Chapter 10
ASSESSMENT OF PHYLOGENETIC INTER-RELATIONSHIPS USING CERTAIN cpDNA SEQUENCES

Systematics has gained a new perspective because of both extensive and intensive use of molecular data in evolutionary studies (Savolainen et al., 2000). Different regions of DNA can be used to examine lineages with different levels of divergence. Among them, chloroplast DNA (cpDNA) sequences have been used extensively for elucidating phylogenetic relationships and taxonomic studies in various plant species (Suh et al., 1993; Baldwin et al., 1995; Ainouche and Bayer, 1997; Lashermes et al., 1997; Goel et al., 2002; van den Berg et al., 2002; 2005; Dkhar et al., 2010). Chloroplasts are vital cell organelles of which plant possess their own genetic material and are believed to have originated from ancient endosymbiotic cyanobacteria (Birky, 1995; Dyall et al., 2004; Ravi et al., 2008). The size of chloroplast genome in angiosperms is usually between 115 and 165 kb (Jansen et al., 2005). Typically the circular quadripartite genome consists of two inverted repeats (IRs) separated by two regions of unique DNA, the large (LSC) and small (SSC) single-copy regions (Jansen et al., 2005; Huotari and Korpelainen, 2012). The lack of recombination, low rates of nucleotide substitutions and usually uniparental inheritance make plant chloroplastDNAs a valuable source of genetic markers for phylogenetic analyses (Wolfe et al., 1987; Provan et al., 2001; Korpelainen, 2004; Ravi et al., 2008).

cpDNA sequences variation are the primary source of characters for phylogenetic studies in plants (Giannasi et al., 1992; Chase et al., 1993; Yong-Pyo et
hybrid zones and long distance gene flow (Soltis et al., 1992). cpDNA has proved to be extensively useful for evolutionary and phylogenetic studies and has been a focus of research in plant molecular evolution and systematics (Brouat et al., 2001; Palmer, 1987; Clegg and Zurawski, 1992; Clegg et al., 1994; Olmstead and Palmer, 1994; Avise, 1994; Palmer and Delwiche, 1998; Kelchner, 2000; Cai et al., 2006; Hansen et al., 2007; Moore et al., 2007; Jansen et al., 2007; Parks et al., 2009; Gao et al., 2010; Li and Zhang, 2010; Moore et al., 2010; Zhang et al., 2011; Wu and Ge, 2012). Compared to the nuclear DNA and animal mitochondrial DNA, chloroplast genes have slow rate of nucleotide substitution with an average synonymous rate almost half that of plant nuclear DNA (Wolfe et al., 1987; Zurawski and Clegg, 1987) and thus evolve very slowly (Clegg et al., 1991; Li, 1997). Certain protein-coding gene sequences have been used to elucidate phylogenetic relationships among higher-level taxa (Chase et al., 1993).

Non-coding regions of cpDNA like intergenic spacers have been successfully used to infer phylogenetic relationships at lower taxonomic levels both at intergeneric (Gielly and Taberlet, 1994; Gielly et al., 1996; Bruneau, 1996; Maguire et al., 1997; Asmussen and Liston, 1998; Cros et al., 1998), inter-specific and intra-specific levels (Taberlet et al., 1991; Olmstead and Palmer, 1994; Demesure et al., 1995, 1996; El Mousadik and Petit, 1996; Dumolin-Lape`gue et al., 1997; Petit et al., 1997; Brouat et al., 2001). These sequences exhibit faster mutation rates and tend to evolve more rapidly than coding loci, both in nucleotide substitutions and in the accumulation of insertion and deletion events (indels), presumably because they are less functionally constrained (Curtis and Clegg, 1984; Zurawski and Clegg, 1987; Palmer, 1991; Clegg et al., 1994). Therefore, these non-coding regions have more
informative characters than coding regions of comparable size and have become popular for phylogenetic studies in a broad range of plant groups among taxa that are recently diverged (Procaccini et al., 1999; Downie et al., 2000) and in seed plants (Shaw et al., 2005), in lycophytes and monilophytes (Kelchner, 2000; Small et al., 2005).

From the phylogenetic point of view, a well-corroborated phylogeny would provide a means for evaluating character evolution (Les et al., 1999; Borsch et al., 2008), molecular evolution (Grimm and Denk, 2007), historical biogeography (Löhne et al., 2008) and global changes (Edwards et al., 2007). Phylogenetic reconstructions may also aid in the discovery of greater plant diversity and assist biologists in choosing areas or species to prioritize in their conservation efforts (Cameron, 2010).

A quick perusal of literature reveals that the there have been no attempts made and hardly any research investigations on the phylogenetic studies in the genus *Abelmoschus* are available. *Abelmoschus*, as a genus is taxonomically difficult and complicated. The aim of this analysis was to clarify the systematic and phylogenetic relationships of *Abelmoschus* species to enhance our knowledge on the phylogenetic and evolutionary pattern of the genus. Thus, a robust phylogenetic tree is still needed for its classification and elucidation of the evolutionary history of this genus. In the present study, an attempt has been made to provide an assessment of species interrelationships among species in *Abelmoschus* and understand the phylogeny through sequence homology of certain targeted sequences of cpDNA *viz.*, *accD* (450 bp) *atpB* (1400 bp) and *psbK-psbI* intergenic spacers (500 bp) among the wild and cultivated representatives of *Abelmoschus* species, to have a better understanding of
the pattern of molecular differentiation and ascertaining the probable origin. Affinity and systematic position of species, including wild ones besides other indigenous cultivars/landraces belonging to genus *Abelmoschus* are investigated.

