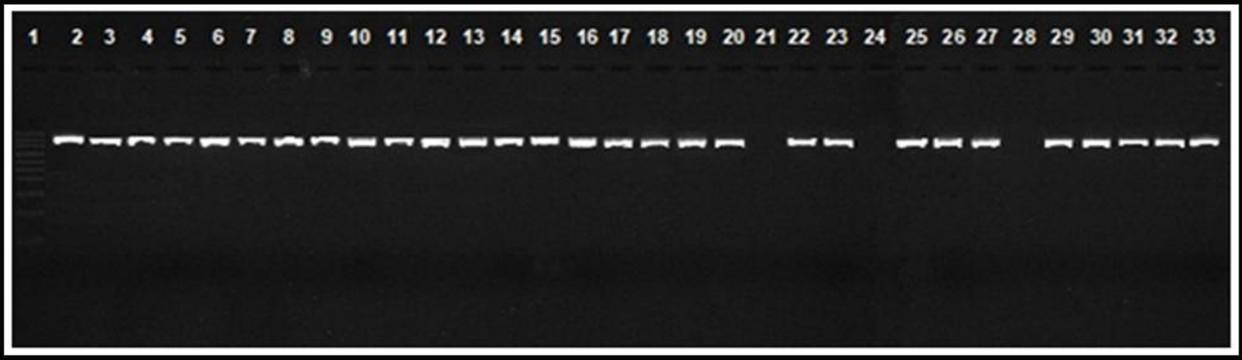
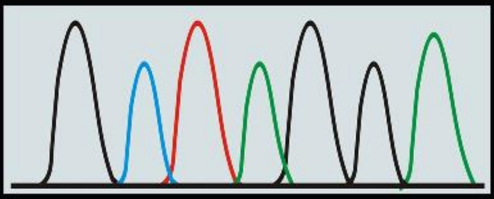
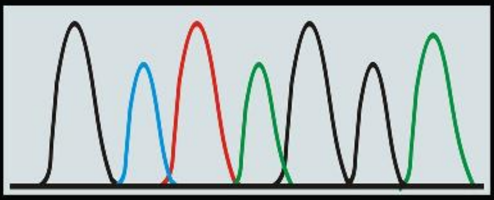
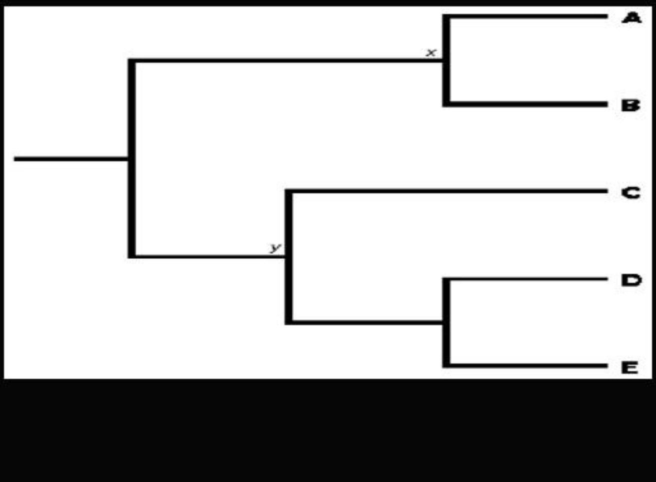






CHAPTER 3
DNA BARCODING



3.1. INTRODUCTION

There are approximately 3,00,000 plant species in the world according to International Union for Conservation of Nature (Singh and Srivastava, 2004; IUCN, 2017). The identification of plants are broadly done by the **conventional method** which includes morphological characterization, microscopic evaluations, chemical composition analysis and **advanced method** which comprised of genetic variations at molecular levels (DNA based methods) (Marcial-quino et al., 2015). Accurate identification using conventional methods requires lots of expertise in the respective area, which is not possible always. Threatened species, whose trade is regulated by the Convention on International Trade of Endangered Species (CITES), correct identification is crucial for the enforcement of the regulations and future conservation of the species. The species identification process is much more complicated, time consuming and challenging globally (Hartvig et al., 2015).

The emergence of DNA barcoding technique is the boon for classification and identification of biodiversity (Gregory, 2005). DNA barcoding is a tool for species identification that uses internationally agreed protocols and regions of DNA to create a global database of living organisms. Barcoding is a highly sophisticated technique for rapid and reliable identification of unknown biological sample using a short DNA sequence of the genome (Hebert et al., 2003). The short DNA sequence either from nuclear or/and cytoplasmic region of the genome are called as **DNA Barcodes**. The identification of species using DNA barcodes has been successfully established in algae (Saunders, 2008), fungi (Seena et al., 2010), plants (Chase et al., 2005; Kress et al., 2005; Fazekas et al., 2012) and various animal groups, such as spiders (Barrett and Hebert, 2005), fish (Ward et al., 2005), birds (Hebert et al., 2004) and rodents (Robins et al., 2007).

Information gathered from DNA barcoding has far reaching applications in several fields of biology other than taxonomic studies, such as, identification of invasive pest species (biosecurity), monitoring of protected species (conservation), rapid biodiversity assessment and food chain analysis (ecology), identification of active compound (pharmacology) and discovering therapeutically important pathogens and their vectors (medicine). The major advantage of DNA barcoding over molecular phylogeny is that it does not require cloning and sequencing of a complete gene, thus making it less laborious and time consuming. Essentially, DNA barcoding is an additional taxonomic tool with high potential of reviving modern taxonomy (Schindel and Miller, 2005).

In Plants, the DNA from chloroplast and nuclear genomes are recognized as appropriate barcodes. Various scientific groups proposed number of genes from chloroplast as a probable barcode for the respective plant group. Consortium for the Barcode of Life considered all the parameters for the barcode analysis and provided seven potential barcode loci including four coding regions (matK, rbcL, rpoC1 and rpoB) and three non-coding regions (psbA-trnH, atpF-atpH and psbK-psbI) (CBOL, 2009). The core validated barcode regions for all land plants from chloroplast genome are rbcL and matK (De Vere et al., 2015). Li et al. (2011) and Fazekas et al. (2012) also recognized rbcL and matK as a core DNA barcode while psbA-trnH as a significant accompanying marker. ITS region has been recommended as a promising barcode for plants, fungi and animals (Kress et al., 2005; Yao et al., 2009; Chen et al., 2010). The China Plant Barcode of Life Group has strongly supported and promoted the addition of ITS as the core barcode for plants along with matK and rbcL (China Plant BOL et al., 2011).

3.1.1. DNA Barcoding of family Orchidaceae

Orchidaceae is one of the largest and complex family of flowering plant comprising a number of taxa which are difficult to identify and classify on the basis of morphology even in the flowering state (Dressler, 1993; Van den Berg et al., 2000; Gravendeel et al., 2001; Cameron, 2004). Orchids are highly promiscuous in nature and hybridize very easily because of which the huge number of hybrids came into existence. The identification and classification of taxa into natural system become intricate as the augmentation of the hybrid population takes place (Soliva et al., 2001).

Orchids possess extensive ornamental as well as healing properties due to which their natural populations have been over exploited, thus rendering these species as threatened and endangered (Jalal et al., 2008). In India, more than 20% orchid diversity *i.e.* 215 species of orchids have been declared as endangered and 14 were nearly extinct (Hegde, 1996). Despite the fact, the illegal market of medicinal and ornamental orchids is flourishing and their identification using conventional methods facing many challenges. DNA barcoding can serve as a magnificent tool for rapid and accurate identification of new or cryptic and polymorphic species (Lahaye et al., 2008; Miwa et al., 2009; Xiao et al., 2010) and also as an effective diagnostic tool for restriction of illegal trade of endangered species *i.e.* biopiracy (Eaton et al., 2010; Jeanson et al., 2011; Muellner et al., 2011; Yesson et al., 2011).

Globally orchids are consumed predominantly however, there are very few and random reports are available on DNA barcoding of orchids. Lahaye et al. (2008) used two datasets, first one contains 71 Costa Rican orchids and another dataset contains 1566 species of orchids from Mesoamerican biodiversity hotspot, Southern Africa. They tested eight potential barcode loci and proposed matK as a universal DNA barcode for flowering plants. Asahina et al. (2010) identified five remedial species of *Dendrobium* using matK and rbcL and interpreted that matK could provide 100% species resolution. Ginibun et al. (2010) screened, eight chloroplast regions (accD, matK, ndhJ, rpoB, rpoC1, ycf5, rbcL-a and trnH-psbA) and two nuclear regions (ITS1 and ITS2) for seven species of *Spathoglottis* and revealed that four loci (matK, rbcL-a, rpoB and rpoC1) from the chloroplast region, yielded 100% amplification. In another investigation by Singh et al. (2012) for the genus *Dendrobium*, matK and ITS resolved 75% and 100% species respectively and matK+rpoB+rpoC1 was the best multilocus combination.

The genus *Holcoglossum* Schltr. was an exceptional case to test DNA barcodes as it contains both long-evolved and recently radiated species. Xiang et al. (2011) used six potential barcode loci *viz.* rbcL, matK, atpF-atpH, psbK-psbI, trnH-psbA and ITS to discriminate 12 species of *Holcoglossum*. The results showed that combination matK+ITS have the greater ability to resolve the species of genus *Holcoglossum*. The work on DNA Barcoding of Indian Orchids carried out by Praveen (2012) concluded ITS (single locus) and the combination of ITS+matK+rbcL as suitable barcode for the discrimination of orchids and other outgroup species. Parveen et al. (2012) tested five barcode loci to discriminate eight endangered species of *Paphiopedilum* from India, and concluded matK as a signature loci to resolve closely related species of Indian *Paphiopedilum* and their interspecific hybrids. Guo et al. (2016) also worked on the same genus *Paphiopedilum* using eight single-loci (accD, matK, rbcL, rpoC2, ycf1, atpF-atpH, atpI-atpH, ITS) and various multilocus combinations. They concluded that nrITS (single loci) and matK+atpF-atpH (multiple locus) are the most efficient barcodes for discriminating the species of the genus. The study of Ghorbani et al. (2017) includes 490 reference sequences of nrITS, trnL-F and matK region to identify 150 individual tubers of 31 batches purchased from 12 cities in Iran to assess species diversity in commerce. The markers enabled unambiguous species identification using tree-based methods showed 67% identity for nrITS, 58% for trnL-F and 59% for matK.

The present study is an approach towards the identification of orchids of Gujarat using three barcode loci *viz.* rbcL, matK and ITS.

3.2. MATERIALS AND METHOD

3.2.1. Collection of Plant sample

The plant tissues (leaf, bulb or tuber) from 32 species were collected from various geographical locations of Gujarat state as described in Table 2.2 of chapter 2 that includes 31 orchidaceous taxa and one out group species (*Holarrhena pubescens*). The 31 orchid taxa include 29 distinct species and two morphological variants of *Nervilia plicata* and *Peristylus constrictus*. During collection, it was ensured that no vegetative link existed between the two different accessions of the same species. The collected tissues were preserved at -20 °C or in silica beads.

3.2.2. Chemicals

All the chemicals used for the study were of analytical grade (AR) and molecular grade (MB). 2-amino-2-(hydroxymethyl) propane-1,3-diol (Tris), hexadecyl trimethyl ammonium bromide (CTAB), sodium acetate, bromophenol blue sodium salt, agarose, ethidium bromide, deoxynucleotide set (dNTPs) and ribonuclease-A were procured from Sigma-Aldrich India. 2-[2-[bis(carboxymethyl)amino] ethyl-(carboxymethyl) amino] acetic acid (Na_2EDTA), chloroform isoamyl alcohol, hydrochloric acid, sodium hydroxide pellets, glycerol, acetic acid (glacial), isopropyl alcohol, sodium chloride, phenol and 2-mercaptoethanol were purchased from Sisco Research Laboratory (SRL), Mumbai, India. Absolute ethanol was purchased from Merck India. Pfu DNA polymerase, exonuclease-I, *E. coli*, shrimp alkaline phosphatise, lambda DNA/HindIII Marker, gene ruler 1000 bp DNA Ladder were procured from Fermentas Inc., USA. MilliQ water was used for all the analysis.

3.2.3. Isolation of Genomic DNA and quantification

The total genomic DNA was isolated from leaves, bulbs or tubers. The isolation was carried out using DNeasy Plant Mini kit (250) (Cat. No.69104) following manufacturer protocol or manual CTAB method (Doyle and Doyle, 1987). The protocol followed for manual DNA isolation was mentioned below;

CTAB Method

- The tissues were wiped with 70% ethyl alcohol; 50 mg of tissue was weighed and homogenized in liquid nitrogen using Polyvinyl Polypyrrolidone (PVPP).
- The homogenate was transferred to 2 mL sterile micro-centrifuge tube (MCT) containing 1 mL CTAB buffer (100 mM Tris-Cl (pH 8.0), 20 mM Na_2EDTA, 2% CTAB, 1.4 M NaCl, 0.2% β-mercaptoethanol) pre-incubated at 65 °C. The sample was mixed gently and incubated in water bath at 65 °C for 35 mins.
- The contents were mixed properly at an interval of 10 mins. After 35 mins MCT was removed and allowed to cool at room temperature (27°C).
- Thereafter equal volume of chloroform and isoamyl alcohol (CI) (24:1) mixture was added to MCT and was mixed gently followed by centrifugation at 10,000 × g for 5 mins. Supernatant was transferred to fresh MCT containing equal volume of CI and again centrifuged at 11,000 × g for 5 mins
- The supernatant was transferred to a new MCT containing 10 µL of RNaseA and tube was incubated in water bath at 37 °C for 30 mins.
- Then 700 µL of CI was added to the sample. Two phases were mixed gently and then centrifuged at 11000 × g for 10 mins.
- The upper aqueous phase was transferred to a fresh MCT and then equal volume of chilled ethanol was added.
- The MCT was shaken gently and centrifuged at 10000 × g for 2 mins. The supernatant was decanted and nucleic acid pellet obtained was washed with 70% ethanol. The pellet, thus obtained, was air dried and resuspended in 100 µL of TE buffer (10 mM Tris-Cl, pH 8.0; 1 mM Na_2EDTA, pH 8.0).
- The isolated DNA was stored at -20°C and used for further analysis.
- The DNA concentration and purity were determined using Qiaxpert system (Qiagen). 2 µL of DNA was loaded on Qiaxpert slide with blank (TE, elution buffer) in the first well.

