Chapter 5:

Discussion

Chapter 5: Discussion

5.1 Sampling

In the present study, collection of Bamboo samples from various places in Southern Assam has yielded about 15 different samples/species of Bamboo. There are a few regions like Dargakona, Singari and Bazaricherra/Shantinagar, which displayed rich species diversity of Bamboo. In Bazaricherra alone, seven species of Bamboo could be collected. The Latitude and Longitude of the places from where bamboo samples were collected was determined by a handheld GPS device by "Garmin".

There is rapid destruction of Bamboo habitat in Southern Assam. What was seen as a Bamboo rich area was found to be cleared and used for other domestic purposes c purpose on the Furthermore, field visit. Also talking with old timers, it was found that the species diversity and richness of Bamboo was much higher previously (even 7 years back). However, the present scenario is appalling due to the rampant destruction of Bamboo habitat. The increased demand for housing land, for constructing commercial building also lead to large clumps being rampantly cut down and the area cleared.

Another reason that is responsible for the wanton destruction of Bamboo habitat is the increasing demand for Bamboo for various purposes. With the ever increasing demand and subsequently increase of price, the landholder feels no qualms about cutting down entire clumps for easy money. They can make a huge profit as some species like Betua can fetch Rs 100-150 per strand.

The Cachar Paper Mill located at Panchgram was established here to the abundance of Bamboo in the region comprising of Southern Assam, Mizoram and Tripura. But despite having such huge catchment area, the paper mill on the verge of shutdown due to lack of Bamboo. With the increasing price of Bamboo, it has also become economically unviable to produce paper from Bamboo. The high input price of raw material, i.e. Bamboo has meant that the Paper Mill incurs huge losses.

In none of the Bamboo clumps that were visited showed any flowers and the local people also do not remember seeing a Bamboo flower in their lifetime. But some of them has mentioned hearing from their father and grandfather about Bamboo flowering. They talked about the mass hysteria and the sense of fear. They said that bamboo flowering is usually associated with famine and extremely low production of food grains.

5.2 Extraction and Sequencing of DNA

The extraction of DNA from Bamboo young leaves was especially difficult for a number of reasons, as mentioned below:

Firstly, the fibre content of the Bamboo leaves is very high (older leaves have much higher fibre content than young ones) thus making it very hard to grind. Excessive friction caused heating of the mortar and pestle, which could lead to improper extraction of DNA. The problem could be easily overcome by chilling the mortar and pestle before grinding. Furthermore, it was ensured that no materials remained from previous grinding. Thorough washing followed by autoclaving was done on the mortar and pestle after every grinding,

Secondly, the RNA content of the leaves was very high, particularly in the very young and young leaves which interfered with the quality of DNA extracted. The high RNA content was clearly visible in the Gel electrophoresis.

Thirdly, the addition of RNAase enzyme solved the problem of RNA content in the extracted DNA samples.

Fourthly, the amplification of *matK* gene, *ITS* and *psbA-trnA* spacer region could be carried out with existing primers cited in literature. There are other variations of primers for the three targeted DNA sequences, but in case of Bamboo, the primers mentioned below showed good result. The primers used for *matK* and *psbA-trnH* have been adapted from Raghupathy (2009) and *ITS* from Kress (2005).

5.3 Genomic analysis of Bamboo DNA Sequences

The analysis of the genomic composition of the three targeted sequences gave an insight into their composition. We know for a fact that higher GC% content was indicative of the fact that the complexity in the genes is higher. Analysis showed that the GC% content (~33%) in *matK* and *psbA-trnH* was half of the AT% content (~66%). But in case of internal transcribed spacer (*ITS*), the opposite is true where the GC% content is twice that of AT% content. So just understanding the fact that species level demarcation and proper phylogenetic relationship can only be derived when there are variations among the species and genus. Two specimens with common ancestry will have lower variations, but variations have to be present for recognition by software. Variations are dependent on sequence complexity and lower GC% content is indicative of low complexity. In this case both *matK* and *psbA-trnH* show much higher AT% content compared to GC%. However, in case of *ITS* sequence, it is clearly indicated by its higher GC% content that the complexity is much more than the other 2 above mentioned sequences.

The Mean pair wise distance analysis was computed using Kimura 2 Parameter to check the interspecies divergence between all species among the members of the Bamboo based on *matK*. A record was made to discriminate the intra and interspecies variation in Bamboo. The maximum mean intra-specific distance was found in *Nardus stricta* i.e. 0.082 and there are innumerable species, which showed a minimum variance of 0.000. A pair wise distance value of 0.000 indicates that there is no variation in the *matK* sequence among two species, although they are clearly different based on morphological data. Although proof of morphological variations and separate species is present, but the variations as not reflected in the *matK* sequence.

The Mean pair wise distance analysis was computed using Kimura 2 Parameter to check the interspecies divergence between all species among the members of the Bamboo based on *psbA-trnH*. A record was made to discriminate the intra and interspecies variation in Bamboo. The maximum mean intra-specific distance was found in *Melocalamus arrectus, Melocalamus compactiflorus var. fimbriatus* and

Melocalamus scandens i.e. 1.152. However, there is many pair of species which showed a minimum variance of 0.000. A pair wise distance value of 0.000 indicates that there are no variations in the two sequence, although they are clearly different based on morphological data. It indicated although there are morphological variations and clear indication of them being separate species, the variations as not reflected in the *psbA-trnH* sequence.