**10.1. Observations:**

Sequence alignment data for elucidating the species inter-relationships among *Abelmoschus* taxa was carried out. Parameters such as sequence length, G+C%, variable sites, conserved sites, parsimony informative sites, number of indels were considered for analyses and the individual sequence data are described below. Because of the different rate of evolution and the different nucleotide composition, phylogenetic analyses were carried out using the sequence data of each region separately as well the cumulative using four different tree building methods (Tables 26 and 27; Figs. 697-712).

**10.1.1. Sequence characteristics of accD and phylogenetic analysis**

In order to conduct phylogenetic analysis, ML, MP, BI and NJ analysis were constructed and the phylogenetic trees were analysed. Sequence length in the 13 *Abelmoschus* accessions ranged from 407 [A. tetraphyllus var. pungens (3M)] to 409 bp (A. tuberculatus, A. enbeepeegearensis, A. ficulneus, A. esculentus and A. caillei) in individual accessions (Table 26 and 27; Figs. 697-700) as compared to 408 bp observed in out-group taxa represented by *Gossypium stocksii*. The final aligned data matrix of the accD, presently analysed shown 409 characters which includes 395 conserved sequences, 14 variable characters and 10 parsimony-informative positions. The number of indels was recorded to be 3 and 409 aligned nucleotide sites were invariant. Percentage of sequence divergence based on substitution plus indels was 2.69% (Table 26). These accessions also showed very little variation not only in
sequence length but also in overall G+C content (%) of the taxa studied (Table 27). In our present study, the chloroplast DNA sequences of the *Abelmoschus* species studied were AT rich as compared to GC content. The G+C content ranged from 37.25% (*A. tetraphyllus* var. *pungens*) to 37.99% (*A. moschatus* ssp. *tuberosus*). The nucleotide frequencies for *accD* gene sequence (partial) were found to be 0.303(A), 0.175(C), 0.204(G) and 0.318(T) while transition/transversion bias (R) was 0.583.

ML method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the *accD* partial sequence (Fig. 697). A clear relationship among subgenera is observed and the phylogenetic tree has been resolved into three clusters with two major clusters and a minor cluster. Cluster 1 comprising of *A. tetraphyllus* var. *pungens*, *A. angulosus* var. *grandiflorus*, *A. tetraphyllus* var. *pungens* (3M), *A. caillei*, *A. tetraphyllus* var. *tetraphyllus*, *A. moschatus* ssp. *tuberosus*, *A. tuberculatus*, *A. moschatus* ssp. *moschatus* and *A. esculentus*. Cluster 2 includes *A. ficulneus*, *Abelmoschus* spp. and *A. enbeepeegearensis*. Cluster 3 consists of only *A. manihot*. *G. stocksii* was separately attached at the base of tree as the diverging *Abelmoschus* relative’s lineage.

MP method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the *accD* partial sequence (Fig. 698). The phylogenetic tree was resolved into two major clusters, with cluster 1 resolving to into distinct clades. Cluster 1 has been further resolved into 4 sub-clusters viz., 1a, 1b, 1c and 1d, wherein cluster 1a consists of *Abelmoschus* spp., *A. ficulneus* and *A. enbeepeegearensis*, 1b consisted of *A. manihot*, *A. esculentus* and *A. tuberculatus*. While 1c and 1d comprised of *A. moschatus* ssp. *tuberosus* and *A. moschatus* ssp. *moschatus*, *A. angulosus* var. *grandiflorus* and *A. tetraphyllus* var. *pungens* respectively.
tetraphyllus var. tetraphyllus and A. caillei formed the cluster 2. G. stocksii was separately attached as an out-group at the base of tree as the diverging Abelmoschus relative’s lineage.

BI method was used to assess the phylogenetic relationship of the genus Abelmoschus based on the accD partial sequence (Fig. 699). The phylogenetic tree based on BI method was resolved into only two major clusters. Cluster 2 consists of A. caillei and A. tetraphyllus var. tetraphyllus while cluster 1 consists of A. tetraphyllus var. pungens, A. angulosus var. grandiflorus, A. tetraphyllus var. pungens (3M), A. moschatus ssp. tuberosus, A. tuberculatus, A. moschatus ssp. moschatus, A. esculentus, A. ficulneus, Abelmoschus spp., A. enbeepeegearensis and A. manihot. G. stocksii was separately attached as an out-group at the base of tree as the diverging Abelmoschus relative’s lineage.

The phylogenetic tree was constructed using NJ method based on the distance matrix for assessing the relationships within the species of the genus Abelmoschus based on the accD partial sequence (Fig. 700). The relationship among the species has been resolved into four major clusters. Cluster 1 has been resolved into two sub-clusters with 1a consisting of A. moschatus ssp. moschatus and A. moschatus ssp. tuberosus while 1b consists of A. esculentus and A. tuberculatus. Cluster has also been resolved into two sub-clusters viz., 2a and 2b comprising of A. enbeepeegearensis and A. manihot, while A. ficulneus and Abelmoschus spp. respectively. Cluster 3 consists of A. caillei and A. tetraphyllus var. tetraphyllus while cluster 4 includes A. tetraphyllus var. pungens (3M), A. tetraphyllus var. pungens and A. angulosus var. grandiflorus which further resolved into sub-cluster 4a consisting of A. tetraphyllus var. pungens and A. angulosus var. grandiflorus.
It is quite intriguing to note that in BI, MP and NJ methods used in deducing phylogenetic relationship in 13 taxa, has revealed an identical pattern in the clustering of the species *A. caillei* and *A. tetraphyllus* var. *tetraphyllus* together. In ML and MP, clustering of *A. ficulneus, Abelmoschus* spp. and *A. enbeepeegearensis* was recorded. While in MP and NJ, *A. moschatus* ssp. *moschatus* and *A. moschatus* ssp. *tuberosus* were clustered together.