3.2.4. PCR Amplification and Sequencing

PCR was performed to amplify the definite region of DNA using specific primers (Table 3.1). The reaction was carried out in Thermal cycler (Applied BiosystemsVeriti®). Amplification was carried out using ReadyMix™ Taq PCR Reaction Mix (Sigma/TopTaq). The PCR reaction mixture composition is shown in Table 3.2. The PCR products were electrophoresed in 2% agarose gel using TAE/TBE buffer.

Table 3.1: Primers and their PCR conditions (CCDB Protocols)

Region	Primers	Predenaturation	Denaturation 35 cycles	Annealing 35 cycles	Extension 35 cycles	Final extension
rbcL	rbcLa-F rbcLa-R	94 °C (4 mins)	94 °C (30 secs)	55 °C (30 secs)	72 °C (1 min)	72 °C (10 mins)
matK	1RKIM-F 3fKIM-R	94 °C (1 min)	94 °C (30 secs)	52 °C (20 secs)	72 °C (1 min)	72 °C (10 mins)
	390F 1326R	94 °C (1 min)	94 °C (30 secs)	50 °C (40 secs)	72 °C (40 secs)	72 °C (5 mins)
		96 °C (4 min)	94 °C (30 secs)	50 °C (1 min)	72 °C (1 min)	72 °C (10 mins)
ITS2	ITS2-S2F ITS4-R	94 °C (5 min)	94 °C (30 secs)	56 °C (30 secs)	72 °C (45 secs)	72 °C (10 mins)
	ITS1-F	94 °C (5 min)	94 °C (30 secs)	56 °C (20 secs)	68 °C (45 secs)	68 °C (10 mins)

Table 3.2: PCR Reaction Mixture

SN	Reagents	Reaction Volume (20µL)
1	2x PCR Ready Reaction Mixture	10 µL
2	Forward and reverse primers (10pM)	1 µL each
3	Additives (0.2 µg/µl BSA, 5% Trehalose)	1 µL
4	DNA template (~100ng)	1 µL
5	Nuclease free water	Made to 20 µL

The amplified products were purified using Exo-SAP method (Mixture of two enzymes - exonuclease and shrimp alkaline phosphatase) (Kress and Erickson, 2007). This treatment degrades the unused primes and removes the dNTPs left over after amplification. A total of 2 µL enzyme mixture was used to clean 5 µL of PCR product. The reaction mixtures were then incubated in a thermal cycler for 15 mins at 35 ºC; 15 mins at 85 ºC to carry out the reaction. The purified products were stored at -20 ºC and later subjected to bi-directional Sanger's di-deoxy sequencing (Sanger et al., 1977) using BigDye Terminator v. 3.1 cycle sequencing kit. The composition and condition of sequencing reaction mixture were given in Table 3.3 & 3.4 respectively.

Table 3.3: Composition of Cycle Sequencing Reaction Mixture

SN	Reagents	Reaction Volume (10µL)
1	BigDye v.3.1 ready reaction mixture	1 µL
2	5x sequencing buffer	2 µL
3	Forward / Reverse sequencing primers (10 pMol)	1 µL
4	Amplified Product	1 µL
5	Nuclease free water	5 µL to make up volume

Table 3.4: Sequencing Conditions for various primers (CCDB Protocols)

SN	Primer	Reaction Conditions
1	rbcL	96 ºC, 4 mins, 96 ºC, 10 secs; 55 ºC, 5 secs, 60 ºC, 4 mins× **25 cycles**
2	matK KIM	94 ºC, 10 secs, 94 ºC, 20 secs; 48 ºC, 20 secs, 60 ºC, 4 mins× **25 cycles**
3	matK _390f & 1326r	94 ºC, 10 secs, 94 ºC, 20 secs; 48 ºC, 20 secs, 60 ºC, 4 mins× **25 cycles**
4	ITS2-S2F & ITS4	96 ºC, 4 mins, 96 ºC, 10 secs; 50 ºC, 5 secs, 60 ºC, 4 mins× **25 cycles**
5	ITS1 & ITS4	94 ºC, 5 mins, 94 ºC, 30 secs; 56 ºC, 30 secs, 68 ºC, 45 secs× **25 cycles**

The sequencing products were purified using BigDye XTerminator® Purification Kit. The purified products were processed for sequencing using Capillary electrophoresis on 3500xL Genetic Analyzer (Applied Biosystems).

3.2.5. Sequence Analysis and Submission of Barcodes to BOLD

The sequences obtained were analyzed using sequencing analysis (v. 5.4) (Applied Biosystems), BioEdit and Codon Code Aligner (v. 6.0.2.). Consensus sequences generated after aligning sequences from forward and reverse primers were subjected NCBI blast analysis. The sequences which showed significant match with NCBI Database were submitted to Barcode of Life Database (BOLD) according to guidelines (http://www.boldsystems.org). For the species where NCBI data was not available were confirmed with detailed taxonomic identification and were submitted to BOLD. The sequences were processed in MEGA 7 (v. 7.0.26) for further analysis.

3.3. RESULTS AND DISCUSSION

The current studies comprised of the selection of three barcode loci *viz*. rbcL and matK from chloroplast genome and ITS from the nuclear genome for identifying 31 orchidaceous taxa and one out group species. All primers were evaluated and compared individually for their amplification and sequencing success rates. The species resolution for each locus was calculated based on three different methods *viz*. genetic distances, phylogenetic tree and BLAST analysis. The species resolution analysis was carried out at generic and specific levels. The interspecific variations were obtained using the distance matrix prepared by aligning all nucleotide sequences belonging to different species with their respective genus. The genetic distance and phylogenetic tree methods were used to discriminate the congeneric species for the respective genus.

3.3.1. Assessment and Quantification of Genomic DNA

The DNA isolation was mainly carried out by CTAB protocol (Doyle and Doyle, 1987) and DNeasy Plant Minikit 250. Using these methods, the present study evidenced 100% genomic DNA isolation.

The isolated DNA showed some scale of shearing in few samples and the concentration of DNA was also low (Table 3.5), in spite of this the high molecular weight of DNA made them suitable for further studies. Orchids are rich in polysaccharides and other secondary metabolites, which inhibit the function of some enzymes like nucleases, polymerase and lipase resulting into low concentration of DNA. Hence the DNA samples were processed further for amplification studies without dilution.

Table 3.5: Concentration of isolated DNA in 32 taxa

SN	Name of Species	DNA concentration ng/µL
1.	*Acampe praemorsa*	46.5

SN	Name of Species	DNA concentration ng/µL
2.	*Aerides maculosa*	95.2
3.	*Aerides ringens*	23.2
4.	*Crepidium mackinnonii*	60.3
5.	*Dendrobium barbatulum*	28
6.	*Dendrobium microbulbon*	55.9
7.	*Dendrobium ovatum*	2.8
8.	*Dendrobium peguanum*	19.2
9.	*Eulophia herbacea*	23.1
10.	*Eulophia ochreata*	118.4
11.	*Geodorum laxiflorum*	10.6
12.	*Habenaria gibsonii*	197.2
13.	*Habenaria furcifera*	28.1
14.	*Habenaria grandifloriformis*	6.3
15.	*Habenaria longicorniculata*	3.6
16.	*Habenaria marginata*	8.0
17.	*Habenaria plantaginea*	21.5
18.	*Nervilia concolor*	215.2
19.	*Nervilia plicata* (Green)	64.6
20.	*Nervilia plicata* (Purple)	47.8
21.	*Oberonia falconeri*	1.2
22.	*Oberonia mucronata*	1.4
23.	*Peristylus constrictus*	376.6
24.	*Peristylus constrictus* (variant)	14.7
25.	*Peristylus lawii*	43.1
26.	*Peristylus plantagineus*	29.9
27.	*Peristylus stocksii*	2.9
28.	*Rhynchostylis retusa*	63.2
29.	*Vanda tessellata*	51.8
30.	*Vanda testacea*	26.1
31.	*Zeuxine strateumatica*	43.1
32.	*Holarrhena pubescens*	68.9

3.3.2. Amplification and Sequencing Success Rates

The relative usefulness of three amplified loci from 32 taxa was analyzed by comparing their amplification and sequencing success rates. Among the tested loci (matK, rbcL and ITS), rbcL and ITS exhibited highest (96.87%) amplification success rate followed by matK 90.625% after a number of PCR trials (Figure 3.1, 3.2 and 3.3).

The high amplification rate for rbcL in the present study is in good agreement with the reported data having amplification success of 95 – 100% (CBOL, 2009, Singh et al., 2012, Parveen 2012). The amplification success rate for matK was 90.625% which is comparatively low and that could be attributed to its high variability exclusively in monocots (Chase et al., 2007; Fazekas et al., 2008). ITS exhibited similar success rate as that of rbcL. A slightly high (98.97%) amplification rate was reported by Singh et al. (2012). *P. stocksii* exhibited negative amplification result for all the studied loci while *P. constrictus* and *N. plicata* (green variant) shown negative result only for matK locus. The negative amplification could be attributed to the low-quality DNA or presence of PCR inhibitors like polyphenols and polysaccharides (Varma et al., 2007).

The amplified products were purified and processed for sequencing. Among the tested loci, rbcL exhibited highest sequencing success followed by ITS and matK (Table 3.6). The sequencing success for rbcL and ITS were in good agreement with the reported one (90 – 100%) (Fazekas et al., 2008; CBOL, 2009; Singh et al., 2012). Significant efforts have been made for matK to achieve the success rate of 84.37%. Fazekas et al. (2008) and Parveen et al. (2012) achieved similar success rate. The total number of barcode sequences generated were 31, 27 and 30 for rbcL, matK and ITS loci respectively (Table 3.6).

Table 3.6: Amplification and sequencing success rates for the three loci (31 accessions)

Locus	**Length of amplicons obtained (bp)**	**No. of amplicons obtained**	**Amplification success (%)**	**No. of Sequences generated**	**Sequencing success (%)**
rbcL	442 – 634	31	96.87	31	96.87
matK	736 – 906	29	90.62	27	84.37
ITS	301 – 422	31	96.87	30	93.75

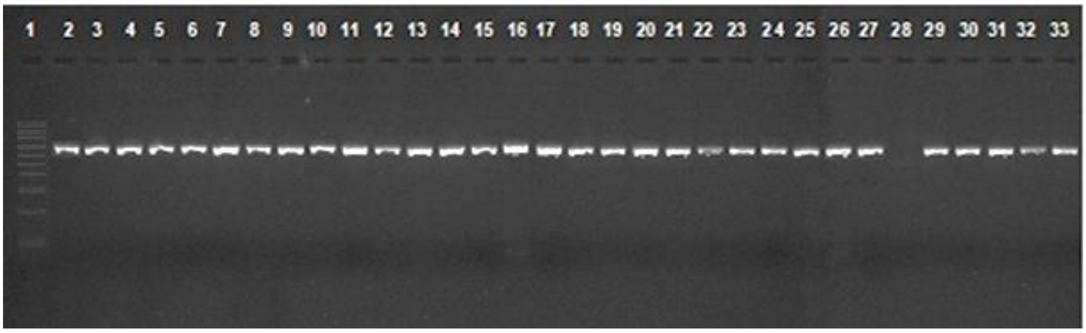


Figure 3.1: PCR amplification using rbcL Primer

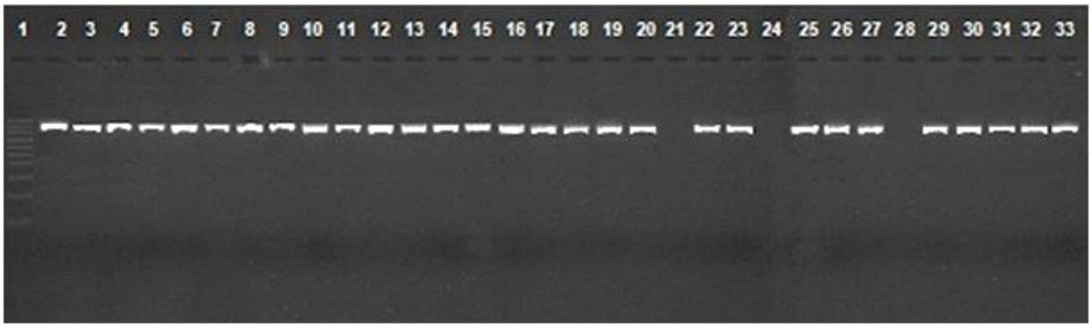


Figure 3.2: PCR amplification using matK Primer

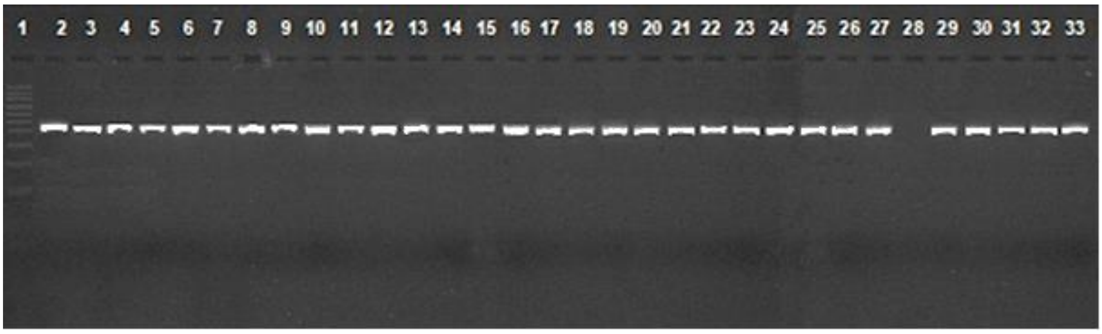


Figure 3.3: PCR amplification using ITS Primer

1	1000 bp ladder	12	*G. laxiflorum*	23	*O. mucronata*
2	*A. praemorsa*	13	*H. furcifera*	24	*P. constrictus*
3	*A. maculosa*	14	*H. gibsonii*	25	*P. constrictus* (variant)
4	*A. ringens*	15	*H. grandifloriformis*	26	*P. lawii*
5	*C. mackinnonii*	16	*H. longicorniculata*	27	*P. plantagineus*
6	*D. barbatulum*	17	*H. marginata*	28	*P. stocksii*
7	*D. microbulbon*	18	*H. plantaginea*	29	*R. retusa*
8	*D. ovatum*	19	*N. concolor*	30	*V. tessellata*
9	*D. peguanum*	20	*N. plicata* (Purple)	31	*V. testacea*
10	*E. herbacea*	21	*N. plicata* (Green)	32	*Z. strateumatica*
11	*E. ochreata*	22	*O. falconeri*	33	*H. pubescens*

3.3.3. Interspecific K2P (Kimura-2-Parameter) distances and Species Discrimination rates for three loci using MEGA 7

The accuracy of species identification were evaluated by using interspecific variations (Hebert et al., 2003; Lahaye et al., 2008). These variations are expressed in terms of K2P distances (Chen et al., 2010).

