CHAPTER 1
INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Overview:

The species has been treated as a fundamental unit in biology (Hull 1977) and, important component of biodiversity (Sites and Crandall 1997). Almost all studies in biology, whether at the level of molecules, cells, individuals or population, are typically referenced to the level of the species. Despite a plethora of works in systematics have been completed through the last 2½ centuries, a large number of the earth's biota yet remained undesirable. Rising extinction rates and the need for increased biological monitoring lend urgency to this task. A reliable and accessible classification of species is fundamental to research in ecology, evolutionary biology and biodiversity and conservation biology. While approximately 1.8 million species have been described to date and this represents only a fraction of the actual diversity on Earth (Wilson 2003). Taxonomic keys are often not good enough to unambiguously identify specimens to the species level. Owing to the constant threat of biodiversity loss, there is an urgent need to accelerate the pace of species discovery and internet accessible taxonomic database (Godfray 2002). Even the routine identification of known species are difficult, often requiring highly specialized knowledge and representing a limiting factor in ecological studies and biodiversity inventories. Recently, several molecular methods have been employed in biodiversity studies owing to the fact of efficient DNA-based methods in the delineation and identification of species (Floyd et al., 2002; Tautz et al., 2003; Hebert et al., 2003a; Blaxter 2004; Hollingsworth et al., 2009;).

The use of DNA-based methods for the delineation and discovery of new species, and thus their broader role in taxonomy, represents a contentious issue in this regard. Unfortunately, much of this debate has remained rhetorical, with a limited empirical assessment of the benefits and limitations of DNA-assisted programs of

species discovery. The objective of any method of species delineation, including DNA-based approaches, is to identify reproductively isolated groups of organisms that warrant classification as distinct species. It is widely acknowledged, and reflected in the Linnaean taxonomic system, that living organism's fall into largely discrete groupings recognizable by differences in morphology or other traits. It is then the role of taxonomy to define and name these groupings. However, their recognition may be difficult because diagnostic traits are lacking, or species divergences are very small. Hence, even in cases of biologically distinct species, their delineation will depend on the thoroughness of study and the interpretation of complex, sometimes variable traits. Until now, DNA barcoding studies have successfully achieved the identification of pre-defined species and thus the barcode tag in the form of nucleotide sequence have been generated and available in the database (Hebert et al., 2003b; Hogg and Hebert 2004; Smith et al., 2005; Vences et al., 2005). The future objective of DNA barcoding is to help identification of species in the field within a moment of time using the DNA barcode machine that is in the process of development. Therefore, it is urgent to generate the barcode data for the known species from different geographical regions. In the context, Indian medicinal plants are the important national resources and popular for its richness. Hence, barcoding of medicinal plants is a major subject of research at the present time.

1.2 Medicinal plant in India and around : History and prospective

The Indian saga has a long heritage of using numerous medicinal and aromatic plants (MAPs) for human health care, and the nation is bestowed with rich resources of plant bio-diversity distributed in various ecological conditions. It is the home of about 17,000 of global plant species and expected to comprise even more which are yet to be fully explored. Several drugs of plant origin are described in the traditional texts of Indian systems of medicine like Ayurveda for their healing properties under the term 'Vranaropaka'. Besides the classical systems of Indian Medicine, the folk and the tribal medicine also employ a number of plant products for treatment of various

diseases. Some of these plants have been screened scientifically for the evaluation of their activity in different pharmacological models and human subjects, but the potential of most of the plants is mostly unexplored. A number of scientific investigations have highlighted the importance and the contribution of many plant families i.e. Asteraceae, Liliaceae, Apocynaceae, Solanaceae, Caesalpinaceae, Rutaceae, Piperaceae, Sapotaceae. According to World Health Organization (WHO), the global market for herbal product is expected to be US$ 5 Trillion by the year 2050 (WHO 2005). The present export of herbal raw material and medicine from India is about US$ 100-114 million approximately per year (Mukherjee and Wahile 2006). Global trend leading to increased demand of medicinal plants for pharmaceuticals, nutraceuticals, cosmetics and other products is an opportunity sector for Indian trade and commerce. Widespread and growing use of botanicals has created public health challenges globally in term of quality, safety and efficiency. Scientifically validated and technologically standardized botanical medicines will play an important role in future advancement in health care. The development of parameters for standardization and quality control of botanicals is a challenging task. Research on identifying medicinal plant by using modern biotechnological tool is one of the developing areas in modern biomedical sciences.

