CHAPTER 3
MATERIALS AND METHODS

CHAPTER 3

MATERIALS AND METHODS

3.1 Samples and sampling site:

Northeast India comprises the states of Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura. These are a part of both Himalaya as well as Indo-Burma biodiversity hotspot. It forms a unique biogeography province encompassing major biomes recognized in the world. It has the richest reservoir of plant diversity in India; about 50% of India's biodiversity. Southern Assam, popularly known as "Barak Valley'', is constituted of three districts *viz.* Cachar, Karimganj and Hailakandi. It is bounded by hills on three sides: Mizoram on the South, N. C. Hills of the North, Manipur on the East and Tripura and Bangladesh on the West. The valley is named after the name of its main river 'Barak' which originates from the Naga and Manipur Hill ranges. The river traverses nearly 160 Km in East-West direction through the central part of the valley and then bifurcates into Surma and Kushiyara near Haritikar of Karimganj district. Southern Assam is situated within 24^0 50' N latitude and 92^0 51' E longitude and covers a total area of about 6951 sq. km (Cachar-3786 sq. Km, Karimganj-1839 sq. Km and Hailakandi-1326 sq. Km). The most part of floristically rich areas of the valley remained unexplored or underexplored at that time probably because of its peculiar political and geographical location along with truncated communication system.

Young leaves of important medicinal plants belonging to 13 different taxa (Table- 3.1) were collected aseptically from different areas of Southern Assam (24°8'N -25°8'N and 92°15'E -93°15'E), India (Figure- 3.1). All the specimens were examined and tentatively identified following the standard literature (Flora of Assam). Voucher specimens were deposited in herbarium of Biotechnology Department, Assam University. Also, 7 different ethnomedicine samples (100g) fresh leaves and 200ml leaf juices) were

