CHAPTER 2
REVIEW OF LITERATURE

CHAPTER 2

REVIEW OF LITERATURE

2.1 Overview :

Traditional medicine refers to health practices, approaches, knowledge and beliefs incorporating plant, animal and mineral-based medicines, spiritual therapies, manual techniques and exercises, applied singularly or in combination to treat, diagnose and prevent illnesses or maintain well-being (WHO 2002). Nearly 25% of modern medicines are derived from plants that were first used in traditional medicine. The global market for herbal medicines currently stands at over US $ 60 billion per annum and is growing (WHO 2005). India is sitting on a gold mine of well-recorded and well practiced knowledge of traditional herbal medicine. Identification of plants at the species level is traditionally achieved by careful examination of the specimen's macroscopic and microscopic morphology. This work usually needs to be performed by a specially trained expert. However, morphological identification is often not accurate when the original plant material undergoes different processing systems. Therefore, additional methods of identification at the species level have been sought after and thus species level DNA barcoding has been developed for the identification of medicinal plants.

2.2 Hitherto documentation of impotant medicinal plants species:

Ayurveda, Siddha and Unani are also known as great traditions of medical systems in the Indian subcontinent and have evolved in a historical period spanning over 3-4 millennia (Ravishankar and Shukla 2007). Ayurveda is the most popular traditional systems of medicine in India. The literal meaning of Ayurveda is “science of life”, because the ancient Indian system of health care focused views of man and his illness. In the oldest medical text of Ayurveda, Caraka samhita is estimated to be written and redacted through various versions

from 1,500 BC-200 AD. Now Ayurveda is popular not only among Indians but in other developed and developing countries' population. In India, Kerala state, Ayurveda is flourishing in the form of health tourism. The Unani traditional medicine is mostly popular among the Muslim community but the propagation is not that wide like Ayurveda. However, The basic philosophy and practices are not much different than Ayurveda. Unani medicine got deep roots and royal patronage during medieval times. It progressed during Indian sultanate and Mughal periods (Ravishankar and Shukla 2007). Another oldest systems of medicine in India is Siddha System which is practiced in the Tamil speaking parts of India. The Indian subcontinent is a vast repository of medicinal plants that are used in traditional medical treatments. The medicinal plants are listed in various indigenous systems such as Siddha (600), Ayurveda (700), Amchi (600), Unani (700) and allopathy which 30 plant species for ailments (Rabe and Staden, 1997).

There are many references to Indian medicinal plants and trade in spices in a number of historical documents. For instance, Indian *Aloe* is very widely used in India for cosmetic, medicinal and nutraceutical purposes (Srivastava and Singh 1996). Despite the global reputation of *Aloe* in dermato-cosmetics, the potential as anti-aging is still untapped. Similarly, the plant *Adhatoda vasica* has been extensively studied for cough and the active principles have been known (Shah and Chauhan 1996). The genus, *Citrus* are notable for its fragrance, partly due to flavonoids and limonoids contained in the rind and most are juice-laden. The juice contains a high quantity of citric acid, giving them their characteristic sharp flavour. They are good sources of vitamin C. *Citrus* plants are also useful in terms of medicinal value and other purposes like Lemon juice is probably the best of all antiscorbutics, being almost a specific in scurvy. Traditionally, it is a good astringent, uterine haemorrhage after delivery, obstinate hiccough, jaundice and hysterical palpitation of the heart, reducing the temperature in typhoid etc. (Ram and Singh 2006).

2.2.1 Anticancer Plants:

Cancer is a general term applied for series of malignant diseases that may affect different parts of the body. These diseases is characterized by a rapid and uncontrolled formation of abnormal cells, which may mass together to form a growth or tumor, or proliferate throughout the body, initiating abnormal growth at other sites. If the process is not arrested, it may progress until it causes the death of the organism. The main forms of treatment for cancer in humans are surgery, radiation and drugs (cancer chemotherapeutic agents). Cancer chemotherapeutic agents can often provide temporary relief of symptoms, prolongation of life, and occasionally cures. In recent years, a lot of effort has been applied to the synthesis of potential anticancer drugs.

***Catharanthus roseus* -** The isolation of the vinca alkaloids, vinblastin and vincristine from the Madagascar periwinkle, *Catharanthus roseus* (Apocynaceae) introduced a new era of the use of plant material as anticancer agents. These are the first agents to advance into clinical use for the treatment of cancer (Cragg and Newman 2005). Vinblastine and vincristine are primarily used in combination with other cancer chemotherapeutic drugs for the treatment of a variety of cancers, including leukemia's, lymphomas, advanced testicular cancer, breast and lung cancers, and Kaposi's sarcoma (Cragg and Newman 2005).

***Taxus brevifolia* -** The discovery of paclitaxel (Taxol) from the bark of the Pacific Yew, *Taxus brevifolia* (Taxaceae) is another evidence of the success in natural product drug discovery. Various parts of *Taxus brevifolia* and other Taxus species (*e.g., Taxus Canadensis*, *Taxus baccata*) have been used by several Native American Tribes for the treatment of some non-cancerous cases (Cragg and Newman 2005); while *Taxus baccata* was reported to be used in the Indian ayurvedic medicine for the treatment of cancer. Paclitaxel is significantly active against ovarian cancer, advanced breast cancer, small and non-small cell lung cancer (Rowinsky et al., 1992).