The interspecific K2P distances and species discrimination rates for rbcL, matK and ITS were analyzed on the basis of genetic distance, phylogenetic tree and BLAST methods.

3.3.3.1. rbcL

3.3.3.1.1. Interspecific K2P Distances

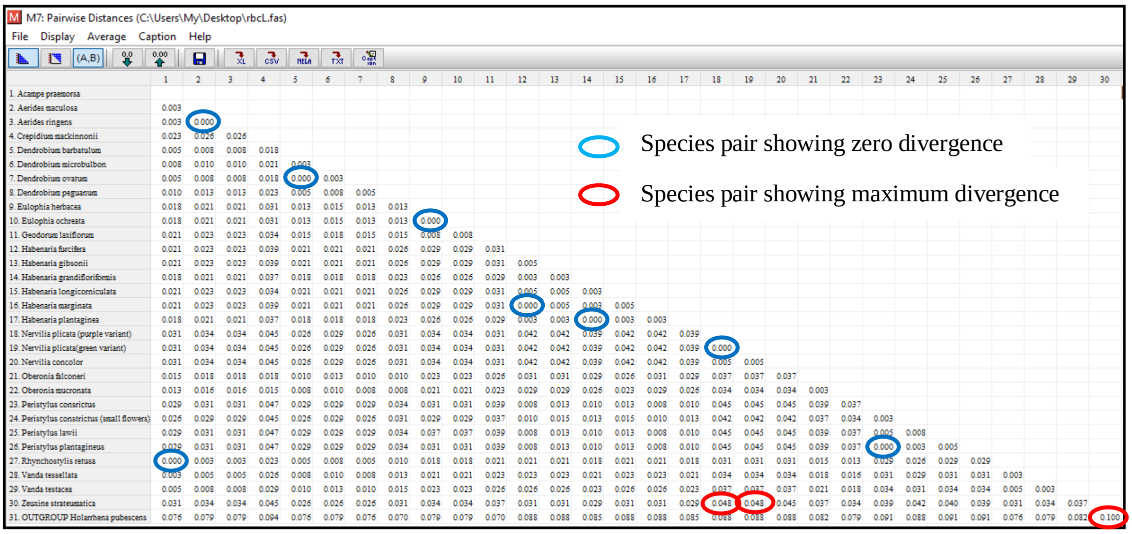
The interspecific divergence value was calculated using the K2P Model (Kimura, 1980). The variation rate among sites was statistically modeled with Gamma distribution (shape parameter = 5). Among 31 taxa, the average interspecific distances were 0.028 with a range of 0.000 – 0.100. Amidst of 31 taxa analyzed, 16 taxa exhibited zero distance with their allied species. The maximum interspecific K2P distance was observed between *H. pubescens* (Out Group) and *Z. strateumatica* (Figure 3.4).

3.3.3.1.2. Species Resolution

The species discrimination rates were computed using three different methods *viz*. genetic distance, phylogenetic tree and BLAST analysis. In the present study, the species resolution calculated using BLAST method was higher compared to the other two methods.

3.3.3.1.2.1. Distance Based Method

The distance matrix was prepared using K2P Model (Kimura, 1980). Species resolution was calculated for all 31 taxa from their aligned sequences. Among 31 taxa, 16 species were well discriminated using this method whereas eight pairs of taxa (15 species) were showing zero distance with their closely allied taxa (Figure 3.4). The eight pairs of orchids exhibiting zero distance were *A. ringens – A. maculosa, R. retusa – A. praemorsa, D. ovatum – D. barbatulum, E. ochreata – E. herbacea, H. marginata – H. furcifera, H. plantaginea – H. grandifloriformis, N. plicata* (green variant) – *N. plicata* (purple variant) and *P. plantagineus – P. constrictus*. Thus, the species discrimination rate using this locus was 51.61%. A similar result was obtained by Parveen et al. (2012). Two morphological variants of *P. constrictus* were differentiated from each other by exhibiting 0.003 divergence value. Among all the analyzed species of orchids both the variants of *N. plicata* showed highest interspecific divergence (0.048) with *Z. strateumatica* (Figure 3.4).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. Acampe praemorsa																														
2. Aerides maculosa	0.003																													
3. Aerides ringens	0.003	0.000																												
4. Crepidium mackinnonii	0.023	0.026	0.026																											
5. Dendrobium barbatulum	0.005	0.008	0.008	0.018																										
6. Dendrobium microbulbon	0.008	0.010	0.010	0.021	0.003																									
7. Dendrobium ovatum	0.005	0.008	0.008	0.018	0.000	0.003																								
8. Dendrobium peguanum	0.010	0.013	0.013	0.023	0.005	0.008	0.005																							
9. Eulophia herbacea	0.018	0.021	0.021	0.031	0.013	0.015	0.013	0.013																						
10. Eulophia ochreata	0.018	0.021	0.021	0.031	0.013	0.015	0.013	0.013	0.000																					
11. Geodorum laxiflorum	0.021	0.023	0.023	0.034	0.015	0.018	0.015	0.015	0.008	0.008																				
12. Habenaria furcifera	0.021	0.023	0.023	0.039	0.021	0.021	0.021	0.026	0.029	0.029	0.031																			
13. Habenaria gibsonii	0.021	0.023	0.023	0.039	0.021	0.021	0.021	0.026	0.029	0.029	0.031	0.005																		
14. Habenaria grandifloriformis	0.018	0.021	0.021	0.037	0.018	0.018	0.018	0.023	0.026	0.026	0.029	0.003	0.003																	
15. Habenaria longicorniculata	0.021	0.023	0.023	0.034	0.021	0.021	0.021	0.026	0.029	0.029	0.031	0.005	0.005	0.003																
16. Habenaria marginata	0.021	0.023	0.023	0.039	0.021	0.021	0.021	0.026	0.029	0.029	0.031	0.000	0.005	0.003	0.005															
17. Habenaria plantaginea	0.018	0.021	0.021	0.037	0.018	0.018	0.018	0.023	0.026	0.026	0.029	0.003	0.003	0.000	0.003	0.003														
18. Nervilia plicata (purple variant)	0.031	0.034	0.034	0.045	0.026	0.029	0.026	0.031	0.034	0.034	0.031	0.042	0.042	0.039	0.042	0.042	0.039													
19. Nervilia plicata(green variant)	0.031	0.034	0.034	0.045	0.026	0.029	0.026	0.031	0.034	0.034	0.031	0.042	0.042	0.039	0.042	0.042	0.039	0.000												
20. Nervilia concolor	0.031	0.034	0.034	0.045	0.026	0.029	0.026	0.031	0.034	0.034	0.031	0.042	0.042	0.039	0.042	0.042	0.039	0.005	0.005											
21. Oberonia falconeri	0.015	0.018	0.018	0.018	0.010	0.013	0.010	0.010	0.023	0.023	0.026	0.031	0.031	0.029	0.026	0.031	0.029	0.037	0.037	0.037										
22. Oberonia mucronata	0.013	0.016	0.016	0.015	0.008	0.010	0.008	0.008	0.021	0.021	0.023	0.029	0.029	0.026	0.023	0.029	0.026	0.034	0.034	0.034	0.003									
23. Peristylus consrictus	0.029	0.031	0.031	0.047	0.029	0.029	0.029	0.034	0.031	0.031	0.039	0.008	0.013	0.010	0.013	0.008	0.010	0.045	0.045	0.045	0.039	0.037								
24. Peristylus constrictus (small flowers)	0.026	0.029	0.029	0.045	0.026	0.029	0.026	0.031	0.029	0.029	0.037	0.010	0.015	0.013	0.015	0.010	0.013	0.042	0.042	0.042	0.037	0.034	0.003							
25. Peristylus lawii	0.029	0.031	0.031	0.047	0.029	0.029	0.029	0.034	0.037	0.037	0.039	0.008	0.013	0.010	0.013	0.008	0.010	0.045	0.045	0.045	0.039	0.037	0.005	0.008						
26. Peristylus plantagineus	0.029	0.031	0.031	0.047	0.029	0.029	0.029	0.034	0.031	0.031	0.039	0.008	0.013	0.010	0.013	0.008	0.010	0.045	0.045	0.045	0.039	0.037	0.000	0.003	0.005					
27. Rhynchostylis retusa	0.000	0.003	0.003	0.023	0.005	0.008	0.005	0.010	0.018	0.018	0.021	0.021	0.021	0.018	0.021	0.021	0.018	0.031	0.031	0.031	0.015	0.013	0.029	0.026	0.029	0.029				
28. Vanda tessellata	0.003	0.005	0.005	0.026	0.008	0.010	0.008	0.013	0.021	0.021	0.023	0.023	0.023	0.021	0.023	0.023	0.021	0.034	0.034	0.034	0.018	0.016	0.031	0.029	0.031	0.031	0.003			
29. Vanda testacea	0.005	0.008	0.008	0.029	0.010	0.013	0.010	0.015	0.023	0.023	0.026	0.026	0.026	0.023	0.026	0.026	0.023	0.037	0.037	0.037	0.021	0.018	0.034	0.031	0.034	0.034	0.005	0.003		
30. Zeuxine strateumatica	0.031	0.034	0.034	0.045	0.026	0.026	0.026	0.031	0.034	0.034	0.037	0.031	0.031	0.029	0.031	0.031	0.029	0.048	0.048	0.045	0.037	0.034	0.039	0.042	0.040	0.039	0.031	0.034	0.037	
31. OUTGROUP Holarrhena pubescens	0.076	0.079	0.079	0.094	0.076	0.079	0.076	0.070	0.079	0.079	0.070	0.088	0.088	0.085	0.088	0.088	0.085	0.088	0.088	0.088	0.082	0.079	0.091	0.088	0.091	0.091	0.076	0.079	0.082	0.100

Species pair showing zero divergence

Species pair showing maximum divergence

Figure 3.4: Estimates of Evolutionary Divergence between rbcL Sequences

3.3.3.1.2.2. Phylogenetic tree Method

The phylogenetic tree is a diagrammatical representation of evolutionary relationship among the different taxa or genes and their ancestors. The phylogeny was constructed in MEGA 7 (version 7.0.26) using best substitution model (Kumar et al., 2016). There are 24 models and for each model, parameters like BIC value (Bayesian Information Criterion), AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value and various other parameters were taken into consideration. The best substitution model selection is based on lowest BIC and AICc score. The lowest BIC and AICc value for rbcL was 2643.37 and 2191.96 respectively. Hence according to the result score, the best substitution model was K2P model (Figure 3.5).

The maximum likelihood tree was constructed using above described model with 1000 bootstrap replicates. The phylogeny involved 31 nucleotide sequences. The gaps and missing data were eliminated resulting into 394 bp sequence length in the final dataset.