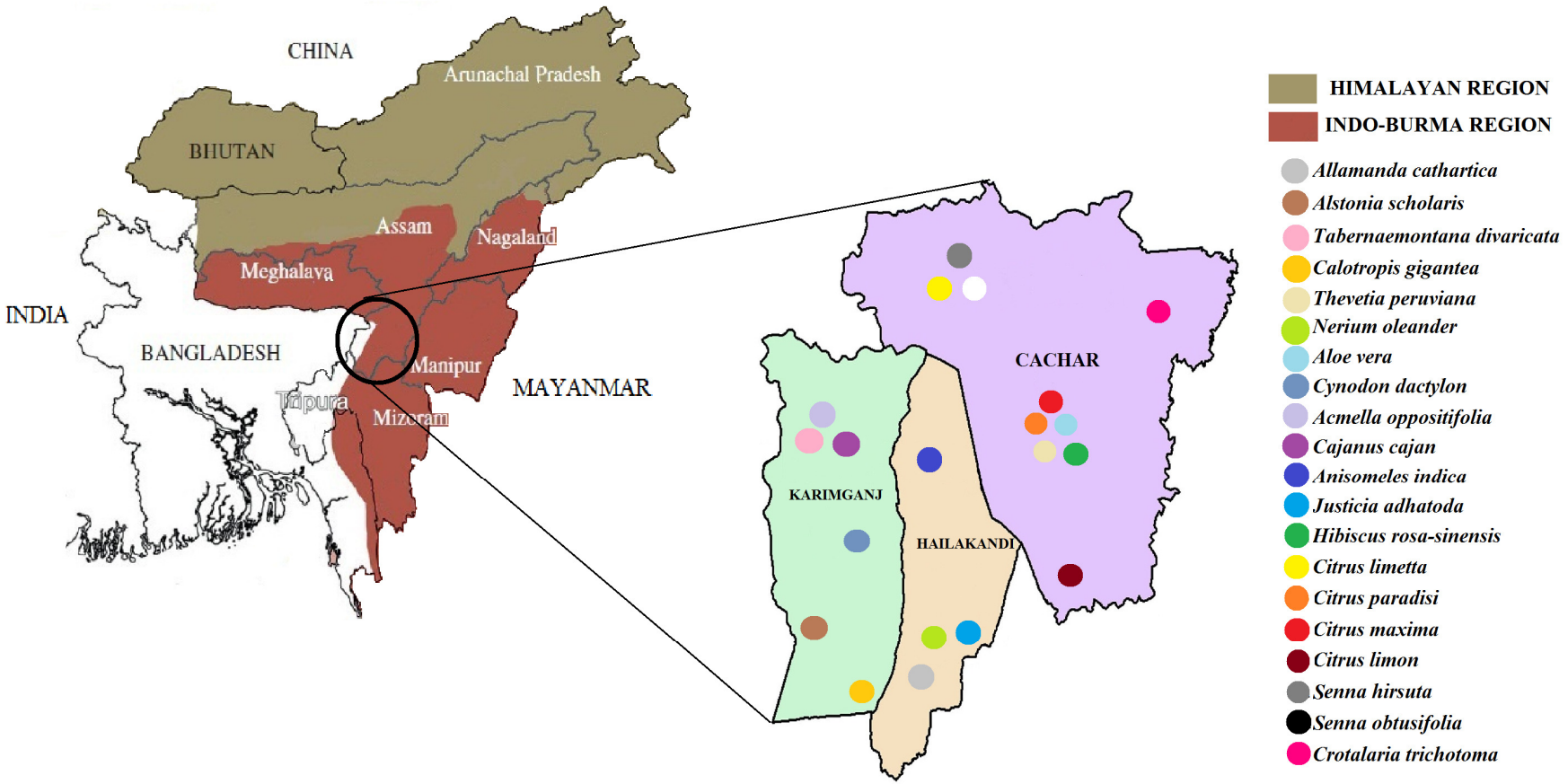

Figure- 3.1 Collection sites of medicinal plants from different district of southern Assam

purchased from herbalists. As leaves and juices were unidentifiable by appearance, we recorded the common vernacular name in local language (Bengali/Assames) and assigned the sample ID to each material. Here, we were avoiding the admixture ethnomedicines samples. Juice and leaf samples were preserved in -80^0 C for re-investigation.

3.2 Chemicals:

All chemicals used in this study were of molecular biology grade. Sucrose, Bromophenol blue, Ethidium bromide, Tris-Base, EDTA (Promega, USA) Magnesium chloride, Sodium chloride, Sodium bicarbonate, Potassium accetate, 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.

3.2.1 Buffer and reagents

i. EDTA (0.5 M, pH 8.0)

EDTA	18.61 g
Distilled water up to	100 ml

The pH is adjusted to 8.0 by NaOH pellets before making to final volume.

ii. SODIUM CHLORIDE (5M)

NaCl	292 g
Distilled water up to	1 lit

Adjusted the volume up to 1liter with distilled water and stored at room temperature.

iii. SODIUM DODECYL SULPHATE (SDS) 20%

SDS	20 gm
Distilled water up to	100 ml

Mixed properly and stored at room temperature.

Sr. No	Scientific Name	Family	Sample ID	Number of Sample
1	***Catharanthus roseus***	Apocynaceae	AUMP03, AUMP36	2
2	***Allamanda cathartica***	Apocynaceae	AUMP29, AUMP33	2
3	***Alstonia scholaris***	Apocynaceae	AUMP5	1
4	***Tabernaemontana divaricata***	Apocynaceae	AUMP32	1
5	***Calotropis gigantea***	Apocynaceae	AUMP22	1
6	***Thevetia peruviana***	Apocynaceae	AUMP1 AUMP92	2
7	***Nerium oleander***	Apocynaceae	AUMP34	1
8	***Aloe vera***	Liliaceae	AUMP16	1
9	***Cynodon dactylon***	Poaceae	AUMP28	1
10	***Acmella oppositifolia***	Asteraceae	AUMP21	1
11	***Cajanus cajan***	Papilinaceae	AUMP2	1
	Anisomeles indica	Lamiaceae	AUMP73	1
12	***Justicia adhatoda***	Acanthaceae	AUMP24	1
13	***Hibiscus rosa-sinensis***	Malvaceae	AUMP4	1
14	***Citrus limetta***	Rutaceae	AUMPT51 AUMP54	2
15	***Citrus x paradisi***	Rutaceae	AUMP49 AUMP55	2
16	***Citrus maxima***	Rutaceae	AUMP45 AUCM3	2
17	***Citrus limon***	Rutaceae	AUMP40 AUMP46	2
18	***Senna hirsuta***	Fabaceae	AUMP61	1
19	***Senna obtusifolia***	Fabaceae	AUMP63	1
20	***Crotalaria trichotoma***	Fabaceae	AUMP68	1