2.2.2 Antioxidants Plants:

Antioxidants neutralize the toxic and 'volatile' free radicals. Antioxidants have many potential applications, especially in relation to human health, both in terms of prevention of disease. Indian medicinal plants provide a rich source of antioxidants. There are over 40 Indian medicinal plants showing antioxidant abilities at various levels of protection. The medicinal plants that show significant antioxidant activity include *Achyranthes aspera, Aegle marmelos, Allium cepa, Allium sativum, Aloe vera, Andrographis paniculata, Asparagus racemosus, Azadirachta indica, Ocimum sanctum, Centella asiatica, Zingiber officinalis* etc (Vaidya and Devasagayam 2007).

2.2.3 Anti-diabetic Plants:

Diabetes Mellitus is a syndrome of disordered metabolism, due to a combination of hereditary and environmental causes, resulting in abnormally high blood sugar levels (hyperglycemia), caused by defects in either insulin secretion or insulin action in the body. A national urban survey in 2005 observed that the prevalence of diabetes in urban India in adults was 15.1% (Kavishankar et al., 2011). Now-a-days natural products and herbal medicines have been recommended for the treatment of diabetes. The ethnobotanical information reports that about 800 plants may possess anti-diabetic potential (Puranik et al., 2010). Even there is the discovery of widely used hypoglycemic drugs, metformin come from the traditional approach of using *Galega officinalis.*

***Cajanaus cajan* -** The methanol extract of *Cajanus cajan* leaves shows significant reduction of fasting blood sugar in alloxan induced diabetic rats related manner (Adaobi et al., 2010).

***Hibiscus rosa-sinensis* -** The ethanol extract of *Hibiscus rosa-sinensis* flower at 250 and 500mg/Kg significantly reduces the blood glucose level in both

acute and sub acute treatments in alloxan induced diabetic rats (Moqbel et al., 2011; Venkatesh et al., 2008).

Aloe vera **-** ethanol extract of *Aloe vera* leaf gel shows significant antihyperlipidae effect in streptozotocin induced diabetic rats (Moqbel et al., 2011).

It is reported that *Tinospora cordifolia, Ocimum sanctum, Mangifera indica, Jatropha curcas, Azadirachta indica* etc have shown varying degree of hyperglycemic and antihyperglycemic activity.

2.2.4 Anti-inflammatory Plants:

Inflammatory diseases including different types of rheumatic diseases are a major cause of morbidity of the working force throughout the world. Gastrointestinal side effect is the major problem associated with the presently available non-steroidal anti-inflammatory agents. Classic examples of herbs traditionally used to treat inflammation in Western medicines are *Matricaria chamomilla* and *Arnica montana* (Asteraceae), *Salix alba* (Salicaceae) and *Glycyrrhiza glabra* (Fabaceae). Other well-known plant products with anti-inflammatory activity are *Hamamelis virginiana*, *Echinacea* species including *Echinacea angustifolia*, *Ananas comosus*, *Abelmoschus esculantus* (Shah et al., 2011). Common examples of Asian anti-inflammatory plants are *Curcuma longa*, *Zingiber officinale* (Shah et al., 2007).

2.2.5 Beneficial Compounds from Plants:

Medicinal plants are also important source of other type of beneficial compounds including the ingredients for functional foods. These functional foods promote better health to prevent chronic illness. Some ingredients that make food functional are dietary fibers, vitamins, minerals, antioxidants, oligosaccharides, essential fatty acids (omega- 3), lactic acid bacteria cultures and lignins. Many of these are present in medicinal plants. Indian systems of medicine believe that

complex diseases can be treated with a complex combination of botanicals unlike in the West, with single drugs. Whole foods are hence used in India as functional foods rather than supplements. Some medicinal plants and dietary constituents having functional attributes are spices such as onion, garlic, mustard, red chili, turmeric, clove, cinnamon, saffron, curry leaves, fenugreek and ginger (Vaidya and Devasagayam 2007).

2.3 Concept of DNA Barcoding:

In 2003, Paul D.N. Hebert from the University of Guelph, Canada proposed the compilation of a public library of DNA barcodes that would be linked to named specimens. This library provided a new master key for identifying species, one whose power will rise with increased taxon coverage and with faster, cheaper sequencing". DNA barcoding provided a standardized method for this process via the use of a short DNA sequence from a particular region of the genome to provide a 'barcode' for identifying species.

2.3.1 Animal DNA Barcoding:

Hebert et al., (2003a) established that the mitochondrial gene *Cytochrome c oxidase I* (*COI*) of mitochondrial genome serves as the core of a global bio-identification system for animals. A model *COI* profile, based upon the analysis of a single individual from each of 200 closely allied species of lepidopterans, was 100% successful in correctly identifying subsequent specimens.

Hebert et al., (2003b) reported that the sequence divergences at *COI* regularly enable the discrimination of closely allied species in all animal phyla except the Cnidaria. This success in species diagnosis reflected both the high rates of sequence change at *COI* in most animal groups and constraints on intra-specific mitochondrial DNA divergence arising, at least in part, through selective sweeps mediated via interactions with the nuclear genome.

Hebert et al., (2004) determined *COI* barcodes for 260 species of North American birds and found that distinguishing species was generally straightforward. All species had a different *COI* barcode(s), and the differences between closely related species were, on average, 18 times higher than the differences within species. Their results identified four probable new species of North American birds. The finding of large *COI* sequence differences between, as compared to small differences within, species confirmed the effectiveness of *COI* barcodes for the identification of bird species.

Ratnasingham and Hebert (2007) developed The Barcode of Life Data System (BOLD) which is an informatics workbench aiding the acquisition, storage, analysis and publication of DNA barcode records. By assembling molecular, morphological and distribution data, it bridges a traditional bioinformatics chasm. BOLD is freely available to any researcher with interests in DNA barcoding. They divided BOLD in three functional units- The managements and Analysis system (MAS), Identification System (IDS), External Connectivity System (ECS).