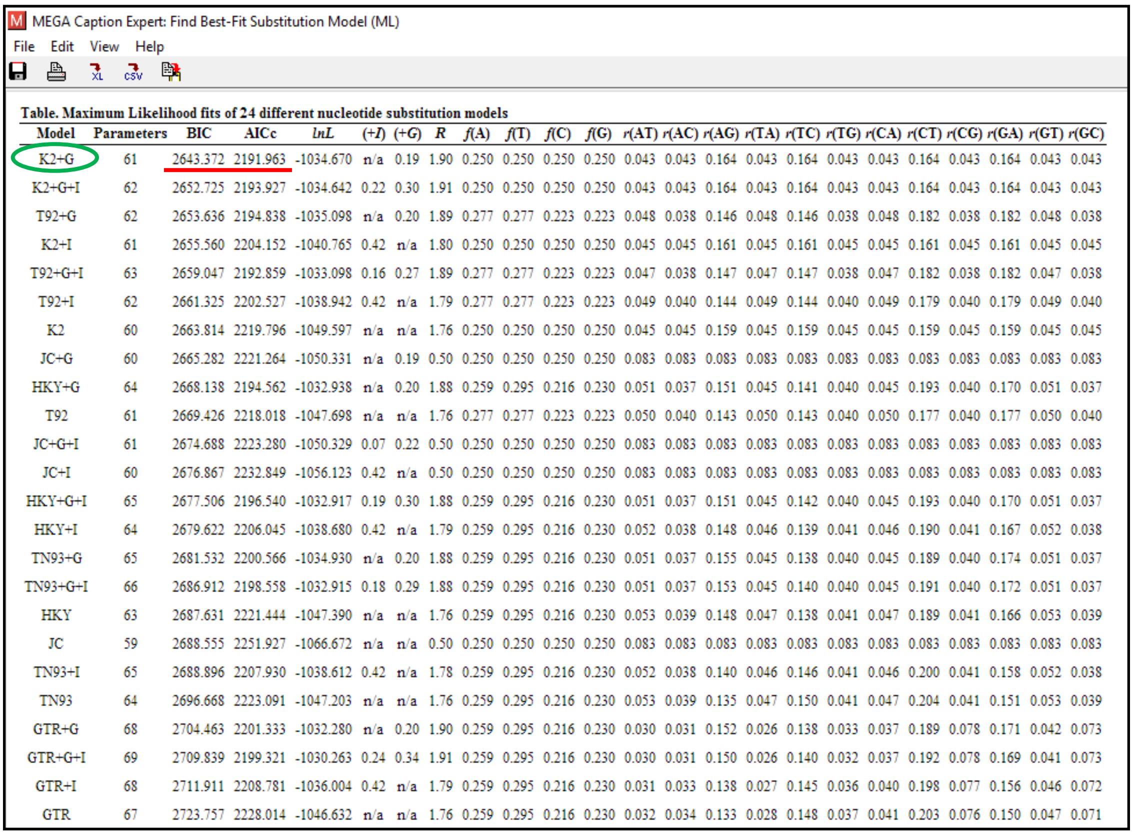


Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	*lnL*	(+*I*)	(+*G*)	*R*	*f*(A)	*f*(T)	*f*(C)	*f*(G)	*r*(AT)	*r*(AC)	*r*(AG)	*r*(TA)	*r*(TC)	*r*(TG)	*r*(CA)	*r*(CT)	*r*(CG)	*r*(GA)	*r*(GT)	*r*(GC)
K2+G	61	2643.372	2191.963	-1034.670	n/a	0.19	1.90	0.250	0.250	0.250	0.250	0.043	0.043	0.164	0.043	0.164	0.043	0.043	0.164	0.043	0.164	0.043	0.043
K2+G+I	62	2652.725	2193.927	-1034.642	0.22	0.30	1.91	0.250	0.250	0.250	0.250	0.043	0.043	0.164	0.043	0.164	0.043	0.043	0.164	0.043	0.164	0.043	0.043
T92+G	62	2653.636	2194.838	-1035.098	n/a	0.20	1.89	0.277	0.277	0.223	0.223	0.048	0.038	0.146	0.048	0.146	0.038	0.048	0.182	0.038	0.182	0.048	0.038
K2+I	61	2655.560	2204.152	-1040.765	0.42	n/a	1.80	0.250	0.250	0.250	0.250	0.045	0.045	0.161	0.045	0.161	0.045	0.045	0.161	0.045	0.161	0.045	0.045
T92+G+I	63	2659.047	2192.859	-1033.098	0.16	0.27	1.89	0.277	0.277	0.223	0.223	0.047	0.038	0.147	0.047	0.147	0.038	0.047	0.182	0.038	0.182	0.047	0.038
T92+I	62	2661.325	2202.527	-1038.942	0.42	n/a	1.79	0.277	0.277	0.223	0.223	0.049	0.040	0.144	0.049	0.144	0.040	0.049	0.179	0.040	0.179	0.049	0.040
K2	60	2663.814	2219.796	-1049.597	n/a	n/a	1.76	0.250	0.250	0.250	0.250	0.045	0.045	0.159	0.045	0.159	0.045	0.045	0.159	0.045	0.159	0.045	0.045
JC+G	60	2665.282	2221.264	-1050.331	n/a	0.19	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
HKY+G	64	2668.138	2194.562	-1032.938	n/a	0.20	1.88	0.259	0.295	0.216	0.230	0.051	0.037	0.151	0.045	0.141	0.040	0.045	0.193	0.040	0.170	0.051	0.037
T92	61	2669.426	2218.018	-1047.698	n/a	n/a	1.76	0.277	0.277	0.223	0.223	0.050	0.040	0.143	0.050	0.143	0.040	0.050	0.177	0.040	0.177	0.050	0.040
JC+G+I	61	2674.688	2223.280	-1050.329	0.07	0.22	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	60	2676.867	2232.849	-1056.123	0.42	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
HKY+G+I	65	2677.506	2196.540	-1032.917	0.19	0.30	1.88	0.259	0.295	0.216	0.230	0.051	0.037	0.151	0.045	0.142	0.040	0.045	0.193	0.040	0.170	0.051	0.037
HKY+I	64	2679.622	2206.045	-1038.680	0.42	n/a	1.79	0.259	0.295	0.216	0.230	0.052	0.038	0.148	0.046	0.139	0.041	0.046	0.190	0.041	0.167	0.052	0.038
TN93+G	65	2681.532	2200.566	-1034.930	n/a	0.20	1.88	0.259	0.295	0.216	0.230	0.051	0.037	0.155	0.045	0.138	0.040	0.045	0.189	0.040	0.174	0.051	0.037
TN93+G+I	66	2686.912	2198.558	-1032.915	0.18	0.29	1.88	0.259	0.295	0.216	0.230	0.051	0.037	0.153	0.045	0.140	0.040	0.045	0.191	0.040	0.172	0.051	0.037
HKY	63	2687.631	2221.444	-1047.390	n/a	n/a	1.76	0.259	0.295	0.216	0.230	0.053	0.039	0.148	0.047	0.138	0.041	0.047	0.189	0.041	0.166	0.053	0.039
JC	59	2688.555	2251.927	-1066.672	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
TN93+I	65	2688.896	2207.930	-1038.612	0.42	n/a	1.78	0.259	0.295	0.216	0.230	0.052	0.038	0.140	0.046	0.146	0.041	0.046	0.200	0.041	0.158	0.052	0.038
TN93	64	2696.668	2223.091	-1047.203	n/a	n/a	1.76	0.259	0.295	0.216	0.230	0.053	0.039	0.135	0.047	0.150	0.041	0.047	0.204	0.041	0.151	0.053	0.039
GTR+G	68	2704.463	2201.333	-1032.280	n/a	0.20	1.90	0.259	0.295	0.216	0.230	0.030	0.031	0.152	0.026	0.138	0.033	0.037	0.189	0.078	0.171	0.042	0.073
GTR+G+I	69	2709.839	2199.321	-1030.263	0.24	0.34	1.91	0.259	0.295	0.216	0.230	0.030	0.031	0.150	0.026	0.140	0.032	0.037	0.192	0.078	0.169	0.041	0.073
GTR+I	68	2711.911	2208.781	-1036.004	0.42	n/a	1.79	0.259	0.295	0.216	0.230	0.031	0.033	0.138	0.027	0.145	0.036	0.040	0.198	0.077	0.156	0.046	0.072
GTR	67	2723.757	2228.014	-1046.632	n/a	n/a	1.76	0.259	0.295	0.216	0.230	0.032	0.034	0.133	0.028	0.148	0.037	0.041	0.203	0.076	0.150	0.047	0.071

Figure 3.5: Best substitution Model for rbcL

31 taxa (30 orchids and one out group) were evaluated in the phylogenetic tree method. The cladistic analysis segregates an out group (*H. pubescens*) from 30 orchidaceous members by forming separate clade. Among the orchidaceae members, the genera under sub-family Orchidoideae (*Zeuxine*, *Habenaria* and *Peristylus*) and Epidendroideae (*Acampe*, *Aerides*, *Crepidium*, *Dendrobium*, *Nervilia*, *Oberonia*, *Rhynchostylis*, *Vanda*) were formed two separate groups which could be attributed to the variation in their pollinial structure. *Eulophia* and *Geodorum*, the genera under sub family Epidendroideae formed a separate clade that could be because of the variation in their habit.

The genus *Zeuxine* formed a separate clade from other Orchidoideae member that could be ascribed to variation in the morphology genus. The genus *Nervilia* also formed a separate branch from other Epidendroideae members that could be attributed to its divergent morphology. The eight pair of taxa showing zero interspecific variation in the distance-based method were clustered together using this method also. Thus, the species resolution using this method was 51.61% (Figure 3.6). Both the variants of *Peristylus constrictus* were well resolved using this method also.

The genera *Acampe* and *Rhynchostylis* were showing zero distance value and clustered together though there is a huge variation in their morphology. Similarly, some species of genus *Habenaria* and *Peristylus* clustered together however, they are very distinct in vegetative and flowering phase. Additionally, a species pair, *A. maculosa – A. ringens, D. barbatulum – D. ovatum; E. ochreata – E. herbacea* had zero distance estimates. These species look quite similar in vegetative phase however, at flowering stage these all species are easily distinguishable. So, the percentage resolution at generic and specific level is poor for this locus.

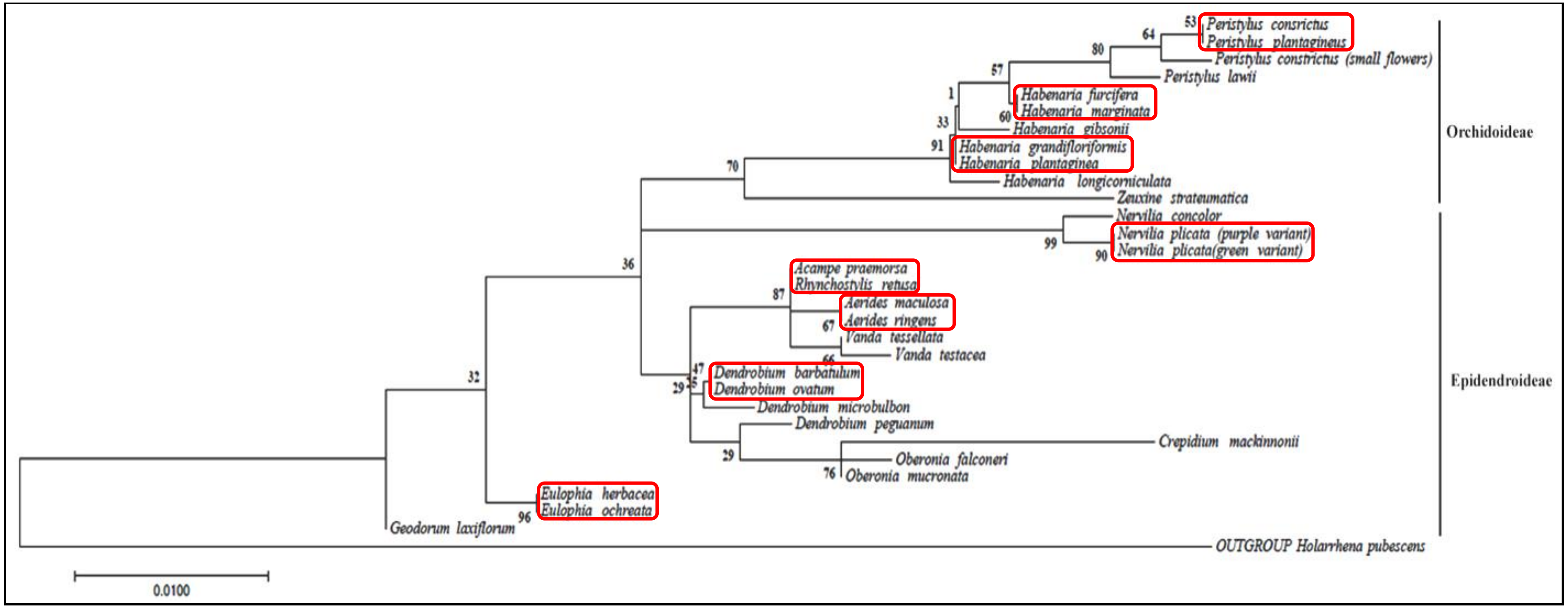


Figure 3.6: Maximum likelihood tree of 32 plant species based on rbcL sequences. (species with zero inter-specific divergence were enclosed in red color box)

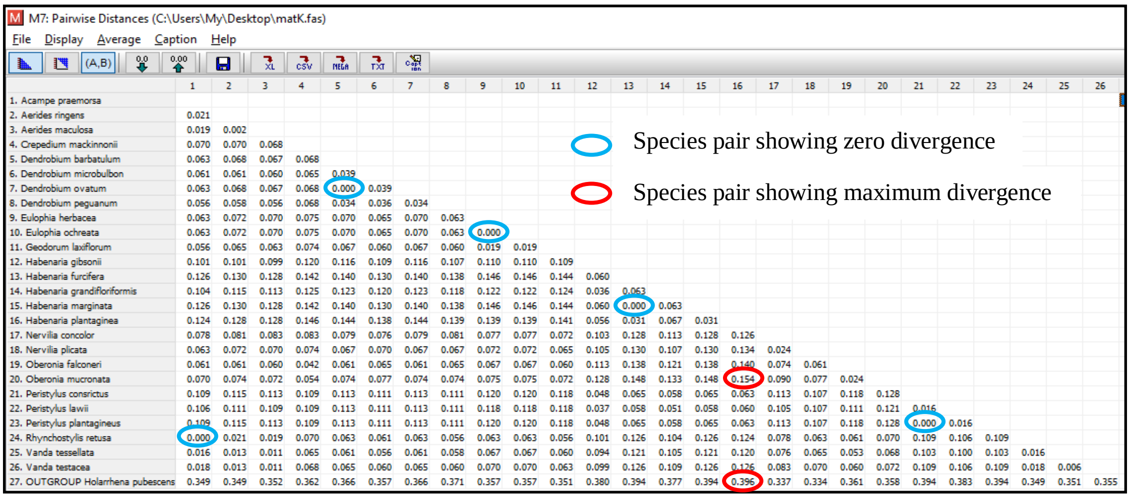
3.3.3.1.2.3. BLAST analysis

The accuracy of species resolution using BLAST method was 83.87% as shown in Table 3.7. It includes BLAST analysis of 31 sequences of which 16 species shows correct match with the available database while 9 species were having unique sequences. Five species sequences were not shown correct match with their own available sequences. The species pair namely *H. grandifloriformis – H. plantaginea, R. retusa – A. praemorsa* and *A. ringens – A. maculosa* which showed 100% similarity in the distance method and phylogenetic analysis were giving exact match with the available sequences.