**10.1.2. Sequence characteristics of *atpB* and phylogenetic analysis**

Phylogenetic analysis based on partial sequence data of (~1409-1413) related to *atpB* using ML, MP, BI and NJ was carried out to assess the species inter-relationship and evolution of various *Abelmoschus* species. Sequence length in the 13 *Abelmoschus* accessions ranged from 1409 (*A. angulosus* var. *grandiflorus, A. moschatus* ssp. *moschatus, A. manihot, A. tetraphyllus* var. *pungens* (3M) and *A. caillei*) to 1413 bp (*A. enbeepeegearensis*) in individual accessions (Table 26 and 27; Figs. 701-704) as compared to 1410 bp observed in out-group taxa represented by *G. stocksii*. Among all *Abelmoschus* species studied, the multiple alignment of *atpB* sequences yielded 1410 characters which includes 1376 conserved, 36 variable and 4 parsimony-informative sites; 1413 aligned nucleotide sites were invariable. The number of indels was recorded to be 5 and percentage of sequence divergence based on substitution plus indels was 2.19 % (Table 26). These accessions also showed very little variation not only in sequence length but also in overall G+C content (%) of the taxa studied (Table 27). In our present study, the chloroplast DNA sequences of the *Abelmoschus* species studied were AT rich as compared to GC with the G+C content ranging from 43.62% (*A. tuberculatus*) to 43.95% (*A. enbeepeegearensis*). The nucleotide frequencies for *atpB* gene sequence (partial) were found to be
0.272(A), 0.238(C), 0.201(G) and 0.289(T) while transition/transversion bias (R) was 2.679.

ML method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the *atpB* partial sequence (Fig. 701). A clear relationship among subgenera is observed and the phylogenetic tree has been resolved into two major clusters. Cluster 1 comprising of *A. tetraphyllus* var. *pungens*, *A. manihot*, *A. tetraphyllus* var. *tetraphyllus*, *A. enbeepeegearensis*, *A. tetraphyllus* var. *pungens* (3M), *A. moschatus* ssp. *moschatus*, *A. moschatus* ssp. *tuberosus*, *A. angulosus* var. *grandiflorus*, *A. caillei* and *A. tuberculatus*. Cluster 2 comprises of *Abelmoschus* spp. and one minor clade comprising of *A. esculentus* and *A. ficulneus*. *G. stocksii* was separately attached at the base of tree as the diverging *Abelmoschus* relative’s lineage.

The MP method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the *accD* partial sequence (Fig. 702). The phylogenetic tree was resolved into four clusters, with cluster 1 comprising of *A. angulosus* var. *grandiflorus*, *A. manihot*, *A. caillei*, *A. tetraphyllus* var. *pungens* (3M) and *A. moschatus* ssp. *moschatus*. Cluster 2 comprises of *A. moschatus* ssp. *tuberosus* and *A. tetraphyllus* var. *tetraphyllus* with minor clade 2a comprising of *A. enbeepeegearensis* and *A. tetraphyllus* var. *pungens*. Cluster 3 comprises of *Abelmoschus* spp. and minor clade 3a which includes *A. ficulneus* and *A. esculentus* while cluster 4 includes *A. tuberculatus*. *G. stocksii* was separately attached at the base of tree as the diverging *Abelmoschus* relative’s lineage.

BI method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the *atpB* partial sequence (Fig. 703). The phylogenetic tree
was resolved into two major clusters. Cluster 1 consists of *A. caillei*, *A. tetraphyllus* var. *pungens* (3M), *A. angulosus* var. *grandiflorus*, *A. moschatus* ssp. *moschatus*, *A. manihot*, *A. tuberculatus*, *A. ficulneus*, *A. esculentus*, *A. moschatus* ssp. *tuberosus* and *A. tetraphyllus* var. *tetraphyllus*. Cluster 2 comprises of one minor clade 2a which includes *A. enbeepeegearensis*, *A. tetraphyllus* var. *pungens* and *Abelmoschus* spp. *G. stocksii* was separately attached at the base of tree as the diverging *Abelmoschus* relative’s lineage.

The phylogenetic tree was constructed using NJ method based on the distance matrix for assessing the relationships within the species of the genus *Abelmoschus* based on the *atpB* partial sequence (Fig. 704). The relationship among the species has been resolved into three clusters with two major and one minor cluster. Cluster consist of *A. tetraphyllus* var. *pungens* (3M), *A. enbeepeegearensis*, *A. manihot*, *A. tetraphyllus* var. *pungens*, *A. tetraphyllus* var. *tetraphyllus*, *A. angulosus* var. *grandiflorus*, *A. moschatus* ssp. *moschatus*, *A. caillei*, *Abelmoschus* spp. and *A. moschatus* ssp. *tuberosus*. Cluster 2 comprises of *A. esculentus* and *A. ficulneus* while cluster 3 consists of only *A. tuberculatus*. *G. stocksii* was separately attached at the base of tree as the diverging *Abelmoschus* relative’s lineage.

It is interesting to note the clustering of *Abelmoschus* spp., *A. esculentus* and *A. ficulneus* in ML and MP and also *A. tuberculatus* has been clustered separately in both ML and NJ. The close genetic relationships of *A. esculentus* and *A. ficulneus* could be observed in all the four phylogenetic trees.