2.3.2 Plant DNA Barcoding:

Kress et al., (2005) proposed the *nuclear internal transcribed* spacer region and the plastid *trnH-psbA* intergenic spacer as potential DNA regions for applying barcoding to flowering plants because the *cytochrome c oxidase I* sequence have a much slower rate of evolution in higher plants than in animals. The *internal transcribed spacer* was the most commonly sequenced locus for plant phylogenetic investigations at the species level and showed high levels of interspecific divergence. The *trnH-psbA* spacer, (~450 bp), is the most variable plastid region in angiosperms and is easily amplified across a broad range of land plants. They compared the total plastid genome of tobacco and closely related species in seven plant families and a group of species sampled from a local flora encompassing 50 plant families (for a total of 99 species, 80 genera and 53

family). They suggested that *trnH-psbA* intergenic spacer is the best plastid option for a DNA barcode that has a good priming site, length and interspecific variation and necessary to employ more than one locus to attain species level discrimination across all flowering plant species. They concluded that *ITS* and *trnH-psbA* serve as good starting points for large scale testing of DNA barcoding across a large sample of angiosperm.

Chase et al., (2007) proposed to use three regions of plastid DNA as a standard for barcoding of all land plants due to the low levels of variation in DNA. There are no plastid regions (coding or non-coding) that evolve as rapidly as mitochondrial DNA (mtDNA) generally does in animals. They outlined two, three-region options, (1) *rpoC1*, *rpoB* and *matK* or (2) *rpoC1*, *matK* and *psbA-trnH* as viable markers for land plant Barcoding.

Lahaye et al., (2008) undertook intensive field collection (more than 1600 samples) in two biodiversity hot spot (Mesoamerica and southern Africa). They compared eight potential barcodes (*matK, trnh-psbA, ycf5, rbcL, rpoB, ndhJ, accD* and *rpoc1*) in all the samples. Based on barcode gap, easy amplification, and alignment, they identified a portion of the plastid *matK* gene as a universal DNA barcode for flowering plants. In addition, analyzing > 1000 species of Mesoamerican orchids, DNA barcoding with *matK* alone revealed cryptic species and proveed useful in identifying species listed in Convention on International Trade of Endangered Species (CITES) appendixes.

Newmaster et al., (2008) reported that *matK* and *trnH-psbA* has significant variation in nutmegs family, is an older group within the angiosperms that contains some recently evolved species. They demonstrated that a two gene approach utilizing a moderately variable region (*matK*) and a more variable region (*trnH-psbA*) were provided resolution among all the *Compsonuera* species.

Fazekas et al., (2008) revealed multiple multilocus DNA barcode from the plastid genome equally well and up to 69-71% level was achieved by several two and three region combinations. In this study, they compared the eight candidates of plant DNA barcoding regions from plastome and one from mt DNA to test the performance among the 92 species in 32 diverse genera of land plants (251 samples). Single locus resolution ranged from 7% (23s rDNA) to 59% (*trnH-psbA*) of species with well supported monophyly. Several loci (*matK, psbK-psbI, trnH-psbA*) were problematic for generating bidirectional sequence.

Selvaraj et al., (2008) proposed that *matK* has a good candidate for DNA barcoding of Zingiberaceae family. The *matK* gene sequence of Zingiberaceae was obtained from genbank for the analysis of variants, parsimony site, pattern, transition/ transversion rate and phylogeny. Their result indicated that the Zingiberaceae genus *Afromonum, Alpinia, Globba, Curcuma* and *Zingiber* showed polyphylogeny. The overall variations between the species were 24% and the transition / transversion rate was 1.54. Phylogenetic tree was designed to identify inter and intra generic relationships.

Gonzalez et al., (2009) examined the eight plant DNA markers (*rbcLc, rpoC1, rpoB, matK, ycf5, trnL, psbA-trnH, ITS*) in two hectares of a tropical forest in French Guiana. *matK* and *ITS* were showed a low rate of sequencing success and none of the plastid markers achieved the rate of correct species identification greater than 70%, either alone or combined. 130 molecular operational taxonomic units including molecular markers increased the identification rate of juveniles from 72% (morphology) to 96% (morphology and molecular). They concluded that while DNA barcoding is an invaluable tool for detecting errors in identifications and for identifying plant at the juvenile stage.

Kress et al., (2009) reported more than 98% correct identifications based on three locus DNA barcodes on 296 species from the tropical forest dynamics plot in Panama. The three locus barcodes data are sufficient for reconstruct

evolutionary relationship among the plant taxa and are congruent with the broadly accepted phylogeny of flowering plants. They concluded that highly resolved phylogenies based on DNA barcode sequence data will enhance research focused on the interface community ecology and evolution.

Hollingsworth et al., (2009) recommended the two locus combination of *rbcL+matK* as the plant barcode. This core 2-locus barcode provides a universal framework for the routine use of DNA sequence data to identify specimens and contribute toward the discovery of overlooked species of land plant. They compared the performance of 7 leading candidate plastid DNA regions (*atpF-atpH spacer, matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK-psbI* spacer and *trnH-psbA* spacer) based on the recoverability, sequence quality and levels of species discrimination. Direct universality assessments using a single primer pair for each locus in angiosperm resulted in 90-98% PCR and sequence success for 6/7 regions. Evaluation of sequence quality and coverage from the candidate loci demonstrated that high quality bidirectional sequences were routinely obtained from *rbcL, rpoC1* and *rpoB*. The remaining 4 loci required more manual editing and *matK* performed best of this group. Among 397 samples, successfully sequenced all 7 loci, species discrimination for single locus barcodes ranked in order: *rpoC1<rpoB<atpF-atpH<rbcL<matK<psbK-psbI< trnH-psbA*. Based on the barcode criteria, four of the candidate loci (*rpoC1, rpoB, atpF-atpH, psbK-psbI*) were excluded in plant DNA barcoding system, although none of the 3 loci fit 3 criteria perfectly.

