Chapter 3:

Materials and Methods

Chapter 3: Materials and Methods

3.1 Materials

3.1.1 Experimental materials

In the current research very young leaves of Bamboo plants of 11 different species (Table- 3.1) were collected from different areas of Southern Assam (24°8'N – 25°8'N and 92°15'E – 93°15'E), India (Figure- 3.1). For DNA extraction, very young shoots and leaves were aseptically collected by first cleaning the surface of the leaves with alcohol and then put in plastic bags. Very young shoots and leaves are chosen as they are tender and thus easier to grind in mortar and pestle for DNA extraction.

Samples were collected from numerous places of Southern Assam but it was seen that a few places showed great diversity of bamboo varieties.

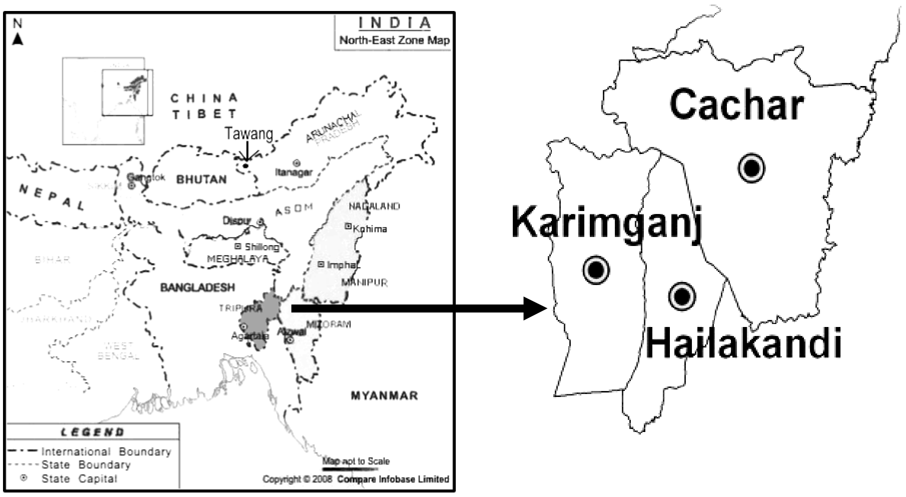


Fig 3.1: Map of Southern Assam in context of North East India

The plants were properly photographed and the morphological characters were duly noted and verified. The specimens were properly marked to avoid any confusion later. All the specimens were examined and identified by experts. The local name in *Bengali*, was also recorded, as informed by local residents. The leaf samples were preserved in -80° C for further investigation.

Sl	Scientific Name	Local Name
1	*Bambusa balcooa*	Sil Borua
2	*Bambusa arundinacea*	Kata Borua
3	*Bambusa vulgaris*	Jai
4	*Bambusa nutans*	Makal
5	*Bambusa chacharensis*	Betua
6	*Bambusa tulda*	Jama Betua
7	*Melocanna baccifera*	Muli
8	*Schizostachyum dullooa*	Dolu
9	*Bambusa assamica*	Mritinga
10	*Dendrocalamus hamiltonii*	Pecha
11	*Bambusa pallida*	Bakal

Table- 3.1 List of samples used in this study with their scientific name.

3.1.2 Chemicals

All chemicals used in this study were of molecular biology grade. Sucrose, Bromophenol blue, Ethidium bromide, Tris-Base, EDTA were procured from Promega, USA. Magnesium chloride, Sodium chloride, Sodium bicarbonate, Potassium Acetate, β-mercaptoethanol, Acetic acid, Methyl Alcohol and Glycerol etc were purchased from SRL Company. Dehydrate Ethanol and Rectified spirits were supplied by Bengal Chemical and Pharmaceuticals Works Ltd. Kolkata. *Taq* DNA polymerase, dNTPs Master mix, Reagents kit, Buffer, etc. were purchased from Fermentas, Germany; Bioline, U.K; Applied Biosystem, USA.

3.1.3 Buffer and Reagents

i. POTASSIUM ACETATE (5 M, pH 9)

Potassium acetate (anhydrous)	98.15 g
Distilled water up to	100 ml

The pH of the solution was brought to 9 by using glacial Acetic Acid.

ii. SODIUM DODECYL SULPHATE (SDS) 20%

SDS	20 g
Distilled water	100 ml

SDS was properly mixed, to ensure that no lumps were present, properly dissolved and then stored at room temperature**.**

iii. TRIS SATURATED PHENOL (pH 7.5)