10.1.3. Sequence characteristics and of *psbK-psbI* phylogenetic analysis

In order to conduct phylogenetic analysis, ML, MP, BI and NJ were constructed and the phylogenetic tree were analysed. Sequence length in the 13
Abelmoschus taxa ranged from 463 (A. enbeepeegearensis) to 466 bp (A. tetraphyllum var. pungens) (Table 26 and 27; Figs. 705-708) as compared to 449 bp observed in out-group taxa represented by G. stocksii. The final aligned data matrix of psbK-psbl sequences resulted in 409 characters including 395 conserved, 18 variable characters and 1 parsimony-informative positions; 467 aligned nucleotide sites were invariant. The number of indels was recorded to be 4. Percentage of sequence divergence based on substitution plus indels was 3.0% (Table 26). These taxa also showed very little variation not only in sequence length but also in overall G+C content (%) of the taxa studied (Table 27). In our present study, the chloroplast DNA sequences of the Abelmoschus species studied were AT rich as compared to GC content. The G+C content ranged from 37.25% (A. tetraphyllum var. pungens) to 37.99% (A. moschatus ssp. tuberosus). The nucleotide frequencies for psbK-psbl sequence (partial) were found to be 0.358(A), 0.160(C), 0.150(G), and 0.332(T) while the rations between transition and transversion bias (R) was 0.551.

The ML method was used to assess the phylogenetic relationship of the genus Abelmoschus based on the psbK-psbl partial sequence (Fig. 705). A clear relationship among species was observed and the phylogenetic tree has been resolved into three major clusters. Cluster 1 comprising of A. angulosus var. grandiflorus, A. tetraphyllum var. pungens (3M) and A. tetraphyllum var. pungens. Cluster 2 includes A. moschatus ssp. moschatus, A. enbeepeegearensis, A. moschatus ssp. tuberosus and A. tuberculatus. Cluster 3 comprises of A. caillei, A. ficulneus, A. esculentus, A. tetraphyllum var. tetraphyllum, A. manihot and Abelmoschus spp. G. stocksii was separately attached at the base of tree as the diverging Abelmoschus relative’s lineage.
MP method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the *psbK-psbI* partial sequence (Fig. 706). The phylogenetic tree was resolved into three clusters, with cluster 1 comprising of *A. manihot*, *Abelmoschus* spp., *A. caillei*, *A. esculentus*, *A. ficulneus* and *A. tetraphyllus* var. *tetraphyllus*. Cluster 2 comprises of *A. angulosus* var. *grandiflorus*, *A. tetraphyllus* var. *pungens* and *A. tetraphyllus* var. *pungens* (3M). Cluster 3 includes *A. tuberculatus*, *A. enbeepeegearensis*, *A. moschatus* ssp. *moschatus* and *A. moschatus* ssp. *tuberosus*. *G. stocksii* was separately attached at the base of tree as the diverging *Abelmoschus* relative’s lineage.

BI method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the *psbK-psbI* partial sequence (Fig. 707). The phylogenetic tree was resolved into two major clusters. Cluster 1 consists of *A. tetraphyllus* var. *pungens*, *A. angulosus* var. *grandiflorus*, *A. tetraphyllus* var. *pungens* (3M), *A. moschatus* ssp. *moschatus*, *A. enbeepeegearensis*, *A. moschatus* ssp. *tuberosus* and *A. tuberculatus*. Cluster 2 comprises of *A. caillei*, *A. ficulneus*, *A. esculentus*, *A. tetraphyllus* var. *tetraphyllus* *A. manihot* and *Abelmoschus* spp. *G. stocksii* was separately attached at the base of tree as the diverging *Abelmoschus* relative’s lineage.

The phylogenetic tree was constructed using NJ method based on the distance matrix for assessing the relationships within the species of the genus *Abelmoschus* based on the *psbK-psbI* partial sequence (Fig. 708). The relationship among the species has been resolved into three clusters. Cluster 1 consists of *A. caillei*, *A. ficulneus*, *A. esculentus*, *Abelmoschus* spp., *A. manihot* and *A. tetraphyllus* var. *tetraphyllus*. Cluster 2 includes *A. moschatus* ssp. *tuberosus*, *A. angulosus* var.
grandiflorus and A. moschatus ssp. moschatus. Cluster 3 includes A. tuberculatus, A. enbeepeegearensis, A. tetraphyllus var. pungens (3M) and A. tetraphyllus var. pungens. G. stocksii was separately attached at the base of tree as the diverging Abelmoschus relative’s lineage.

It is interesting to note that clustering of A. caillei, A. ficulneus, A. esculentus, Abelmoschus spp., A. manihot and A. tetraphyllus var. tetraphyllus in all the four tree building methods differing only in the position of species within clusters. Likewise, A. angulosus var. grandiflorus, A. tetraphyllus var. pungens and A. tetraphyllus var. pungens (3M) clustered together in both ML and MP. Similarly, A. tuberculatus, A. enbeepeegearensis, A. moschatus ssp. moschatus and A. moschatus ssp. tuberosus also clustered together in both ML and MP.

10.1.4. Cummulative phylogenetic analysis based on accD, atpB and psbK-psbI sequence data

Various methods like ML, MP, BI and NJ have been used to analyze phylogeneny by taking into account the three cpDNA sequences, presently analyzed.

