Chapter 2:

Review of Literature

Chapter 2: Review of Literature

2.1 History of Bamboo taxonomy

The earliest attempt at describing bamboos appear to be by Rumphius (1741) who described and named some bamboos in his publication entitled Herbarium Amboinense. In the first edition of Species Plantarum by Linnaeus (1800), only one species is mentioned under the name *Arundo bambos*, now referable to *Bambusa arundinacea*. It was in the year 1789 that the first bamboo genus was scientifically described under the name Bambos on the basis of a single species *Bambos arundinacea*, now known as *Bambusa arundinacea*. Roxburgh (1814) the father of Indian Botany, in his *Hortus Bengalensis* enumerated seven species under the name *Bambusa arundinacea, B. striata, B. tulda, B. balcooa, B. baccifera, B. spinosa* and *B. nana*.

The first mention of bamboo is found in works of Ctesius, in a letter from Alexander the Great to Aristotle and in the Natural History of Pliny (Bahadur 1979). Munro (1868) in his Monograph of Bambusaceae laid the foundation of our modern knowledge of bamboos where he enumerated 220 species with description of 170 species under 20 genera of bamboos of the world and classified them under three divisions. His work also includes 70 species of bamboo from Indo-Malayan region.

Beddome et al. (1873) dealt with 18 South Indian bamboos. Many of the species of bamboos found in India were also studied by Kurz (1877) in his Forest Flora of British Burma, where he described 30 species.

In India the first taxonomic study on bamboos was initiated by Von Rheede, Dutch Governor of Malabar, who in his *Hortus Malabaricus* in 1678 described and illustrated two kinds of bamboos, which are now known as *Bambusa arundinacea* and *Ochlandra scriptoria* (Chatterji et al. (1963); Bahadur (1979)). Grass family is monophyletic and the early diverging lineages recognized within the family are Anomochlooideae, Pharoideae and Puelioideae (Barker et al. 2001). Anomochlooideae lacks a true spikelet and is sister to the rest of the family

members. Pharoideae is the earliest lineage from the true spikelet-bearing group and was followed by Puelioideae. Traditionally, the members of the group share some common features that include rhizomatous habit, hollow segmented culms, petiolate blade with tessellate venation, flowers with three or more lodicules, usually with six stamens, and fruit possess small embryo and linear hilum (Soderstrom 1981). Few synapomorphic features which are unique for Bambusoideae were reported by (Barker et al. 2001). Leaf blade is mainly constituted of mesophyll tissue with asymmetrically invaginated arm cells, while pseudo-petiole structures are secondary gain for the sub-family. It is broadly divided into two tribes, that is Bambuseae/woody bamboos and Olyreae/herbaceous bamboos depending on the presence (Bambuseae) or absence (Olyreae) of the abaxial ligule (Zhang et al. 2000); (Barker et al. 2001).

The taxonomic study of bamboo has undergone numerous changes over the decades. Because of its complexity and diverse nature, time and again the taxonomy of bamboo has been studied with the latest study by the Grass Phylogeny Working Group (GPWG) (Barker et al. 2001).

Bamboos are members of the sub-family Bambusoideae within the grass family Poaceae. Traditionally, the members of the group share some common features that include rhizomatous habit, hollow segmented culms, petiolate blade with tessellate venation, flowers with three or more lodicules, usually with six stamens, and fruit possess small embryo and linear hilum. A few synapomorphic features which are unique to Bambusoideae, have been reported. Leaf blade is mainly constituted of mesophyll tissue with asymmetrically invaginated arm cells, while pseudo petiole structures are secondary gain for the sub-family.

Bamboo is one of the most productive and fastest growing plants on the planet. The fastest-growing species may grow up to 1.2 m a day. The unique growing capacity makes bamboo a valuable sink for carbon storage. Below ground bamboo biomass makes up 25-50% of the total stock. Carbon content comprises usually about 50% of the total biomass.

