CHAPTER 5
DISCUSSION

CHAPTER 5

DISCUSSION

5.1 Efficiency of Chloroplast *matK* as well as *trnH-psbA* intergenic spacer as DNA passport for common medicinal plants:

The *matK* gene was selected as plant DNA barcoding system on the basis of its extensive variation and ability to distinguish species from one another. Inherently, this variation also extends across the primer binding sites (Dunning and Savolainen 2010). Hence some researchers had reported that the amplification of *matK* was problematic by using single universal set primer (Chase et al., 2007; Hollingsworth et al., 2011). In this study, the previously published primer (*matK* X-F and *matK* 5 r) is certainly successful in the development of DNA barcode from important medicinal plants only exception in the *Citrus* species. Luo et al., (2010) also reported inability of the *matK* primer in the Rutaceae family due to the instability and the uniqueness of the primer's 3' end. Therefore the family specific *matK* primer (RUT *matK* F and RUT *matK* R) was designed for family Rutaceae to test its efficiency as DNA barcode for *Citrus* species and successful PCR amplification had been achieved. We determined the 5 *matK* sequences from young leaves of different *Citrus* species. **Our newly designed family specific primer set is an example of broad range primer, which is one of the criteria for ideal barcode gene and *matK* do possess the same**. Further, the published primer of *trnH-psbA* successfully amplified the target region. The sequences of the primers used for PCR amplification can have a major effect on the specificity and sensitivity of the reaction. Since, specificity and the temperature as well as time of annealing are at least partly dependent on primer length, this parameter is critical for successful PCR. One of the most important applications of barcoding is to facilitate the identification of unknown specimens. If the previously published universal barcoding primers fail, **the family-specific primers will be**

useful. If no taxonomic information about the sample is available, species identification can still be achieved through a two tiered approach, like at the very outset a sample can be categorized to its family by the use of family specific primer on the *trnH-psbA*.

One of the conditions for DNA identification of species is representation of relevant taxa in the reference database (GenBank) and there must be sequence differences that discriminate among closely-related species. A more precise indicator of species representation is whether the recovered sequences are similar to in the database. We can determine how well this condition is met for our specimens by comparing the closest match and nearest neighbor with same or different species in each haplotype. Sample AUMP4 produced *matK* sequence closest to both *Helicteropsis microsiphon* and *Hibiscus rosa-sinensis.* Similarly, *matK* sequences of *Citrus maxima* (AUMP45 and AUCM3) and *Crotalaria trichotoma* (AUMP68) showed representations of two different species and remained inconclusive. It may be due to short sequence length in comparison to other barcode sequences. Plant genera with the possible occurrence of natural hybridization and gene introgression may be quite challenging in search of universal loci for plant DNA barcode (Fazekas et al., 2009; Roy et al., 2010). Plant DNA barcoding in these cases may be problematic and demanding. Longer sequence or additional barcode markers may provide a better resolution in case of *Hibiscus rosa-sinensis, Citrus* species. Large sample sizes are required to increase the power of the test with such members, but the limited number of *matK* sequences in the database poses a limitation to flags diagnostic nucleotide positions (details discussed in 5.3). The *matK* gene showed the relatively small number of position's differences that distinguish many closely related plant species by comparing the closest match with identical species and NN of same or different genus (Stoeckle et al., 2011). About one-third of the important medicinal plant species lacked GenBank records for *matK, trnH-psbA* at the time of the

study. This indicates that many plant species are either not represented or have undocumented intraspecific variation. In compressions to *matK,* only few *trnH-psbA* sequences were found in the database.

We authenticated the species name of ethnomedicinal products not only based on BLAST result but also by cross checking the common vernacular name according to the Flora of Assam (Kanjilal et al., 1934-1940). Single loci *matK* show enough discrimination power among the studied medicinal plants' species which reckoned the other findings (Asahina et al., 2010; Gao et al., 2011). **This study further proved that DNA barcoding can provide the means for studying and identifying medicinal plant products that are unidentifiable by morphology alone** (Srirama et al*.,* 2010; Kool et al., 2012; Mahadani and Ghosh 2013). In the present study, some ethnomedicine from southern Assam were successfully identified using *matK* sequence based barcode. This study reckoned the significance of *matK* gene having a powerful discrimination capacity at the species level (Lahaye et al., 2008). Similarly, different plant barcoding studies have proven the evidence of *matK* gene as a possible marker of Plant DNA barcode. **Thus, the *matK* gene would be a potential marker in ethnobotany research** (Mahadani et al., 2013).