ML method was used to assess the phylogenetic relationship of the genus Abelmoschus based on the cumulative cpDNA sequences of accD, atpB and intergeneric spacer psbK-psbI partial sequences (Fig. 709). A clear relationship among species was observed and the phylogenetic tree has been resolved into five major clusters. Cluster 1 comprising of A. moschatus ssp. moschatus and A. moschatus ssp. tuberosus. Cluster 2 consists of A. caillei and A. tetraphyllus var. tetraphyllus. Cluster 3 consists of A. ficulneus and A. manihot with one minor clade consisting of Abelmoschus spp. and A. enbeepeegearensis cluster 4 comprises of A. esculentus and A. tuberculatus. Cluster 5 comprises of A. tetraphyllus var. pungens
(3M) and a minor clade 5a which includes *A. tetraphyllus* var. *pungens* and *A. angulosus* var. *grandiflorus*.

The MP method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the cumulative cpDNA sequences of *accD*, *atpB* and intergeneric spacer *psbK-psbI* partial sequences (Fig. 710). The phylogenetic tree was resolved into three clusters, with cluster 1 comprising of one minor cluster 1a which consists of *A. enbeepeegearensis* and *Abelmoschus* spp. and also includes *A. ficulneus*, *A. manihot*, *A. esculentus*, *A. tuberculatus*, *A. moschatus* ssp. *tuberosus* and *A. moschatus* ssp. *moschatus*. Cluster 2 comprises of *A. angulosus* var. *grandiflorus*, *A. tetraphyllus* var. *pungens* and *A. tetraphyllus* var. *pungens* (3M). Cluster 3 comprises of *A. caillei* and *A. tetraphyllus* var. *tetraphyllus*.

The BI method was used to assess the phylogenetic relationship of the genus *Abelmoschus* based on the cumulative cpDNA sequences of *accD*, *atpB* and intergeneric spacer *psbK-psbI* partial sequences (Fig. 711). The phylogenetic tree was resolved into three major clusters and one minor cluster. Cluster 1 consists of *A. angulosus* var. *grandiflorus*, *A. moschatus* ssp. *moschatus* and *A. manihot*. Cluster 2 consists of only *A. tetraphyllus* var. *pungens* (3M) while cluster 3 is resolved into one sub-cluster 3a which comprises of *A. esculentus*, *A. ficulneus* and *Abelmoschus* spp. and also *A. tuberculatus* and *A. enbeepeegearensis*. Cluster 4 includes *A. caillei*, *A. tetraphyllus* var. *tetraphyllus*, *A. tetraphyllus* var. *pungens* and *A. moschatus* ssp. *tuberosus*. *G. stocksii* was separately attached as an outgroup at the base of tree as the diverging *Abelmoschus* relative’s lineage.

The phylogenetic tree was constructed using NJ method based on the distance matrix for assessing the relationships within the species of the genus *Abelmoschus*.
based on the cumulative cpDNA sequences of accD, atpB and intergeneric spacer psbK-psbl partial sequences (Fig. 712). The relationship among the species has been resolved into five clusters. Cluster 1 has been resolved into 1a and 1b, with 1a consisting of A. caillei and A. tetraphyllus var. tetraphyllus while 1b of A. esculentus and A. manihot. Cluster 2 comprises of A. enbeepeegeearensis and one minor clade which includes A. ficulneus and Abelmoschus spp. Cluster 3 includes A. tetraphyllus var. pungens (3M) and one minor clade 3a with A. tetraphyllus var. pungens and A. angulosus var. grandiflorus. Cluster 4 consists of A. moschatus ssp. moschatus and A. moschatus ssp. tuberosus while cluster 5 of A. tuberculatus.

10.2. Discussion:

The chloroplast genome of plants has been used extensively for understanding the mechanism of evolutionary change and systematics. Due to its small genome size, relatively stable gene content and slow rates of nucleotide substitution (Palmer, 1985b; Perl-Treves and Galun, 1985; Hiratsuka et al., 1989; Clegg et al., 1994; Nakamura et al., 1997; Provan et al., 1999), they have been used for tracing the evolutionary history of plant species focused on species relationships within a genus. The chloroplast genome generally shows maternal inheritance in higher plants (Reboud and Zeyl, 1994) and has a similar gene arrangement among plant species. Variation in cpDNA is sufficient to provide a high degree of phylogenetic resolution, yet limited enough to be readily assayed by simple techniques of restriction-fragment analysis and mapping analysis. Moreover, parallelisms and convergences occur so rarely at this level that unambiguous phylogenetic trees can be easily built using cladistic approaches (Palmer and Zamir, 1982; Palmer et al., 1983; Palmer et al., 1985; Sytsma and Gottlieb, 1986a, b).
Coding genes exhibit different patterns of codon bias that appear to violate the equilibrium assumptions of some evolutionary models. While non-coding regions of cpDNA diverge through insertion/deletion changes that are sometimes site dependent (Taberlet et al., 1991; Golenberg et al., 1993; Morton and Clegg, 1993; Bohle et al., 1994; Ham et al., 1994; Maner et al., 1994; Mes and Hart, 1994). Only comparative studies of molecular sequences have the resolution to reveal this underlying complexity.

cpDNA variation has been studied for phylogenetic relationship in some 25 different genera of flowering plants. Most of these genera include one or more important crop species with the focus on the origin and evolutionary relationships of these crop plants like Lycopersicon (Palmer and Zamir, 1982; Hosaka et al., 1984), Nicotiana (Kung et al., 1982; Salts et al., 1984), Triticum-Aegilops (Bowman et al., 1983; Tsunewaki and Ogihara, 1983; Terachi et al., 1984), Brassica-Raphanus (Palmer et al., 1983; Erickson et al., 1983), Pisum (Palmer et al., 1985a), Cucumis (Perl-Treves and Galun, 1985), Brassica (Erickson et al., 1983; Palmer et al., 1983), Coffea (Berthou et al., 1983) and Linum (Coates and Cullis, 1987). The most detailed and informative studies on non-crop plant genera are those by Sytsma and coworkers on Lisianthus, Clarkia and Heterogaura (Sytsma and Schaal, 1985; Sytsma and Gottlieb, 1986a, b).