Solid phenol was initially liquefied at 68° C in a water bath. A solution was made taking equal volume of liquid phenol and 1 M Tris HCL (pH 7.5) (1:1 ratio). The upper aqueous phase was discarded and the whole extraction procedure was repeated until the pH of the aqueous phase became 7.5. An equal volume of TE was added, after the final extraction and stored in a bottle. The bottle was covered carefully with Aluminium foil to ensure that the solution was not exposed to any light. The solution was stored at 4° C.

iv. TRIS (1M, pH 8):

Tris	12.11 g
HCL	7 ml

Distilled water was used to bring the volume upto 100 ml and the pH adjusted with NaOH buffer.

v. EDTA (0.5 M, pH 8.0) :

EDTA	18.61 g
Distilled water up to	100 ml

The final pH was adjusted to 8.0 using NaOH buffer tablets before reaching the final volume.

vi. ALCOHOL GRADES: 70%, 90% and 100%

vii. SODIUM CHLORIDE, (5M) :

NaCl	292 g
Distilled water up to	1 lit

The volume was brought to exactly 1 litre with distilled water and stored at room temperature.

viii. CHLOROFORM – ISOAMYL ALCOHOL (24:1)

A mixture was made where 1 part of isoamyl alcohol was mixed with 24 parts of chloroform and thoroughly mixed and stored at 4° C.

ix. DNA EXTRACTION BUFFER (for 100 ml)

1M Tris- HCl (pH 7.5)	10 ml
0.5 M NaCl	10 ml
0.5 M EDTA	10 ml

The volume was brought to exactly 100ml with milliQ, autoclaved water and stored at room temperature.

x. ETHIDUM BROMIDE (10 mg/ml):

Ethidium bromide	10 mg
Distilled water	1 ml

The solution was thoroughly mixed and the container was covered carefully with Aluminium foil to ensure that the solution was not exposed to any light. The solution was stored at room temperature.

xi. TAE (TRIS – ACETATAE – EDTA) BUFFER (20X)

Tris base	9.68 g
Glacial acetic acid	2.284 ml

0.5 M EDTA (pH 8.0)	4 ml
Distilled water up to	500 ml

The solution was thoroughly mixed and stored at room temperature**.**

3.1.4 Primers

Locus	Primer Name	Sequences (5' – 3')	Source
matK	*matK X F*	TAA TTT ACG ATC AAT TCA TTC	Ragupathy et al.,(2009)
	matK 5r	GTT CTA GCA CAA GAA AGT CG	
trnH-psbA	*trnH-F*	CGC GCA TGG TGG ATT CAC AAT CC	Ragupathy et al.,(2009)
	psbA-R	GTT ATG CAT GAA CGT AAT GCT C	
ITS	*ITS-5a*	CCT TAT CAT TTA GAG GAA GGA	Kress et al., (2005)
	ITS-4	TCC TCC GCT TAT TGA TAT GC	

Table- 3.2 List Primer used in this study

3.2 Methods

3.2.1 Isolation & purification of DNA from young leaf of Bamboo

1. 40 mg of the young leaf tissue sample were crush thoroughly in 600 µl of DNA Extraction Buffer (1 M Tris-Cl, pH-8, 5 M NaCl and 0.5 M EDTA pH-8) and taken in a sterile micro centrifuge tube.
2. Immediately 10% SDS and 2 µl of β- mercaptoethanol were added and incubated at 65° C for 40 mins. The tube was inverted every 10 min to ensure adequate mixing.
3. Add 200 µl of 5 M potassium acetate (pH-9) in micro centrifuge and keep at -20° C for 20 min.
4. Centrifuge at 12000 rpm for 15 mins
5. Top Aqueous phase is removed very carefully into a new centrifuge tube.
6. RNAse is added in 1 µl per 10 µl concentration and incubates at 37° C for 60 - 90 mins.
7. After incubation, equal volume of Phenol: Chloroform: Isoamylalchol: (25:24:1) was added carefully to the tube and shake the tube gently and centrifuge at 12000 rpm for 10 mins.
8. The upper aqueous phase was taken into a new centrifuge tube very carefully, keeping in mind not to disturb the debris in interphase.
9. Add equal volume of Chloroform: Isoamylalchol (24:1), shake the tube gently and repeat the centrifuge step.
10. Supernatant is taken into new centrifuge and ads double volume of chilled ethanol (absolute) and kept in -20° C for 2 hr for precipitation.
11. It is then centrifuged at 10000 rpm for 10 min.
12. The supernatant is discarded gently and the pellet is retained. To it, add 1ml 70% ethanol for washing the pellet and repeat the centrifugation step. Subsequently, the pellet is kept for air dry until smell of the alcohol was removed.