Bamboo is fast and effective sequesters of carbon, and will provide a harvest of woody biomass for durable uses on an annual basis. Given the strong potential of bamboo for poverty alleviation in rural areas and in helping communities adapt to climate change, the expansion of bamboo, coupled with the production and development of a broader range of durable products, could address mitigation without compromising development objectives (Nath et al. 2009).

Terrestrial ecosystems constitute a major Carbon sink owing to the photosynthesis and storage of CO_2 in live and dead organic matter. Owing to its numerous ancillary benefits (e.g. improved soil and water quality, restoration of degraded ecosystems, increased crop yield), terrestrial Carbon sequestration is often termed as win–win or no-regrets strategy. It offers multiple benefits even without the threat of global climate change (Song et al. 2011).

2.2 Classical taxonomy

Plant morphology is basically a study of the structure of plants or rather its phenotypic characters. Bamboo identification through morphological features has always been a challenge for Botanists and Taxonomist. Bamboo very rarely flowers and few species have no record of flowering at all. But the morphological characters for identification of higher plants are heavily skewed towards flowers and floral characteristics.

One of the most important and seminal work on morphological with a view of characterizing and identifying plants was carried out by Joseph Dalton Hooker from 1875 to 1897 and published the "Flora of British India" in 7 volumes (Hooker 1897). This work by Hooker is a path breaking work on plant identification and for the first time a study on the plants in India was carried out with a scientific approach.

Morphological studies have been carried out in various plants and plant parts. The study of Foster et al. (1959) have done a comprehensive and detailed study in their book "Comparative morphology of vascular plants." Keng (1962) has done a

comparative study on leaf, petiole, stem, wood, floral, fruit, seed, seedling anatomy of Theaceae. Chatterji et al. (1963) in their study entitled Culm sheaths as Aid to Identification of Bamboos has provided an insightful work on Bamboo identification. Knowing fully well the lack of flowering in Bamboo, they have provided an alternative but effective means for easy identification of Bamboo.

Bardenas et al. (1965) has done a very important study on the morphology and varietal characteristics of the rice plant. Takhtajan (1980) in his study on Classification of flowering plants (Magnoliophyta) has considered growth habit, leaves and their arrangement, stomatal apparatus, nodal arrangement etc. in his study. The above two studies was extremely useful in later studies on Poaceae family. The studies revisited some characters and showed their importance in the taxonomic study of grasses.

Soderstrom et al. (1988) has done an in-depth Morphological - anatomical study of the woody bamboo of Sri Lanka. They have given all the morphological and anatomical characteristics which are and can be used in bamboo taxonomical studies.

Stapleton (1997) in his study on the morphology of woody Bamboo has provided numerous keys for Bamboo identification. But also mentions that it's very easy to make mistakes in identifying a Bamboo species as 1 or 2 mistake is enough to completely misidentify a species. The evolutionary placement of the different genus and species also proves problematic due to lack of flowering. Stapleton (1997) has studied the morphology of bamboo rhizome, branching, inflorescence and synflorescence to elucidate the Bamboo group. He has not only done a comprehensive study of Bamboo morphology, but has also pointed out the various difficulties in species identification and its pitfalls.

These are some of the reasons that as early as 2001, the Grass Phylogeny Working Group has worked on Phylogeny and subfamilial classification of the grasses (Poaceae) (Barker et al. 2001) and has made a number of recommendations.

A study by Machine vision or image analysis for morphological identification of Oats was carried out by Sumathi et al. (2013) and could successfully identify 11 cultivars of Oats by comparing their morphological characters. In many cases only one morphological important character was chosen for study as indicated by the study on pollen morphology. Erdtman (1986) in his book entitled “Pollen Morphology and Plant Taxonomy: Angiosperms” have discussed in great details how pollen structures can be used in plant identification.