The accD is a class III type of gene encoding one of four subunits of the acetyl-CoA carboxylase beta (carboxyltranferase) subunit enzymes of the type found in prokaryotes and in most chloroplasts (Wakasugi et al., 2001). They are approximately 367 bp long and is large single copy having a high evolutionary rate which is used for phylogenetic constructions. In all vascular plants, except for
Gramineae, the *accD* gene is situated downstream of the *rbcL* gene (Sasaki *et al*., 1993; Konishi *et al*., 1996).

The clustering of *A. moschatus* ssp. *moschatus* and *A. moschatus* ssp. *tuberosus* could be observed in both MP and NJ, they were closely related in BI and ML. *A. caillei* and *A. tetraphyllus* var. *tetraphyllus* were clustered together in all the four topology of the trees. Moreover, the clustering of *A. tetraphyllus* var. *pungens* and *A. angulosus* var. *grandiflorus*, along with *A. tetraphyllus* var. *pungens* (3M) is recorded in all the trees methods. *A. esculentus* was closely related to *A. tuberculatus*, both by MP and NJ tree analysis. However, in ML and BI, *A. esculentus* was closely related to *A. moschatus* ssp. *moschatus* than to *A. tuberculatus*. From all the tree building methods, *A. ficulneus* was found to be closely related to the unidentified taxa of *Abelmoschus* spp., having 2n=130. *A. enbeepeegearensis* was found to be closely related to *A. ficulneus* and *Abelmoschus* spp. in ML, MP and BI. However, the close clustering was also recorded of *A. manihot* to *A. enbeepeegearensis* in both BI and NJ.

*atpB* is a chloroplast gene coding for β subunit of ATP synthase. They are located in the large single-copy region of the plastid genome, 900 bp downstream from *rbcL* in flowering plants, but transcribed in the opposite direction to *rbcL* (Downie and Palmer, 1992) and approximately 1497 bp long. They are highly conserved structure and can be used for phylogenetic studies (Hoot *et al*., 1997, 1999; Hoot and Douglas, 1998; Bayer *et al*., 1999; Chase *et al*., 1999).

Phylogenetic analyses were performed by the ML, MP, BI and NJ. Form the clustering pattern, it is easily it could be noted the grouping of *A. esculentus* with *A. ficulneus* and *Abelmoschus* spp. in both ML and MP. This could suggest that the
other probable genome of allopolyploid, *A. esculentus* is *A. ficulneus* and the unidentified taxa of *Abelmoschus* spp., with somatic chromosome number of 2n=130 could therefore suggest that this might be another accessions belonging to *A. esculentus*. It may be noted that *A. tuberculatus* is found at the base of the tree in both MP and NJ. However, the rest of the species clustered separately in all the four trees topology.

The *psbK-psbI* intergenic spacer showed high levels of discriminatory power, but lower sequence quality and universality (Lahaye et al., 2008). The genes *psbK* and *psbI* encode two low molecular mass polypeptides, K and I, respectively, of the photo-system II component. They are localized in the large single copy (LSC) of the plastid genome. They are approximately 641 bp long. The molecular phylogenetic analysis of intergenic spacer, *psbK-psbI* revealed the unique clustering of *A. caillei*, *A. ficulneus*, *A. esculentus*, *Abelmoschus* spp., *A. manihot* and *A. tetraphyllus* var. *tetraphyllus* in all the four tree building methods. The clustering of *A. angulosus* var. *grandiflorus*, *A. tetraphyllus* var. *pungens* and *A. tetraphyllus* var. *pungens* (3M) was observed in both ML and MP. Similarly, *A. moschatus* ssp. *tuberosus*, *A. moschatus* ssp. *moschatus*, *A. enbeepeegearensis* and *A. tuberculatus* also clustered by both the above methods. However, BI and NJ had shown the different clustering patterns for the above mentioned species. The ratios between transitions and transversions obtained from MEGA version 5 were recorded as 0.551. Based substitution patterns toward transitions, with deviation from random mutation (with an expected transitions/transversions ratio of 0.5), have been found in several fast evolving genes, such as primate mtDNA control region (Kocher and Wilson, 1991; Vigilant et al., 1991; Tamura and Nei, 1993) and nuclear satellite DNA (Wu et al., 1999) both of
which are generally subject to very weak selective constraints (Li, 1997). Low transition/transversion ratio suggests the conserved nature of this chloroplast spacer and thus the low evolutionary rates.