Table- 3.1 List source materials used in this study, with their scientific name (according to Flora of Assam), family and number of sample also given.

iv. POTASSIUM ACETATE (5 M, pH 9)

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

pH is adjusted to 9 using glacial acetic acid.

v. TRIS-HCL (1M, pH 8)

Tris	12.11 gm
HCL	7 ml

Distilled water up to the volume 100 ml and adjusted the pH with NaOH pellets.

vi. ETHIDUM BROMIDE (10 mg/ml)

Ethidium bromide	10 mg
Distilled water	1 ml

Mixed properly and wrap the container in Aluminum foil and store at room temperature.

vii. ALCOHOL GRADES 100%, 90%, 70%

viii. DNA EXTRACTION BUFFER (for 100 ml)

1M Tris- HCl (pH-8)	10 ml
0.5 M EDTA (pH-8)	10 ml
0.5 M NaCl	10 ml

Made the volume to 100ml with milli Q water, autoclaved and stored at room temperature.

ix. GEL LOADING DYE (6X)

Bromophenol Blue	0.2%
Xylene Cyanol	2%
Glycerol	25%
Tris-HCl, pH 8.0	50 mM
EDTA	5 mM

This loading dye was used for non-denaturing ds DNA

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

Tris base	9.68 gm
Glacial acetic acid	2.284 ml
0.5 M EDTA (pH 8.0)	4 ml
Distilled water up to	500 ml

Mixed properly and stored at room temperature.

xi. TRIS SATURATED PHENOL (pH 7.5)

Solid phenol was liquefied at 68° C in water bath and equal volume of 1 M Tris HCL (pH 7.5) was added. Discarded the upper aqueous phase and repeated the extraction procedure until the pH of the aqueous phase became 7.5. After final extraction an equal volume of TE was added to it and stored in a bottle wrapped with aluminum foil at 4° C.

xii. CHLOROFORM – ISOAMYL ALCOHOL (24:1)

24 parts of chloroform were added with 1 part of isoamyl alcohol and mixed properly and stored at 4° C.

3.2.2 Primers:

Plastid locus	Primer Name	Sequences	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**	
	RUT F	**TCA GAG GTA TTT GCT GCT GTG GTG**	Newly designed for this study
	RUT R	**GAC CAA GTC GAC CTA CTG ATA GG**	
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**	

Table- 3.2 List of primers used in this study.

3.3 Methods

3.3.1 Isolation & purification of DNA from plant materials

3.3.1.1 Isolation of DNA from young leaves:

1. 40 mg of the young leaves 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 65^0 C for 40 mins. The tube was inverted at every 10 mins interval to ensure adequate mixing.
3. 200 µl of 5 M potassium acetate (pH-9) was added in micro centrifuge and kept -20^o C for 20 min.
4. Centrifuged at 12000 rpm for 15 mins
5. Top Aqueous phase was removed very carefully into a new centrifuge tube.
6. RNase was added in 1 µl per 10 µl concentrations and incubated at 37^0 C for 1-1.30 hour.
7. After incubation, equal volume of Phenol:Chloroform:Isoamylalchol: (25:24:1) was added carefully to the tube and shook the tube gently and centrifuged 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 of interphase.
9. Equal volume of Chloroform: Isoamylalchol (24:1) was added, shooke the tube gently and centrifuged.
10. Supernatant was taken into new centrifuge and added double volume of chilled ethanol (absolute) and kept in -20^0 C for 2 hour for precipitation.
11. It was then centrifuged at 10000 rpm for 10 min.
12. The supernatant was discarded gently and the pellet was retained. To it, added 1ml 70% ethanol for washing the pellet and centrifuged step. Subsequently, the pellet was kept for air dry until smell of the alcohol was removed.