Apart from the pure morphological study with an aim to identify plant species, morphological studies has also been carried out to enumerate the morphological changes brought about by environmental factors or chemicals. Funayama-Noguchi et al. (2014) has studied the difference in root morphology in response to Phosphorus deficiency in *Lupinus albus* and *Lupinus angustifolius*.

Thorne (1992) in his study on floral, seed and fruit morphology has revisited their importance in plant taxonomy. Reynolds et al. (1994) studied the physiological and morphological traits associated with hot irrigated condition in Wheat. Floral morphology has also showed the shift in reproductive strategy in Solanum as indicated by the study of Vallejo-Marin et al. (2014).

Wei et al. (2014) carried out Quantitative trait loci (QTL) studies in plants and have linked up certain morphological and physiological characters which have the potential to enable plant breeders to work on desirable characters.

The problems of Bamboo taxonomy and identification can be understood by the fact that there have been many studies to redefine and rethink about the classification of bamboo. The bamboo being a complex group and the lack of flowers (in most cases) have always proved to be a challenge to taxonomist for a long time. Studies on Bamboo taxonomy by (Bahadur 1979), (Stapleton 1997), (Zhang et al. 2000) and the Grass Phylogeny Working Group (Barker et al. 2001) clearly showed that the problems have not been solved till date.

2.3 Molecular Marker

Review of literature has showed that studies on the identification, genetic variation, diversity has been carried out using techniques like Amplified Fragment Length Polymorphism (AFLP) (Loh et al. 2000) and Polymorphic expressed sequence tag - simple sequence repeat (EST-SSR) markers (Barkley et al. 2005). The former was done with samples from the Singapore Botanic Gardens and the latter was done on 44 species from 11 genera, samples which were collected from the United States Department of Agriculture – Agricultural Research Station (USDA–ARS) germplasm collection. It is maintained by the Plant Genetic Resources Conservation Unit, Byron, Georgia, USA. Taxonomic information on each of these samples can be found on the Germplasm Resource Information Network.

Review of literature also showed that different groups working on elucidation of Poaceae and Bamboo taxonomy had an inclination of using RAPD based molecular marker and had faith that RAPD could indeed provide a solution as indicated by the following studies.

Nayak et al. (2003) have worked to evaluate the genetic variability in bamboo using RAPD Markers. Das et al. (2005) have worked on generation and characterisation of SCARs and use of RAPD as species specific markers to study bamboo biodiversity. They had limited success in their effort as Bamboo being a complex group; it could not be explained in a simple approach.

Das et al. (2008) in a paper have discussed in great details about the Taxonomy and diversity of Bamboo in molecular markers era. In this paper the group has looked at Relevance of Molecular Taxonomy in Bamboo. Also most of the DNA and sequence based methods for studying the taxonomy and diversity of Bamboo has been studied. The study included DNA Fingerprinting Based techniques like RFLP, RAPD, SCARs, AFLP, Microsatellites (SSRs), Expressed Sequence Tag Derived Microsatellites (EST-SSR) and Transposon. Ramanayake et al. (2007) used RAPD to check the genetic diversity and relation between 9 species of Bamboo in Sri Lanka.

Mukherjee et al. (2010) have also worked on ISSR and EST-Based Random Primers on 22 different taxa of Bamboo and concludes that "This calls for correct taxonomic delineation at the genus and species level using both vegetative and reproductive characters and correlation of molecular data with morphologically definable taxonomic groupings at the proper taxonomic level."

Pervaiz et al. (2010) have used RAPD to assess the genetic variability of Rice germplasm of Pakistan, with limited success. Dong et al. (2012) have worked on sixteen novel microsatellite markers developed for *Dendrocalamus sinicus* with an aim to extend the work to encompass other bamboo species and develop identification markers. Ruiz-Sanchez et al. (2010) understanding the limitations of molecular data in phylogenetic study of Bamboo undertook a study using morphological, ecological and molecular data with a view to delimiting species boundaries within the Neotropical bamboo *Otatea* (Poaceae: Bambusoideae). Yang et al. (2012) have worked on Genetic diversity and differentiation of *Dendrocalamus membranaceus* (Poaceae: Bambusoideae) using ISSR.