Asahina et al., (2010) investigated the species identification of five *Dendrobium* plants (*Dendrobium fimbriatum, D. moniliforme, D. nobile, D. pulchellum, and D. tosaense*) which have long been used in traditional medicines. Based on the proposal of CBOL plant working group, they chose the *matK* and *rbcL* to conduct the phylogenetic analyses for assessing the intra and interspecies relationship of *Dendrobium* species. The 3'-half of *matK* sequence (Nt.944-1616)

was sufficient to distinguish among these five species. The phylogenetic analysis using full length *rbcL* sequences showed no species discrimination. *matK* is better for identifying medicinal *Dendrobium* species.

Bruni et al., (2010) evaluated the universal application of the DNA barcoding approach to univocally identify toxic plants (A total 50 land plant species) starting from different plant portions, based on five DNA barcode regions (*trnH-psbA*, *rpoB*, *matK*, *At103*, and *sqd1*). They found that the *trnH-psbA* ranked first in divergence value in the analyzed species but the high variability of this DNA spacer did not allow to align properly. *matK* gene showed easy amplification and alignment in the analyzed species and high level of discrimination values. Among the nuclear sequences, they suggested *At103* as the most suitable candidate. They recommend the combination of plastidial and nuclear markers to identify toxic plants concerning plastidial markers, *matK* and *trnH-psbA* showed consistent genetic variability.

Dunning and Savolainen (2010) recommended 26 new primers for *matK* to increase the amplification success rate that was possible by introducing degeneracy or a deoxyinosine base to the second position from the 3' end of the primer. This study mainly focused on designing order specific primer for monocot and eudicots.

Wang et al., (2010) proposed the *atpF-atpH* non coding spacer as a universal DNA barcoding marker for duckweeds. They compared the seven potential barcoding markers (*rpoB, rpoC1, rbcL, matK, atpF-atpH, psbK-psbI* and *trnH-psbA*) from 31 species of the Lemnaceae family. They examined the extent of barcoding gap between intra and inter specific variation by pair wises comparision and found *atpF-atpH* has sufficient interspecific (0.0633) but relatively low interspecific divergence (0.0008) in comparison to other six markers.

Shao et al., (2010) identified the 17 plants from Huperziaceae based on the phylogenetic analysis of *trnH-psbA* gene. Their result showed that Huperziaceae divided into two genera Huperzia and phlegmariurus and bootstrap value reached 91%. It suggested that *trnH-psbA* can be useful for DNA barcoding to identify plants.

Chen et al., (2010) proposed that the *ITS2* region are potential for a standard DNA barcode to identify medicinal plant parts and their closely related species. Based on the DNA barcoding criteria, they compared seven candidates (*psbA-trnH, matK, rbcL, rpoC1, ycf5, ITS2* and *ITS*) from medicinal plant species and found *ITS2* of nuclear ribosomal DNA represented the most suitable region for barcoding applications. They also tested the discrimination ability of *ITS2* in more than 6600 plant samples belonging to 4800 species from 753 distinct genera and found 92.7% successful identification at the species level.

Luo et al., (2010) suggested that *ITS2* locus have highest species level identification efficiency compare to *psbA-trnH, matK, ycf5, rpoC1, rbcL* and *ITS* in Rutaceae family. Their results showed that inter-specific divergence of *ITS2* was significantly higher than the inter species variation in the DNA barcode gap. Among the 197 samples tested, 21 samples were not identified, of which 18 samples belong to *Citrus*. There was many arguments about the classification of the *Citrus* genera, especially the division of species. *Citrus* are easy to bud mutation and the long history of artificial cultivation and there are many morphological types of *Citrus*. The unsuccessful identified species in this study were mainly cultivars.

Lou et al., (2010) developed an integrated DNA barcode multimedia information platform - Medicinal Material DNA barcode Database (MMDBD) for data retrieval and similarity search. Their work provides a centralized medicinal material DNA barcode database and bioinformatics tools for data storage analysis and exchange for promoting the identification of medicinal material.

Jeanson et al., (2011) first time tested the DNA barcoding technique (*matK, rbcL, nrITS, trnH-psbA*) on palms. This study conducted on 40 out of the 48 species of the Southeast Asian tribe Coryoteae (Subfamily: Coryphoideae). 92% species discrimination was possible by using the combination of three markers – *matK, rbcL, and nrITS*. DNA barcoding can be a useful tool to identify species within this ecological importance tropical plant family.

Xiang et al., (2011) recommended a tiered or multilocus method for barcoding plants species. They investigated the utility of *matK, rbcL, trnH-psbA, ITS* on 196 individuals from 9 genera and 54 species of family Juglandaceae and found that *ITS* has the most variable information and *rbcL* has the least. *matK* has enough efficient to discriminate the seven of nine genera of Juglandaceae. *ITS* has higher interspecific p-distance than the *trnH-psbA* region. But, *ITS* appeared to have limited power for species identification within the Carya and Engelhardia complex, and has no power for Juglans or Pterocarya. They proposed as the first tier DNA region for genus discrimination and second locus at species level.