Medicinal plants play a vital role in the development of new drugs. During 1950-1970 approximately 100 plants based new drugs were introduced including deserpidine, reseinnamine, reserpine, vinblastine and vincristine which are derived from higher plants. From 1971 to 1990 new drugs such as ectoposide, E-guggulsterone, teniposide, nabilone, plaunotol, Z-guggulsterone, lectinan, artemisinin and ginkgolides appeared all over the world. 2% of drugs were introduced from 1991 to 1995 including paciltaxel, toptecan, gomishin, irinotecan etc (Heinrich and Gibbons 2001). Plant based drugs provide an outstanding contribution to modern therapeutics; for example: serpentine isolated from the root of Indian plant *Rauwolfia serpentina* in 1953, was a revolutionary event in the treatment of hypertension and

lowering of blood pressure. Vinblastine isolated from the *Catharanthus rosesus* is used for the treatment of Hodgkins, choriocarcinoma (Farnsworth et al., 1967), non-hodgkins lymphomas, leukemia in children, testicular and neck cancer. Vincristine is recommended for acute lymphocytic leukemia in childhood, advanced stages of Hodgkins, lymophosarcoma, small cell lung, cervical and breast cancer (Farnsworth and Bingel 1977). The anticancer activity of flavonoids found in *Citrus* juices has been reported worldwide (Gaiotto et al., 1997). Plant derived drugs are also used to cure mental illness, skin diseases, tuberculosis, diabetes, jaundice, hypertension, cancer etc. Plant derived drugs came into use in the modern medicine through the uses of plant material as indigenous cure in folklore or traditional systems of medicine. More than 64 plants have been found to possess significant antibacterial properties; and more than 24 plants have been found to possess anti diabetic properties (Arcamone et al., 1980).

1.3 Need of species level identification of important medicinal plants:

The principal issues in ethnobotany emphasized the importance of correct species identification and the deciphering of indigenous and traditional knowledge of medicinal plant usage and their transfer for the promotion of prospect in human health care. It is reported that above 2000 species of medicinal plants have been utilized by various traditional herbal medicine providers in the Northeast (NE) India (Kala et al., 2006). There are cases of the commercial harvest of medicinal plants from wild in NE, which render threats to biodiversity from large scale harvest. Concomitant loss of biodiversity from NE is no less than any other region in India (Vaidya and Devasagayam 2007). Further, notwithstanding the anthropogenic threats, many flora and fauna have declined in the region; subsequently, the major part of the region has been categorized as hotspot of biodiversity. This suggest the need of exploration of traditional herbal medicine from NE. Amidst the galaxy of rich traditional knowledge of herbal medicine in use by the majority of tribal people in NE India, there is an urgent need of correct taxonomic inventorization vis-à-vis species

level molecular characterization of medicinal plants from this region. Such a research work in the selected study area is the need of the hour as well.

1.4 Important medicinal plant group (Family: Apocynaceae) a candidate for current study:

Apocynaceae is one of the ten largest angiosperm families (including Asclepiadaceae), and comprise of several important medicinal plants. Apocynaceae has grown to 395 genera and some 5100 species of tropical tree, shrub and vine. Robert Brown was one of the influent people in the classification of Apocynaceae. He described more than 40 genera in the family, the great majority of which are still valid today. Rauvolfioideae subfamily of Apocynaceae is known for the rich source of typical laticiferous tissues, which produce various alkaloids and cardenolides being used in traditional medicines for stomach ulcer, fever, asthma, whooping cough, etc (Mahadani et al., 2013).

More than 200 alkaloids were derived from Apocynaceae through scientific investigation (Raffauf and Flagler 1960) much of this progress has been made in recent years as a result of the economic importance of *Rauvolfia serpentine, Catharanthus roseus* (van Der Heijden et al., 2004). *Calotropis gigantea* is also a potential candidate source for anticancer drugs (Wong et al., 2011) and *Allamanda cathartica* possesses a remarkable wound healing function (Nayak et al., 2006). The recent popularity of the Apocynaceae as a source of a number of interesting drugs has led to extensive chemical studies of the family as a whole. Therefore, molecular taxonomy may be of valuable assistance in the examination of new plants species.