RuiHua et al. (2012) has carried out a study on ISSR marker based study on genetic diversity of turf bamboo. Yeasmin et al. (2014) have undertaken a study taking a number of molecular markers viz. AFLP, ISSR, RAPD, RFLP, SNP and SSR to look at Bamboo genetic diversity and characterisation. It is seen that molecular marker whether alone or in combination is capable of Genetic diversity assessments among the identified bamboo species, either independently or in combination with morphological traits. But the group has also pointed out that although they have studied 6 molecular markers, RAPD and ISSR has showed the maximum potential for Bamboo diversity study.

2.3.1 DNA sequence based marker

With the advancement of technology, DNA sequencing has progressed rapidly, thus becoming cheaper, faster and enhanced fidelity. This has led to a spurt of studies based on short segments of Chloroplast, Mitochondrial and/or Nuclear DNA. A few successful attempts at whole Chloroplast sequencing have also been noted. While

reviewing literature, we have observed the following DNA sequences available in public databases like NCBI for Bamboo biodiversity and genetic variability studies.

Sl	Genes/ Introns	Sl	Genes/ Introns
1	*atpB-rbcL* intergenic spacer,	17	*trnD-trnT* intergenic spacer,
2	retrotransposon *Ty1-copia*,	18	*psbA-trnH* intergenic spacer,
3	*GBSSI* gene	19	*trnK-rps16* intergenic spacer,
4	maturase K (*matK*) gene,	20	Rps16 (*rps16*) gene,
5	*ndhF* gene	21	*trnT-trnL* intergenic spacer,
6	AtpI (*atpI*) gene,	22	*trnV-ndhC* intergenic spacer,
7	*atpI-atpH* intergenic spacer,	23	*ORF170* gene,
8	*psaA-ORF170* intergenic spacer,	24	*rps16* gene, intron
9	Rpl32 (*rpl32*) gene,	25	PsaA (*psaA*) gene,
10	*rpl32-trnL* intergenic spacer,	26	transposon Pong-Like,
11	tRNA-Leu (*trnL*) gene,	27	tRNA-His (*trnH*) gene
12	*rps16-trnQ* intergenic spacer,	28	tRNA-Leu (*trnL*) gene
13	tRNA-Gly (*trnG*) gene, intron	29	*trnL-trnF* intergenic spacer
14	*trnC*-rpoB intergenic spacer,	30	tRNA-Phe (*trnF*), region
15	ribosomal protein L16 (*rpl16*) gene,	31	*rbcL* gene
16	transposon mariner-like element,	32	*trnK*-trnI

Table 2.1: List of genes/introns used for study of Bamboo diversity (till date).

DNA Sequence based techniques are used on Organellar genes and Nuclear genes. All these techniques have indicated that there is no single molecular technique or single gene which can be used in identification of different Bamboo species. But on the positive side it has been proved that all the above mentioned modern molecular techniques can be used in the study and elucidation of Bamboo genetic diversity and derivation of its phylogenetic relationships. Many groups have used varied DNA sequences which showed different success rates among plants. A DNA segment which could establish the phylogenetic relationship in a specific plant group showed its inability to do so in another plant group.

Zhang et al. (2000) has tried to construct the phylogeny of the grass family (Poaceae) using *rpl16* intron sequence data and found that some plant groups are clearly resolved and some are poorly supported and demarcated.

