uncontrolled observation, but when observation takes place according to definite pre-arranged plans, involving experimental procedure, the same is then termed as controlled observation.

(ii) Interview method: It involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. This method includes personal interviews and telephonic interviews. Personal interview method requires a person known as the interviewer asking questions generally in a face to face contact to the other person or persons. The method of collecting information through personal interviews can be done with structured or unstructured interviews. Focused interview is meant to focus attention on the given experience of the respondent and its effects. Telephonic interview includes contacting respondents on telephone itself. It is not very widely used method, but plays important part in industrial surveys.

(iii) Through questionnaire: This is the most popular method of data collection particularly for big enquiries among private individuals, research workers, private and pubic organizations and even by governments. In this method a questionnaire is sent to the persons concerned with a request to answer the questions and return the questionnaire.

(iv) Through schedules: This method of data collection is very much like the collection of data through questionnaire, with little different lie in the fact that schedules i.e. Performa containing a set of questions are being filled in by the enumerators who are specially appointed for the purpose. This method is very useful in extensive enquiries and can lead to fairly reliable results. However, it is very expensive method and is usually adopted in investigations conducted by governmental agencies or by some big organizations. Population census all over the world is conducted through this method.

(v) Other modules like – (a) Warranty cards
(b) Distributor audits
(c) Pantry audits
(d) Consumer panels
(e) Using mechanical devices
(f) Through projective techniques
(g) Depth interviews
(h) Content analysis

Secondary data are the data that are already available. When the researcher utilizes secondary data, then he has to look into various sources from where he can obtain them. In this case he is certainly not confronted with the problems that are usually associated with the collection of original data. Secondary data may either be published data or unpublished data. Usually published data are available in form of various publications of the central, state and local governments. Also, various publication of foreign governments or of international bodies and their subsidiary organizations, technical and trade journal, books, magazines and newspapers, reports and publications of various associations connected with business and industry, reports prepared by research scholars, universities, economists etc in different fields, public records and statistics, historical documents and other sources of published information. The sources of unpublished data can be in form of diaries, letters, unpublished

biographies and autobiographies which can be available with research workers, trade associations, labour bureaus and other public or private individuals and organizations. If the secondary data are reliable, suitable, adequate and readily available from authentic sources then it will be economical to use these data instead of spending time and energy in field surveys for collecting information. There may be wealth of usable information in the already available data which must be used by an intelligent researcher with due precaution (Kothari, 2003).

Plenty of work is going on in this field. However, we don't have single media to access this enormous information at one desk. Now a days many institutions and experts are preparing database in their respective fields in digital form. NBRI Bioinformatics Centre is maintaining a World Database of 19,000 taxa. International Legume Database and Information Service (ILDIS) created database on Leguminosae. The world biodiversity database CD-ROM was set up by ETI (Expert Centre for Taxonomic identification) at Amsterdam to provide easily accessible high quality taxonomic and ecological information. A CD-ROM named as 'Sasya Sahyadri' has been prepared for the trees of Western Ghats by Dr. K. N. Ganeshaiah and Dr. R. Uma Shannkar, University of Agriculture, Bangalore in 2004. The PROSEA Foundation, the University of Wageningen and an international network developed a CD-ROM containing detailed information on 400 genera and 2900 timber yielding species. Flora Europaea is available as CD-ROM for. Digital flora of Texas is available on www.botany.cs.tamu.edu/FLORA/perdeck. The scope of plant taxonomy has greatly expanded on account of our better understanding of the plant wealth with the help of Geographic Information System (GIS) and Remote Sensing. The electronic revolution, with the development of websites, has brought the plant systematic to become dominant component of plant sciences. Outline on the system of classification proposed by Thorne (2003) is available at www.rsab.org/publications/angiosp.htm. Many searchable databases such as Index Nominum Genericorum (ING) for all published generic names of plants, World Biodiversity database (WBD) for published scientific names of plants

and animals under species 2000 project, International Plant Names Index (IPNI) for database of names and associated basic bibliographical details of all seed plants incorporating data from Index Kewensis and Gray Cards Index, are already operative. Missouri Botanical Garden maintains TROPICOS which is the world's largest botanical database and is estimated to have 9,20,000 plant names and more than 18,00,000 specimen records. We can get the information online by just few clicks through online journals, books and paper uploaded on World Wide Web. But, the information about plants on international websites is more related to the respective area and environment. And also the information is so wide and sometimes complicated that even on the internet; it becomes tedious to gather the useful information about single plant species. Biological data is complex and interrelated. Hence the vast amount of biological information needs to be stored, organized and indexed so that the information can be retrieved and used. Therefore there should be some media to connect all the information on one site itself. The main purpose of the study is to prepare such a programme so that one can get the information about the single species at one site only which will be easy to use and is less time consuming. As a starting effort the present study has been done to generate a database for major tree forms of Gujarat state, and to produce a retrieval system as per the requirement. The database will cover almost all available data for each tree species, so that it will help not only in correct identification of a species but provide all available information about that species. The study is restricted to tree forms available in Gujarat state including various aspects of bioinformatics. As tree species are perennial so their phylogeny is easy to study and the time available to study their characters is more. They last for years so revision or monitoring would be easier. Thus, assessment of data will become simple. In fact, the project is to prepare a soft program on trees of Gujarat with relevant scientific information including their identification based on general plant characteristics. Based on records, important trees of Gujarat were selected for the study.