3.3.3.2. matK

3.3.3.2.1. Interspecific K2P Distances

The average interspecific K2P distance among 27 species was 0.106 with a range of 0.000 – 0.396. During analysis of 27 species (26 were of orchids and one was out group), 10 species exhibited zero interspecific distance value while *H. pubescens* (out group) and *H. plantaginea* exhibited maximum interspecific K2P distance among all (Figure 3.7).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1. Acampe praemorsa																										
2. Aerides ringens	0.021																									
3. Aerides maculosa	0.019	0.002																								
4. Crepedium mackinnonii	0.070	0.070	0.068																							
5. Dendrobium barbatulum	0.063	0.068	0.067	0.068																						
6. Dendrobium microbulbon	0.061	0.061	0.060	0.065	0.039																					
7. Dendrobium ovatum	0.063	0.068	0.067	0.068	0.000	0.039																				
8. Dendrobium peguanum	0.056	0.058	0.056	0.068	0.034	0.036	0.034																			
9. Eulophia herbacea	0.063	0.072	0.070	0.075	0.070	0.065	0.070	0.063																		
10. Eulophia ochreata	0.063	0.072	0.070	0.075	0.070	0.065	0.070	0.063	0.000																	
11. Geodorum laxiflorum	0.056	0.065	0.063	0.074	0.067	0.060	0.067	0.060	0.019	0.019																
12. Habenaria gibsonii	0.101	0.101	0.099	0.120	0.116	0.109	0.116	0.107	0.110	0.110	0.109															
13. Habenaria furcifera	0.126	0.130	0.128	0.142	0.140	0.130	0.140	0.138	0.146	0.146	0.144	0.060														
14. Habenaria grandifloriformis	0.104	0.115	0.113	0.125	0.123	0.120	0.123	0.118	0.122	0.122	0.124	0.036	0.063													
15. Habenaria marginata	0.126	0.130	0.128	0.142	0.140	0.130	0.140	0.138	0.146	0.146	0.144	0.060	0.000	0.063												
16. Habenaria plantaginea	0.124	0.128	0.128	0.146	0.144	0.138	0.144	0.139	0.139	0.139	0.141	0.056	0.031	0.067	0.031											
17. Nervilia concolor	0.078	0.081	0.083	0.083	0.079	0.076	0.079	0.081	0.077	0.077	0.072	0.103	0.128	0.113	0.128	0.126										
18. Nervilia plicata	0.063	0.072	0.070	0.074	0.067	0.070	0.067	0.067	0.072	0.072	0.065	0.105	0.130	0.107	0.130	0.134	0.024									
19. Oberonia falconeri	0.061	0.061	0.060	0.042	0.061	0.065	0.061	0.065	0.067	0.067	0.060	0.113	0.138	0.121	0.138	0.140	0.074	0.061								
20. Oberonia mucronata	0.070	0.074	0.072	0.054	0.074	0.077	0.074	0.074	0.075	0.075	0.072	0.128	0.148	0.133	0.148	0.154	0.090	0.077	0.024							
21. Peristylus consrictus	0.109	0.115	0.113	0.109	0.113	0.111	0.113	0.111	0.120	0.120	0.118	0.048	0.065	0.058	0.065	0.063	0.113	0.107	0.118	0.128						
22. Peristylus lawii	0.106	0.111	0.109	0.109	0.113	0.111	0.113	0.111	0.118	0.118	0.118	0.037	0.058	0.051	0.058	0.060	0.105	0.107	0.111	0.121	0.016					
23. Peristylus plantagineus	0.109	0.115	0.113	0.109	0.113	0.111	0.113	0.111	0.120	0.120	0.118	0.048	0.065	0.058	0.065	0.063	0.113	0.107	0.118	0.128	0.000	0.016				
24. Rhynchostylis retusa	0.000	0.021	0.019	0.070	0.063	0.061	0.063	0.056	0.063	0.063	0.056	0.101	0.126	0.104	0.126	0.124	0.078	0.063	0.061	0.070	0.109	0.106	0.109			
25. Vanda tessellata	0.016	0.013	0.011	0.065	0.061	0.056	0.061	0.058	0.067	0.067	0.060	0.094	0.121	0.105	0.121	0.120	0.076	0.065	0.053	0.068	0.103	0.100	0.103	0.016		
26. Vanda testacea	0.018	0.013	0.011	0.068	0.065	0.060	0.065	0.060	0.070	0.070	0.063	0.099	0.126	0.109	0.126	0.126	0.083	0.070	0.060	0.072	0.109	0.106	0.109	0.018	0.006	
27. OUTGROUP Holarrhena pubescens	0.349	0.349	0.352	0.362	0.366	0.357	0.366	0.371	0.357	0.357	0.351	0.380	0.394	0.377	0.394	0.396	0.337	0.334	0.361	0.358	0.394	0.383	0.394	0.349	0.351	0.355

Figure 3.7: Estimates of Evolutionary Divergence between matK Sequences

3.3.3.2.2. Species Resolution

3.3.3.2.2.1. Distance Based Method

The distance matrix prepared from 27 species using their aligned nucleotide sequences exhibited zero distance value for five pairs of species (Figure 3.7). The five pair of species showing zero divergence value were *R. retusa – A. praemorsa, D. ovatum – D. barbatulum, E. ochreata – E. herbacea, H. marginata – H. furcifera* and *P. plantagineus – P. constrictus*. Thus, the species discrimination rate for matK was 62.96%

which was comparatively higher than that of rbcL. Some of the species like *A. ringens – A. maculosa, H. plantaginea – H. grandifloriformis* showing zero divergence value in rbcL were well resolved using matK. In comparison with other orchidaceous members, the interspecific divergence value was higher (0.154) between *O. mucronata – H. plantaginea* (Figure 3.7).

3.3.3.2.2.2. Phylogenetic tree Method

The best substitution model for matK was General Time Reversible (GTR) as it shows lowest BIC and AIC value *i.e.* 6691.475 and 6227.031 respectively (Figure 3.8). The phylogeny imparts Maximum Likelihood tree which was constructed using the GTR model (Kimura, 1980), with 1000 bootstrap replicates. The analysis involved 27 nucleotide sequences of 634 bp in the final dataset.

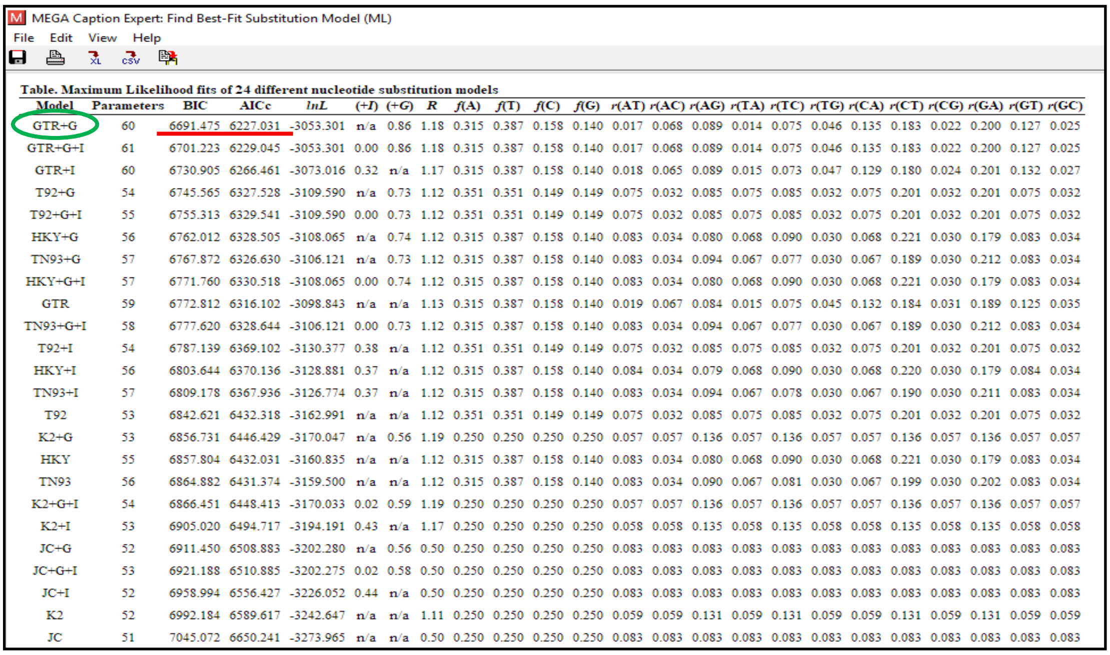


Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
GTR+G	60	6691.475	6227.031	-3053.301	n/a	0.86	1.18	0.315	0.387	0.158	0.140	0.017	0.068	0.089	0.014	0.075	0.046	0.135	0.183	0.022	0.200	0.127	0.025
GTR+G+I	61	6701.223	6229.045	-3053.301	0.00	0.86	1.18	0.315	0.387	0.158	0.140	0.017	0.068	0.089	0.014	0.075	0.046	0.135	0.183	0.022	0.200	0.127	0.025
GTR+I	60	6730.905	6266.461	-3073.016	0.32	n/a	1.17	0.315	0.387	0.158	0.140	0.018	0.065	0.089	0.015	0.073	0.047	0.129	0.180	0.024	0.201	0.132	0.027
T92+G	54	6745.565	6327.528	-3109.590	n/a	0.73	1.12	0.351	0.351	0.149	0.149	0.075	0.032	0.085	0.075	0.085	0.032	0.075	0.201	0.032	0.201	0.075	0.032
T92+G+I	55	6755.313	6329.541	-3109.590	0.00	0.73	1.12	0.351	0.351	0.149	0.149	0.075	0.032	0.085	0.075	0.085	0.032	0.075	0.201	0.032	0.201	0.075	0.032
HKY+G	56	6762.012	6328.505	-3108.065	n/a	0.74	1.12	0.315	0.387	0.158	0.140	0.083	0.034	0.080	0.068	0.090	0.030	0.068	0.221	0.030	0.179	0.083	0.034
TN93+G	57	6767.872	6326.630	-3106.121	n/a	0.73	1.12	0.315	0.387	0.158	0.140	0.083	0.034	0.094	0.067	0.077	0.030	0.067	0.189	0.030	0.212	0.083	0.034
HKY+G+I	57	6771.760	6330.518	-3108.065	0.00	0.74	1.12	0.315	0.387	0.158	0.140	0.083	0.034	0.080	0.068	0.090	0.030	0.068	0.221	0.030	0.179	0.083	0.034
GTR	59	6772.812	6316.102	-3098.843	n/a	n/a	1.13	0.315	0.387	0.158	0.140	0.019	0.067	0.084	0.015	0.075	0.045	0.132	0.184	0.031	0.189	0.125	0.035
TN93+G+I	58	6777.620	6328.644	-3106.121	0.00	0.73	1.12	0.315	0.387	0.158	0.140	0.083	0.034	0.094	0.067	0.077	0.030	0.067	0.189	0.030	0.212	0.083	0.034
T92+I	54	6787.139	6369.102	-3130.377	0.38	n/a	1.12	0.351	0.351	0.149	0.149	0.075	0.032	0.085	0.075	0.085	0.032	0.075	0.201	0.032	0.201	0.075	0.032
HKY+I	56	6803.644	6370.136	-3128.881	0.37	n/a	1.12	0.315	0.387	0.158	0.140	0.084	0.034	0.079	0.068	0.090	0.030	0.068	0.220	0.030	0.179	0.084	0.034
TN93+I	57	6809.178	6367.936	-3126.774	0.37	n/a	1.12	0.315	0.387	0.158	0.140	0.083	0.034	0.094	0.067	0.078	0.030	0.067	0.190	0.030	0.211	0.083	0.034
T92	53	6842.621	6432.318	-3162.991	n/a	n/a	1.12	0.351	0.351	0.149	0.149	0.075	0.032	0.085	0.075	0.085	0.032	0.075	0.201	0.032	0.201	0.075	0.032
K2+G	53	6856.731	6446.429	-3170.047	n/a	0.56	1.19	0.250	0.250	0.250	0.250	0.057	0.057	0.136	0.057	0.136	0.057	0.057	0.136	0.057	0.136	0.057	0.057
HKY	55	6857.804	6432.031	-3160.835	n/a	n/a	1.12	0.315	0.387	0.158	0.140	0.083	0.034	0.080	0.068	0.090	0.030	0.068	0.221	0.030	0.179	0.083	0.034
TN93	56	6864.882	6431.374	-3159.500	n/a	n/a	1.12	0.315	0.387	0.158	0.140	0.083	0.034	0.090	0.067	0.081	0.030	0.067	0.199	0.030	0.202	0.083	0.034
K2+G+I	54	6866.451	6448.413	-3170.033	0.02	0.59	1.19	0.250	0.250	0.250	0.250	0.057	0.057	0.136	0.057	0.136	0.057	0.057	0.136	0.057	0.136	0.057	0.057
K2+I	53	6905.020	6494.717	-3194.191	0.43	n/a	1.17	0.250	0.250	0.250	0.250	0.058	0.058	0.135	0.058	0.135	0.058	0.058	0.135	0.058	0.135	0.058	0.058
JC+G	52	6911.450	6508.883	-3202.280	n/a	0.56	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G+I	53	6921.188	6510.885	-3202.275	0.02	0.58	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	52	6958.994	6556.427	-3226.052	0.44	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2	52	6992.184	6589.617	-3242.647	n/a	n/a	1.11	0.250	0.250	0.250	0.250	0.059	0.059	0.131	0.059	0.131	0.059	0.059	0.131	0.059	0.131	0.059	0.059
JC	51	7045.072	6650.241	-3273.965	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

Figure 3.8: Best substitution Model for matK

The species belongs to both the subfamilies Orchidoideae and Epidendroideae were clustered separately and well resolved. The species pairs with zero divergence value in distance method were also clustered together using this method. Thus, species resolution was 62.96% using phylogenetic tree method. An out group forms a different clade from orchidaceous members (Figure 3.9). The genus *Nervilia* branched separately from other Epidendroideae members because of its divergent morphology. Gale et al. (2015) also got similar result wherein they considered *Nervilia* in separate subtribe Nerviliinae.