Ge et al. (2002) has tried to build the phylogeny of rice tribe Oryzeae (Poaceae) based on *matK* sequence data. Yang et al. (2008) using nuclear *ITS*, *GBSSI* gene and plastid *trnL-F* DNA sequences conducted a molecular phylogenetic and fruit evolutionary analysis of the major groups of the paleotropical woody bamboos (Gramineae: Bambusoideae)

Sungkaew et al. (2009) conducted a study on woody bamboo undertaking a combined analysis of five plastid DNA regions, *trnL* intron, *trnL*-F intergenic spacer, *atpB-rbcL* intergenic spacer, *rps16* intron and *matK*. But the study is not entirely successful as the authors conclude that the results indicate a need to revise the classification of Bambuseae. Goh et al. (2010) undertook a study on Phylogenetic relationships among Southeast Asian climbing bamboos using the plastid DNA non-coding regions *rps16-trnQ*, *trnC-rpoB*, *psbA-trnH* and *trnD-T*, and a partial nuclear *GBSSI* gene.

Zhou et al. (2011) have carried out study on the transposable elements, particularly mariner-like elements (MLEs). They showed that MLEs are widespread, abundant and diverse in the Bambusoideae subfamily. In this study, the group isolated 79 full-length MLE transposase genes from 63 bamboo species representing 38 genera. Zeng et al. (2012) has taken just 2 regions viz. *atpB-rbcL* and *ndhF* DNA sequences to study Phylogeny of Oryzoideae species and related taxa of Poaceae.

The review of literature of various studies, time and again points to the problem that DNA based studies, both marker based and DNA sequence based, fails to agree in innumerable cases with the traditional morphology based phylogenetic findings. Zhang et al. (2012) in their study has found incongruence between plastid and nuclear *GBSSI* gene phylogenies of Arundinarieae (the temperate woody bamboos).

Kelchner (2013) conducted a study on Five plastid DNA regions: one gene (*ndhF*), two group II introns (*rpl16* intron, *rps16* intron), and two Intergenic spacers (*trnD*–

trnT, *trnT–trnL*) to determine the Higher level phylogenetic relationships within the bamboos, with limited success.

Liu et al. (2014) in their study with sequence data of two low-copy nuclear (LCN) genes, Phosphoenolpyruvate carboxylase 4 (*Pepc4*) and granule-bound starch synthase I (*GBSSI*) on *Sorghum* has successfully differentiated the species and has showed that it is also capable of resolving poorly resolved species.

Wynns et al. (2014) has conducted a study taking 16 DNA regions which were sequenced. The regions are nuclear *ITS* and *26S*, and plastid *rps4*, *rps4-trnL*, *trnL*-F, *trnK, matK-psbA*, *psbA-trnH*, *trnM-V*, *trnD-T*, *rbcL*, *atpB-rbcL*, *psbT-H, rpoC1* exon 2 (partial), the *trnG* intron, the *rpl16* intron and the plastid ribosomal spacer DNA (cp*ITS*). As discussed previously the current research trends points towards multi locus approach towards genetic diversity and phylogenetic studies.

2.4 DNA barcoding

Hebert et al. (2003) of University of Guelph, Canada, for the first time proposed the formation of a public domain of small, specific regions of DNA (called barcodes) that would be unique to each species. This domain shall endeavour to provide a new way for identifying species, one whose capability will increase as more specimen and taxa are added with the decreasing cost of DNA sequencing. DNA barcoding shall provide a standard protocol for identifying species.

2.4.1 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 *psbA-trnH* 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 *psbA-trnH* 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 *psbA-trnH* serve as good starting points for large scale testing of DNA barcoding across a large sample of angiosperm.

Newmaster et al. (2006) 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.

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 as viable markers for land plant Barcoding:

(1) *rpoC1*, *rpoB* and *matK* or

(2) *rpoC1*, *matK* and *psbA-trnH*.

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 more than 1000 species of Mesoamerican orchids, DNA barcoding with *matK* alone revealed cryptic species and proved useful in identifying species listed in Convention on International Trade of Endangered Species (CITES) appendixes.

