5.1 Introduction

Phylogenetic and taxonomic difficulties are common within the woody bamboos, due to their unique life cycle, which severely limits the availability of floral characters. Information on genetic diversity or population structure of the bamboo germplasm is limited, as are published reports of the ploidy level and the use of molecular markers in the various bamboo accessions. Therefore an approach of DNA barcoding in Bamboos will assists the classical taxonomist in identification and classification of the same.

DNA barcoding can serve a dual purpose as a new tool in the taxonomists toolbox supplementing knowledge as well as being an innovative device for non-experts who need to make a quick identification. DNA barcoding is a technique for characterizing species of organisms using a short universal gene sequence taken from a standard and agreed-upon position in the genome. DNA barcode serve as a unique species identifier and exploits the property of being ideal sequence that is constrained by overall conservation (preserve gene function), but still has substantial sequence variation which differentiates species.

5.2 DNA isolation and PCR amplification

The technique of DNA barcode involve the steps from DNA isolation, followed by PCR amplification using the barcode primers, sequencing of the amplicons for
generation of sequence data, data analysis using available software, interpretation of the result.

Starting from the DNA isolation, CTAB method (Sambrook and Russell, 2001) was followed for extraction of DNA. Generally, all methods of DNA extraction involves the disruption and lysis of cells followed by the removal of RNA by RNase or salts and isolation from proteins by several methods including digestion of proteins by the enzyme proteinase K. Isolating high quality DNA is essential for molecular research. In the present study the common DNA isolation procedures viz. the CTAB, SDS and extraction using phenol have been used on bamboo species to check for optimum genomic DNA output. The quantity and quality of DNA extracted by these methods have been sufficient enough to perform PCR-based reactions with optimum output using CTAB DNA isolation protocol.

PCR amplification is the next immediate step which needs further standardization. The concentration of the reagents such as the template DNA, primer concentration, working units of Taq polymerase requires standardization. As per the protocol given in the source articles from where the primer sequence with their respective PCR condition was undertaken, the concentration of the template DNA was proposed to be 25-500 ng, for the primer concentration the amount suggested was 1µM (Bafeel et al., 2012). Standardization of the reagent concentration was followed by finding the right combination of the forward and the reverse primer. rbcL and trnH2-psbA primer sets did not require any such trials and were used as published in the source article.
In the case of *matK* primer forward primer sequence included *matK*390, *matK*X, *matK*2.1, *matK* 2.1a and the reverse included *matK*1326 and *matK*5. The best amplification was observed in *matK*X and *matK*5. Primer *rpoC*1 had 2 forward primers as *rpoC*11, *rpoC*12 and reverse as *rpoC*13, *rpoC*14. The four combinations *rpoC*11 and 3, *rpoC*11 and 4, *rpoC*12 and 3, *rpoC*12 and 4 were tested for best amplification. The primer pair *rpoC*12-3 was found to give best amplification. Negative amplification of combination of two primer sequences as reverse and forward may be attributed to the fact that they lack complementary binding sequences in the target template DNA or sometimes the forward and the reverse primer lie outside the amplification range due to presence of INDELS.

Apart from the combination of forward and reverse primer set, the annealing temperature plays an important role in generation of amplicons. Therefore, it is required to be standardized by performing gradient PCR. The primer set *rbcL*, *matK* and *rpoC*1 gave unambiguous result at the reported annealing temperature, except *trnH2-psbA* primer set which required gradient PCR for determination of appropriate amplification. In addition to the annealing temperature it (*trnH2-psbA*) a required changes in extension temperature from 72°C to 68°C (which was 72°C for other PCR condition). Su et al. (1996) reported that reduction in the extension temperature from 72 to 60°C was required for amplification of A/T-rich regions of DNA obtained from *Plasmodium falciparum* (Su et al., 1996).
5.3 DNA Barcodes

Combination of primers has been used for species discrimination in several molecular approaches. Thus, the present study was also planned to analyze discriminatory power, of the combination of primers, taken from the coding as well as non-coding region, in species studied.

The combination of the primers from coding (rbcL, matK and rpoC1) and non-coding region (trnH2-psbA) exhibited better discrimination among the species studied. Kress and Erickson (2007) performed a study in which they examined the performance of 9 loci (ITS1, trnH-psbA, rbcL-a, matK, rpoC1 ycf5, rpoB2, ndhJ and accD) in pairs of species from 48 phylogenetically diverse plant genera. Hollingsworth et al. (2009) also evaluated the seven main candidate plastid regions (rpoC1, rpoB, rbcL, matK, trnH-psbA, atpF-atpH, psbK-psbI) in three divergent groups of land plants. Both the results suggested that the combination of primers obtained from coding and non-coding region results into better discrimination of species. The results obtained from Kress and Erickson (2007) and Hollingsworth et al. (2009) favoured the suitability of non-coding trnH-psbA intergenic spacer as the most viable candidate for a single-locus barcode for land plants (Kress and Erickson, 2007).