The Mean pair wise distance analysis was computed using Kimura 2 Parameter to check the interspecies divergence between all species among the members of the Bamboo based on *ITS*. A record was made to discriminate the intra and interspecies variation in Bamboo. The maximum mean intra-specific distance was found in *Otatea glauca* i.e. 0.121, but there are very few species, which showed a minimum variance of 0.000. Maximum data of species is seen to be in the range of 0.030 – 0.040 which are adequate for differentiation at species level.

The transition/transversion rate ratios for *matK* sequences are $k_1 = 1.936$ (purines) and $k_2 = 2.442$ (pyrimidines). The overall transition/transversion bias is $R = 1.021$. The *psbA-trnH* values are $k_1 = 2.041$; $k_2 = 2.009$ and $R = 0.958$. Similarly, $k_1 = 2.915$; $k_2 = 1.231$ and $R = 0.944$ values are seen in case of *ITS* sequences.

5.4 Phylogenetic Relationship of Bamboo Sequences

Among the sequence of our own and those which have been downloaded from the database, the Subtribe Bambuseae was the best represented. This is expected as the species diversity of the bamboo is very high in the Asian region.

The analysis of Neighbour-Joining Phylogenetic tree of Bamboo showed the different result for the three bamboo sequences. The tree was constructed using the Kimura 2 Parameter with a bootstrap value of 1000.

The *matK* sequences show the capability to easily resolve the Subtribes and correctly resolve upto the genus level. But *matK* failed to resolve at the species level and give a clear idea about the ancestry and hierarchy of the species. Other ambiguous mixing up of genus like *Bambusa* and *Dendrocalamus* also showed that

it has failed in the discriminating at the Genus level but has successfully grouped them together in a Subtribe clade. However, in the case of the species *Olmeca fulgor* the *matK* sequences are unable to put them together in a single clade. Instead it is observed, it is spread all over the phylogenetic tree at 7 different locations in Fig 5.1

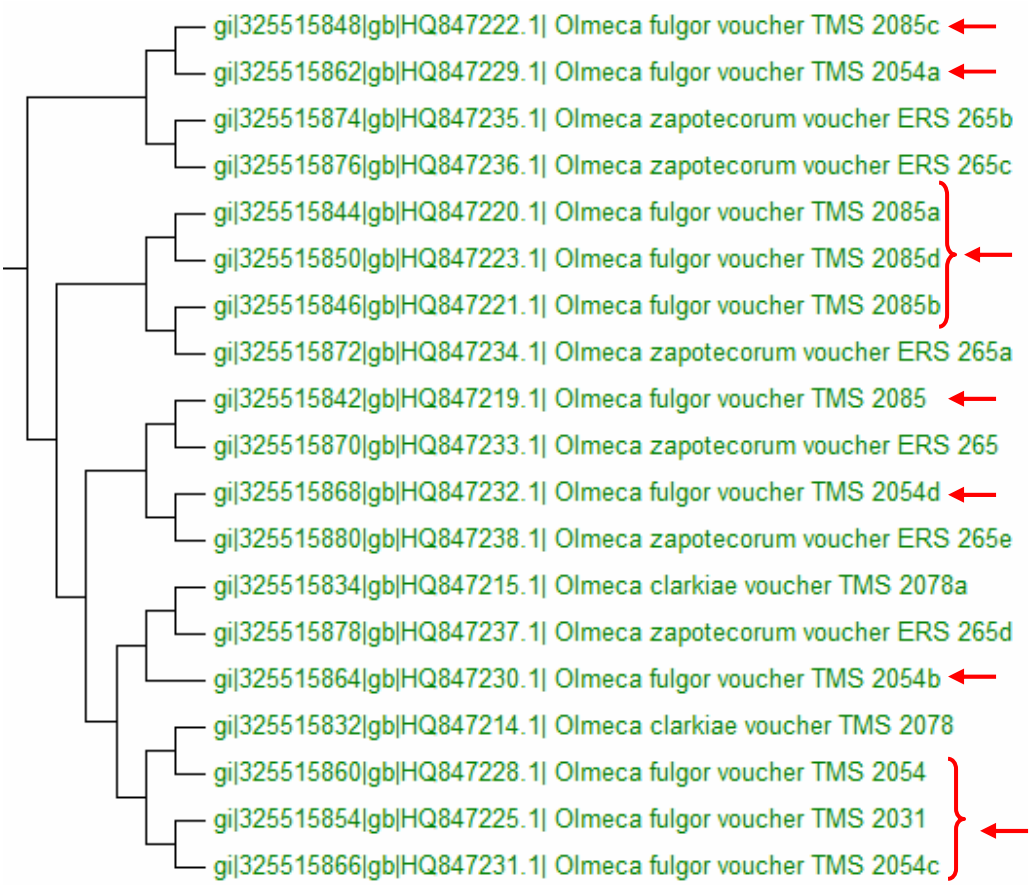


Fig 5.1: Error of species demarcation in *matK* phylogenetic tree.