Fu et al., (2011) studies the *rbcL+matK+ITS* barcode for the genus Tetrastigma. *ITS* as a barcode showed significant inter-specific genetic variability but multilocus provided a greater ability to distinguish species than single loci.

Xue and Li, (2011) revealed that *ITS* region discrimination significantly between *Gentianopsis paludosa* and all nine adulterate species. Their finding also showed that short length of *ITS* is a advantages on the amplification DNA and interspecific divergence of *ITS* is higher than the intraspecies divergence.

Gu et al., (2011) compared the four barcoding markers (*matK, rbcL, trnH-psbA, nrITS*) to differentiate species within Ligustrum. Their result showed that *ITS* sequence has the most variable followed by *trnH-psbA*, *matK*, and *rbcL*. All the species were differentiated using *ITS* when combining the NJ tree method with character based or MP tree method.

Sun et al., (2012) tested the applicability of three candidate DNA barcodes (*rbcL, matK* and *trnH-psbA*) to identify species within 148 individual plant samples encompassing of 38 species of *Dioscorea*. *matK* successfully identified 23.36% of all species compared with 9.30% for *rbcL* and 11.63% for *trnH-psbA*. They found that the combination of two or three loci achieved a higher success rate of species discrimination than one locus alone. They conclude that *matK* is a strong, although not a perfect candidate as a DNA barcode for *Dioscorea* identification.

Wallinger et al., (2012) presented a PCR based approach to identify a variety of plant taxa commonly occurring in Central European Agriculture land. Based on the *trnT-F* CpDNA region, PCR assays was developed to identify two plant families (Poaceae and Apiaceae), the genera Trifolium and Plantago, and nine plant species: *Achillea millefolium, Fagopyrum esculutum, Lolium perenne, Lupinus angustifolius, Phascolus coccines, Sinapis alba, Taraxacum officinale, Triticum sativum* and *Zea mays*. These assays allowed identification of plant based on size-specific amplicons ranging from 116 bp to 381 bp. They concluded that this molecular assay will be applicable manifold, such as for root and leaf litter identification, botanical trace evidence and, analysis of herbivores.

Singh et al., (2012) compared seven loci of plant DNA barcoding among multiple accessions of 36 *Dendrobium* species. The *trnH-psbA* spacer showed problematic in sequence quality and *ITS* provided 100% species identification. Another locus *matK* resolved 80.56% of 36 species. They recommended combination of *matK*, *rpoB* and *rpoC1* to resolve the maximum number of species. They also discussed the problem for *ITS* as a barcode.

Dong et al., (2012) scanned the entire chloroplast genomes of 12 genera to search the most variable region for the molecular studies on angiosperm at lower taxonomic level, and for DNA barcoding of species. They identified nearly 5% of the most variable loci from all variable loci in the chloroplast genomes for each

genus and selected 23 loci including 4 coding regions, 2 introns and 17 intergenic spacers. Three loci, *trnS*^{UGA} *-trnG*^{UCC}, *trnF-psbD* and *trnW-psaJ* showed very high nucleotide diversity per site (Π values) across three genera. Other loci may have strong potential for resolving phylogenetic and species identification problem at the species level. The loci *accD-psal, rbcL-accD, rpl32-trnL, rps16-trnQ* and *ycf1* are absent from some genera. They also designed primers from their conserve flanking regions and tested the applicability of the primers to amplify the target sequence. They concluded that a chloroplast genome sequence contain region that are highly variable such regions are the consideration when screening the suitable loci to resolve closely related species or genera in phylogenetic analysis and for DNA barcoding.

Bruni et al., (2012) examined the core DNA barcoding (*rbcL* and *matK*), plus the additional *trnH-psbA* region in the identification of vascular plants belonging to a local flora of a few hundred species, that of Mt. Valerio (Trieste, NE Italy). They found that the core barcode markers univocally identify most species of our local flora (96%). The *trnH-psbA* data improve the discriminating power of DNA barcoding among closely related plant taxa. In their simulation, The DNA barcoding approach was compared to the use of a digital identification key based upon morphological features. DNA barcoding added value in absence of some morphological feature, reaching a correct identification for 100% of the species. They proposed to use the combination of morphological feature and molecular data as well as create an integrated identification system for plant biodiversity surveys.

2.2.3 Hurdles in plant DNA Barcoding:

Fazekas et al., (2009) revealed the higher monophyletic species in animals (>90%) than plants (~70%) using barcode marker. Both animals and plant species pairs have variable size gaps between intra and interspecific genetic distances but animal species tend to have larger gaps than plants, even in relatively densely

sampled genera. An analysis of 12 plant genera suggested that hybridization contributes significantly to the variation in genetic discontinuity of plants. However, overall fine scale species discrimination in plants relative are more difficult than animals due to greater levels of gene tree paraphyly.

Spooner (2009) tested barcoding with the most variable and frequently suggested plant barcoding regions (*ITS, trnH-psbA, matK*) on complicated plant group, *Solanum* sect. *Petota* (wild potatoes). These DNA regions were failed to provide species specific marker in *Petota* because the *ITS* has too much intraspecific variation and the plastid markers lack sufficient polymorphism. Barcode regions are often useful to detect new species, but often not due to biological complications by interspecific hybridization, introgression, allopolyploidy, a mixture of sexual and asexual reproduction and possible recent species divergence.