5.2 Phylogenetic relationship of important medicinal plants with special reference to Apocynaceae family:

All the selected important medicinal plants were clustered cohesively with the congeneric Genbank sequence. But in case of *Citrus* species, both the loci (*matK* and *trnH-psbA*) showed insufficient intraspecific variation in the genus that hinders the species level differentiation and phylogenetic study. Luo et al., (2010) failed to identify the *Citrus* species among Rutaceae family based on seven DNA barcode markers viz., *psbA-trnH*, *matK*, *ycf5*, *rpoC1*, *rbcL*, ITS2 and ITS (Luo et al., 2010). However, they did not include *matK* in analysis due to the efficiency of PCR amplification. Plant barcodes generally do not exhibit the strong clustering

pattern in such groups even when additional markers are sampled (Stoeckle et al., 2011). Relationship among *Citrus* species is complex due to several factors such as high frequency of bud mutation, long history of cultivation and wide cross compatibility. The identification of citrus at species level through molecular marker has been problematic due to the occurrence of natural hybridization (Fazekas et al., 2009). Character based method was applied for species delineation in this group, as an alternative to the genetic divergence based approach and to be discussed in the successive section 5.3.

The high rate of substitution of *matK* contributed a considerable number of characters for resolving the phylogeny of the ethnomedicinal plants of the Apocynaceae. The partial amino acid sequences of represent the whole RT domain and partially belongs to N terminal and X domain. The *matK* has chemically conserved amino acid replacement properties in this amino acid sequences which may not change the polarity or structural framework of the protein in consequence (Barthet and Hilu 2008). In the present study the high non-synonymous mutation mainly with A+T bias in all codon positions strongly influence conserved amino acid replacement. It can be explained as A+T rich codons encode FYMINK amino acids and out of which 18.39% nonpolar amino acid (FMI) compared to 9.13% that of (AGP) aroused from G+C rich codon contributing integrity of structural framework. It is known that transversion result in more dramatic changes than do transitions. That is, transversion are more likely than transitions to be non-synonymous in protein coding regions and non-synonymous transversion are more likely to result in chemically conserve amino acid replacement than non-synonymous transitions. It is, therefore, possible that differences in chemical conservative amino acid replacement may be caused by mutation factors, such as the transition-transversion ratio, A+T bias rather than selection forces (Dagan et al., 2002; Zhang 2000).

The sequence divergence (K2P) among the studied ethnomedicinal plants of Apocynaceae revealed the highest divergence (0.119) between *Catharanthus roseus* and *Calotropis gigantea*. Moreover, *Calotropis gigantea* being a member of subfamily Asclepiadoideae always consistently high rate of divergences with other 5 studied members of subfamily Rauvolfioideae. Thus, following the notional DNA barcode concept, it can be justifiably inferred that the use of the partial *matK* sequence having reliable barcode gap as characterized in the study would be appreciably applicable to the species level discrimination of the important ethnomedicinal plants belonging to the family Apocynaceae.

Furthermore, NJ tree showed that the members of Rauvolfioideae subfamily Apocynaceae formed one clade where different species clustered into different subclade. The generated sequences of *Allamanda cathartica* is found closely related to *Allamanda schottii*. It is also close to genera *Thevetia* and *Plumeria*. Although *Alstonia microphylla* and *Alstonia scholaris* are the congeners but placed in different clades, which may be due to polyphyletic nature of Alstonieae (Simoes et al., 2007). Two sequences from *Neruim oleander* of subfamily Apocynoideae, and two members, *viz. Calotropis gigantea* and *Asclepias curassavica* of subfamily Asclepiadoideae, formed two distinct clades at the basal position of phylogenetic tree. Large sample sizes are required to increase the power of the test in Asclepiadaceae subfamily members, but the poor number of *matK* sequences of Asclepiadaceae in the database remained a limitation of the study, which entails the study using large sample sizes from different geographical location.