Analysis of Fig 5.2, clearly showed the various genus and species have been badly mixed up not only at the Genus level but also at the Sub tribe level. Whereas the species *Nardus strictus* belongs to the Subtribe Arundinariinae and the species *Lygeum spartum* belongs to the Subtribe Neurachninae. But as indicated by arrows, they have been brought together to form a clade which is absolutely wrong.

Fig 5.2: Error of genus demarcation in *matK* phylogenetic tree.

The ambiguities mentioned above indicate a mixed result from *matK* sequences. On one hand it showed capability to clearly demarcate at genus level but it mixes up the genus in other cases. The phylogenetic tree fails to provide us a clear picture as to the grouping of species into distinct clades. It also fails to show the lineage and the evolutionary relationship among them.

The *psbA-trnH* sequences of Bamboo showed the least capability to resolve the Phylogenetic relationship. The NJ tree did not form clades or superclades and there was arbitrary mixing of various genus and species. On analysis of Figure 5.3, it showed *psbA-trnH* failed to resolve the phylogenetic relationship in Bamboo at the Species, Genus or Subtribe level, as seen below. These set of sequence showed the worst result among the 3 sequences.

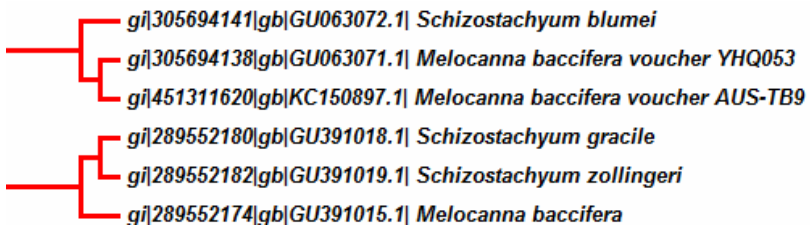


Fig 5.3: Error of genus demarcation in *psbA-trnH* phylogenetic tree.

Analysis of Fig 5.4 showed the massive failure of *psbA-trnH* to resolve the phylogenetic relationship in Bamboo. In this case, *Dinochloa* and *Kinabaluchloa* belong to the Subtribe Bambusinae. *Chusquea* belongs to the Subtribe Chusqueinae and Olmeca belongs to the Subtribe Guaduinae. However, analysis of the Fig 5.4 clearly showed that *psbA-trnH* sequences have placed 3 different Subtribes under a single clade. It is a well known fact that Subtribes have different ancestry, and as such they tend to form separate clades in a phylogenetic tree. Subtribes are quite distant evolutionarily compared to Genus and species, which have closer ancestry. However, in above case it is seen that *psbA-trnH* sequences have mixed up the whole evolutionary part and have brought together different Subtribe and separated similar Genus and species, as is seen below.

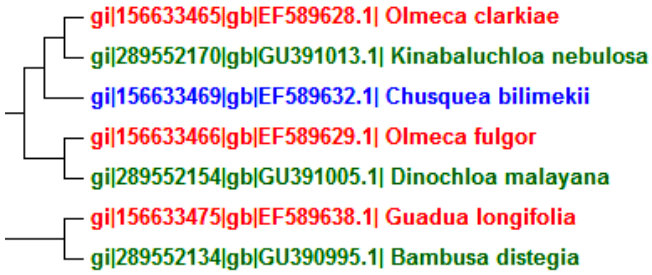


Fig 5.4: Error of Subtribe demarcation in *psbA-trnH* phylogenetic tree.

Analysis of Fig 5.5 the part of the *psbA-trnH* phylogenetic tree, two portions of which is given below, showed no clade formation. Also there is no clustering of the sequences into the correct subtribes. The tree showed that the phylogenetic relationship, whether the hierarchy or ancestry is not formed and *psbA-trnH* sequences of Bamboo is incapable of resolving the phylogenetic relationship.

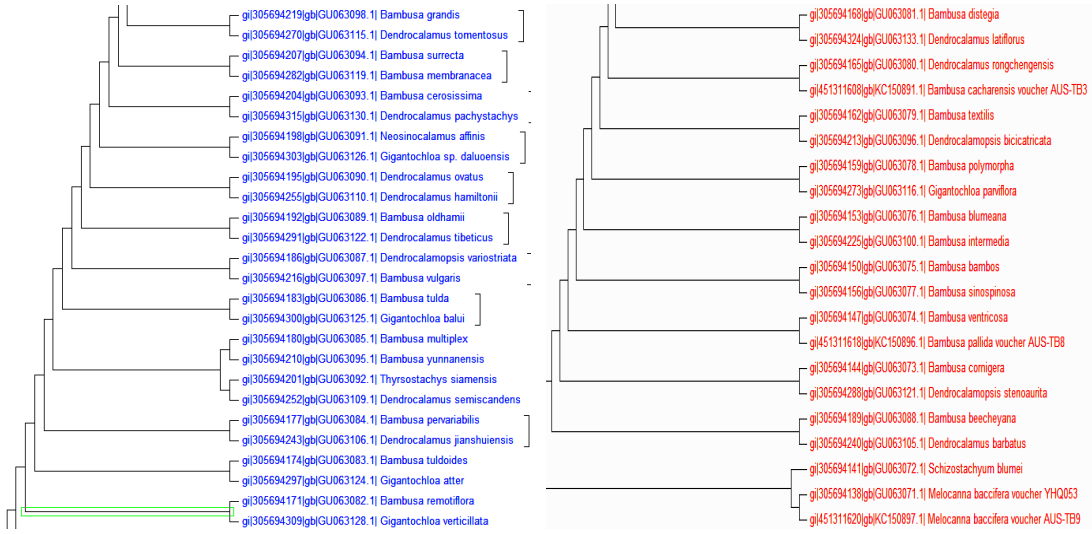


Fig 5.5: Non formation of taxonomic hierarchy in *psbA-trnH* phylogenetic tree.