5.4 Sequencing and software analysis

In the present study, the amplicons generated by the primer pair rbcL, matK and rpoC1 2-3 resulted into 100% PCR success and sequencing to generate 102 sequences. However, the primer pair trnH2-psbA resulted into amplification of 30 out of 34 species studied thereby, accounting for 88.23% PCR success rate. A total of 132 sequences were
generated in the study. The relative PCR success obtained by Kress and Erickson, (2007) for the primer sets trnH-psbA and rbcLa, were 95.8% (46 of 48 genera) and 92.7% (43 of 48 genera). A low PCR success was observed using matK primer set (Kress and Erickson, 2007).

Once the amplicons were obtained they were subjected to sequencing and the sequence data was analyzed using software MEGA 5.2.2. The sequences were aligned using MUSCLE as it is claimed to achieve both better average accuracy and better speed than ClustalW2 or T-Coffee, depending on the chosen options. total sequence length of 666 bp for rbcL, 834 bp for matK, 529bp for trnH2-psbA and 409 for rpoC1 were generated, in which no of variable sites found were 114, 621, 111 and 41 respectively. Similarly, Selvaraj et al. (2008) reported that maximum number of indels were found in the gene matK and on alignment of matK gene of combined nucleotide sequence resulted into 497 variable sites and 251 parsimony sites, the overall mean distance is 0.027.

Alignments of the sequences were followed by computing the genetic diversity existing within the population. The values for genetic diversity were calculated as per the parameters selected through the ‘Best substitution model’. The overall mean distance was highest in matK (0.049), followed by trnH2-psbA (0.016)>rpoC1 (0.012) and the least overall mean distance was found in rbcL (0.006). Similar result is being reported by (Kress and Erickson, 2007) in a study to test the performance of 9 loci in 48 species. They found highest degree of divergence in ITS1 (5.7%), followed by trnH-psbA (2.69%), then rpoB2 (2.05%), rpoC1 (1.38%) and rbcL (1.29%).
The maximum genetic divergence was obtained with $matK$ primer set, therefore, it could be considered as potent candidate for the barcoding purpose. But in general, one limitation associated with this primer set is that it is not easy to amplify in some group and that additional work is required for primer development (Kress and Erickson, 2007).

Keeping the above mentioned difficulty of universal amplification of $matK$, better alternative for barcoding is $trnH2-psbA$ primer set. This primer set showed an average genetic diversity of 0.016 and exhibited high sequence divergence among the species. From the present study despite exhibiting sufficient genetic diversity this locus lack 100% PCR success rate therefore, acquire certain limitation for suitability of its usage as barcode.

As far as the 100% PCR success rate is considered, $rbcL$ primer set worked up to the mark by amplifying in all of the 34 species of bamboo used. High PCR success rate of $rbcL$ is also reported by Kress and Erickson (2007).

Alignment of the sequences was followed by the generation of dendrogram. The dendrogram generated from the three primer pair $rbcL$, $matK$ and $rpoC1$ 2-3 showed greater number of branching therefore exhibiting their discriminating power. The morphologically similar species such as $B. tulda$ and $B. nutans$, occupied their position in separate division. Similarly, $B. polymorpha$ and $P. aurea$ is placed under one division in three primer pairs ($rbcL$, $matK$ and $rpoC1$ 2-3), which accounts for their closeness at the genetic level and warrants confirmation by further study.

The four considered barcode primer sets exhibited enough potential to distinguish the closely related species. The primers $rbcL$ coding for larger subunit of ribulose 1-5 bi-
phosphate carboxylase/oxygenase (RUBISCO) enzyme, *matK* coding for maturase and *rpoC1* coding for one of the subunit of chloroplast RNA polymerase, are taken for the coding region and therefore, largely comprises of conserved region resulting into low degree of variation. The primer set *rbcL* showed 100% PCR success rate and was suitable to distinguish closely related species as *B. tulda* and *B. nutans*. The other primer sets used in the study were also capable enough to differentiate between the 34 species studied.

On the contrary, primer set *trnH2-psbA* belonging to the intergenic spacer region, is the most variable genome segment, shows greater degree of variation and can prove itself to be a better option for the barcode purpose. The distinguishing power of primer set *trnH2-psbA* was found to be maximum as the greatest number of branches was found from the dendrogram generated but one of the problem associated with this locus was that it doesn’t exhibited 100% PCR success rate.

5.5 Relative efficacy of primers

5.5.1 Barcoding

CBOL, the Consortium of Barcode of Life has proposed two loci, *matK* and *rbcL* as the BARCODE regions for Land Plants. The amplification success rate of both the primer sets are 100% therefore it is proposed to create backbone of tree with *rbcL* as the ‘Anchor’; then separate individuals species in smaller groups with *trnH-psbA* as the ‘Identifier’. Result suggests that the *trnH-psbA* intergenic spacer is the best plastid option for a DNA barcode sequence that has good priming sites, length, and interspecific variation. If the problem of amplification is managed for the primer *trnH2* and *psbA*, it will prove itself to be a better choice for the generation of universal barcode.
5.5.2 Phylogenetic analysis

The chloroplast genome is also well suited for evolutionary and phylogenetic studies particularly above the species level, because cpDNA (Liang, 1997):

1) is a relatively abundant component of plant total DNA, thus facilitating extraction and analysis;

2) contains primarily single copy genes;

3) has a conservative rate of nucleotide substitution; and

4) extensive background for molecular information on the chloroplast genome is available.