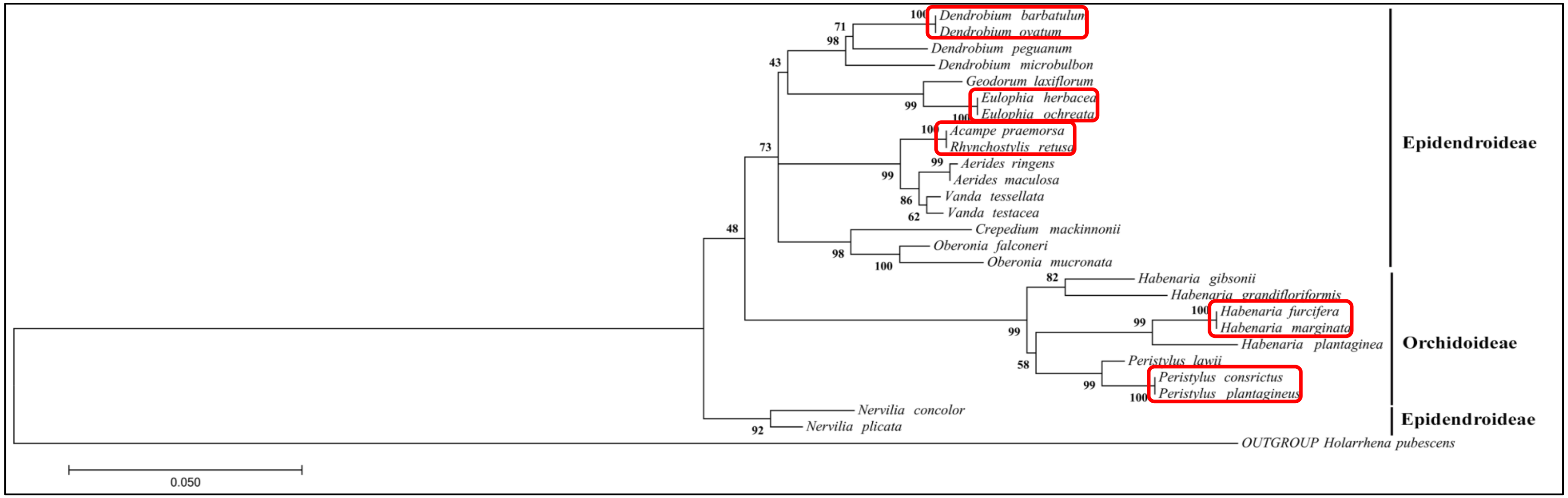


Figure 3.9: Maximum likelihood tree of 27 plant species based on matK sequences. (species with zero inter-specific divergence were enclosed in red color box)

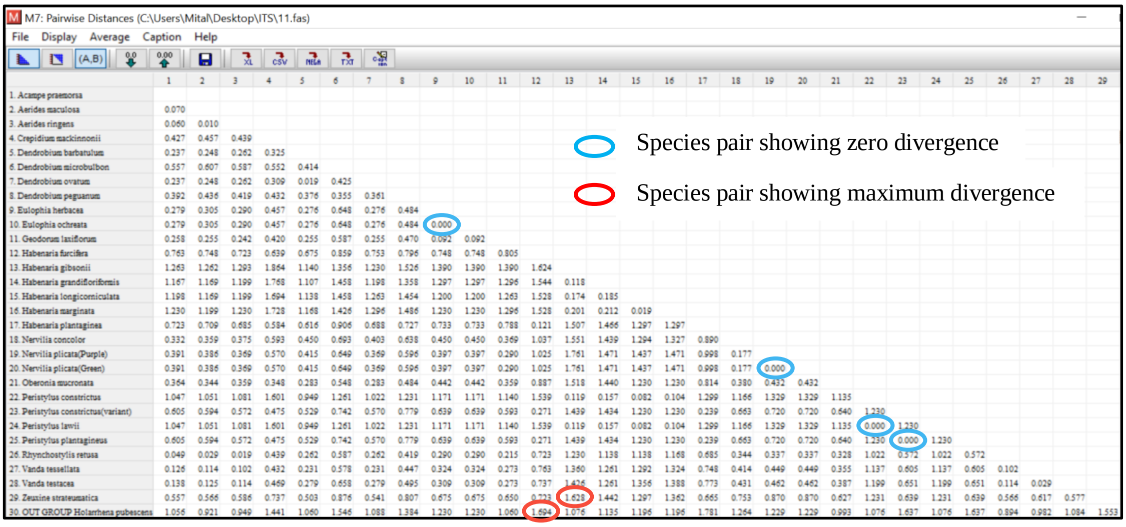
3.3.3.2.2.3. BLAST analysis

The BLAST analysis of 27 species revealed correct identification of 17 species and seven were unique sequences. Only three species were not correctly identified using this method. Thus, the species resolution was 88.88% (Table 3.7). All the species which were clustered together in the above described methods were showing exact match with their own species resulting into maximum percentage species identification.

3.3.3.3. ITS

3.3.3.3.1. Interspecific K2P Distances

Among 28 species (30 accessions), the average interspecific distance value was 0.777 with a range of 0.000 – 1.781. Seven species (four pairs) of orchids exhibited zero interspecific distance value. The divergence value between *H. pubescens* and *H. furcifera* was maximum among all (Figure 3.10).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1. Acampe praemorsa																													
2. Aerides maculosa	0.070																												
3. Aerides ringens	0.060	0.010																											
4. Crepidium mackinnonii	0.427	0.457	0.439																										
5. Dendrobium barbatulum	0.237	0.248	0.262	0.325																									
6. Dendrobium microbulbon	0.557	0.607	0.587	0.552	0.414																								
7. Dendrobium ovatum	0.237	0.248	0.262	0.309	0.019	0.425																							
8. Dendrobium peguanum	0.392	0.436	0.419	0.432	0.376	0.355	0.361																						
9. Eulophia herbacea	0.279	0.305	0.290	0.457	0.276	0.648	0.276	0.484																					
10. Eulophia ochreata	0.279	0.305	0.290	0.457	0.276	0.648	0.276	0.484	0.000																				
11. Geodorum laxiflorum	0.258	0.255	0.242	0.420	0.255	0.587	0.255	0.470	0.092	0.092																			
12. Habenaria furcifera	0.763	0.748	0.723	0.639	0.675	0.859	0.753	0.796	0.748	0.748	0.805																		
13. Habenaria gibsonii	1.263	1.262	1.293	1.864	1.140	1.356	1.230	1.526	1.390	1.390	1.390	1.624																	
14. Habenaria grandifloriformis	1.167	1.169	1.199	1.768	1.107	1.458	1.198	1.358	1.297	1.297	1.296	1.544	0.118																
15. Habenaria longicorniculata	1.198	1.169	1.199	1.694	1.138	1.458	1.263	1.454	1.200	1.200	1.263	1.528	0.174	0.185															
16. Habenaria marginata	1.230	1.199	1.230	1.728	1.168	1.426	1.296	1.486	1.230	1.230	1.296	1.528	0.201	0.212	0.019														
17. Habenaria plantaginea	0.723	0.709	0.685	0.584	0.616	0.906	0.688	0.727	0.733	0.733	0.788	0.121	1.507	1.466	1.297	1.297													
18. Nervilia concolor	0.332	0.359	0.375	0.593	0.450	0.693	0.403	0.638	0.450	0.450	0.369	1.037	1.551	1.439	1.294	1.327	0.890												
19. Nervilia plicata(Purple)	0.391	0.386	0.369	0.570	0.415	0.649	0.369	0.596	0.397	0.397	0.290	1.025	1.761	1.471	1.437	1.471	0.998	0.177											
20. Nervilia plicata(Green)	0.391	0.386	0.369	0.570	0.415	0.649	0.369	0.596	0.397	0.397	0.290	1.025	1.761	1.471	1.437	1.471	0.998	0.177	0.000										
21. Oberonia mucronata	0.364	0.344	0.359	0.348	0.283	0.548	0.283	0.484	0.442	0.442	0.359	0.887	1.518	1.440	1.230	1.230	0.814	0.380	0.432	0.432									
22. Peristylus constrictus	1.047	1.051	1.081	1.601	0.949	1.261	1.022	1.231	1.171	1.171	1.140	1.539	0.119	0.157	0.082	0.104	1.299	1.166	1.329	1.329	1.135								
23. Peristylus constrictus(variant)	0.605	0.594	0.572	0.475	0.529	0.742	0.570	0.779	0.639	0.639	0.593	0.271	1.439	1.434	1.230	1.230	0.239	0.663	0.720	0.720	0.640	1.230							
24. Peristylus lawii	1.047	1.051	1.081	1.601	0.949	1.261	1.022	1.231	1.171	1.171	1.140	1.539	0.119	0.157	0.082	0.104	1.299	1.166	1.329	1.329	1.135	0.000	1.230						
25. Peristylus plantagineus	0.605	0.594	0.572	0.475	0.529	0.742	0.570	0.779	0.639	0.639	0.593	0.271	1.439	1.434	1.230	1.230	0.239	0.663	0.720	0.720	0.640	1.230	0.000	1.230					
26. Rhynchostylis retusa	0.049	0.029	0.019	0.439	0.262	0.587	0.262	0.419	0.290	0.290	0.215	0.723	1.230	1.138	1.138	1.168	0.685	0.344	0.337	0.337	0.328	1.022	0.572	1.022	0.572				
27. Vanda tessellata	0.126	0.114	0.102	0.432	0.231	0.578	0.231	0.447	0.324	0.324	0.273	0.763	1.360	1.261	1.292	1.324	0.748	0.414	0.449	0.449	0.355	1.137	0.605	1.137	0.605	0.102			
28. Vanda testacea	0.138	0.125	0.114	0.469	0.279	0.658	0.279	0.495	0.309	0.309	0.273	0.737	1.426	1.261	1.356	1.388	0.773	0.431	0.462	0.462	0.387	1.199	0.651	1.199	0.651	0.114	0.029		
29. Zeuxine strateumatica	0.557	0.566	0.586	0.737	0.503	0.876	0.541	0.807	0.675	0.675	0.650	0.723	1.628	1.442	1.297	1.362	0.665	0.753	0.870	0.870	0.627	1.231	0.639	1.231	0.639	0.566	0.617	0.577	
30. OUT GROUP Holarrhena pubescens	1.056	0.921	0.949	1.441	1.060	1.546	1.088	1.384	1.230	1.230	1.060	1.694	1.076	1.135	1.196	1.196	1.781	1.264	1.229	1.229	0.993	1.076	1.637	1.076	1.637	0.894	0.982	1.084	1.553

Figure 3.10: Estimation of Evolutionary Divergence between ITS Sequences

3.3.3.3.2. Species Resolution

3.3.3.3.2.1. Distance Based Method

The distance matrix was prepared from 30 taxa. Among 30 taxa, four pair of taxa exhibited zero distance value (Figure 3.10). These four pairs showing zero divergence value were, *E. ochreata* – *E. herbacea*, *N. plicata* (green variant) – *N. plicata* (purple variant), *P. lawii* – *P. constrictus* (variant) and *P. plantagineus* – *P. constrictus*. Thus, the species discrimination rate for ITS was 80%, which was higher compared to rbcL and matK. Some of the closely allied species namely *A. ringens* – *A. maculosa*, *R. retusa* – *A. praemorsa*, *H. furcifera* – *H. marginata*, showing zero divergence value in rbcL and matK were well resolved with ITS. In comparison with other members of Orchidaceae, interspecific divergence value was higher (1.628) between *Z. strateumatica* – *H. gibsonii* (Figure 3.10).

3.3.3.3.2.2. Phylogenetic tree Method

The best substitution model was K2P model as it shows lowest BIC and AIC value *i.e.* 3440.735 and 3085.191 respectively (Figure 3.11). The Maximum Likelihood tree was constructed using the K2P model (Kimura, 1980), with 1000 bootstrap replicates. The analysis involved 30 nucleotide sequences of 103 bp in the final dataset.

30 taxa were evaluated, among which an out group formed a separate clade from the Orchidaceae member. The members of both subfamily *viz.* Orchidoideae and Epidendroideae formed two separate groups using the locus. The species with zero divergence value in distance method were clustered together using this method also. Thus, the species resolution was 80.00% (Figure 3.12).