13. Dissolved in Nuclease free water or 1X TE for long term preservation and stored in -86^0 C.

3.3.1.2 Isolation of DNA from leaves juice:

The extraction method was based on Ng et al., (2006) with slight modifications.

1. 3 mL of juice was centrifuged 14 000 g for 1 min and the supernatant was discarded.
2. 1 ml of DNA extraction, 20 µL / mL β- mercaptoethanol, 10% SDS were added into precipitation and kept in water-bath under 65^0 C for 1 hour.
3. Later, centrifuged at 14 000 rpm for 10 min
4. Supernatant was collected and mixed well with Potassium acetate (pH-9) and kept it in -20^0 C for 30 min.
5. Equal volume of PCI (Phenol:Chloroform:Isoamyl alcohol, 25:24:1) was added and mixed well, then centrifuged at 14 000 rpm for 10 min.
6. 300 µl of supernatant was collected and equal volume of isopropanol was added and left under -20^0 C for overnight, later, followed by centrifugation at 14 000 rpm, for 10 min.
7. Precipitate was washed twice using 70% alcohol and air dried.

3.3.2 Determination of the yield and purity of DNA

3.3.2.1 Spectrophotometric determination:

The isolated DNA stock solution was taken at different dilution rate *viz*. 200, 50 and 10 times by adding milli Q water. 50 µl of dd water was taken in a cuvette and calibrated the spectrophotometer at 260 nm as well as 280 nm. 2 µl of each DNA sample was added to 48 µl of dd water in a cuvette and mix well. 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 must 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. In order to determine the quality and molecular weight of isolated DNA.

3.3.2.2 Agarose gel electrophoresis for DNA quantification and quality analysis:

Required amount of agarose (highly purified) was mixed with electrophoresis buffer (1× TAE), and heated on a boiling water bath to allow all of the grains of agarose to dissolve, the resulting solution was cooled to about 50° C and poured on a clean glass plate after setting with a suitable comb. After the gel combs completely set (30 – 40 min at room temp), the comb was removed carefully, and the gel was placed in the electrophoresis tank, and covered with electrophoresis buffer (1 × TAE) to a depth of all out 2 -4 mm with or without EtBr (Sambrook and Russell 2001).

The DNA samples (usually 0.5 to 2 µg) was mixed with desired amount of gel loading buffer, followed by slow loading into the slots of the submerged gel. Usually a voltage of 5 – 6 V/cm (measured as distance the between the two electrodes) was applied. For better resolutions, sometimes a voltage of 2- 3 V/cm was applied. The DNA bands were visualized directly by ultraviolet light (UV-Transilluminator) and gels photograph were captured by using a Gel Documented system (BioRed XR).

3.3.3 PCR amplification of Chloroplast DNA:

Two sets 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 (Table- 3.2). Amplification was performed in thermal cycles (ABI system) in the following conditions.

3.3.3.1 Chloroplast 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

3.3.3.2 Chloroplast DNA-PCR cycling condition:

***matK* amplification**

The PCR reaction was set with an initial denaturation temperature of 94^0 C host start for 3 min and subsequently, 94^0 C for 1 min for denaturation, 46^0 C for 45 sec 72^0 C for 45 sec for extension primer annealing for 40 cycles followed by 72^0 C for 10 min for final extension using gradient thermal cycler (Applied Biosystem, Inc. USA).

***trnH-psbA* amplification**

The PCR reaction was set with an initial denaturation temperature of 94^0 C host start for 3 min; and subsequently, 94^0 C for 1 min for denaturation, 51^0 C for 45 sec 72^0 C for 45 sec for extension primer annealing for 40 cycles followed by 72^0 C for 10 min for final extension.

Aliquots for 10 µl of DNA products from PCR amplification were loaded in 0.8-1.2% agarose gel for electrophoresis in 1 x TAE. Gel was stained with ethidium bromide and observed under UV transillumior.