Therefore, most phylogenetic reconstructions in plant systematics conducted so far is based on molecular data from the cpDNA genes. An attempt was made to study the phylogenetic analysis of 34 species of bamboo using rbcL, matK, trnH2-psbA and rpoC1. The dendrogram generated from the three primer set exhibited that the species B. polymorpha and P. aurea were genetically similar and showed a common ancestor with greater bootstrap value.

The most common gene which has been used to provide sequence data for plant phylogenetic analyses is the plastid-encoded rbcL gene (Chase et al., 1993; Donoghue et al., 1993) in which the success rate of PCR and sequencing is higher compared with the other selected plant characterization genes such as matK (maturase K) (CBOL Plant Working Group, 2009; Parveen et al., 2012). The phylogenetic tree generated from rbcL resulted formation of 6 subdivisions, in which clumping of B. multiplex, B. striata, B.
vulgars, B. wamin under on subdivision with other members P. japonica, S.palmata, D. andamanica and O. scriptoria were observed. G. angustifolia, B. polymorpha and P. aurea were closely related as they lie separately in one subdivision. A mixed distribution of Dendrocalamus genera is observed. But in the first subdivision Dendrocalamus genera lie with C. capitatum and P. polymorphum, all belonging to Subtribe III-Dendrocalameae. This kind of result may be attributed to highly conserved nature of rbcL locus, therefore lacking in resolving the phylogenetic relationships among the closely related species (Doebley et al., 1990). An alternative to this locus will be required for phylogenetic studies.

Elsaied and Naganuma (2001) used rbcL gene for study of phylogenetic diversity existing in Deep sea microorganisms. Bafeel et al. (2012) utilized rbcL gene sequences representing from 90 taxa of 12 genera and 10 angiospermic plant families (dicot and monocots) and unambiguously resolved the relationships, as well as provided a good indication of major supra-generic groupings among the selected angiospermic plant families.

The sequence data of the rbcL gene despite being widely used in the reconstruction of phylogenies throughout the seed plants it is apparent that the ability of rbcL to resolve phylogenetic relationships below the family level is often poor (Doebley et al., 1990). Thus, interest exists in finding other useful DNA regions that evolve faster than does rbcL to facilitate lower-level phylogenetic reconstruction. The matK gene is a promising gene in this regard.
matK, primer set were used to generated dendrogram to show the phylogenetic relationship among the studied population. On alignment of the sequence data obtained from matK primer set greater columns of misalignment were observed due to presence of INDELS owing to gaps. Greater number of variation was observed at this locus. Similarly, Selvaraj et al. (2008) generated a Phylogenetic tree to identify the ideal regions that could be used for defining the inter and intra-generic relationships. They concluded that the matK gene is a good candidate for DNA barcoding of plant family Zingiberaceae. Komatsu, et al., (2001) worked on phylogenetic analysis based on 18S rRNA gene and matK gene sequences of Panax vietnamensis and five related species. The phylogenetic tree reconstructed by the combined 18S rRNA-matK gene analysis using the maximum parsimony method showed that P. vietnamensis was sympatric with other Panax species and had a close relationship with P. japonicus var. major and P. pseudo-ginseng subsp. himalaicus.

trnH2-psbA was the next locus studied in the present study. The dendrogram generated from the primer set was much more reliable as the bootstrap values were high in comparison to the other primer sets used in the study. Members of the same genera were clustered together, thus showing their common background. Andrade et al. (2010) utilized evidences from nuclear (ITS) and plastid (psbA-trnH and trnL-F) DNA sequences for a comprehensive phylogenetic analysis of Eriocaulaceae. Yan et al. (2014) utilized intergenic spacer psbA-trnH and single-copy nuclear gene Acc1 for phylogenetic analysis of the genus Avena based on chloroplast. Two uncorrelated nucleotide sequences, chloroplast intergenic spacer psbA-trnH and acetyl CoA carboxylase gene (Acc1), were used to perform phylogenetic analyses in 75 accessions of the genus Avena, representing
13 diploids, seven tetraploid, and four hexaploids by maximum parsimony and Bayesian inference. Studies suggested greater utility of \( trnH2-psbA \) in studying phylogenetic relationship among the population.

\( rpoC1 \) was the fourth primer set used in the study. Mixed distribution of the species of bamboo were exhibited. Seo et al. (2003) worked on Phylogenetic analysis of cyanobacteria out using the small subunit rRNA (16S rRNA), DNA gyrase subunit B (\( gyrB \)), DNA-dependent RNA polymerase gamma subunit (\( rpoC1 \)) and a principal sigma factor of E. coli sigma (70) type for DNA-dependent RNA polymerase (\( rpoD1 \)) gene sequences of 24 strains. Long et al. (2004) worked on phylogenetic analysis of the DNA sequence of the non-coding region of nuclear ribosomal DNA and chloroplast of \( Ephedra \) plants in China. A specific primer set was designed to classify the 3 groups by routine PCR. Combined analysis of \( ITS \) and \( trnL/ trnF \) differentiated the 8 \( Ephedra \) species.