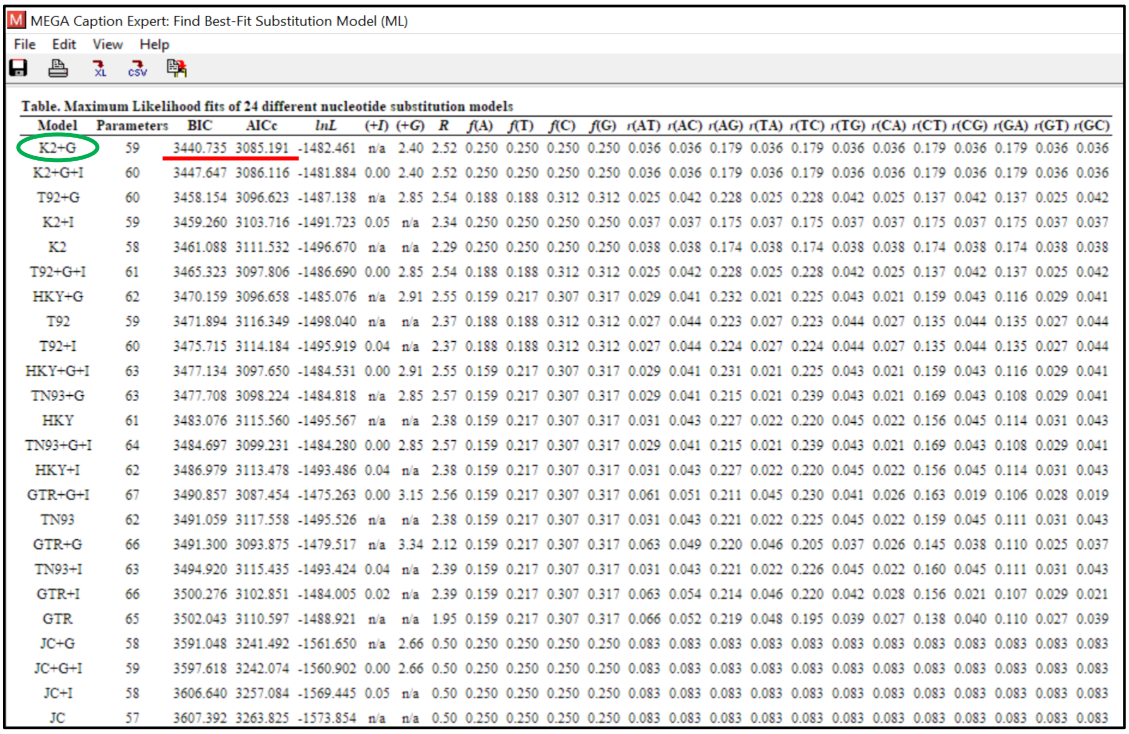


Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	*lnL*	(+*I*)	(+*G*)	*R*	*f*(A)	*f*(T)	*f*(C)	*f*(G)	*r*(AT)	*r*(AC)	*r*(AG)	*r*(TA)	*r*(TC)	*r*(TG)	*r*(CA)	*r*(CT)	*r*(CG)	*r*(GA)	*r*(GT)	*r*(GC)
K2+G	59	3440.735	3085.191	-1482.461	n/a	2.40	2.52	0.250	0.250	0.250	0.250	0.036	0.036	0.179	0.036	0.179	0.036	0.036	0.179	0.036	0.179	0.036	0.036
K2+G+I	60	3447.647	3086.116	-1481.884	0.00	2.40	2.52	0.250	0.250	0.250	0.250	0.036	0.036	0.179	0.036	0.179	0.036	0.036	0.179	0.036	0.179	0.036	0.036
T92+G	60	3458.154	3096.623	-1487.138	n/a	2.85	2.54	0.188	0.188	0.312	0.312	0.025	0.042	0.228	0.025	0.228	0.042	0.025	0.137	0.042	0.137	0.025	0.042
K2+I	59	3459.260	3103.716	-1491.723	0.05	n/a	2.34	0.250	0.250	0.250	0.250	0.037	0.037	0.175	0.037	0.175	0.037	0.037	0.175	0.037	0.175	0.037	0.037
K2	58	3461.088	3111.532	-1496.670	n/a	n/a	2.29	0.250	0.250	0.250	0.250	0.038	0.038	0.174	0.038	0.174	0.038	0.038	0.174	0.038	0.174	0.038	0.038
T92+G+I	61	3465.323	3097.806	-1486.690	0.00	2.85	2.54	0.188	0.188	0.312	0.312	0.025	0.042	0.228	0.025	0.228	0.042	0.025	0.137	0.042	0.137	0.025	0.042
HKY+G	62	3470.159	3096.658	-1485.076	n/a	2.91	2.55	0.159	0.217	0.307	0.317	0.029	0.041	0.232	0.021	0.225	0.043	0.021	0.159	0.043	0.116	0.029	0.041
T92	59	3471.894	3116.349	-1498.040	n/a	n/a	2.37	0.188	0.188	0.312	0.312	0.027	0.044	0.223	0.027	0.223	0.044	0.027	0.135	0.044	0.135	0.027	0.044
T92+I	60	3475.715	3114.184	-1495.919	0.04	n/a	2.37	0.188	0.188	0.312	0.312	0.027	0.044	0.224	0.027	0.224	0.044	0.027	0.135	0.044	0.135	0.027	0.044
HKY+G+I	63	3477.134	3097.650	-1484.531	0.00	2.91	2.55	0.159	0.217	0.307	0.317	0.029	0.041	0.231	0.021	0.225	0.043	0.021	0.159	0.043	0.116	0.029	0.041
TN93+G	63	3477.708	3098.224	-1484.818	n/a	2.85	2.57	0.159	0.217	0.307	0.317	0.029	0.041	0.215	0.021	0.239	0.043	0.021	0.169	0.043	0.108	0.029	0.041
HKY	61	3483.076	3115.560	-1495.567	n/a	n/a	2.38	0.159	0.217	0.307	0.317	0.031	0.043	0.227	0.022	0.220	0.045	0.022	0.156	0.045	0.114	0.031	0.043
TN93+G+I	64	3484.697	3099.231	-1484.280	0.00	2.85	2.57	0.159	0.217	0.307	0.317	0.029	0.041	0.215	0.021	0.239	0.043	0.021	0.169	0.043	0.108	0.029	0.041
HKY+I	62	3486.979	3113.478	-1493.486	0.04	n/a	2.38	0.159	0.217	0.307	0.317	0.031	0.043	0.227	0.022	0.220	0.045	0.022	0.156	0.045	0.114	0.031	0.043
GTR+G+I	67	3490.857	3087.454	-1475.263	0.00	3.15	2.56	0.159	0.217	0.307	0.317	0.061	0.051	0.211	0.045	0.230	0.041	0.026	0.163	0.019	0.106	0.028	0.019
TN93	62	3491.059	3117.558	-1495.526	n/a	n/a	2.38	0.159	0.217	0.307	0.317	0.031	0.043	0.221	0.022	0.225	0.045	0.022	0.159	0.045	0.111	0.031	0.043
GTR+G	66	3491.300	3093.875	-1479.517	n/a	3.34	2.12	0.159	0.217	0.307	0.317	0.063	0.049	0.220	0.046	0.205	0.037	0.026	0.145	0.038	0.110	0.025	0.037
TN93+I	63	3494.920	3115.435	-1493.424	0.04	n/a	2.39	0.159	0.217	0.307	0.317	0.031	0.043	0.221	0.022	0.226	0.045	0.022	0.160	0.045	0.111	0.031	0.043
GTR+I	66	3500.276	3102.851	-1484.005	0.02	n/a	2.39	0.159	0.217	0.307	0.317	0.063	0.054	0.214	0.046	0.220	0.042	0.028	0.156	0.021	0.107	0.029	0.021
GTR	65	3502.043	3110.597	-1488.921	n/a	n/a	1.95	0.159	0.217	0.307	0.317	0.066	0.052	0.219	0.048	0.195	0.039	0.027	0.138	0.040	0.110	0.027	0.039
JC+G	58	3591.048	3241.492	-1561.650	n/a	2.66	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G+I	59	3597.618	3242.074	-1560.902	0.00	2.66	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	58	3606.640	3257.084	-1569.445	0.05	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC	57	3607.392	3263.825	-1573.854	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

Figure 3.11: Best substitution Model for ITS locus

3.3.3.3.2.3. BLAST analysis

In the BLAST analysis of 30 species, 18 species matched correctly and 10 species were having unique sequences. Only two species were not correctly identified. Thus, the species resolution rate using this method based on this locus was 93.33% (Table 3.7). The species resolution using BLAST method was comparatively higher than the formerly described methods.

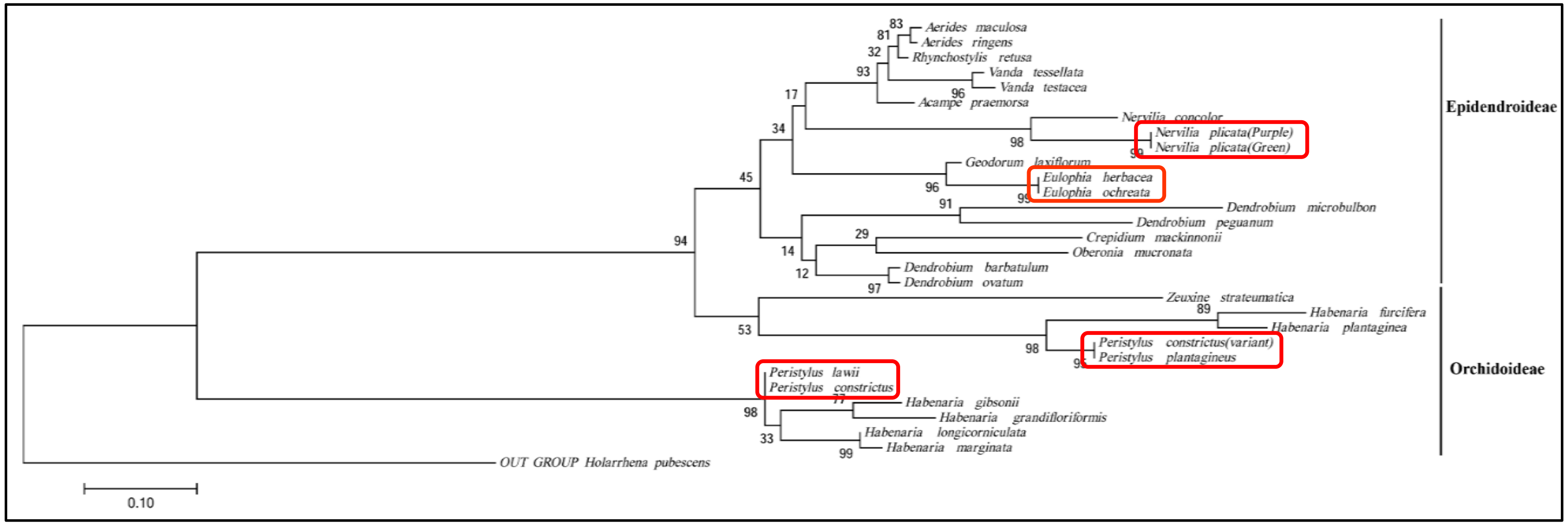


Figure 3.12: Maximum likelihood tree of 30 plant species based on ITS sequences. (species with zero inter-specific divergence were enclosed in red color box)

Table 3.7: Average inter-specific K2P distance and percent species resolution using three loci

Locus	No. of species analyzed	Average interspecific K2P distance	Species discrimination rates		
			Distance based method (%)	Phylogeny method (%)	BLAST method (%)
rbcL	31	0.027	51.61	51.61	83.87
matK	27	0.106	62.96	62.96	88.88
ITS	30	0.777	80.00	80.00	93.33

3.3.3.4. Multilocus Combinations

As none of the locus (rbcL, matK and ITS) yielded 100% species discrimination for all the analyzed species of orchids, different multilocus combinations were too experienced. The species resolution rates were calculated using single as well as in combination with two and/or three loci. The analysis was carried out for only those species, for which all the three locus nucleotide sequences were obtained. The resolution rate was calculated using genetic distance method. Out of 31 orchidaceous taxa, 25 species were analyzed for multilocus combinations. The analysis shows maximum percent species resolution *i.e.* 92% (23 species) using ITS as a single locus followed by matK with 80% (20 species) and rbcL with 72% (18 species). The result obtained by the combinations of two locus *i.e.* rbcL+matK, rbcL+ITS and matK+ITS, were well resolved 21, 23 and 24 species respectively. The three locus combination of rbcL+matK+ITS resulted in 100% species resolution (Figure 3.13).

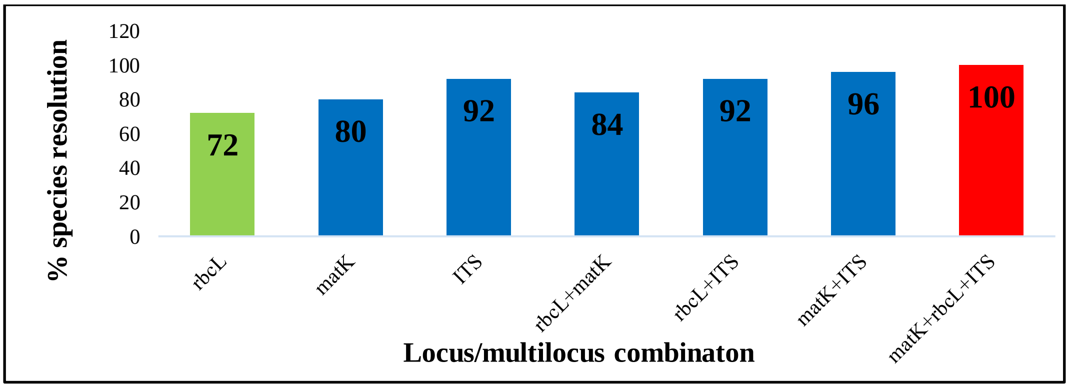


Figure 3.13: Percentage Species resolution for single locus and multiple locus

3.3.3.4.1. Species Resolution Rate for congeneric taxa using Multilocus Combinations

The inter-specific variations and species discrimination rates among the congeneric species were calculated using multilocus combinations. The multilocus combinations tried for the analysis were rbcL+matK, rbcL+ITS, matK+ITS and matK+rbcL+ITS. The species resolution rate was calculated using genetic distance and phylogenetic tree method. The genera having minimum four accessions (*Dendrobium* and *Habenaria*) were analyzed using these methods. The maximum likelihood tree was constructed using GTR model with 1000 bootstrap replicates. The genus wise interspecific variations recorded and percent species resolutions yielded are mentioned in the figure 3.13.

3.3.3.4.1.1. Dendrobium

A total of four species of genus *Dendrobium* namely *D. barbatulum*, *D. microbulbon*, *D. ovatum* and *D. peguanum* were analyzed for multilocus combinations. The average interspecific distances for rbcL+matK, rbcL+ITS, matK+ITS and rbcl+matK+ITS were 0.021, 0.143, 0.112 and 0.077 respectively. The multilocus combinations of rbcL+ITS, matK+ITS and rbcl+matK+ITS well resolved all the species of *Dendrobium*. The combination of matK+rbcL failed to differentiate between *D. barbatulum* and *D. ovatum*, by showing 100 % similarity in phylogenetic tree. (Figure 3.14, 3.15, 3.16 and 3.17).