13. Dissolve in Nuclease free water or 1X TE for long term preservation and store in -86° C

3.2.2 Determination of the yield and purity of DNA

3.2.2.1 Spectrophotometric determination

The isolated DNA was dissolved in milliQ water to make a stock solution; it was further diluted to different concentration by adding milliQ water e.g. 200, 50 and 10 times the stock solution. The spectrophotometer was calibrated at 260 nm as well as 280 nm using 50 µl of dd water in a cuvette. The yield of DNA was checked by taking 48 µl of dd water in a cuvette to which was added 2 µl of each DNA sample and thoroughly mixed. Optical densities (OD) were measured at 260 (OD_{260}) and 280 (OD_{280}) in UV spectrophotometer (Biophotometer, Eppendrof) against sterile distilled water as blank. The yield and purity of DNA samples were estimated as follows:

Concentration of DNA stock solution (µg / ml) = OD_{260} X 100(dilution factor) X 50 µg/ml/1000

Purity of DNA stock solution = OD_{260}/OD_{280} (for pure DNA sample this ratio should be in the range of (1.75 – 1.80)

From the concentration of DNA stock solution, the total yield of DNA was calculated and recorded. This helped us to determine the isolated DNA samples molecular weight and quality.

3.2.2.2 Agarose gel electrophoresis For DNA quantification and quality analysis

Required amount of agarose (highly purified) was mixed with electrophoresis buffer (1X TAE), and heated on a boiling water bath or in a Microwave oven after sealing the mouth of the container and ensuring that all the agarose is completely dissolved. The agarose solution is the allowed to cool to about 50° C and poured onto a clean glass plate suitable for making Gel slab for Electrophoresis. A suitable plastic comb was added to make the sample loading wells. The comb's size was chosen based on the number of samples and amount of materials to be loaded. The gel takes about an hour at room temp, to properly set. Upon setting the comb was carefully removed and the gel was placed in the electrophoresis tank. Electrophoresis buffer (1X TAE) is carefully poured to completely cover the Gel upto a height of about 2 - 4 mm above the Gel, with or without EtBr (Sambrook et al. 2001).

The DNA sample (usually 0.5 to 2 µg) was mixed with desired amount of gel loading buffer to which EtBr is added. This can also be done without the EtBr, if desired. A micropipette is used to slowly load the mixture into the wells of the submerged gel, ensuring that the side of the well is not torn by the tip of the micropipette. A voltage of 5 – 6 V/cm (measured as distance the between the two electrodes) was applied. A voltage of 2- 3 V/cm was applied when there was a need of high resolution bands. The DNA bands were visually checked by placing the gel in ultraviolet light (UV-Transilluminator) and Gels were photographed using a Gel Documentation system (BioRad XR).

3.2.3 PCR amplification of Sample DNA

One set of forward and reverse primer was used for Internal Transcribed Spacers (*ITS*) amplification, *ITS* is nuclear DNA (Table- 3.2). One set of Forward and reverse primer were used to amplify the *matK* gene and one set of Forward and reverse primer were used for the amplification of *trnH-psbA intergenic* spacer of the cp DNA. Amplification was performed in thermal cycles (ABI system) in the following conditions.

3.2.3.1 Sample DNA PCR Reaction settings

Each 50 µl PCR reaction mixers contain:

Genomic DNA (100-200 ng)	: variable
dNTPs Mastermix (10 mM)	: 25µl
10X PCR Buffer	: 5 µl
Forward primer (20 pmole / µl)	: 1 µl

Reverse primer (20 pmole / µl)	: 1 µl
High fidelity DNA polymerase (5 Unit/ µl)	: 1 µl
Nuclease free water	: Up to 50 µl

3.2.3.2 Sample DNA-PCR cycling condition

(A) *ITS* amplification

The PCR reaction was set with an initial denaturation temperature of 94° C hot start for 3 min; and subsequently, 94° C for 1 min for denaturation, 42° C for 45 sec 72° C for 45 sec for extension primer annealing for 40 cycles followed by 72° C for 10 min for final extension.

(B) *matK* amplification

The PCR reaction was set with an initial denaturation temperature of 94° C hot start for 3 min and subsequently, 94° C for 1 min for denaturation, 46° C for 45 sec 72° C for 45 sec for extension primer annealing for 40 cycles followed by 72° C for 10 min for final extension using gradient thermal cycler (Applied Biosystem, Inc. USA).

(C) *trnH-psbA* amplification

The PCR reaction was set with an initial denaturation temperature of 94° C hot start for 3 min; and subsequently, 94° C for 1 min for denaturation, 51° C for 45 sec 72° C for 45 sec for extension primer annealing for 40 cycles followed by 72° C for 10 min for final extension.