3.3.4 Purification of PCR product:

The product generated from amplification cycles subjected to purification procedure, initially the amplification mix was resolved in LMA gel. A band of interest, which was a **s**pecific amplified amplicons for *matK* and *trnH-psbA*,

were excised and DNA was extracted by Bioline, Isolate PCR and Gel Kit (BIOLINE; Cat: BIO-52029). A procedure is given below.

1. The DNA fragments were excised from the agarose gel with a clean, sharp scalpel.
2. Transferred the gel slice to a 1.5 ml or 2.0 ml tube.
3. Added 650 µl Gel Solubilizer.
4. Incubated the tube for 10 minutes at 50º C in a water bath until the gel slices were completely dissolved.
4. 50 µl Binding Optimizer were added to the sample.
5. Transferred 750 µl of the sample to Spin Column, a placed in a 2ml Collection tube.
6. Centrifuged at 12,000 rpm for 1 minute. Discarded the filtrate and reused the collection tube by placing the spin column back in the collection tube.
7. Loaded the residual solution and repeat the centrifugation step.
8. Added 700 µl Wash Buffer A and centrifuged at 12,000 rpm for 1 minute.
9. Discarded the filtrate and reused the collection tube by placing the Spin Column back in the collection tube.
10. Repeated the step 8.
11. Centrifuged at maximum speed for 2 minutes to remove all traces of ethanol and discarded the collection tube.
12. Spin Column A were placed into a 1.5 ml Elution Tube and added 30-50µl Elution Buffer directly to the Spin Column membrane, Incubated at room temperature for 1 minute and Centrifuged at 12,000 rpm for 1 minute to elute the DNA.
13. The isolated DNA was kept at -20º C for storage.

3.3.5 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 96˚ C for 10 sec, 50˚ C for 5 sec and 60˚ C for 4 min. 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 3700 automated DNA Analyzer.

3.3.6 Bioinformatics analysis

3.3.6.1 Format of sequences:

FASTA - The format begins with single line sequence description proceed by ">". On the next line, the actual sequence is represented in the standard IUB\IUPAC amino acid or nucleic acid code. Most of the programs can read or import fasta format.

PHYLIP - PHYLIP is the standard input file format for programs in the PHYLIP package and subsequently been implemented as a input file format for many other programs. The first line of the input file contains the number of taxa and the number of characters (in this case alignment sites), separated by blanks. PHYLIP files frequently have the ".phy" extension.

CLUSTAL - CLUSTAL is not widely supported as input format in phylogenetic programs but since it is the standard output format of popular alignment software. The format is recognized by the word CLUSTAL at the beginning of the file. The sequence alignment output from CLUSTAL software is usually gives the default extension ".aln". CLUSTAL also indicates conserved residues in the alignment using a "*" for each block.

MEGA - "# Mega" keyword indicates that the data is prepared for analysis using MEGA. It must present on the very fast line in the data file. On the second line, the word "Title" must be written, which can be followed by some description of data on the same line. Each taxon label must be written on a new line starting with #. The sequences can be formatted in both sequential and interleaved format.

GenBank -This format consists of a mixture of compressed and uncompressed ASCII text files, containing sequence data and indices that cross reference author names, journal citations, gene names and keywords to individual GenBank records. For convenience, the GenBank records are

partitioned into divisions according to source organism or type of sequence. Records within the same division are packaged as a set of numbered files so that records from a single division may be contained in a series of many files.

3.3.6.2 Sequence search:

BLAST - Bioinformaticians have developed so called 'heuristic' algorithms, which allow searching a database in considerably less time. The most popular one is Basic Local Alignment Search Tool (BLAST). The Percentage similarity of the resulting DNA or protein sequences was analyzed through BLAST (Altschul et al., 1997); http://www.ncbi.nlm.nih.gov/blast/), a choice is offered between the different BLAST programs through different hyperlinks (nucleotide blast, protein blast, blastx, tblastn, tblastx etc). In this study, we were searched nucleotide (nucleotide blast) and protein sequences (protein blast).

matK and *trnH-psbA intergenic* sequences were used in FASTA format or Accession number as query in nucleotide BLAST (BLASTN). We specified the database because our sequences were from chloroplast DNA. So, we choice the "others" database option. We also selected megablast under the program selection header, which optimized the search for highly similar sequences and clicked on the Blast button to initiate the search. The output of the megabalst search contains a table with sequences producing significant alignments. We also used specialized BLAST in bl2seq for the alignment two (or more) sequence and primer blast to make specific primer.