The primer sets \( rbcL, matK \) and \( rpoC1 \) were obtained from the coding region of chloroplast. The coding regions codes for certain protein therefore flexibility in the nucleotide sequence experience constraints resulting into less variation. On the other hand, the primer \( trnH2-psbA \) is taken from non-coding regions which can accommodate larger variation in terms of nucleotide sequences.

**5.6 DNA document generated from the study**

The technique of DNA barcoding involves generation of sequence data for the different species under study, at a particular target region of interest, followed by comparison of those sequence data to the available database at NCBI. Therefore, the sequence database is greatly required for comparison and analysis of the generated sequence data. A total of
132 sequence data from 34 species of bamboo using four primer sets were generated which can further act as basis for similar works. The DNA document available in the public domain will lay the foundation to generate a universal barcode for identification and classification of different bamboo species.

In agreement with Kress and Erickson (2007) (working over land plants) a two-locus barcode that combines a subunit of the coding locus \textit{rbcL} (\textit{rbcL}-a) with the non-coding \textit{trnH-psbA} spacer is recommended for barcoding in eastern Indian bamboos. \textit{rbcL}-a provides a strong recognition anchor that will place an unidentified specimen into a family, genus, and sometimes species; the highly variable \textit{trnH-psbA} spacer will further narrow the correct species identification where \textit{rbcL}-a lacks discriminating power, especially in species-rich genera of angiosperms. In the expanded sampling of loci and taxa the \textit{trnH-psbA} spacer continued to successfully address the trade-off between universal application and high sequence divergence. PCR priming sites within highly conserved flanking coding sequences combined with a non-coding region that exhibits high sequence divergence among species as well as diagnostic insertion/deletion mutations makes the \textit{trnH-psbA} spacer highly suitable as a plant barcode.

5.7 Conclusion

In conclusion, the findings of the present study results presented in this study demonstrate the utility of DNA barcoding as a good method for identifying different species of bamboo. With the use of four primer set in 34 species of bamboo the following conclusion were drawn:

1. 100% PCR success and sequencing data rate were obtained for the primer sets \textit{rbcL}, \textit{matK} and \textit{rpoC1}. \textit{trnH2-psbA} primer set gave 88.23% PCR success rate resulting into generation of 30 sequence data out of 34 species considered for the study. Negative
amplification was observed in *Bambusa multiplex, Dendrocalamus sikkimensis, Dendrocalamus strictus* and *Schizostachyum pergracile*.

2. All the four barcoding primer sets could distinguish 30 of 34 species of bamboo studied. The barcoding sequences could distinguish even morphologically similar species such as *B. nutans* and *B. tulda*.

3. Employing the barcoding primers, phylogenetic analysis was possible within the studied population. The primer sets derived from the non-coding regions showed greater variation in comparison to the primer designed from coding region.

4. Estimation of mean diversity in entire population was computed which resulted into overall mean distance. Maximum variation was observed at the locus *matK* followed by *trnH2* and *psbA*. Therefore the first choice of primer being utilized for barcoding purpose is *matK* followed by *trnH2-psbA*.

5. The coding region represents highly conserved region on DNA, resulting into exhibition lesser degree of variation in order to maintain the quality and quantity of proteins coded. On the other hand, the non-coding regions offer variable sequences without affecting the quality and the quantity of protein. To conclude a two-locus barcode that combines a subunit of the coding locus such as *rbcL (rbcL-a)* with the non-coding *trnH-psbA* spacer is recommended. *rbcL-a* provides a strong recognition anchor that will place an unidentified specimen into a family, genus, and sometimes species; the highly variable *trnH-psbA* spacer will further narrow the correct species identification where *rbcL-a* lacks discriminating power, especially in species-rich genera of angiosperms.
6. DNA document in the form of 132 sequence data has been generated which can act as a foundation for the upcoming studies not only related to DNA barcoding but also other Molecular Biology research.

The results presented, adequately address to the concerns about the applicability of DNA barcoding to plants. However, quest for a perfect universal barcode for plants providing 100% species resolution across the plant kingdom appears to be unrealistic, as DNA barcoding, like any other technology, is not expected to be 100% perfect. However, within a taxonomic group 100% species resolution could possibly be obtained by taxa specific barcodes. Thus, the projection that DNA barcodes, once available for all the described species, would be able to provide a correct identity up to species level to any unknown sample, whether available in vegetative, fragmented or DNA form, or would indicate the discovery of a new species does not hold true. Nevertheless, more than 90% success in species identification with single locus or two-/three-locus combinations emphatically demonstrates the efficacy of the technique.