Checking of the PCR products for correct amplification of target sequence was checked by taking aliquots for 10 µl of DNA PCR product and was loaded in 0.8-1.2% agarose gel for electrophoresis in 1X TAE buffer solution. Gel was stained with Ethidium bromide in this case. The gel checked under UV transilluminator and photographed using a Gel Documentation system (BioRed XR).

3.2.4 Purification of PCR product

The product generated from amplification cycles subjected to purification procedure, initially the amplification mix is resolved in LMA gel. A band of interest, which is a specific amplified amplicons for *matK*, *trnH-psbA* and *ITS* were excised and DNA was extracted by Bioline, Isolate PCR and Gel Kit (BIOLINE; Cat: BIO-52029). The procedure is given below:

1. The DNA fragments were excised from the agarose gel with a clean, sharp scalpel.
2. The gel slices were transferred to a 1.5 ml or 2.0 ml tube.
3. 650 µl Gel Solubilizer was added to the excised Gel slices in the same tube.
4. The tubes were incubated for 10 minutes at 50° C in a water bath. The time can be extended or shortened on checking whether the gel slices were completely dissolved or not.
5. 50 µl Binding Optimizer were added to the mixture.
6. 750 µl of the sample was transferred to a Spin Column, a placed in a 2ml Collection tube.
7. The spin columns along with the collection tube were centrifuged at 12,000 rpm for 1 minute. The filtrate is discarded and the collection tube is reused by placing the spin column back in the collection tube.
8. The residual solution is collected and loaded and the centrifugation step (6) is repeated.
9. 700 µl Wash Buffer A is added to the residual solution and centrifuged at 12,000 rpm for 1 minute.
10. The filtrate is discarded and the collection tube is reused by placing the Spin Column back in the collection tube.
11. Step 8 is repeated.
12. The Spin column along with the collection tube is centrifuged at maximum speed for 2 minutes to remove all traces of ethanol. The collection tube is discarded.
13. Spin Column A were placed into a 1.5 ml Elution Tube and to it 30-50µl Elution Buffer was added directly to the Spin Column membrane. The

elution tube is incubated at room temperature for 1 minute and centrifuged at 12,000 rpm for 1 minute to elute the DNA.

14. The isolated DNA was kept at -20° C for storage.

3.2.5 Sequencing of amplicons

Cycle sequencing was carried out in an automated DNA sequencer (ABI 3700 DNA Analyzer; Applied Biosystem, Inc. USA) employing 30 cycles at 96ºC for 10 sec, 50ºC for 5 sec and 60ºC for 4 min. The extended products were purified by alcohol precipitation followed by washing with 70% alcohol. Purified samples were dissolved in 10 µl of 50% Hi-Di formamide and analyzed in an ABI 13700 automated DNA Analyzer.

3.2.6. Bioinformatics analysis

3.2.6.1 Format of sequences

FASTA: A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line (defline) is distinguished from the sequence data by a greater-than (">") symbol at the beginning. It is recommended that all lines of text be shorter than 80 characters in length. Blank lines are not allowed in the middle of FASTA input. Sequences are expected to be represented in the standard IUB/IUPAC amino acid and nucleic acid codes, with these exceptions: lower-case letters are accepted and are mapped into upper-case; a single hyphen or dash can be used to represent a gap of indeterminate length; and in amino acid sequences, U and * are acceptable letters. (http://www.ncbi.nlm.nih.gov/blast/blastcgihelp.shtml)

PHYLIP: There are only two kinds of information provided in Phylip file format: Line 1 provides the number of Taxa and Characters in the matrix; Line 2 and subsequent lines provide data in the following rigid format: a Taxon identifier (up to 10 characters), followed by characters for that taxon. PHYLIP package uses this format as the standard input file format. This format is presently used by other

Bioinformatics tools as an accepted format. PHYLIP files usually have the “.phy” extension. (www.phylo.org/tools/obsolete/phylip.html)

CLUSTAL: Various programs in the MEME Suite allow as input a file containing a multiple alignment of protein or DNA sequences. These input files must be in CLUSTAL format (usually identified with the suffix ".aln"). The format is very simple:

1. The first line in the file must start with the words "CLUSTAL". Other information in the first line is ignored.
2. One or more empty lines.
3. One or more blocks of sequence data. Each block consists of:
    - One line for each sequence in the alignment. Each line consists of:
        1. The sequence name
        2. White space
        3. Up to 60 sequence symbols.
        4. Optional - white space followed by a cumulative count of residues for the sequences
    - A line showing the degree of conservation for the columns of the alignment in this block.
    - One or more empty lines.