So it can be easily seen that *psbA-trnH* fails not in resolving species and genus diversity, it also fails at a higher hierarchy level, i.e. Subtribe. As such *psbA-trnH* is best not used alone in elucidating the phylogenetic relationship of Bamboo.

Internal transcribed spacer (*ITS*) sequences proved to be the best performer in elucidating the Phylogenetic relationship in Bamboo. Not only *ITS* show the correct

hierarchy and ancestry of Bamboo upto the Subtribe and Genus level, but it is the only set of Sequences that could resolve the Species level relationship in Bamboo.

Close examination of Fig 5.6, it apparently showed the inclusion of wrong genus and species in the phylogenetic tree. The names of the plants were checked in the website "The Plant List". The Plant List is a working list of all known plant species. *The Plant List* provides the Accepted Latin name for most species, with links to all Synonyms by which that species has been known. *Oxynanthera abyssinica* which seems to be wrongly included in the same clade as *Bambusa chungii* proves to be correct when it is seen that *Oxynanthera abyssinica* is a synonym of *Bambusa abyssinica.* Another remarkable example is also seen below where *Dendrocalamus strictus* seems to be an aberration to be put in a superclade of different *Bambusa* species. But upon knowing that *Bambusa stricta* is the synonym of *Dendrocalamus strictus*, the whole phylogenetic tree makes much more sense and coherent. It now showed that the phylogenetic tree that is formed is correct.

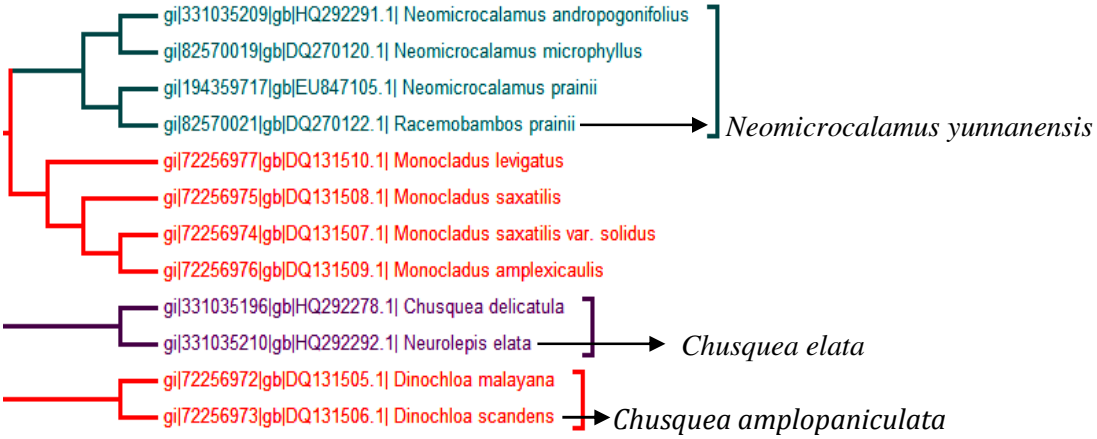


Fig 5.6: Enlargement of Fig 4.56 with synonyms.

Close observation of Fig 5.7 and Fig 5.8, showed the inclusion of wrong genus and species in the phylogenetic tree, but checking the specimen name at the website "The Plant List" provided a much clearer picture about the correct clade formation. "The Plant List" website is at `http://www.theplantlist.org/tpl/record/kew-438976.` The entry marked # was checked at the website `http://www.bambus-lexikon.de/fargesia-boliana.html`