The instances of failure of DNA barcodes to correctly assign the species should encourage taxonomist to re-consider or reinvestigate such taxa. Future developments in the techniques and technologies underpinning DNA barcoding could improve the success rate and practicalities of identification of species. However, the most important development will be the continued expansion of the Barcode of Life Database.
5.8 Future Prospects

In the longer term, it is desirable to increase the levels of species discrimination beyond those achievable by combining even all four of the markers tested here. Options for simultaneously sequencing both strands of the amplicons are desired to obtain the full length sequence of target locus. Further authentication by comparison of sequences of multiple individuals of the same species is also required. However, these criteria have yet to be satisfied, and ultimately these approaches still do not address the crux challenge, which is obtaining a universal barcode primer set for discrimination of all the species under study. The technical and analytical framework to deliver on this problem remains a pressing challenge.

There are several imperfections in the plant barcodes suggested for species level resolution. Thus, perfect identification and accurate classification of bamboo species will require sufficient DNA database. At present, there is no universal database populated with DNA sequence data to provide a robust genus-level identification system across bamboos supported by links to high-quality digitized reference specimens of the samples that were sequenced.

The development of next generation sequencing will expedite building of a high quality reference library of DNA sequences of bamboos and their subsequent re-sequencing for additional loci in light of technical improvements will be relatively straightforward.
SUMMARY

The merit of the problem “Studies on DNA barcoding in eastern Indian bamboos” has been considered briefly. In view of lacunae in the understanding of problem, rationale for undertaking the present work has been marked out in Chapter 1.

The review of existing relevant literature on the problem has been incorporated in Chapter 2. The review accentuates the problem of availability of a universal barcode for identification of plant species.

Chapter 3 contains the description of materials used for the study and methodology followed in an attempt to generate the DNA database for phylogenetic study and barcoding purpose.

Sequence database and its analysis using MEGA 5.2.2 software with detailed results have been presented in Chapter 4 and are summarized below:

The quality and quantity of genomic DNA of all 34 species of bamboo was assessed by 0.7% agarose gel electrophoresis. Though some degree of shearing was observed, the DNA was largely high molecular weight and suitable for the present study. The yield of DNA was also found to be fairly uniform ranging from 85.2 in Cephalostachyum capitatum to 1570.8 in Dendrocalamus hamiltonii. The DNA yield in 34 target species was found suitable for further molecular work.

Four primer pairs were tested for their amplification in 34 species of bamboo used. In case of rbcL 1F and 724R, reported PCR condition gave sharp, intense, unambiguous band of 738 bp which were monomorphic in nature and present in all studied species of bamboo. Thus, no standardization of PCR condition was required for this primer pair.
Different forward and reverse primers for the *matK* region were available, such as *matK*390, *matK* X, *matK* 2.1, *matK* 2.1a as forward primer and *matK* 1326, *matK* 5 as reverse primer, which were tested for unambiguous and sharp amplification. The primer combination *matK* X as forward and *matK* 5 as reverse yielded sharp, single, unambiguous and monomorphic band of 930 bp in all 34 species of bamboo.

In case of *trnH2* and *psbA* gradient PCR was performed to standardize the annealing temperature and best amplification was obtained at 55°C. The extension temperature was also required to be decreased from 72°C to 68°C for amplification. *trnH2-psbA* primer pair resulted into 88.23% PCR success rate which corresponds to positive amplification in only 30 species. Amplicons considered for further sequencing was of 635 bp molecular weight.

For *rpoC1* 2 and 3 two forward and two reverse primers were tested for amplification in four combinations. Primer pair yielding single, sharp and intense band was considered which resulted into amplicon of 450bp in all 34 species of bamboo.

The amplicons thus obtained were subjected to sequencing. 100% sequencing were attained for the primer set *rbcL1F*-724R, *matK* X-5 and *rpoC1* 2-3. The primer set *trnH2-psbA* showed no amplification in *Bambusa multiplex, Dendrocalamus sikkimensis, Dendrocalamus strictus* and *Schizostachyum pergracile* resulting into sequencing of rest of the amplicons of 30 species.

Once the sequences have been obtained they were subjected to MEGA 5.2.2 software. Molecular Evolutionary Genetics Analysis version 5 (MEGA5), enables mining online databases, building sequence alignments and phylogenetic trees, and using methods of evolutionary bioinformatics in basic biology, biomedicine and evolution.
An input file in the form of FASTA format was prepared and subjected for multiple sequence alignment using MUSCLE (MUltilple Sequence Comparison by Log-Expectation). All the 34 sequences obtained from each primer were aligned. Once alignment was successfully completed, the primer pairs rbcL yielded sequences ranging from 584 bp for B. nana to 653bp for T. oliveri; matK ranging from 624-772 bp (G. angustifolia to B. striata); trnH2-psbA ranging from 505-524 bp (M. baccifera to D. maclellandii) and rpoC1 2-3 ranging from 398-403 bp (B. baloooa, B multiplex, G. atroviolacea to B. striata).

Alignment of the sequences was further proceeded by analyzing for the ‘Best Substitution Model’. It determines the best criteria appropriate for generation of phylogenetic tree. According to the software, models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also determined.