Some rules about representing sequences:

- Case doesn't matter.
- Sequence symbols should be from a valid alphabet.
- Gaps are represented using hyphens ("-").
- The characters used to represent the degree of conservation are
- * -- All residues or nucleotides in that column are identical
- : -- Conserved substitutions have been observed
- . -- Semi-conserved substitutions have been observed

  -- No match.

(http://web.mit.edu/meme_v4.9.0/doc/clustalw-format.html)

MEGA: Input files for MEGA software usually have the file extension ".meg". "# Mega" at the beginning of the first line indicates that sequence data is ready for analysis. "Title" must be present in the second line, which can be followed by some description of data on the same line. Each taxon label starts on a new line starting with # as a prefix. The sequences can be formatted in both sequential and interleaved format. (http://www.megasoftware.net/mega4/WebHelp/helpfile.htm#part_iii___input_data_types_and_file_format/mega_input_data_formats/rh_mega_format.htm)

GenBank: Genbank is currently the standard in sequence file formats, the Genbank format is widely used by public databases such as NCBI. The Genbank file format is quite flexible and allows annotations, comments, and references to be included within the file. The file is plain text and thus can be read with a text editor. Genbank files often have the file extension '.gb' or '.genbank'. (http://www.algosome.com/articles/bioinformatics-sequence-file-formats.html)

3.2.6.2 Sequence search

BLAST: The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. Altschul et al. (1996) worked out the local alignment statistics and created the BLAST program (Altschul et al. 1990). The BLAST program along with its various flavours is available at (http://blast.ncbi.nlm.nih.gov/Blast.cgi). From the initial effort, a wide variety of BLAST programs are available, which is suitable for various tasks of choice. Programs like blastn, megablast, discontiguous megablast, blastp, psi-blast, phi-blast, delta-blast etc are available at the website mentioned above.

The *matK*, *psbA-trnH* and *ITS* sequences were first converted into FASTA format or Accession number of submitted sequence was put as a query in nucleotide BLAST (BLASTN). Since plant DNA was being queried the “others” database option was chosen along with the Megablast variety, as it optimized for searching sequences with high similarity. A BLAST variety called bl2seq was also used for the alignment two (or more) sequence and primer blast for primers.

GenBank database was searched using megablast during April-May 2014 with default search parameters. In most of the case, the program retrieved sequences with the high BLAST score. In many cases, the optimized match was sequences with high level of identity but with shortened sequence length. Ambiguous bases in target sequence were considered as matching.

3.2.6.3 Sequence Alignment

CLUSTAL: ClustalW and ClustalX are both freely available and were downloaded from the EMBL/EBI file server (*ftp://ftp.ebi.ac.uk/pub/software/*) or from ICGEB in Strasbourg, France (*ftp://ftp-igbmc.u strasbg.fr/pub/*ClustalW/ and *ftp://ftp-igbmc.u-strasbg.fr/pub/*ClustalX/). ClustalX provides a graphical user interface with colourful display of alignments and CLUSTALW provides a simple text interface (ClustalW) suitable for high- throughput tasks. CLUSTALW (Thompson et al. 1994) and CLUSTALX (Thompson et al. 1997) are the most commonly used programs for progressive alignment of sequences. ClustalX and ClustalW takes input files in FASTA format and automatically iterate the entire progressive alignment procedure. The sequences are initially aligned pairwise which in turn generates a distance matrix which is subsequently used to make a simple initial tree of the sequence. Finally, the multiple sequence alignment is carried out using the progressive approach.

After starting the ClustalX program, the sequence file is loaded using File → Load Sequences. The graphical interface allows selection and viewing of the unaligned sequences. “Do complete Alignment” option is selected from the Alignment menu and started the computing process. ClustalX performs the progressive alignment and automatically creates an output guide tree file and an output alignment file in the default Clustal format if not mentioned otherwise before the program is run. ClustalX also allows variations in the alignment parameters (from Alignment Parameters in the Alignment menu). The alignment may show too many large gaps, which indicates wrong parameter selection. The gap-opening penalty and/or gap extension penalty can be increased and the alignment rerun to get optimized result. ClustalX indicates the degree of conservation at the bottom of the aligned sequences, which is very useful for evaluating a given alignment.