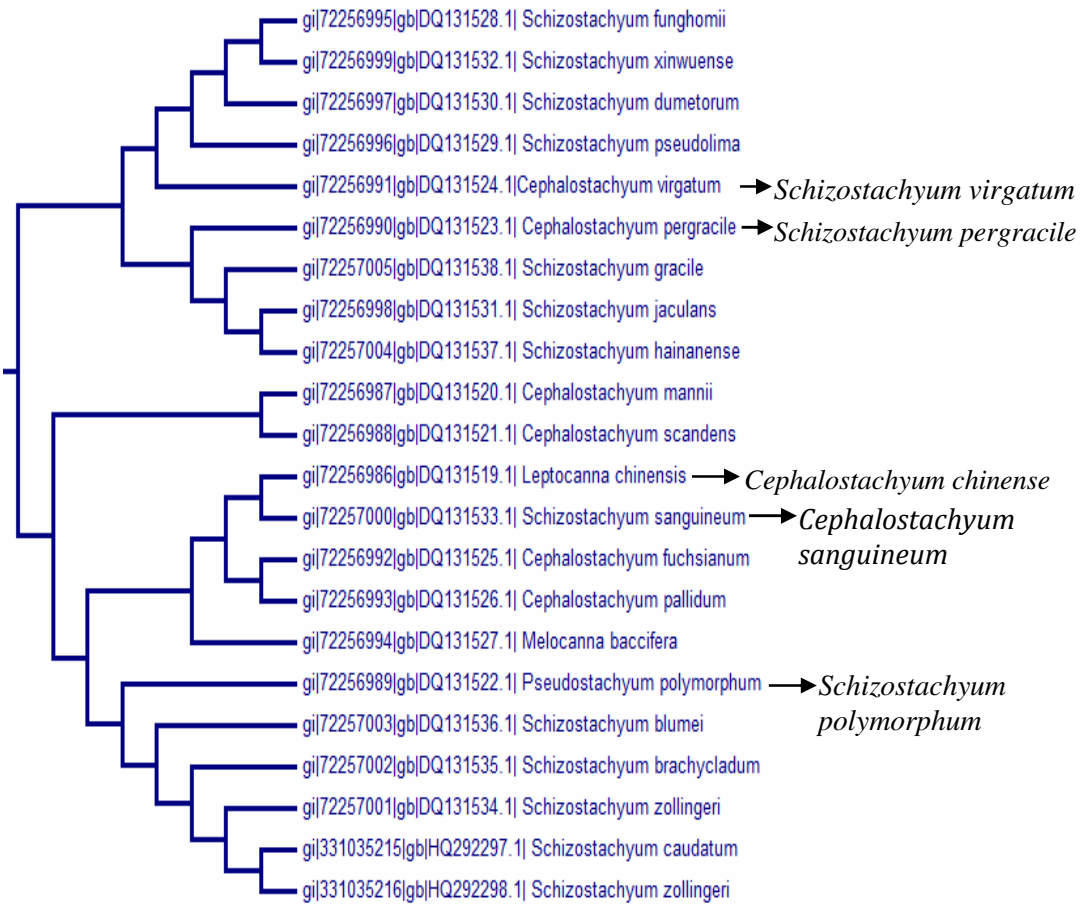


Fig 5.7: Enlargement of Fig 4.57 with synonyms.

Fig 5.8: Enlargement of Fig 4.57 with synonyms.