Twenty four different nucleotide substitution models were tested to obtain the lowest BIC and AIC value. The primer pair rbcL had lowest BIC value 2649.449 and lowest AIC value 2130.355, matK had lowest BIC value 4176.833 and lowest AIC value 3677.898, trnH2-psbA had lowest BIC value 1021.958 and lowest AIC value 673.067 and rpoC1 had lowest BIC value 2442.442 and lowest AIC value 1940.272 respectively. On the basis of these lowest BIC and AIC value best model was selected which was utilized to generate phylogenetic tree.

On the basis of lowest BIC scores (Bayesian Information Criterion), best substitution model selected for rbcL was Jukes-Cantor+ Gamma (rates among sites), matK was Kimura 2-parameter, trnH2-psbA was Tamura 3-parameter+gamma and rpoC1 2-3 was Kimura 2-
parameter + evolutionarily invariable respectively. These selected models with their rates were utilized for generation of dendrogram.

Four dendrograms were generated for each of the primer pair. In the dendrogram generated from \textit{rbcL} primer pair, 34 number of species were divided into six divisions. The members belonging to subtribe III namely \textit{Cephalostachyum capitatum}, \textit{Dendrocalamus asper}, \textit{Dendrocalamus brandisii}, \textit{Dendrocalamus membranaceus}, \textit{P. polymorphum} were grouped together along with the members of subtribe II such as \textit{B. balcooa}, \textit{B. nana}, \textit{G. atroviolacea}. One member of subtribe IV (\textit{D. macclellandii}) was also present in the same division. In the second division \textit{S.palmata and P. japonica} belonging to subtribe I were present along with other members of subtribe I namely \textit{B.wamin}, \textit{B. multiplex}, \textit{B. striata}, \textit{B.vulgaris}. Mixed kind of distribution of different species of bamboo was observed at this locus.

Dendrogram generated from \textit{matK} showed that the two species belonging to subtribe I (\textit{O. scriptoria} and \textit{M. baccifera}) occupies their position under same division. \textit{B. nutans}, \textit{B. striata}, \textit{B. vulgaris} were nearby to one another occupying same subdivision. The boot strap values for different nodes were found to be highest as compared to other dendrograms therefore the reliability of this locus is greater as compared to the other studied loci. According to the \textit{matK} dendrogram \textit{T. oliveri}, \textit{B. balcooa}, \textit{D. asper} and \textit{G. angustifolia} stand separately from the other species. Dendrogram shows greater number of branches which shows that the bamboo species considered are genotypically different from each other and does not fall into one group. The members belonging to on genera are also located under different divisions. Closely similar \textit{B.tulda} and \textit{B. nutans} were present under different subdivisions. While \textit{B. striata} and \textit{B. vulgaris} were present under different subdivision.
The dendrogram generated from the primer pair *trnH2-psbA* showed the greatest number branches among the four primer pair studied. Eighteen of the 34 species of bamboo occupied separate division which shows that these all species are independent from each other. The dendrogram was divided into two divisions with one division containing *B. polymorpha*, *B. balcooa* and *O.scriptoria*. The other division contained members of *G.macrostachya*, *G. nigrociiliata* and *G. atroviolacea* positioned nearby followed by 5 members of *Dendrocalamus* genera such as *D. asper*, *D. brandisii*, *D. giganteus*, *D. hamiltonii* and *D. membranaceus* clumped together. A clear cut division is exhibited genera-wise, resulting into the analysis of phylogenetic evolution. The primer sets proves to be efficient in phylogenetic analysis.

The dendrogram generated from *rpoC1* primer pair also yielded greater number of branches. Closely similar *B.tulda* and *B. nutans* species were present under different division. Dendrogram generated from rpoC1 showed a common ancestry of *B. nana*, *B. multiplex* and *B. vulgaris*. The two species *B.nutans* and *B. striata*; and *B.tulda* and *B. balcooa* occupied very close position to each other, exhibiting their common ancestor respectively. The overall dendrogram generated showed a mixed distribution of different species considered, thus exhibiting low resolvability of phylogenetic background.

The sequence data generated from four primer pair were concatenated using the software “SEAVIEW” and a combined dendrogram has been generated representing all the sequences. The four species *T. oliveri*, *B. balcooa*, *D. asper* and *G. angustifolia* occupied a separate position in the dendrogram.

Estimation of mean diversity in entire population was computed which resulted into overall mean distance of 0.006 for *rbcL*, 0.049 for *matK*, 0.016 for *trnH2-psbA* and 0.012 for *rpoC1*. Standard error estimate(s) are shown in the second column and were obtained by a
bootstrap procedure (100 replicates). The numbers of base substitutions per site from averaging over all sequence pairs are shown. Sequences were submitted to NCBI through ‘BANKIT’ in order to obtain the accession number.