1.5 Advanced molecular tool for species level identification:

There is a worldwide effort to revolutionize the way scientists identify species in the laboratory and in the field with a technique called DNA Barcoding, similar to the barcode that identifies an item at the grocery store, a DNA Barcode is used to identify and distinguish biological species. Each of the world's estimated 1.8 million species is genetically unique- its unique identity is carried in its DNA molecules (Hebert et al., 2003a). DNA barcoding rapidly sequences the DNA from a single, standardized

gene on the DNA molecule. It uses a short genetic marker in an organism's genome to identify it as belonging to a particular species. It is based on a relatively simple concept that most eukaryotic cells contain mitochondria and mitochondrial DNA has a relatively fast mutation rate, which results in significant variations in mtDNA sequences between species and, in principle, a comparatively small variation within species.

A remarkably short DNA sequence can contain more than enough information to resolve 10 or even 100 million species. For example, a 600-nucleotide segment of a protein-coding gene contains 200 nucleotides that are in the third position within a codon. At these sites, substitutions are (usually) selectively neutral and mutations accumulate through random drift. Even if a group of organisms was completely biased to either adenosine or thymine (or alternatively, to either guanidine or cytosine) at third nucleotide positions, there would still be 2^{200} or 10^{60} possible sequences based on third-position nucleotides alone. DNA sequence analysis of a uniform target gene to enable species identification has been termed DNA barcoding, by analogy with the Uniform Product Code barcodes on manufactured goods (Ghosh et al., 2012). Proof of principle for DNA barcoding has been provided by analysis of a 650 base pair fragment of *Cytochrome c oxidase subunit I* (*COI*) sequences among closely related species and across diverse phyla in the animal kingdom (Hebert et al., 2003b). An important and perhaps unexpected finding is the congruence between morphological taxonomy and DNA barcode analysis. Several studies have now established that sequence diversity in a ~650bp region near the 5' end of the mitochondrial *Cytochrome c oxidase subunit I* gene provides strong species level resolution for varied animal groups including birds (Hebert et al., 2004), fishes (Ward et al., 2005), springtail (Hogg and Hebert 2004), spiders (Barret and Hebert 2005), moths (Hebert et al., 2003a) etc. But in case of plants equivalent has proved difficult. The generally low rate of nucleotide substitution in plant mitochondrial genomes precludes the use of *COI* as a universal plant barcode (Fazekas et al., 2009). Instead,

the search for a plant barcode has involved looking out with the mitochondrial genome and from the outset many researchers have accepted that plastids have become the primary source of DNA barcodes in plants and multiple markers will be required to obtain adequate species discrimination. Agreement on a common barcode is necessary for plant barcoding to progress towards the creation of a shared community resource. To facilitate and formalize the selection of a plant barcode, the Consortium for the Barcode of Life (CBOL) instigated the formation of a working group with representation from the different research groups/research consortia from the systematics community. They had tested the seven leading candidate plastid DNA regions (*atpF–atpH*, *psbK–psbI*, *trnH–psbA spacers* and *matK, rbcL, rpoB, rpoC*1 genes). The Plant Working Group of the Consortium for the Barcode of Life (CBOL) recommended the two-marker combination *rbcL*/*matK* as the standard DNA barcode for plants to be supplemented with additional markers as required (Hollingsworth et al., 2009). *Maturase K* gene (*mat*K*)* of chloroplast DNA is highly conserved in plant systematic which is involved in Group II intron splicing. It is another emerging gene with a potential contribution to plant molecular systematics and evolution (Barthet and Hilu 2008). Kress et al., (2005) first proposed that the plastid *trnH-psbA* intergenic spacer as a potential DNA regions for applying barcoding to flowering plants (Kress et al., 2005). The *trnH-psbA* spacer (~450bp) is the most variable plastid region in angiosperms and is easily amplified across a broad range of land plants. This region is one of the most variable non coding regions of the plastid genome in angiosperm in term of having the highest percentage of variable sites (Strochova and Olson 2007). It appears that even within closely related taxa, great length differences exist, such that at greater taxonomic distances no shared sequence remains like a document, issued by a national government, which certifies, for the purpose of international travel, the identity and nationality of its holder (http://en.wikipedia.org/wiki/Passport)

1.6 Why DNA passport is mandatory for the important Medicinal Plants?

Under the World Trade Organization (WTO) regime for the export of medicinal plants, it is important that plants to be exporting plant should have precise "passport data" in the way of their authentic characterization. Data gathering is an important aspect of collecting operation. The important observations on samples and their collection sites are called the passport data. The availability of such data would help in planning the exploration route in a future collection/recollection programmer and to decide ecological regions and analogous agro-climatic conditions where such materials are to be evaluated. Further passport data are an asset to understand nature, cause and even consequences of variation among plant populations. The passport data are very useful to curators/breeders and other users to draw valid conclusions about the utility of the material. With the implementation of the Conventional on Biological Diversity (CBD) and patenting of genetic resources, the passport data assume additional significance. To ensure fair benefit sharing and to prevent unwarranted bio-piracy of indigenous plant genetic resources, it is important to authentically characterize and precisely evaluate the medicinal plant diversity of nation wise using modern biotechnological tools that allow precession in species of the plants.