Roy et al., (2010) reconfirmed the concept of universal barcode in plants may not work in some cases. Most promising plant DNA barcode loci (one from nuclear genome - *ITS*, and three from plastid genome - *trnH-psbA, rbcL*, and *matK*) were failed to resolve the species identification in Indian *Berberis* and two other genera, *Ficus* and *Gossypium*. Finally, they employed AFLP test in species of *Berberis* to determine the relationship. Morphological, geographical and molecular marker analyses suggested probable reticulate evolution of Indian species of *Berberis* and thus barcode marker were not work in this case.

Maia et al., (2012) tested the effectiveness of the *rbcL* and *matK* barcoding marker for identifying 46 endemic *Bromeliad* species from endangered Brazilian Atlantic Rainforest. They were obtained only 43.48% species discrimination, the addition of a third marker, *trnH-psbA* did not show significant improvement. Bromeliaceae's sequence divergence was almost three times lower than the observed for Asteraceae and Orchidaceae. This low variation rate tends to fail in taxonomy complicated and recently diverged plant groups, such as Bromeliaceae.

Arca et al., (2012) investigated the performance of DNA barcoding (*matK, rpo13, rpoC1* and *trnH-psbA*) for identifying the large worldwide sample of *Fraxinus* species (56 species). The chloroplast intergenic spacer *rpl32-trnL* was further assessed in search for a potentially variable and useful locus. Their result suggested that the proposed CpDNA loci, alone or in combination, were not successful among *Fraxinus* species due to the low rates of substitution in the chloroplast genome. The intergenic spacer *trnH-psbA* was best performed locus up to the subgenus level. The current CpDNA barcodes are inadequate to fully discriminate *Fraxinus* species. They proposed supplementary barcoding loci of the nuclear genome for alternative solution.

2.3.4 Application of Plant DNA barcoding:

Guo et al., (2011) distinguished authentic Radix Astragali from its adulterants based on DNA barcode sequences (*ITS*, *matK*, *rbcL* and *COXI*). The roots of *Astragalus membranaceus* and *A. membranaceus* var. mongholica are commonly used as Radix Astragali. However, in additions to the two species, 23 species of genera *Astragalus*, *Oxytropis*, *Hedysarum* and *Glyoyrhiza* have been used as adulterants not only in trading markets but also by the herbal medicine industry. *ITS* showed sufficient divergence at intra and interspecies level. Moreover, two Indels detected in the *matK* sequences are useful for distinguishing Radix Astragali from its adulterants. They suggested that the combination of *ITS* and *matK* are superior barcode for Radix Astragali.

Li et al., (2010) authenticated the *Taxillus chinensis* from adulterants by using DNA barcoding technique. The amplification and sequencing efficiency of *rbcL* and *trnH-psbA* were 100% and efficiency of *matK*, *ITS* and *ITS2* regions was very low. *psbA-trnH* has a suitable DNA barcode in authenticating *T. Chinensis* and its related species.

Srirama et al., (2010) assessed species admixture in raw drug trade of *Phyllanthus* using morphological and DNA barcoding tool. *Phyllanthus* species are well known for their hepatoprotective activity and are used as ethnomedicine in India. They collected 25 raw drug trade samples of *Phyllanthus* from different shops of southern India. First, species were identified using morphological taxonomic keys and reveled six different species of *Phyllanthus* in the market samples. 66% of samples contained *Phyllanthus amarus* as the predominant Species and found another five different species of *Phyllanthus*. Specific DNA barcode signatures were developed based on *psbA-trnH* for further confirmation of species. *psbA-trnH* region of the chloroplast effectively discriminated *Phyllanthus* species from admixture in the raw drug trade of *Phyllanthus*.

Stoeckle et al., (2011) first time applied DNA barcodes concept on large array of commercial tea products and analyzed their performance to identify the constituents using existing databases. 90% tea products were yielded *rbcL* and *matK* barcode using a standard protocol. Matching DNA identification to listed ingredients was limited by incomplete databases for the two markers, shared or nearly identical barcode among some species and lack of standard names for the plant Species. About 1/3 of herbal tea generated DNA identification were not found on levels. Broad scale adoption of plant DNA barcoding required algorithms that searched results in context of standard plant names and characters based keys for distinguishing closely related species. This study demonstrated the importance of plant barcoding.

Sui et al., (2011) investigated the identification of six *Sabia* species and their seven adulterants through DNA barcodes (*trnH-psbA, rbcL-a, matK*). Based on sequence alignments, they concluded that not only *trnH-psbA* spacer sequence distinguished *S. parviflora* from other *Sabia* species, but the *matK+rbcL-a* sequence also differentiated it from the substitute and adulterants. The three

candidate barcodes identificated *S. parviflora* and distinguished it from common substitutes or adulterants.

Kool et al., (2012) tested the functionality, efficiency and accuracy of the use of barcoding for identifying 110 medicinal plant roots. They collected root samples from herbalist in Marrakech, Morocco and generated four barcode region (*rpoC1, psbA-trnH, matK* and *ITS*) to search against a tailored reference database Moroccan medicinal plants and their closest relative using BLAST. Sequencing success was high for *rpc1, trnH-psbA*, *ITS* but low for *matK*. Searching using *rpc1* alone were resulted in a number of ambiguous identification and combining *rpc1, trnH-psbA* and *ITS* were identified the majority market sample up to the genus level. Their study suggested that unknown samples were more difficult to identify, especially if the reference sequences were obtained from different population and required a global database that should therefore contain references sequence from different population of the same species.