In the case of $matK$ primer forward primer sequence included $matK390$, $matK$ X, $matK$ 2.1, $matK$ 2.1a and the reverse included $matK$ 1326 and $matK$ 5. The best amplification was observed in $matK$ X and $matK$ 5. Primer $rpoC1$ had 2 forward primers as $rpoC11$, $rpoC12$ and reverse as $rpoC13$, $rpoC14$. The four combinations $rpoC11$ and 3, $rpoC11$ and 4, $rpoC12$ and 3, $rpoC12$ and 4 were tested for best amplification. The primer pair $rpoC12$-3 was found to give best amplification. Negative amplification of combination of two primer sequences as reverse and forward may be attributed to the fact that they lack complementary binding sequences in the target template DNA or sometimes the forward and the reverse primer lie outside the amplification range due to presence of INDELS.

Apart from the combination of forward and reverse primer set, the annealing temperature plays an important role in generation of amplicons. Therefore, it is required to be standardized by performing gradient PCR. The primer set $rbcL$, $matK$ and $rpoC1$ gave unambiguous result at the reported annealing temperature, except $trnH2$-$psbA$ primer set which required gradient PCR for determination of appropriate amplification. In addition to the annealing temperature it ($trnH2$-$psbA$) a required changes in extension temperature from 72°C to 68°C (which was 72°C for other PCR condition).

The combination of the primers from coding ($rbcL$, $matK$ and $rpoC1$) and non-coding region ($trnH2$-$psbA$) exhibited better discrimination among the species studied. In the present study, the amplicons generated by the primer pair $rbcL$, $matK$ and $rpoC1$ 2-3 resulted into 100% PCR success and sequencing to generate 102 sequences. However, the primer pair $trnH2$-$psbA$
resulted into amplification of 30 out of 34 species studied thereby, accounting for 88.23% PCR success rate. A total of 132 sequences were generated in the study.

Once the amplicons were obtained they were subjected to sequencing and the sequence data was analyzed using software MEGA 5.2.2. The sequences were aligned using MUSCLE as it is claimed to achieve both better average accuracy and better speed than ClustalW2 or T-Coffee, depending on the chosen options. total sequence length of 666 bp for rbcL, 834 bp for matK, 529 bp for trnH2-psbA and 409 for rpoC1 were generated, in which no of variable sites found were 114, 621, 111 and 41 respectively.

Alignments of the sequences were followed by computing the genetic diversity existing within the population. The values for genetic diversity were calculated as per the parameters selected through the ‘Best substitution model’. The overall mean distance was highest in matK (0.049), followed by trnH2-psbA (0.016) > rpoC1 (0.012) and the least overall mean distance was found in rbcL (0.006).

The maximum genetic divergence was obtained with matK primer set, therefore, it could be considered as potent candidate for the barcoding purpose. But in general, one limitation associated with this primer set is that it is not easy to amplify in some group and that additional work is required for primer development. Keeping the above mentioned difficulty of universal amplification of matK, better alternative for barcoding is trnH2-psbA primer set. This primer set showed an average genetic diversity of 0.016 and exhibited high sequence divergence among the species. From the present study despite exhibiting sufficient genetic diversity this locus lack 100% PCR success rate therefore, acquire certain limitation for suitability of its usage as barcode.
As far as the 100% PCR success rate is considered, \(rbcL\) primer set worked up to the mark by amplifying in all of the 34 species of bamboo used. Alignment of the sequences was followed by the generation of dendrogram. The dendrogram generated from the three primer pair \(rbcL, matK\) and \(rpoC1\) 2-3 showed greater number of branching therefore exhibiting their discriminating power. The morphologically similar species such as \(B.\) \(tulda\) and \(B.\) \(nutans\), occupied their position in separate division. Similarly, \(B.\) \(polymorpha\) and \(P.\) \(aurea\) is placed under one division in three primer pairs \((rbcL, matK\) and \(rpoC1\) 2-3), which accounts for their closeness at the genetic level and warrants confirmation by further study.

The four considered barcode primer sets exhibited enough potential to distinguish the closely related species. The primers \(rbcL\) coding for larger subunit of ribulose 1-5 bi-phosphate carboxylase/oxygenase (RUBISCO) enzyme, \(matK\) coding for maturase and \(rpoC1\) coding for one of the subunit of chloroplast RNA polymerase, are taken for the coding region and therefore, largely comprises of conserved region resulting into low degree of variation. The primer set \(rbcL\) showed 100% PCR success rate and was suitable to distinguish closely related species as \(B.\) \(tulda\) and \(B.\) \(nutans\). The other primer sets used in the study were also capable enough to differentiate between the 34 species studied.

On the contrary, primer set \(trnH2-psbA\) belonging to the intergenic spacer region, is the most variable genome segment, shows greater degree of variation and can prove itself to be a better option for the barcode purpose. The distinguishing power of primer set \(trnH2–psbA\) was found to be maximum as the greatest number of branches was found from the dendrogram generated but one of the problem associated with this locus was that it doesn’t exhibited 100% PCR success rate.
CBOL, the Consortium of Barcode of Life has proposed two loci, *matK* and *rbcL* as the BARCODE regions for Land Plants. The amplification success rate of both the primer sets are 100% therefore it is proposed to create backbone of tree with *rbcL* as the ‘Anchor’; then separate individuals species in smaller groups with *trnH-psbA* as the ‘Identifier’. Result suggests that the *trnH-psbA* intergenic spacer is the best plastid option for a DNA barcode sequence that has good priming sites, length, and interspecific variation. If the problem of amplification is managed for the primer *trnH2* and *psbA*, it will prove itself to be a better choice for the generation of universal barcode.