The practice of ethnobotany is very old in Southern Assam. In the region, the traditional knowledge of using plant for treatment of various ailments is also rich due to a high diversity of the tribes as well as a rich diversity of plants. Such rich knowledge is least explored, and remained fragmentary. Tribal healers in most of the countries, where ethnomedical treatment is frequently used to treat cuts, wounds, skin infection, swelling, aging, mental illness, cancer, asthma, diabetes, jaundice, scabies, eczema, venereal diseases, snakebite and gastric ulcer, provide instructions to local people as how to prepare medicines from herbs (Perumal Samy et al., 2006). They keep no records, and the information is mainly passed on verbally from generation to generation. Moreover, there is a trend of not revealing the knowledge by the herbal medicine provider to common people. In both the ways, such rich knowledge is dying

by the time. The sources of ethno-medicinal plants are mostly wild. World Health Organization (WHO) has shown great interest in documenting the use of medicinal plants used by tribes from different parts of the world. Many developing countries have intensified their efforts in documenting the ethnomedical data on medicinal plants. Research to find out the scientific evidence of claims by tribal healers on Indian herbs has been intensified. Once these local ethnomedical preparations are scientifically evaluated and disseminated properly, people will be better informed regarding efficacious drug treatment and improved health status. Furthermore, the conventional morphological taxonomic techniques involve difficulties in species level identification from any unstructured plant part. Thus, the ethnomedicinal resources of southern Assam seem least explored and found fragmentary, which entail the need of intervention of modern tools to characterize the molecular marker of the important and vulnerable medicinal plants for correct species level identification and their inventorization.

A large number of molecular techniques have been used to authenticate medicinal plants based on species-specific variations in the sequences of various chloroplast and nuclear DNA regions. Using PCR-based methods, species identification have been achieved using DNA that was isolated from fresh and dried plant parts, plant extracts, processed herbal drugs, as well as finished products such as herbal teas, tablets and capsules. In fact, molecular taxonomists now envision cataloging all living species on earth using DNA barcoding. The generation of molecular “barcodes” of medicinal plants and deposition of sequence data in publicly accessible databases will be worth by the concerted effort of the medicinal plant research community and contribute to the ongoing effort of defining barcodes for every (plant) species on earth. Along these lines, future studies aimed at the authentication of medicinal plants using genomic methods should focus on genetic loci that have been found useful for barcoding of plants. DNA barcode based authentication of medicinal plants is a work in progress that offers powerful new tools

and entry points for measures aimed at quality control and quality assurance in medical plant research as well as the production, clinical use, and forensic examination of herbal medicines and may use as a species level DNA passport for medicinal plants.

DNA barcoding is a powerful technique for species identification and exemplified with its wide application in monitoring and documentation of bio resources. The DNA barcoding is rapidly evolving but it is yet to provide full agreement on which region(s) of DNA should be universally used for plants. The application of the technique emphasizes some thrust areas, like documentation of the important and vulnerable ethno-medicinal plant bio resources dealing with which is recently defined as the subject "Ethnobotany Genomics" (Newmaster and Ragupathy 2010). These genetic signatures of species, which at the minimum ensure accurate identifications, have the potential to provide much more information about a plant, including insights into its geographic origin, placement in the Tree of Life, maternal and paternal lineages, as well as its genomic structure (Kress and Erickson 2012). This should aim to explore the effectiveness of *matK* and *trnH-psbA* spacer in differentiation of ethnomedicinal plants with special reference to the Apocynaceae family.

1.7 Objectives of the research:

On the backdrop of the research problem, following objectives were set

1. Development of species specific molecular marker based on *matK* sequence of Chloroplast DNA for common medicinal plants.
2. Establishment of *matK* and *trnH-psbA* intergenic sequences as DNA passport for Common medicinal Apocynaceae plants.
3. Sequence analysis and phylogenetic relationship of common medicinal plants with special reference to Apocynaceae family.
4. Identification of characteristic species specific Single Nucleotide Polymorphism (SNP) / Insertions and deletions (Indels) in common medicinal plants for the development of DNA passport.