Hence it may be mentioned that the different clustering patterns were recorded in all the four tree building methods for all the three cpDNAs sequences. However, the clustering pattern of accD and atpB were somewhat similar, though with some exceptions in the positions of the species may differ in the tree. It was recorded that atpB had the highest percentage of conserved sites of 97.58% and the highest percentage of variable sites of 3.88 were recorded in intergenic spacer, psbK-psbI. The lowest percentage of sequence divergence substitution was recorded as 2.19 in atpB, confirming that nucleotide substitutions occur at a relatively slow rate in cpDNA (Clegg and Curtis, 1984). These sequence comparisons have further revealed that, although the genome as a whole changes slowly, certain genes change either more rapidly or more slowly than the average. However, the reduced rate of evolution for any particular gene appears to be fairly constant in different lineages (Palmer, 1985a). The conservative evolution of the chloroplast genome provides the distinct advantages in phylogenetic studies compared to the much more dynamic and diverse mitochondria (Palmer, 1985c) and nuclear (Flavell, 1980) genomes of plants. The region is highly conserved in Abelmoschus and relatively few sites in the aligned data matrix are parsimony informative, a variety of relationships among the species are revealed by the analyses, some of which are congruent with the proposed species relationship from cytogenetical studies. As expected, the number of variable and parsimony informative sites in a given data set is associated with the overall sequence length of the data set. In an analysis of cpDNA sequence variations in seed
plants showed that sequence length accounted for anywhere from 22 to 83% of the variation in the number of variable characters observed in a data set. This analysis indicates that 81% of the variation in the number of variable sites and 79% of the variation in the number of parsimony informative sites is explained by sequence length (Shaw et al., 2005).

We could observe that in all the three cpDNAs sequences, the nucleotides A and T were rich, and it doubles its amount of A and T as compared to G and C in the spacer \textit{psbK-psbI}, which is consistent with the nucleotide composition of most non-coding spacers and pseudogenes due to low functional constraints (Li, 1997) and thus evolve more rapidly (Wolfe et al., 1987; Zurawski and Clegg, 1987; Clegg and Zurawski, 1992; Clegg et al., 1994; Gielly and Taberlet, 1994; Kelchner, 2000). Comparisons of homologous sequence data for non-coding regions of the chloroplast genome also show that microstructural changes, primarily those involving insertion/deletion events, or “indels”, are especially frequent in these regions (Graham et al., 2000; Simmons et al., 2001; Geiger, 2002; Vogt, 2002; Hamilton et al., 2003; Müller and Borsch, 2005) and are more frequently associated with short repeats of 10 bp or shorter (Takaiwa and Sugiura, 1982; Zurawski et al., 1984; Graham et al., 2000). Chloroplast genomes are essentially free of any of those evolutionary processes, such as gene duplication and deletion, concerted evolution and pseudogene formation, which are common among nuclear genes and which can dangerously distort the evolutionary history of DNA sequences relative to that of organisms (Palmer, 1985a, b).

Kelchner and Clark (1997) had stated that the alignment programs developed for coding regions may be generally inadequate for non-coding data. This is
problematic since sequence alignment is the first and most important assessment of nucleotide homology (de Pinna, 1991). In addition, it has been argued that computer-generated alignments are more objective than those produced manually because they result from the application of explicit rules (Giribet and Wheeler, 1999). However, any alignment can only be as good as the rules embodied in the alignment algorithm, and these are likely to be only rough approximations of the actual processes of molecular evolution responsible for generating DNA sequence diversity. Most workers end up adjusting computer-generated alignments “by eye”, to a greater or lesser degree, prior to formal phylogenetic analysis and no single alignment program was ever satisfactory. Indels are generated by a variety of molecular processes, and there is a growing recognition that this should be considered during alignment (Thorne et al., 1992; Gu and Li, 1995; Benson, 1997; Kelchner, 2000).

Since the entire chloroplast genome is considered as one linkage group, individual regions are expected to exhibit the same phylogenetic pattern (Doyle, 1992). Therefore, in the present study, the coding gene accD and atpB, and intergenic spacer region psbK-psbl has been combined into a single matrix for a collective approach analysis to obtain greater phylogenetic resolution. The concatenate topologies of the trees constructed by ML, MP, BI and NJ for the consensus sequence were consistent with the clustering of A. caillei and A. tetraphyllus var. tetraphyllus in all the methods. From the inter-specific F1 hybrids also, we propose A. tetraphyllus var. tetraphyllus to be one of the probable progenitor species of A. caillei, and with this clustering together it also supports our previous hypothesis. The bootstrap probabilities calculated based on 1000 replications were high, even in the analysis of separate sequences.
A feature which has provided unique insights into the directionality of the hybridization events of the specific parentage and timing of origin of inter-specific hybrid species and polyploid complexes is inheritance of cpDNA is clonal i.e. the predominantly maternal mode of inheritance of cpDNA that are so common among flowering plants (Sears, 1980; Kung et al., 1982; Palmer et al., 1983; Erickson et al., 1983; Bowman et al., 1983; Terachi et al., 1984; Salts et al., 1984; Hosaka et al., 1984; Corriveau and Coleman, 1988; Harris and Ingram, 1991; Mogensen, 1996). It may be noted that the origins of amphidiploid species could be analyzed with further support from the cpDNAs. The cpDNAs of *A. esculentus* and *A. caillei* were essentially identical with those of diploid, *A. tuberculatus* and allopolyploid, *A. tetraphyllus* var. *tetraphyllus* respectively. Since cpDNA is maternally inherited, one can conclude that these latter two species served as the maternal parents in the inter-specific hybridizations that gave rise to the amphidiploids, *A. esculentus* and allohexaploid, *A. caillei*. The origin of amphidiploid, *A. esculentus* is further supported from the inter-specific hybridization studies where complete pairing of 29 chromosomes *A. tuberculatus* was observed with that of *A. esculentus* (see chapter 8). Further the clustering of *A. esculentus* with *A. ficulneus* along with *A. tuberculatus* and *Abelmoschus* spp. (unidentified taxa with 2n=130) in BI also gives further support that *A. ficulneus* to be the other probable parents of *A. esculentus*. *A. esculentus* and *A. caillei* were judged to result from recent hybridization events since their cpDNAs were identical at all the 2,289 base pairs. However, these recent hybridizations must have taken place after a substantial period of separation and diversification of the parents. Such observations were recorded in amphidiploid species of *Brassica* viz., *B. carinata*, *B. juncea* and *B. napus* (Erickson et al., 1983;
Palmer et al., 1983) and Paeonia viz., P. xinjiangensis, P. japonica, P. obovata, P. wittmanniana, P. arietina, P. humilis, P. officinalis, and P. parnassica (Sang et al., 1997). Allopolyploidy has been considered to be a primary mode for the formation of fertile and stable hybrid species (Grant, 1981). Frequency of natural hybrid speciation at the diploid level, however, has been controversial (Rieseberg et al., 1990; Rieseberg, 1991; Wolfe and Elisens, 1994; Rieseberg et al., 1995; Sang et al., 1997).