Sun and Chen (2013) discriminated the cortex herbs of the Chinese pharmacopoeia (51 sample belongs to 19 species) using the *Second internal transcribed spacer* (*ITS2*) of ribosomal DNA. They found that the *ITS2* sequence (207 bp to 256 bp) were easy to be amplification. The intraspecific genetic distance of the cortex herb was between 0 - 0.073, which was lower than their interspecific genetic distance (Mean 0.868; minimum 0.101). The cluster dendrogram showed study materials are monophyletic. They concluded that *ITS2* barcode is suitable for the identification of cortex herb of Chinese pharmacopeia and will a play important role in the field of identification of traditional Chinese medicine.

2.4 Maturase K:

Hilu and Liang (1997) examined the rates, patterns and types of nucleotide substitutions in *matK* gene. GenBank and new sequences of monocot, dicot,

gymnosperm, and liverwort families were analyzed. Their results showed that the 3' region was most useful in resolving phylogeny. The presence of a relatively conserved 39 region and the less conserved 59 region provided two sets of characters that are used at different taxonomic levels from the tribal to the division levels.

Hilu et al., (2003) tested the angiosperm phylogeny based on partial *matK* sequences using parsimony (MP) and Bayesian inference (BI) approaches. They found that the *matK* gene evolves approximately three times faster than the widely used plastid genes *rbcL* and *atpB* and the MP and BI trees are highly congruent. With highest internal support yet for basal eudicots and within edicots. A crown group comprising berberidopsidaceae/ Aextoxicaceae, Santalales and Caryophylles + Asterids were resolved. Moreover, *matK* sequences provided good resolution within many angiosperm orders.

Barthet and Hilu (2008) explored the underlying factors behind the mode and tempo of *matK* evolution that allowed this protein coding gene to accommodate such elevated rate of substitution and yet maintain functionality. They examined putative amino acid sequences for *matK* from across green plants to determine constraint on this protein as indicated by variation in their side chain category composition. Amino acids in the *matK* ORF were divided into six categories based on the chemical properties of their side chain. The amount of standard deviation (SD) in the side chain composition was used as a measure of variation and constrain, where a low SD implied high evolutionary constrain and a higher SD implied low constraint. Further, secondary structure prediction showed evolutionary constraint on *matK* and identified three functional important regions. Their study supported the putative function of *matK* as a group II intron maturase.

Hao et al., (2010) examined the molecular adaptation and evolutionary dynamic of gene divergence in *matK* at various evolutionary levels. Selective

influences were investigated using standard dN/dS ratio methods. They revealed the presence of positive selective sites in the N-terminal region than in the RT domain and domain X of *matK*. Moreover, amino acid sites under positive selection were signified the effect the bootstrap value in phylogenetic reconstruction. Their results suggested that *matK* evolved under positive selection in some lineages of land plants and reflected the continuous fine-tuning of maturase performance under varying physiological conditions.

2.5 *trnH-psbA* spacer:

Strochova and Olson (2007) investigated the architecture of the *psbA-trnH* intergenic region including the 3' UTR stem-loop region of angiosperm. They compared the *psbA-trnH* spacer at two hierarchical levels. First, *psbA* 3' UTR sequences were compared across all flowering plants and secondly, focused on *psbA-trnH* non transcribed intergenic spacer in the genus *Silene*. Align the sequences across the families was impossible because insertions, deletions and simple sequence repeats were common and often more numerous than single base pair substitutions. They recognized that the end of this spacer nearest to *psbA* was highly conserved and exhibited large inversions where as the end nearest to *trnH* appeared more variable. From the *psbA* 3' UTR stem-loop structure, they found conserved motif and a consensus TTAGTGTATA box. They suggested that indel hot spots at the intergenic locus should be used extensively to assess population structure within the plant species.

Whitlock et al., (2010) examined the effect of intra specific inversions in the *trnH-psbA* region for DNA barcoding and found polymorphic inversions within six species of Gentianaceae. Typical simple methods of sequence alignment were lead to miss assignment of conspecifics and incorrect assessment of the relationship. Thus, the alignment of the *trnH-psbA* spacer region will need careful attention if it is used as marker for DNA barcoding.

2.6 Insertion and Deletion (Indels) and their utility:

Hilu and Alice (1999) examined the systematic of Poaceae using the plastid *matK* sequences. They identified insertion and deletion (indels) and nucleotide substitution at the extreme 3' end of the chloroplast *matK* gene that distinguished certain major lineages of grass. A 1bp deletion and a point mutation support the positions of *Streptochaeta* and *Anomochloa* as the two most basal lineage in Poaceae. Another 1bp deletion resulting in early termination of the ORF is unique to *Ehrharta*. A 6bp insertion supported monophyly of subfamilies Panicoideae, Arundinoideae, Centothecoideae and Chloridoideae (PACC). Indels and nucleotide substitution at the 3' end of the *matK* provided valuable phylogenetic markers in Poaceae. They concluded that these characters also may useful in phylogeny reconstruction of other plant taxa.

Ingvarsson et al., (2003) studied the molecular evolution of Indels in three interagenic spacers and one intron from the CpDNA of *Silene latifolia* and *Silene vulgaris*. They showed that Indels in these regions of *Silene* evolved at slightly higher rates than base pair substitutions and found coded Indels were highly informative for phylogenetic analysis and concluded that indel are useful for phylogenetic relationships at the intraspecific level.

Vischi et al., (2006) characterized four sunflower species (*Helianthus annuus*, *H. argophyllus*, *H. debilis* and *H. tuberosus*) by using *trnH-psbA* sequences. No sequence variation was detected within each of the species *H. annuus*, *H. debilis* and *H. tuberosus*. On the other hand, they found two groups of sequences within the species *H. argophyllus* with slight differences. They recognized approximately 60bp region (between the base 100-160) with SNPs, Indels, and SSR length that were sufficient for an unambiguous identification of each species.