GenBank database was searched using megablast during November-December 2011 with default parameter adjusted to retrieving 5000 sequences. In most of the case, this corresponded to the sequence with the high BLAST score. In other cases, the closest match was a shorter target with a higher percent identity. Ambiguous bases in target sequence were considered as matching.

3.3.6.3 Sequence Alignment:

CLUSTAL- The most commonly used software for progressive alignment is ClustalW (Thompson et al., 1994) and ClustalX (Thompson et al., 1997). These programs are identical to each other in term of alignment method but offer either a simple text interface (ClustalW) suitable for high- throughput tasks or a graphical interface (ClustalX). ClustalX and ClustalW were taken a set of input sequences and carry out the entire progressive alignment procedure automatically. The sequences were aligned in pairs in order to generate a distance matrix that was used to make a simple initial tree of the sequence. Finally, the multiple sequence alignment is carried out using the progressive approach.

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/). In each case, ClustalX (X stands for X windows) provides a graphical user interface with colorful display of alignments. Open ClustalX and open the sequence file using File → Load Sequences. The graphical display allowd to slide over the unaligned sequences. Selected to Do complete Alignment from the Alignment menu. ClustalX performed the progressive alignment (progress were follow up in the lower left corner), and created an output guide tree file and an output alignment file in the default Clustal format. It was, however, possible to choose a different format in the Output Format Options from the Alignment menu. ClustalX also allows the user to change the alignment parameters (from Alignment Parameters in the Alignment menu). If an alignment showed, too many large gaps, increased the gap-opening penalty and redo the alignment. ClustalX indicated the degree of conservation at the bottom of the aligned sequences, which were used for evaluate a given alignment.

3.3.6.4 Sequence editing:

BioEdit - BioEdit is a mouse-driven, 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 (Hall 1999).

Sequence Manipulations Suit - The Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/) 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). The first version of the ECMA standard is documented in the ECMA-262 specification. The ECMA-262 standard is also approved by the ISO (International Organization for Standards) as ISO-16262. JavaScript 1.5 is fully compatible with ECMA-262, Edition 3. Sequences submitted to the Sequence Manipulation Suite instead manipulated by the web browser, which executes the JavaScript. The Sequence Manipulation Suite was written by Paul Stothard (University of Alberta, Canada). Short descriptions of the programs which were used in this study:

Reverse Complement - converted a DNA sequence into its reverse, complement, or reverse-complement counterpart. The entire IUPAC DNA alphabet was supported, and the case of each input sequence character was maintained.

ORF Finder - searched for open reading frames (ORFs) in the DNA sequence. The program returned the range of each ORF, along with its protein translation. ORF Finder supports the entire IUPAC alphabet and several genetic codes. Here, bacterial genetic code was selected in chloroplast *matK* sequences. ORF Finder was used to search newly sequenced DNA for potential protein encoding segments.

Pairwise Align DNA - accepted two DNA sequences and determined the optimal global alignment. Pairwise Align DNA was used to look for conserved sequence regions.

3.3.6.5 Pre-sequence Analysis:

Trace file were assembled in Applied Biosystems Sequence Scanner v1.0 (Applied Biosystem, Inc. USA) and sequence with greater than 2% ambiguous bases were discarded, using quality Value of 40 for bidirectional reads. Manual editing of raw traces and subsequent alignments of forward and reverse sequences enabled us to assign edited sequences for most species. In some cases of discrepancy, both the sequences were reviewed and quality values of the sequences were considered to determine the most likely nucleotide. Then the 3' and 5' terminals were clipped to generate consensus sequences for each sample.