3.2.6.4 Sequence editing

BIOEDIT: BioEdit is a graphic, easy-to-use sequence alignment editor and sequence analysis program. BioEdit is intended to supply a single program that can handle most simple sequence and alignment editing and manipulation functions that researchers are likely to do on a daily basis, as well as a few basic sequences analyses. (http://www.mbio.ncsu.edu/bioedit/page2.html)

Sequence Manipulations Suite 2: The Sequence Manipulation Suite is a collection of JavaScript programs for generating, formatting, and analyzing short DNA and protein sequences. It is commonly used by molecular biologists, for teaching purposes, and for program and algorithm testing. The program suite can be easily accessed online at http://www.bioinformatics.org/sms2/

The Sequence Manipulation Suite is written in JavaScript 1.5, which is a lightweight, cross-platform, object-oriented scripting language. JavaScript is now standardized by the ECMA (European Computer Manufacturers Association) and the ISO (International Organization for Standards). Each program in the Sequence Manipulation Suite generates HTML output. It can print and save the output, and can be edited using an HTML editor or a text editor. The Sequence Manipulation Suite was written by Paul Stothard (University of Alberta, Canada). Short descriptions of the programs which were used in this study are given below and have been derived from the official website.

Combine FASTA - converts multiple FASTA sequence records into a single sequence. Combine FASTA can be used to determine the codon usage for a collection of sequences using a program that accepts a single sequence as input.

EMBL to FASTA - accepts an EMBL file as input and returns the entire DNA sequence in FASTA format. Using this program can remove all of the non-DNA sequence information from an EMBL file.

EMBL Feature Extractor - accepts an EMBL file as input and reads the sequence feature information described in the feature table. The program extracts or highlights the relevant sequence segments and returns each sequence feature in FASTA format. EMBL Feature Extractor is particularly helpful when deriving the sequence of a cDNA from a genomic sequence that contains many introns.

EMBL Trans Extractor - accepts an EMBL file as input and returns each of the protein translations described in the file in FASTA format. EMBL Trans Extractor can be used when the interest is the predicted protein translations of a DNA sequence than the DNA sequence itself.

Filter DNA - removes non-DNA characters from text. Using this program can remove digits and blank spaces from a sequence to make it suitable for other applications.

Filter Protein - removes non-protein characters from text. Using this program can remove digits and blank spaces from a sequence to make it suitable for other applications.

GenBank to FASTA - accepts a GenBank file as input and returns the entire DNA sequence in FASTA format. This program is particularly useful to quickly remove all of the non-DNA sequence information from a GenBank file.

One to Three - converts single letter translations to three letter translations.

Range Extractor - accepts a DNA sequence along with a set of positions or ranges. The bases corresponding to the positions or ranges are returned, either as a single new sequence, a set of FASTA records, or as uppercase text inside of the original lowercase sequence. Range Extractor is used to obtain sub sequences using position information.

Reverse Complement - converts a DNA sequence into its reverse, complement, or reverse-complement counterpart. The entire IUPAC DNA alphabet is supported, and the case of each input sequence character is maintained. You may want to work with the reverse-complement of a sequence if it contains an ORF on the reverse strand.

Multiple Align Conservation - accepts a group of aligned sequences (in FASTA or GDE format) and colours the alignment. The program examines each residue and compares it to the other residues in the same column. Residues that are identical among the sequences are given a black background, and those that are similar among the sequences are given a grey background. The remaining residues receive a white background. The percentage of residues, can be specified, that must be identical and similar for the colouring to be applied. It is useful for enhancing the output of sequence alignment programs.

Primer Map - accepts a DNA sequence and returns a textual map showing the annealing positions of PCR primers. Restriction endonuclease cut sites, and the protein translations of the DNA sequence can also be shown.

Restriction Map - accepts a DNA sequence and returns a textual map showing the positions of restriction endonuclease cut sites. The translation of the DNA sequence is also given, in the reading frame you specify. The output of this program can be used as a reference when planning cloning strategies. Restriction Map uses the standard genetic code and supports the entire IUPAC alphabet.

Translation Map - accepts a DNA sequence and returns a textual map displaying protein translations. The reading frame of the translation can be specified (1, 2, 3, or all three), or it can choose to treat uppercase text as the reading frame. Translation Map uses the standard genetic code and supports the entire IUPAC alphabet.

3.2.6.5 Pre-sequence Analysis

Applied Biosystems Sequence Scanner v1.0 (Applied Biosystem, Inc. USA) was used to assemble the Trace files. Sequence with greater than 2% ambiguous bases was discarded, using quality value of 40 for bidirectional reading of sequences. Manual editing was necessary in many of raw traces and subsequent alignments of forward and reverse sequences ensured the assignment of sequences for most species. In case of ambiguity, both the sequences were thoroughly checked and quality values of the sequences were considered while determining the most likely nucleotide. To generate consensus sequences for each sample, the 3' and 5' terminals were clipped. The sequences was checked in NCBI through BLASTN to examine the complete alignment with the partial coding sequence of chloroplast *matK* gene, *ITS* and *trnH-psbA*. The coding sequences were translated using the online software ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and aligned through BLASTP (Altschul et al. 1990). The result was used to determine whether the partial amino acid codes were similar to the chloroplast *matK* gene frame and presence of any stop codon in middle of the sequence. The sequences generated consensus was confirmed to be the desired loci. All the analyzed sequences were then deposited in GenBank (Mahadani et al. 2012).