The CBOL Plant Working Group (2009) confirmed and suggested the combination of *matK* with *rbcL* as a universal plant DNA barcode (Hollingworth et al., 2009) though the low discriminating power of *rbcL* gene is severally reported (Asahina et al., 2010; Sun et al., 2012). On the contrary, insertions, deletions, and short sequence repeats were common and often more numerous

than single base pair substitution that has been the limitation on the part of *trnH-psbA,* hence remained unable to fulfill the criteria of plant DNA barcoding (Bruni et al., 2010). Nevertheless, in the present study, intergenic spacer *trnH-psbA* also exhibited persistent problem in obtaining constant bidirectional sequences. **This study showed that species identification of Rauvolfioideae subfamily is possible using phylogenetic analyzes constructed from partial *matK* sequences (Nt. 520-1278), which is comparable to that of the full-length sequences, also had species discrimination power** (Mahadani et al., 2013). The observed divergences among the studied species using the partial *matK* sequences maintained a reliable gap, which holds good to the concept of species discrimination through DNA barcoding. Furthermore, the NJ phylogenic tree, based on K2P model, also efficiently distinguished the species under study using the partial *matK* sequence. This gene has been identified as a universal DNA barcode for flowering plants (Lahaye et al., 2008). This result suggests that *matK* sequence information could help in correct species identification of medicinal plants of Rauvolfioideae and in providing diagnostics for rapid and easier identification of fallacious species forensics in herbal formulation, which bear the insights of similar application in the family Apocynaceae.

5.3 Utility of Single Nucleotide polymorphism (SNP) / Insertion and deletion (indels) for the development of DNA passport:

Analyzes of the targeted single loci *matK* (~ 750 bp, Nt. 520-1278) sequences of selected medicinal plants of Apocynaceae and depicted repeat structures with AT-rich regions possessing indels in multiple of three. **Occurrences of indels in *matK* sequences have also been explored to the extent of their applicability as qualitative molecular markers depending upon the size, position, and influence of open reading frame.** Several molecular processes are known to create indels, *viz*., polymerase slippages during DNA replication so called slipped-strand mispairing, (Kelchner 2000) and due to

addition or subtraction of short repeat sequences, which are primarily AT rich (Hilu and Alice 1999). In general, microstructural changes in DNA, such as, insertions and deletions (indels), and inversions in introns and intergenic spacers, have been importantly used both for resolving phylogenetic relationships among the angiosperms (Graham et al., 2000; Ingvarsson et al., 2003) and for inferring relationships among more closely related taxa (Golenberg et al., 1993). Imperatively, these changes in protein coding gene are very rare phenomenon, because these changes would lead into non-synonymous mutation. But, the observed indels in the presumed barcode region of *matK* happened in multiple of 3 nucleotides, thereby reduced the chances of frame shift mutation and did not interrupt the site of maturase activity in X domain. So, *matK* indels could be utilized as a qualitative molecular marker for studies both at the intra-specific and shallow inter-specific levels like the intergenic spacers of CpDNA. **To evaluate the indel containing regions, a more powerful algorithm is needed to calculate the intra and inter-species comparisons**.

We examined the use of indels polymorphism as an alternative to distance-based approach for species identification. **We identified a unique 54 bp large inverted repeat within the *trnH-psbA* sequences in *Citrus* only.** Single residue indels were ignored in this analysis because mononucleotide indels and shorts sequence repeats are common and often more numerous than single base pair substitution in *trnH-psbA* (Strochova and Olson 2007) and to avoid the possible sequencing error. We recognized the important diagnostic characters from multiple residue indels in *trnH-psbA* that were successful in indentifying different species (5 out of 7) within the Citrus except *Citrus limon* and *Citrus medica*. These two species not only share identical sequence but also indels in the same positions. **Therefore, this study provides an insight of using indel polymorphism as a species level marker in C*itrus* and the same may be applied in other complex groups**. Likewise, other indels occurring chloroplast regions viz. *trnL-trnF* (Liu et al., 2012) may be tested to resolve species identification in *Citrus medica* and *Citrus limon*.