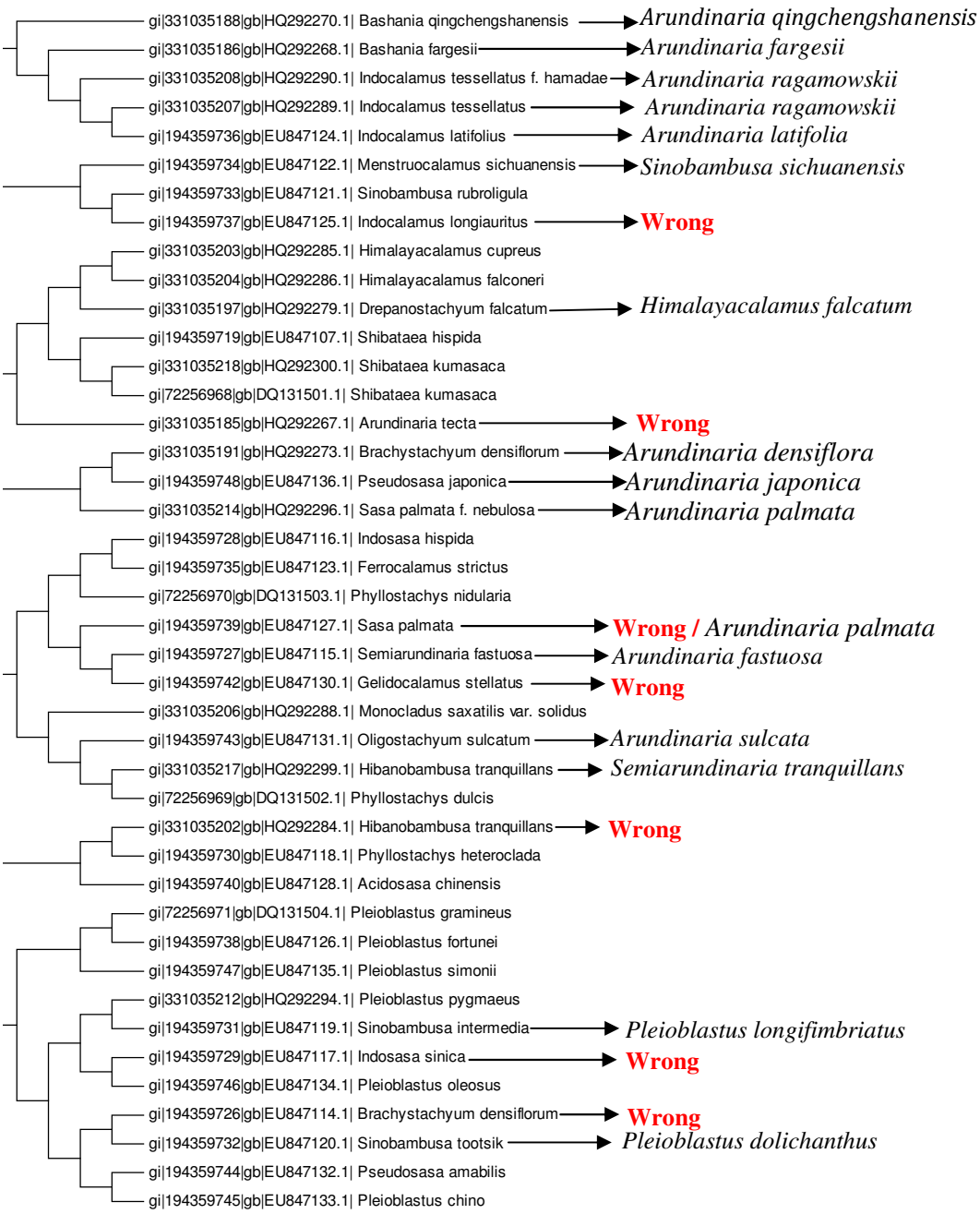
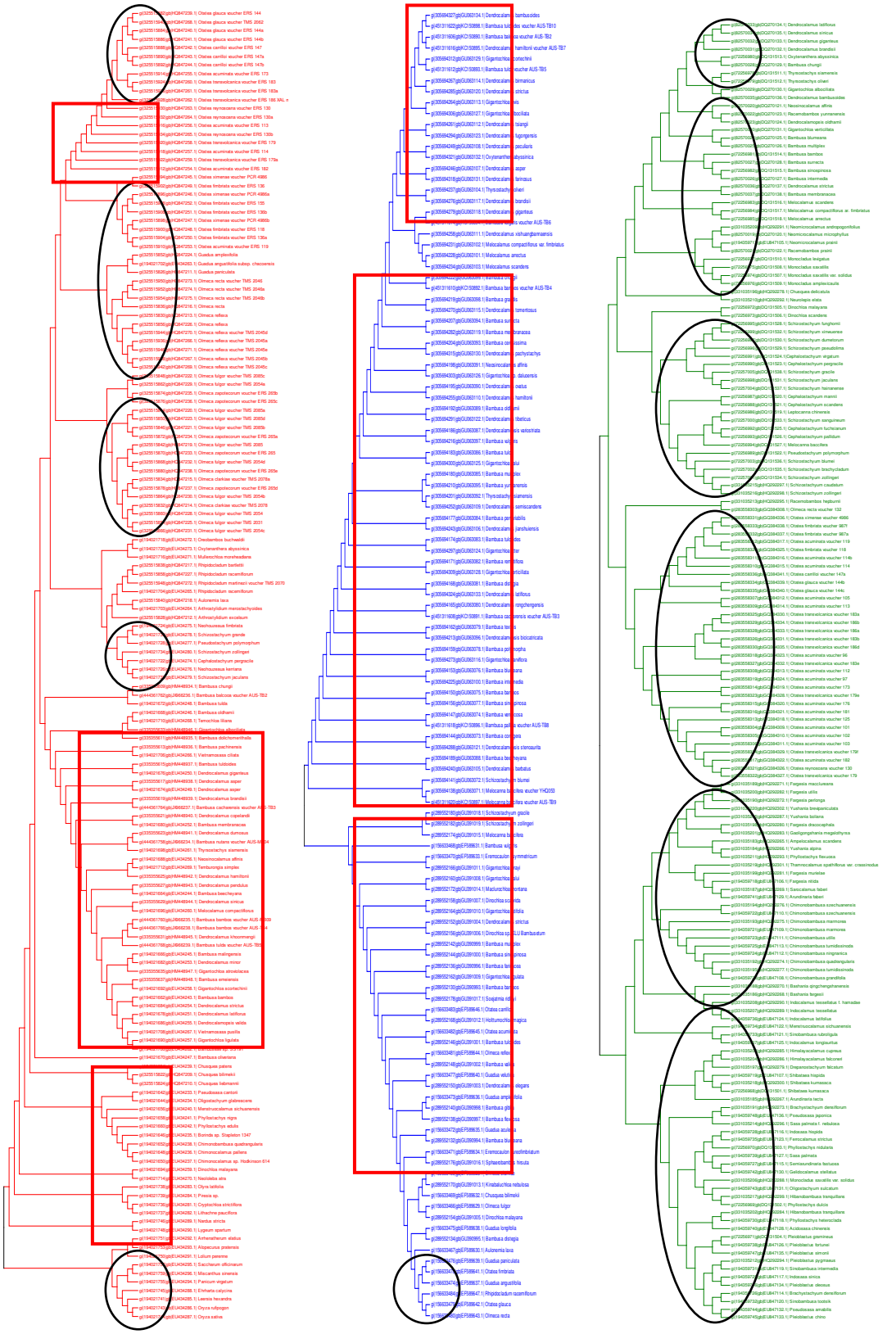


Fig 5.9: Enlargement of Fig 4.57 with synonyms.