Chromosome numbers have been treated as an important evolutionary feature in plant systematic studies. Cytologically Abelmoschus is heterogeneous species with variable chromosomes numbers and indistinct small size chromosomes. The clustering of Abelmoschus spp., 2n=130 with A. ficulneus, 2n=72 in BI may also direct that this unidentified taxa might be the cultivated species A. esculentus and also their close affinity towards A. enbeepeegearensis, 2n=72. It may be also noted the close relationships of A. esculentus with A. manihot in all the tree building methods except in BI. This may also correlate that both have the very close somatic chromosome numbers of 2n=130 and 2n=132 respectively. However, chromosome number as a whole did not really relate to the clustering pattern of the species having the same chromosome numbers.

In order to carefully evaluate the genomic diversity and to get a better resolution of the phylogenetic relationships of the taxa studied, RAPD analysis can be also be combined with the comparative study of nucleotide sequences (Blattner et al., 2001; Gehrig et al., 2001). The coding sequences of the plastid genome (accD and atpB) have been shown to have a lower rate of nucleotide substitutions than non-coding sequences of the plastid (psbK-psbl) and were more successfully used in
phylogenetic studies (Freshwater et al., 1994; Fritsch, 2001; Lewis and Doyle, 2001; Lia et al., 2001). The distribution of RAPDs and cpDNAs revealed the close relationship of A. tetraphyllus var. pungens and A. angulosus var. grandiflorus, as also revealed by ML, MP and NJ. The closeness of A. esculentus to A. moschatus ssp. moschatus was also revealed by both RAPD and NJ. Such studies of close relationships between species revealed by both RAPDs and cpDNAs were also reported in Monanthes (Mes et al., 1997). It may be observed that the close clustering of A. tetraphyllus var. pungens and A. angulosus var. grandiflorus with A. tetraphyllus var. pungens (3M) recorded both in ML and MP. These observations might be related as both the species A. tetraphyllus var. pungens and A. angulosus var. grandiflorus occurs at high altitude from 400 up to 3000 masl and below 900 m respectively.

The robustness of a hypothesis can be tested by assessing its congruence with phylogenetic hypothesis developed from different methods of cytogenetical techniques like mitotic and male meiotic studies as well the crossability relationships within the species and the meiotic studies of F₁ hybrids. In the genus Abelmoschus, such cytogenetical investigations supported by classical taxonomic data contributed to the establishment of inter-genomic relationships among the species. The phylogeny established from molecular data strongly supported by the higher bootstrap suggested the close relationship among the species within these clades. However, cpDNA phylogeny was found to be congruent with the inference based on BI method.

Phylogenies of the genus Abelmoschus based on certain cpDNAs sequences region of the chloroplast genome are presented. Although the region is highly
conserved in plants and relatively few sites in the aligned data matrix are parsimony informative, a variety of relationships among members of the family are revealed by the analyses, some of which are congruent with the current classification and others which are not. However, consensus trees contain high levels of ambiguity, partly due to the inadequate numbers of informative characters in the dataset. Such studies were also reported in palm (Baker et al., 1999).

Various alternative methods have been used to analyze phylogenetic relationships of plants which includes similarity (Little and Stevenson 2007), phylogenetic (UPGMA, NJ, MP, ML and BI), multiple-character based (DeSalle et al., 2005), genealogical (Nielsen and Matz, 2006), pure statistical (Lou and Golding, 2010) approaches, but it is still controversial as to which method is the best. Little and Stevenson (2007) pointed out that none of the methods performed particularly well, and preferred to simply select one of the similarity methods. Frezal and Leblois (2008) concluded that the ML approach always seems to be more accurate than the distance-based (NJ) phylogenetic inference, but needs more computational time. According to our present study, the ML analysis does not show a higher rate of species discrimination than other analytical methods except MP and the NJ method has a similar level of resolution as ML.

From the information generated by molecular phylogenetic studies based on sequence data of three important chloroplast DNA sequences (\textit{accD}, \textit{atpB} and \textit{psbK-psbI}), which are considered more authentic as well as reliable, we have tried to deduce inter-species relationships among various \textit{Abelmoschus} taxa. They are in conformity with the scheme of phylogeny, earlier proposed on the basis of meiotic and crossability data, \textit{though} to some extent. In the present studies while species
belonging to group A did not resolve into one group as per the earlier proposed scheme, however *A. enbeepeegearensis* and *A. ficulneus* of group B were clustered together as expected. The remaining two basal species, *A. moschatus* ssp. *moschatus* and *A. moschatus* ssp. *tuberosus* of group B on the other hand did group separately. Further, the hybrid taxa namely *A. tetraphyllus* var. *tetraphyllus* and *A. caillei* were resolved into same clade, while *A. esculentus* resolved differently. Therefore it is concluded that more number of loci from nuclear as well as organellar genomes, need to be investigated before making taxonomical conclusions of various *Abelmoschus* species.