Finally, each of the sequences was compared in NCBI through BLASTN to examine the complete alignment with the partial coding sequence of chloroplast *matK* gene and *trnH-psbA*. The 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) to examine whether the partial amino acid codes were coherent with the chloroplast *matK* gene frame and without any stop codon (Mahadani and Ghosh 2013). In this way, the generated sequences were confirmed to be the fragments of chloroplast gene. All the analyzed sequences were then deposited in GenBank.

3.3.6.6 Primer design:

OligoCalc - (*http://basic.northwestern.edu/biotools/OligoCalc.html.*) is a web-accessible, client-based computational engine. Molecular weight, solution concentration, melting temperature, estimated absorbance coefficients, inter-molecular self-complementarities estimation and intra-molecular hairpin loop formation were calculated through OligoCalc.

PCR Products - (*http://www.bioinformatics.org/sms2/pcr_products.html*) is one of the most important programs of Sequence Manipulation Suite which is online available. The program searched for perfectly matching primer annealing sites that generated a PCR product. PCR Products determined the expect product sizes in the lab.

3.3.7 Phylogenetic analysis:

The phylogenetic studies were performed using the molecular evolutionary genetic analysis (MEGA) 4.2 software in accordance with the Kimura 2-Parameter (K2P) model. DNA Barcoding 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 and Nei (1987) and modified by Studier and Keppler (1988). The Kimura model is an extension of the Jukes and Cantor (JC) basic model (Saitou and Nei 1987; Studier and Keppler 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, the best metric when distances were low as in DNA barcode sequence.

3.3.7.1 MEGA4:

Mega4 performed for tree inference and the input file were in '.meg' format. The files contained align DNA sequences in mega format. Using Mega4, 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, open the align file in Mega4 and selected the

submenu Bootstrap Test of Phylogeny > Neighbor -Joining from the Phylogeny menu in the Mega4 main window. The Analysis Preferences window was appeared. Clicked on 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). Similarly, selected the Model row Nucleotide >Kimura-2-parameter. To set the number of bootstrap replicates, clicked on the Test of Phylogeny tab on the top of the window and entered 1000 in the Replications cell. After a few seconds the NJ tree with bootstrap values appeared in the Tree Explorer window. By default, the tree is midpoint rooted. If the location of the root needs to be placed on any other branch of the tree, this was done by selecting the top button on the left side of the window.

3.3.7.2 PHYLIP:

PHYLIP is a public domain package written by *Joe Felsenstein* and composed of a large number of programs, which makes it versatile and very powerful.

Preparing input data:

PHYLIP programs read their data from a file that were named “infile”. The first line contained the number of species, the number of characters and, possibly, one or more program options. Next comes, the species and character data in separate lines. Each line started with 10 letters or symbols reserved for the species name and were followed by the characters to analyze. These were built by the use of other programs like, e.g. *Clustal* or *TreeAlign* (select PHYLIP output format for the data file, inspected it for correctness and renamed it to infile). When run, each program was display about its role, its version and a menu. To change any option, type the corresponding letter from the left column in the menu.

Constructing phylogenies:

DNADIST built a matrix using either of Jukes-Cantor, Kimura, Jin-Nei or maximum likelihood methods.

Displaying results:

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

DRAWGRAM had drawn a cladogram or phenogram of a *rooted* tree.

DRAWTREE had drawn *unrooted* phylogenies on a variety of output devices.

Bootstrapping

SEQBOOT allowed resampling of data sets by the bootstrap, jackknife or permutation methods.

Bootstrap analysis was done by following these steps:

1) Run SEQBOOT on the input dataset, selected a shuffling method and specified at least 1000 replicates.

2) Phylogeny analysis program was depending on the data type, desired analysis and preferred method with the desired options. Selected option "M" (multiple data sets) and entered the number of replicates generated with SEQBOOT.

3) Distance matrices were generated in the previous step, needed to rename outfile to infile and started one of the Distance Matrix programs (FITCH, KITSCH or NEIGHBOR) now to generate the trees.

4) Renamed the treefile to infile and started CONSENSE to evaluate the significance of the analysis.

3.3.8 Data presentation:

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 given a good general impression of relative rates of change across a tree. Bootstrap values were displayed as percentages on each branch (Mahadani et al., 2013).