3.2.6.6 Primer design

OligoCalc: Oligo Calc is a free online calculator to obtain properties of single stranded and double stranded DNA or RNA sequences. These properties include oligonucleotide melting temperature, sequence molecular weight, %GC content of the entered sequence and absorbance coefficients. (http://simgene.com/OligoCalc)

PCR Products: PCR Products accepts one or more DNA sequence templates and two primer sequences. The program searches for perfectly matching primer annealing sites that can generate a PCR product. Any resulting products are sorted by size, their position in the original sequence, and the primers that produced them. (http://www.bioinformatics.org/sms2/pcr_products.html)

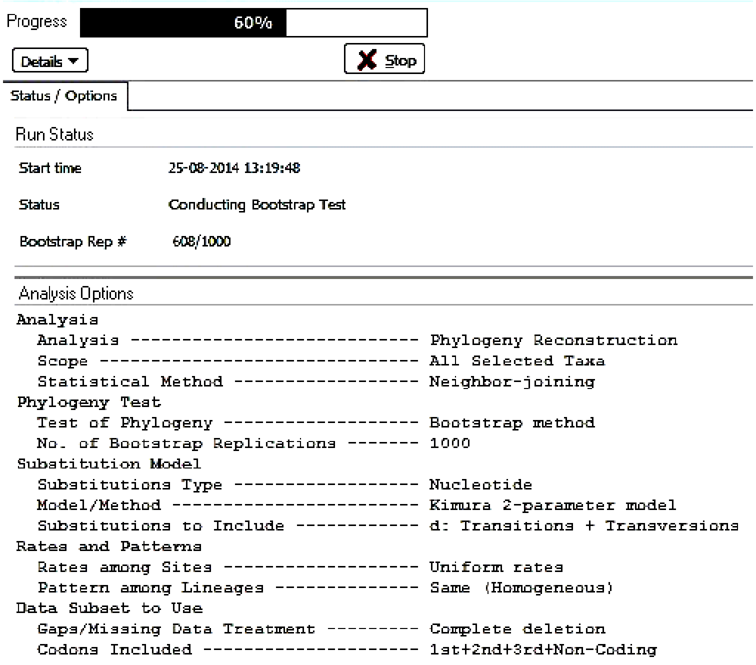
3.2.7 Phylogenetic analysis

The phylogenetic studies were performed using the molecular evolutionary genetic analysis (MEGA6) software in accordance with the Kimura 2-Parameter (K2P) model. DNA sequences were analyzed by using the phylogenetic tree reconstruction methods such as Neighbor-joining (NJ) which is a heuristic method for estimating the minimum evolution tree originally developed by Saitou et al. (1987) and modified by Studier et al. (1988). The Kimura model is an extension of the Jukes and Cantor (JC) basic model (Saitou et al. 1987; Studier et al. 1988). This model distinguishes between two types of substitutions: transitions, where a purine is replaced by another purine (A<-->G) or a pyrimidine is replaced by another pyrimidine (C<-->T), and transversions, where a purine is replaced by a pyrimidine or vice versa (A or G <--> C or T). The model assumes that the rate of transitions is different from the rate of transversions. For the species-level analysis, nucleotide sequence divergences were calculated using the Kimura-2-Parameter (K2P) model, this is the best parameter when distances were low as in DNA barcode sequence.

3.2.7.1 MEGA 6

MEGA 6 also was use for Phylogenetic tree construction and inference. The input file for MEGA is '.meg' format, which can be easily generated from FASTA files. The file contains aligned DNA sequences in MEGA format. Using MEGA6, it was possible to estimate a NJ tree and performed the bootstrap test in an automated fashion. The program displayed the tree in a new window and superimposed bootstrap support values along each branch of the tree. To estimate a NJ tree using K2P corrected distances and performed bootstrap analysis on 1000 replicates, The alignment file is input in MEGA6 and the submenu Bootstrap Test of Phylogeny > Neighbor-Joining is selected from the Phylogeny menu in the MEGA6 main window. The Analysis Preferences window will drop down. The green square to the right of the Gaps/missing data row and selected pair wise deletion (specifying that for each pair of sequences only gaps in the two sequences being compared), is selected. Similarly, Model row Nucleotide >Kimura-2-parameter is also selected. The number of bootstrap replicates is set at 1000 in the Test of Phylogeny tab on the top of the window. The NJ tree with bootstrap values is given as an output in the Tree Explorer window. By default, the tree is midpoint rooted. The location of the root can be placed on any other branch of the tree. The distance matrices were then checked for level of divergence in intra- and inter- species level. A value of 0.000 indicates that two sequences are identical or very similar and will not be helpful in diversity or phylogenetic analysis.