All through the study of phylogenetic relationship, the use of various synonym and basionym tends to give false impression of failure and misalignment. But closer examination along with study of each and every scientific name of Bamboo species showed innumerable synonyms and basionym being used. This fact is clearly illustrated in Fig 5.9

The following observations were also made:

1. As the Figures 4.45, 4.46, 4.48, 4.50 showed, the *ITS* could fully demarcate the 161 *ITS* sequence into its respective Tribes and Subtribes.
2. It could easily resolve the different genus of the sequences.
3. It showed great capability to differentiate species and assign them to correct clades. In the Subtribe Melocanninae, it might appear that genus *Schizostachyum* and *Cephalostachyum* is incorrectly resolved and there is an ambiguity in resolution. These 2 genuses are nomenclatural synonyms and *Cephalostachyum* is the basionym of *Schizostachyum*. So although what may appear to be uncertainty in correct assignment of hierarchy, it arises due to taxonomic complications. Many more such examples are given in the figures above.
4. Clade assignment was also correct as it maintained the correct hierarchy and ancestry of the species.
5. It is observed in Fig 4.48 that the different species under the genus *Otatea* has been properly demarcated and the clade formation is correct. The evolutionary relationship and the ancestry are also clearly indicated.
6. It is also observed that in Fig 4.50, the species under the genus *Chimonobambusa* has been properly demarcated and the clade formation is correct. The evolutionary relationship and the ancestry are also clearly indicated.
7. Only 7 out of 166 sequences were wrongly assigned or not assigned at all, which showed a 95.78% accuracy of *ITS* in resolving the phylogenetic relationship of Bamboo. In biological sciences, most of the studies are carried out at 95% confidence level (p value = 0.05) and in this case of *ITS*, it also attains the same significance level.



MatK tree ***psbA-trnH tree*** ***ITS tree***

Fig 5.10: Comparison of all phylogenetic trees

Upon analyzing the bird's eye view phylogenetic tree in Fig 5.10, it can be easily seen which of the three trees provides relatively more coherent information. The black oval indicates those specimens which have been correctly assigned and red rectangle depict the area which has been wrongly depicted.

It is seen that the *matK* is a mixture of success and failure. It could form small clades correctly but failed in superclades. In two instances large areas show no superclades formation. This failure is also seen in *Calligonum* species by Li et al. (2014). The *psbA-trnH* phylogenetic tree could not form clades correctly except for one instance, but 90% of the tree is without proper clade formation. Two large areas marked in a red rectangle showed non formation of any taxonomic hierarchy or ancestry. The work of Spooner (2009) on *Solanum* sect. *Petota* showed that *psbA-trnH* sequence failed as the plastid markers lack sufficient polymorphism. In case of Bromeliaceae, it is proved that both *matk* and *psbA-trnH* failed due low amounts of plastid variation Maia et al. (2012).

The *ITS* phylogenetic tree showed correct clade and superclade formation. It also correctly depicts the ancestry and phylogenetic relationship between the species. The work of Roy (2010) showed that *rbcL, matK, psbA-trnH* and *ITS* all failed to discriminate the species of *Berberis.* The work also showed that *ITS* resolved all the tested species of *Ficus* and *Gossypium* and *psbA-trnH* resolved 82% of the tested species in *Ficus*. The recommended *matK* and *rbcL* sequences could not resolve all the species. Sequences of *ITS* has also showed its capability to discriminate adulterants mixed with medicinal plants, *Gentianopsis paludosa* as shown by Xue (2011). The work of Dong et al. (2011) using *rbcL*, *matK* and *ITS* to evaluate the five species of *Pterygiella* of family Orabanchaceae found that only *ITS* could successfully identify all species of this genus. Studies to evaluate the plastid DNA markers and ITS on species of *Ligustrum* of family Oleaceae by Gu et al. (2011), *Hedyotis* of family Rubiaceae by Guo et al. (2011), *Primula* of family Primulaceae by Yan et al (2011) and *Tetrastigma* of family Vitaceae by Fu et al. (2011) showed that ITS was the only sequence capable of discriminating at the genus and species level and also correctly depict the phylogenetic relationship.

While analyzing the pairwise distance matrices and the phylogenetic tree a number of discrepancies were glaringly evident. In many cases where it seemed that two different Species, Genus or even sub families have been brought together in a clade, which is incorrect. There can be a number of reasons for such result:

1. The species has been misidentified at the very beginning. The identification of Bamboo has always been a very difficult task due to lack of floral characters.

2. Some of the characters are seasonal in nature and are not available in every season, e.g. the culm sheath of Bamboo was present in spring and falls off with the start of the rainy season / summer.

3. Many of the morphological characters are environmentally influenced, and it's very easy to misidentify the species based characters, e.g. the tuft of hairs on the culm sheath can be blown away in high windy conditions.

4. Misidentification of just one or two characters in Bamboo is enough to determine the wrong taxonomic name.

5. The sequences may have been incorrectly labelled at any of the stages of processing.

6. The sequence may have been poorly resolved during sequencing.

7. Mixing up of the sequences at any stage.

8. The original taxonomic hierarchy based on morphological characters may be wrongly established. A multiple of author is of the opinion that the hierarchical taxonomy of Bamboo needs to be revisited. Even as late as 2001, the latest study by the Grass Phylogeny Working Group (GPWG) took place, and they have made radical changes in the taxonomy of Bamboo and tried to correct flaws.

9. The parameters for the bioinformatics study may have been incorrect, which led to wrong placement of hierarchical position.