Liu et al., (2012) investigated the effect of sampling design and utility of insertion and deletion (indels). For these purposes, they sampled 39 populations for nine *Taxus* species. Two or three individuals were randomly sampled from each population and sequenced one core DNA barcode (*matK*) and three supplementary regions (*trnH-psbA, trnL-trnF* and *ITS*). They revealed that P-distance was zero for most of the population in all regions and nodal support of monophyletic group was significantly increased when Indels were considered. They also found that Indels in the chloroplast *trnL-trnF* and *trnH-psbA* regions were informative to discriminate among closely related taxa. They concluded that one individual per population was adequate to represent the within population variation for DNA barcoding. They also proposed that indel coding methods should be considered in DNA barcoding of closely related plant species.

Garcia-Lor et al., (2013) clarified the phylogenetic relationships between 'true Citrus fruit trees'. These species are sexually compatible and are belived to be intra specific origin. They studied the polymorphism of 27 nuclear genes (18 genes involve in metabolite biosynthesis pathways and nine putative genesof salt tolarence) from 45 genotypes of *Citrus* and also analyzed 55 nuclear simple sequence repeats (SSRs). A total 16,238 kb of DNA was sequenced for each genotypes and revealed 1097 single nucleotide polymorphisms (SNPs) and 50 insertion and deletion (indels). These poly-morphism are potentially useful for the analysis of interspecific genetic structures. This study also presents new insights into the origin of *Citrus sinensis*.

2.7 Molecular taxonomy of Apocynaceae and *Citrus* :

Sennblad et al., (1998) studied the monophyly of the tribe wrightieae and its position in Apocynaceae family. They identified the morphological synapomorphies and sequenced *rbcL* chloroplast DNA and also combined floral morphology for the classification of the group. Their result recognized that

wrightieae was largely paraphyletic and the constituent taxa dispersed among four monophyletic clades.

Simoes et al., (2007) examined the deeper relationships within subfamily Rauvolfioideae (Apocynaceae) based on molecular as well as morphological evidences. They sequenced five DNA regions of the chloroplast genome (*matK, rbcL, rp116 intron, rps16 intron* and *3'trnK* intron) and performed bayesian and parsimony analysis among the 50 taxa of Rauvolfioideae and 16 taxa from Apocynoideae. In case of *matK* sequence, few gaps in multiple of three were found without exception. Neither subfamily Rauvolfioideae nor Apocynideae showed monophyletic. Rauvolfioideae recognized as a grade and the subfamily Periplocoideae, Secamonoideae and Asclepiadoideae nested within the Apocynideae. Three of the nine currently recognized tribes of Rauvolfioideae (Alstonieae, Melodineae and Vinceae) are polyphyletic. Based on the phylogeny, they proposed that simple style-heads, syncarpous ovaries, indehiscent fruits, and winged seeds evolved in parallel numerous times. They also proposed a revised classification for the subfamily Rauvolfioideae.

Jena et al., (2009) studied the taxonomy and phylogeny of Indian *Citrus* by using PCR-RFLP of the *trnD-trnT* and *rbcL* ORF regions as well as analyzed the *trnL-trnF* intergenic Spacer sequences of Cp-DNA. Eight major clusters were formed based on PCR-RFLP data and showed genetic distance ranging from 0 to 0.79 among 50 accessions of *Citrus* genotypes. *trnL-trnF* sequences from 23 representative accessions formed five clusters in the phylogenetic tree. Their study showed that the *trnL-trnF* spacer sequence have enough genetic variation in Indian *Citrus* genotype, but the utility of data at intra and inter specific level is limited probably by factors such as hybridization, bud mutation, apomixes and polyploidy. PCR-RFLP and *trnL-trnF* data supported the recognition of *C. maxima*, *C. medica, C. reticulate* as the basal species of edible *Citrus*. Based on

that work, the species status of *Citrus* in India as well as Northeast India became clear.

Kumar et al., (2010) examined inter simple sequence repeats (ISSR) polymorphism in *Citrus indica* along with three closely related wild taxa (*Citrus. medica, Citrus latipes* and *Citrus* sp.) from Northeast India. All the accession of *Citrus* formed four major clustres and 34 accession of *C. indica* cluster into one clade in UPGMA dendrogram. They found a significantly very low level of genetic variation within *C. indica*. Based on the molecular taxonomy they could state that destruction of natural habited pose serious threat to these group even in the *Citrus* gene Sanctuary in Nokrek Biosphere Reserve (NBR) in Meghalaya.

Penjor et al., (2010) sequenced the *rbcL* gene of Cp DNA from 24 genera of *Citrus* and their relatives. They revealed that both the NJ and MP trees supported the subfamily Aurantioideae as monophyletic but did not support the tribe and subtribe of the classification of Swingle and Reece. The subgenera *Citrus* and *Pepada* were not clustered clearly. Furthermore, Atalantia species showed polytomy. They proposed the rivision of the Swingle and Reece's classification. It signifying the poor performance of *rbcL* in molecular taxonomy of *Citrus* and need to be studied with other markers.

Mahmood et al., (2011) studied the level and pattern of genetic variability among selected species of Apocynaceae family by using randomly amplified polymorphic DNA (RAPD) marker. Genetic diversification was low both at interspecific and intra specific level. At the molecular level, *Alstonia scholaris*, and *Catharanthus roseus* appeared in a group and *Thevetia peruviana* formed another group and confirmed the classification based on morphological characters.