The phylogenetic tree generated by NJ method with 1000 bootstrap replicates is a very robust method of visualizing the phylogenetic relationship among different species, genus or taxa. The tree is checked to see if any species is misplaced or there is wrong formation of clades. The detail of parameters used for construction of Phylogenetic tree by Neighbour-Joining method is given below.



Run Status

Start time 25-08-2014 13:19:48

Status Conducting Bootstrap Test

Bootstrap Rep # 608/1000

Analysis Options

```
Analysis
  Analysis ---------------------------- Phylogeny Reconstruction
  Scope ------------------------------- All Selected Taxa
  Statistical Method ------------------ Neighbor-joining
Phylogeny Test
  Test of Phylogeny ------------------- Bootstrap method
  No. of Bootstrap Replications ------- 1000
Substitution Model
  Substitutions Type ------------------ Nucleotide
  Model/Method ------------------------ Kimura 2-parameter model
  Substitutions to Include ------------ d: Transitions + Transversions
Rates and Patterns
  Rates among Sites ------------------- Uniform rates
  Pattern among Lineages -------------- Same (Homogeneous)
Data Subset to Use
  Gaps/Missing Data Treatment --------- Complete deletion
  Codons Included --------------------- 1st+2nd+3rd+Non-Coding
```

Fig 3.2: Parameters used for constructing Phylogenetic Tree using NJ method.

3.2.7.2 PHYLIP

Joe Felsenstein is the creator of PHYLIP and is a General Public License software package and is composed of a large number of programs, which is useful, capable and very accurate.

Preparing input data:

PHYLIP software takes input of data specifically from a file named “infile”. The first line contained the number of species, the number of characters and the program options. The species and character data are next presented to the software, in separate lines. The line contains 10 letters or symbols reserved for species nomenclature and followed by data to be analyzed. These characteristics are built in into the software for ease of use with other programs like, e.g. *Clustal* or *TreeAlign.* When run, each program was display about its role, its version and a menu.

Phylogeny Construction:

DNADIST builds a matrics using either of Jukes-Cantor, Kimura, Jin-Nei or maximum likelihood method.

Displaying results: Various options available.

RETREE allows the interactive construction and manipulation of trees (topology, branch lengths, labels etc).

DRAWGRAM is capable of displaying a rooted tree as a cladogram or phenogram.

DRAWTREE is capable of displaying *unrooted* phylogenies on a variety of output devices.

Bootstrapping options:

SEQBOOT allows resampling of the data sets by any of the options, like, bootstrap, jackknife or permutation methods.

Bootstrap analysis is carried out as follows:

1) Run SEQBOOT on the input dataset after selecting a shuffling method and specified Bootstrap value of 1000 replicates.

2) Phylogeny analysis program is dependent on the data set and type, analysis and method with variable parameters. Multiple data sets can be analyzed easily by selecting the option “M” and feeding the parameter for number of desired replicates to be generated with SEQBOOT.

3) Distance matrices were generated as outfile in step 2, is renamed as infile (which is the default input file) for the various Distance Matrix programs viz. FITCH, KITSCH or NEIGHBOR. Distance Matrix programs now generates the trees.

4) The output treefile is again renamed as “infile” which becomes the input file for CONSENSE, a subprogram capable of evaluating the significance of the analysis.

3.2.8. Data presentation

The data generated by the various softwares is predominantly a simple flat file or text file, which ensures that the files are much smaller in size, easily portable and can be read across cross platform operating systems like Windows, Linux, Sun Solaris or Mac. But the numbers without visual representation does not make much sense at first glance.

To ensure easy understanding of immense amount of data various programs now come with visual output. For example CLUSTALW is Command Line Interface whereas CLUSTALX is a Windows/ visual based program. In molecular phylogenetic trees, branch lengths were almost always drawn to scale; that was, proportional to the amount of evolution estimated to have occurred along them. Although the relationship between branch lengths and real time was far from straightforward and probably unreliable for any single gene, lengths still gives a good general impression of relative rates of change across a tree. Bootstrap values were displayed as percentages on each branch (Mahadani et al. 2012).