10. The species mentioned may be synonymous. The nomenclature of Bamboo is a labyrinth of synonyms, basionym etc. This brings about a great deal of confusion as apparently similar species may be interrupted by other species, but more often than not, the interrupting species may be a synonym s or basionym of each other. A few of such taxonomical instances are given below:

Acidosasa fujianensis

Synonym: *Acidosasa notata, Indosasa longoligula, Pseudosasa notata*

Bambusa aculeata

Synonym: *Guadua aculeate, Guadua aculeata var. liebmanniana, Guadua inermis, Guadua intermedia*

Bambusa agrestis

Synonym: *Arundinaria kokantsik, Arundinaria marmorea, Arundinaria matsumurae, Bambusa marmorea*

Dendrocalamus affinis

Synonym: *Bambusa emeiensis, Lingnania affinis, Neosinocalamus affinis, Sinocalamus affinis*

Dendrocalamus albociliata

Synonym: *Gigantochloa albociliata, Oxytenanthera albociliata*

5.5 Conclusion

The present study was carried out in Southern Assam and field studies to collect different species of Bamboo were carried out. In the study, 11 species of Bamboo were collected from different places. The 11 species are: *Bambusa balcooa, Bambusa arundinacea, Bambusa vulgaris, Bambusa nutans, Bambusa chacharensis, Bambusa tulda, Bambusa assamica, Bambusa pallida, Melocanna baccifera, Schistachyum dullooa* and *Dendrocalamus hamiltonii.*

The leaf samples were used for DNA extraction and sequencing. Three sequences of Bamboo were targeted in the present study for determining the genomic variation in them. They are *matK, psbA-trnH* and *ITS* (Internal transcribed spacer). Six sequences of *matK* were submitted to the NCBI database of which five species were submitted for the first time. Nine sequences of *psbA-trnH* were also submitted to the NCBI database out of which five species were submitted for the first time anywhere. *In silico* study of genomic variations in Bamboo was carried out on 152 sequences (including our sequences) belonging to 103 different species of *matK*. Furthermore, 120 sequences belonging to 109 different species of *psbA-trnH* and 166 sequences belonging to 124 different species of *ITS* were downloaded from database.

The sequences for *matK* and *psbA-trnH* showed higher AT% content compared to GC% content. Only *ITS* sequences among the three targeted sequences showed high GC% content. The *matK* sequences have GC content = 35.0 and AT content = 65.0. The *psbA-trnH* showed AT content = 63.5 while the GC content = 36.5. The *ITS* sequences showed GC content = 71.4 and AT content = 28.6. Codon usage analysis of *matK* sequences showed UUU and AUU show count of over 20 while CCG and GUC showed the least count is an average of 431 codons. Analysis of the Amino acid composition of *matK* sequence indicates that Serine, Leucine, Isoleucine, Phenylalanine and Lysine are much more abundant than other amino acids. Methionine and Alanine showed up least in the *matK* protein. A nucleotide pair frequency of sequences is indicated by si (transitional pairs) and sv (transversional pairs). The nucleotide pair frequency of the three targeted sequences showed the following values. The si value for *matK* gene is 1049, while that of sv is 15; si value for *psbA-trnH* is 399, while that of sv is 57 and the si value for *ITS* is 469, while that

of sv is 26. Mean pairwise distance matrices of the three targeted sequence showed that different species have very fewer variations among *matK* and *psbA-trnH* sequences. There are innumerable pairs which have distance of "0.000" which indicates the sequences are identical. Only in *ITS* sequences it is shown that there are very few cells containing distance value of "0.000", most of them are in the 0.030 to 0.040 range, which is capable of species level demarcation as well as elucidating the phylogenetic relationship. The higher the distance values, the easier it is to gain useful information about the phylogenetic relationship among species. The transition/transversion ratio for purine is k_1, for pyrimidine is k_2 and the overall transition/transversion bias is R. The values for *matK, psbA-trnH* and *ITS* are k_1 = 1.936, 2.041, 2.915; k_2 = 2.442, 2.009, 1.231 and R = 1.021, 0.958, 0,944 respectively.

The phylogenetic tree analysis of the three targeted DNA sequences showed the *matK* sequences are able to demarcating Subtribes, Family upto genus level but are incapable of identifying at the species level. The *psbA-trnA* sequences are incapable of correctly identifying Subtribes let alone genus and species level. It showed the worst performance among the three targeted sequences. The Internal Transcribed Spacer (*ITS*) sequences are capable of clearly identifying specimens at Subtribe, Family, genus and species level. It showed the best performance. Out of 166 *ITS* sequences undertaken in the present study, there was an in only 7 entries which showed success rate or accuracy of 95.78%. The 7 errors seen cannot be attributed alone to *ITS* sequence, as a number of errors may occur in any stage of species identification, DNA contamination, sample labeling, sequence retrieval, etc.

The present study showed that *ITS* sequence has got great potential for correct identification of bamboo specimens at the species level and also elucidate the phylogenetic relationship among them. The *ITS* sequence showed accuracy of over 95%, and so it can be used for species level identification and can be relied upon to correctly depict the phylogenetic relationship and ancestry in Bamboo. The *ITS* sequences are easy to amplify using PCR and the primers work correctly, particularly in Bamboo. So it can be concluded from the DNA sequence based study of Bamboo that *ITS* sequences should be used as a standard.