Fazekas et al. (2008) revealed multiple multilocus DNA barcode from the plastid genome worked equally well and up to 69-71% level of discrimination 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% (*psbA-trnH*) of species with well supported monophyly. Several loci (*matK, psbK-psbI, psbA-trnH*) were found 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 (R) 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. The *matK* and *ITS* sequences 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 in combination. 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 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 candidates of plastid DNA regions (*atpF-atpH* spacer*, matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK-psbI* spacer and *psbA-trnH* 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 in this group. Among 397 samples successfully sequenced, for single locus barcodes, the species discrimination capability were ranked in order (all 7 loci): *rpoC1* < *rpoB* < *atpF-atpH* < *rbcL* < *matK* < *psbK-psbI* < *psbA-trnH*. Based on the barcode criteria, four of the candidate loci (*rpoC1, rpoB, atpF-atpH, psbKpsbI*) were excluded in plant DNA barcoding system; although none of the 3 loci fit all the criteria perfectly.

Asahina et al. (2010) investigated the species identification of five *Dendrobium* plants (*D. 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 (Hollingsworth et al. 2009). 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. The *matK* gene 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 *psbA-trnH* ranked first in divergence value in the analyzed species but the high variability of this DNA spacer

did not allow proper alignment. The *matK* gene showed easy amplification and alignment in the analysed species and high level of discrimination values. Among the nuclear sequences, they suggested *At103* as the most suitable candidate. They recommend the combination of plastid and nuclear markers to identify toxic plants. The plastid markers, *matK* and *psbA-trnH* showed consistent genetic variability.

Dunning et al. (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 *psbA-trnH*) from 31 species of the Lemnaceae family. They examined the extent of barcoding gap between intra and inter specific variation by pair wises comparison 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 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 were many arguments about the classification of the *Citrus* genera, especially the division of species. *Citrus* easily undergo bud mutation and has a long history of artificial cultivation and there are many morphological types of *Citrus*. The unsuccessful identified species in this study were mainly cultivars.

Jeanson et al. (2011) tested the DNA barcoding technique (*matK, rbcL, nrITS, psbA-trnH*) on palms, for the first time. This study was 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, psbA-trnH, ITS* on 196 individuals from 9 genera and 54 species of family Juglandaceae and found that *ITS* sequence has the most variable information and *rbcL* gene has the least. The *matK* has efficient enough to discriminate the seven of nine genera of Juglandaceae. The *ITS* sequence has higher interspecific p-distance than the *psbA-trnH* 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 *ITS* as the first tier DNA region for genus discrimination and second locus at species level.

Fu et al. (2011) studied 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 et al. (2011) revealed that *ITS* region correctly identified between *Gentianopsis paludosa* and all nine adulterate species. Their finding also showed that short length of *ITS* is a

advantages on the DNA amplification and interspecific divergence of *ITS* is higher than the intraspecies divergence. Gu et al. (2011) compared the four barcoding markers (*matK, rbcL, psbA-trnH, nrITS*) to differentiate species within Ligustrum. Their result showed that *ITS* sequence has the most variable sites followed by *psbA-trnH*, *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 *psbA-trnH*) to identify species within 148 individual plant samples encompassing of 38 species of *Dioscorea*. The *matK* gene successfully identified 23.36% of all species compared with 9.30% for *rbcL* and 11.63% for *psbA-trnH*. 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 *psbA-trnH* spacer showed problems 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.

Apart from the studies cited below, there has been tremendous effort from the Indian purview regarding Plant DNA Barcoding as clearly evident from the works of Parveen et al. (2012) on endangered Indian *Paphiopedilum* species; Sharma et al. (2012) work on *Cymbidium* using *ITS* region and Tripathi et al. (2013) on tropical tree species of India using the *ITS* region.

2.4.2 Problems in Plant DNA Barcoding

Plant Barcoding has proved to be a difficult task from the very beginning. The huge size of the Plant genome and characteristics like self fertilization, asexual reproduction, polyploidy, clonal varieties had indicated that DNA Barcoding for plants is going to be difficult. In animal kingdom Cytochrome oxidase I (*COI*) has proved itself to be the universal Barcode sequence. However, *CO1* and other mitochondrial genes have not proven suitable as a barcode for plants because of their low mutation rate and the rapidly changing structure of this genome (Cho et al. 1998); (Adams et al. 2002); (Cho et al. 2004).