Most phylogenetic reconstructions in plant systematics conducted so far is based on molecular data from the cpDNA genes.

However, an attempt was made to study the phylogenetic analysis of 34 species of bamboo using *rbcL*, *matK*, *trnH2-psbA* and *rpoC1*. The dendrogram generated from the three primer set exhibited that the species *B. polymorpha* and *P. aurea* were genetically similar and showed a common ancestor with greater bootstrap value.

The most common gene which has been used to provide sequence data for plant phylogenetic analyses is the plastid-encoded *rbcL* gene in which the success rate of PCR and sequencing is higher compared with the other selected plant characterization genes such as *matK* (maturase K). The phylogenetic tree generated from *rbcL* resulted formation of 6 subdivisions, in which clumping of *B. multiplex*, *B. striata*, *B. vulgaris*, *B. wamin* under on subdivision with other members *P. japonica*, *S.palmata*, *D. andamanica* and *O. scriptoria* were observed. *G. angustifolia*, *B. polymorpha* and *P. aurea* were closely related as they lie separately in one subdivision. A mixed distribution of *Dendrocalamus* genera is observed. But in the first
subdivision *Dendrocalamus* genera lie with *C. capitatum* and *P. polymorphum*, all belonging to Subtribe III- Dendrocalameae.

The technique of DNA barcoding involves generation of sequence data for the different species under study, at a particular target region of interest, followed by comparison of those sequence data to the available database at NCBI. Therefore, the sequence database is greatly required for comparison and analysis of the generated sequence data. A total of 132 sequence data from 34 species of bamboo using four primer sets were generated which can further act as basis for similar works. The DNA document available in the public domain will lay the foundation to generate a universal barcode for identification and classification of different bamboo species.

In conclusion, the findings of the present study results presented in this study demonstrate the utility of DNA barcoding as a good method for identifying different species of bamboo. With the use of four primer set in 34 species of bamboo the following conclusion were drawn: 100% PCR success and sequencing data rate were obtained for the primer sets *rbcL*, *matK* and *rpoC1*. *trnH2-psbA* primer set gave 88.23% PCR success rate resulting into generation of 30 sequence data out of 34 species considered for the study. Negative amplification was observed in *Bambusa multiplex, Dendrocalamus sikkimensis, Dendrocalamus strictus* and *Schizostachyum per gracile*.

All the four barcoding primer sets could distinguish 30 of 34 species of bamboo studied. The barcoding sequences could distinguish even morphologically similar species such as *B. nutans* and *B. tulda*.

Employing the barcoding primers, phylogenetic analysis was possible within the studied population. The primer sets derived from the non-coding regions showed greater variation in comparison to the primer designed from coding region.
Estimation of mean diversity in entire population was computed which resulted into overall mean distance. Maximum variation was observed at the locus matK followed by trnH2 and psbA. Therefore the first choice of primer being utilized for barcoding purpose is matK followed by trnH2-psbA.

The coding region represents highly conserved region on DNA, resulting into exhibition lesser degree of variation in order to maintain the quality and quantity of proteins coded. On the other hand, the non-coding regions offer variable sequences without affecting the quality and the quantity of protein. To conclude a two-locus barcode that combines a subunit of the coding locus such as rbcL (rbcL-a) with the non-coding trnH-psbA spacer is recommended. rbcL-a provides a strong recognition anchor that will place an unidentified specimen into a family, genus, and sometimes species; the highly variable trnH-psbA spacer will further narrow the correct species identification where rbcL-a lacks discriminating power, especially in species-rich genera of angiosperms.

DNA document in the form of 132 sequence data has been generated which can act as a foundation for the upcoming studies not only related to DNA barcoding but also other Molecular Biology research.

The results presented, adequately address to the concerns about the applicability of DNA barcoding to plants. However, quest for a perfect universal barcode for plants providing 100% species resolution across the plant kingdom appears to be unrealistic, as DNA barcoding, like any other technology, is not expected to be 100% perfect. However, within a taxonomic group 100% species resolution could possibly be obtained by taxa specific barcodes. Thus, the projection that DNA barcodes, once available for all the described species, would be able to provide a correct identity up to species level to any unknown sample, whether available in
vegetative, fragmented or DNA form, or would indicate the discovery of a new species does not hold true. Nevertheless, more than 90% success in species identification with single locus or two-/three-locus combinations emphatically demonstrates the efficacy of the technique.

The instances of failure of DNA barcodes to correctly assign the species should encourage taxonomist to re-consider or reinvestigate such taxa. Future developments in the techniques and technologies underpinning DNA barcoding could improve the success rate and practicalities of identification of species. However, the most important development will be the continued expansion of the Barcode of Life Database.