DNA Barcode is in use since the last decade and has proved itself to be very effective, efficient and accurate in identification of different species at any stage of life. But many plant scientists had different opinion as regards to which segment of DNA to be used in Plant Barcode. Newmaster et al. (2006) proposed a single locus *rbcL* for plant Barcode. Lahaye et al. (2008) proposed that *matK* locus be used universally for Plant Barcode.

In the DNA Barcoding (Consortium of Barcode of Life) conference held in Mexico, a plant study group finally came to consensus on two genes which can be effectively used as Plant Barcode. Based on assessments of recoverability, sequence quality, and levels of species discrimination, they recommend the 2-locus combination of

rbcL+matK as the plant barcode. This core 2-locus barcode will provide a universal framework for the routine use of DNA sequence data to identify specimens and contribute toward the discovery of overlooked species of land plants (Hollingsworth et al. 2009).

The passport or Barcode provides easy, effective and accurate method for identification of species. This process does not require the sequencing of large DNA sequence but just two locus of the Chloroplast DNA.

A team in Instituto de Ecología, A.C. (Biología Evolutiva), Veracruz, México have started a project with an objective to barcode the Mexican Bamboos species to allow easy identification. Such a project is yet to be undertaken in North East India even though the species richness of bamboo is similar or greater than in Mexico. In this project they will be following the CBOL protocol and will use the 2-locus combination of *rbcL+matK* as the plant barcode.

There are a few more genes or stretch of genome which has been considered as possible barcoding gene. Apart from *matK* and *rbcL* genes like *rpoB, rpoC1, trnH-psbA, psbK-psbI, atpF-atpH,* and internal transcribed spacer (*ITS*) is also being considered. Recently a series of work seems to indicate that the non-chloroplast region of *ITS* has a lot of promise as a possible Barcode. But it does not mean that *matK* is defunct, it simply means that more work needs to be done. Recently new universal primers for amplifying *matK* in Gymnosperms and Angiosperms was created.

Fazekas et al. (2008) showed that the number of species that were correctly identified by barcode marker was lower in plants (~70%) than animals (>90%). A study with 12 genera of plants showed that hybridization has a large contribution to variation in genetic diversity in plants.

Despite the success of Plant Barcoding, a number of failures have also been reported in diverse groups. In complicated plant groups like *Solanum* sect. *Petota* (wild potatoes) Spooner (2009) showed that despite using various potential DNA Barcoding sequences, it was incapable of elucidating and clarifying the phylogenetic

relationship among the species. Roy et al. (2010) showed that Universal plant DNA barcode loci may not work in *Berberis* and Maia et al. (2012) showed the failure of DNA Barcoding to correctly differentiate the plant group Bromeliaceae, the Barcode sequence chosen could not properly discriminate the phylogenetic relationship. Recently, few researchers have called for evolution, adaptation of new technologies and an overhaul of the Barcoding strategy, thinking and philosophy (Taylor et al. 2012).

The *ITS* sequences have proved itself by successfully discriminating Green Algae, in the family of Primulaceae and Juglandaceae. The *ITS* also worked in the genus *Pterygiella, Hedyotis* and *Ligustrum.* But as in the case of genus *Tetrastigma*, it has been proved that *ITS* alone cannot be used as a barcode sequence. A combination of 3 sequences *matK+rbcL+ITS* has been suggested for full discrimination and divergence study. *ITS* also has showed its capability to discriminate adulterants mixed with medicinal plants, *Gentianopsis paludosa* as showed by Xue (2011). So, form the above information and many other examples, *ITS* has proved its capability as a Barcode sequence with high success rate.