Figure 3.14: ML tree for *Dendrobium* **rbcL+matK**

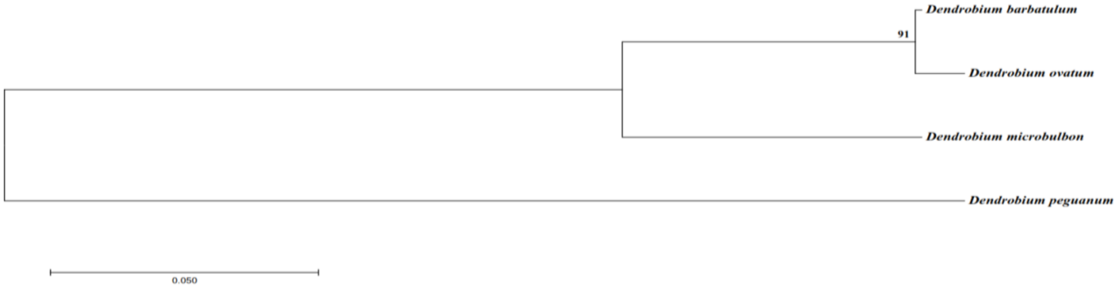


Figure 3.15: ML tree for *Dendrobium* **rbcL+ITS**

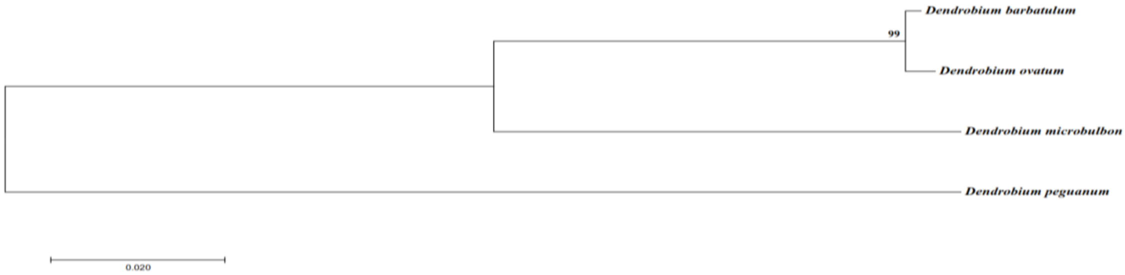


Figure 3.16: ML tree for *Dendrobium* **matK+ITS**

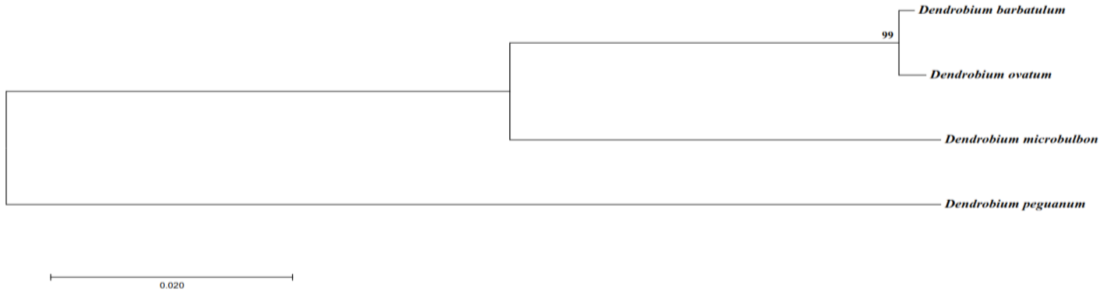


Figure 3.17: ML tree for *Dendrobium* **rbcL+matK+ITS**

3.3.3.4.1.2. Habenaria

Six species of *Habenaria* namely *H. furcifera*, *H. gibsonii*, *H. grandifloriformis*, *H. marginata*, *H. longicorniculata* and *H. plantaginea* were evaluated. As the nucleotide sequence of *H. longicorniculata* for matK locus was not obtained, therefore it does not qualify for these multilocus sequencing. So only five species of *Habenaria* were used for these multilocus studies *i.e.* rbcL+matK, rbcL+ITS, matK+ITS and rbcl+matK+ITS. The average interspecific distances for rbcL+matK, rbcL+ITS, matK+ITS and rbcl+matK+ITS were 0.029, 0.117, 0.180 and 0.109 respectively. The multilocus combinations of matK+ITS and rbcl+matK+ITS resolved all the species of *Habenaria* (Figure 3.20 and 3.21) while matK+rbcL and rbcL+ITS shown zero interspecific distance between *H. furcifera* – *H. marginata*, thus failed to differentiate these species (Figure 3.18, 3.19, 3.20 and 3.21).

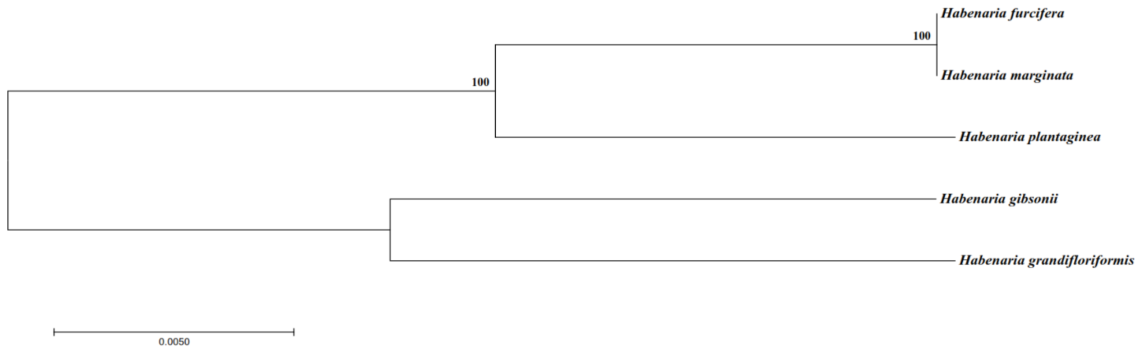


Figure 3.18: ML tree for *Habenaria* **rbcL+matK**

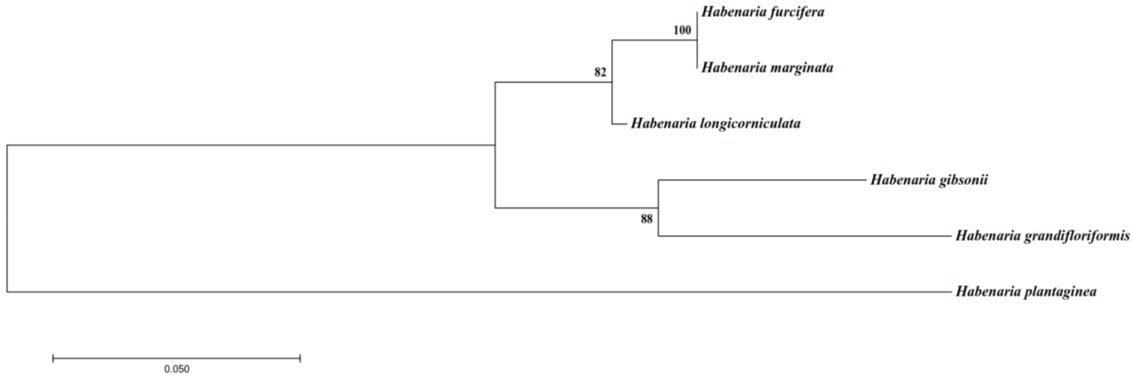


Figure 3.19: ML tree for *Habenaria* **rbcL+ITS**

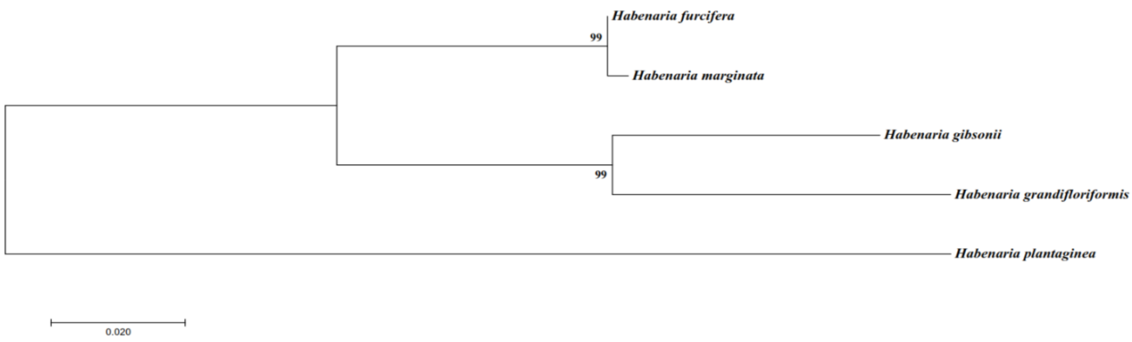


Figure 3.20: ML tree for *Habenaria* **matK+ITS**

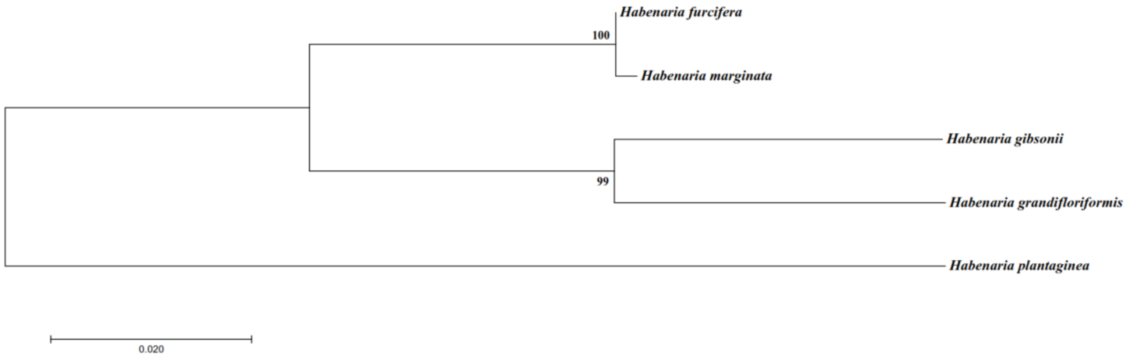


Figure 3.21: ML tree for *Habenaria* **rbcL+matK+ITS**

3.4. CONCLUSION

The current study concludes that, individually ITS is the best locus to identify different species of orchids, as the species discrimination rate for this locus was higher compared to matK and rbcL. The multilocus combination of rbcL+matK+ITS successfully resolved all the 25 species of orchids, thus, it could be used to differentiate the orchidaceous taxa. At congeneric level, closely allied species of the genus *Dendrobium* and *Habenaria* were well resolved using the combination of matK+ITS and rbcL+matK+ITS. Thus, these multilocus combinations could be used to identify the species belonging of these genera. This study was a basic approach to identify some species of family Orchidaceae using DNA Barcoding.

3.5. REFERENCES

Asahina, H., Shinozaki, J., Masuda, K., Morimitsu, Y., Satake, M., 2010. Identification of medicinal *Dendrobium* species by phylogenetic analyses using matK and rbcL sequences. J. Nat. Med. 64, 133 – 138.

Barrett, R.D.H., Hebert, P.D.N., 2005. Identifying spiders through DNA barcodes. Can. J. Zool. 83, 481 – 491. https://doi.org/10.1139/z05-024

Cameron, K.M., 2004. Utility of plastid psaB gene sequences for investigating intrafamilial relationships within Orchidaceae. Mol. Phylogenet. Evol. 31, 1157 – 1180. https://doi.org/10.1016/j.ympev.2003.10.010

CBOL, P.W.G., 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. 106, 12794 – 127947. https://doi.org/10.1073/pnas.0905845106

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., Berg, C. Van Den, Madriñán, S., Petersen, G., Seberg, O., Jørgsensen, T., Cameron, K.M., Pedersen, N., Hedderson, T.A.J., Conrad, F., Salazar, G.A., James, E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L., Wilkinson, M., 2007. A Proposal for a Standardised Protocol to Barcode All Land Plants. Taxon 56, 295 – 299.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N., Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long-term goals. Philos. Trans. R. Soc. B Biol. Sci. 360, 1889 – 1895.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., Leon, C., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5, 1 – 8. https://doi.org/10.1371/journal.pone.0008613

China Plant BOL, G., Li, D.Z., Gao, L.M., Li, H.T., Wang, H., Ge, X.J., Liu, J.Q., Chen, Z.D., Zhou, S.L., Chen, S.L., Yang, J.B., Fu, C.X., Zeng, C.X., Yan, H.F., Zhu, Y.J., Sun, Y.S., Chen, S.Y., Zhao, L., Wang, K., Yang, T., Duan, G.W., 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc. Natl. Acad. Sci. 108, 19641 – 19646. https://doi.org/10.1073/pnas.1104551108

De Vere, N., Rich, T.C.G., Trinder, S.A., Long, C., 2015. DNA Barcoding for Plants, in: Batley, J. (Ed.), Plant Genotyping: Methods and Protocols. Springer New York, New York, NY, 101 – 118. https://doi.org/10.1007/978-1-4939-1966-6_8

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11 – 15.

Dressler, R.L., 1993. Phylogeny and Classification of the Orchid Family. Cambridge University Press, 1 – 314.

Eaton, M.J., Meyers, G.L., Kolokotronis, S.O., Leslie, M.S., Martin, A.P., Amato, G., 2010. Barcoding bushmeat: Molecular identification of Central African and South American harvested Vertebrates. Conserv. Genet. 11, 1389 – 1404.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M., Barrett, S.C.H., 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3, 1 – 12. https://doi.org/10.1371/journal.pone.0002802

Fazekas, A.J., Kuzmina, M.L., Newmaster, S.G., Hollingsworth, P.M., 2012. DNA Barcoding Methods for Land Plants, in: Kress, W.J., Erickson, D.L. (Eds.), DNA Barcodes: Methods and Protocols. Humana Press, Totowa, 223 – 252.

Gale, S.W., Li, J., Kinoshita, A., Yukawa, T., 2015. Studies in Asian Nervilia (Orchidaceae) V: *Nervilia futago*, a Cryptic New Species from Southwest Japan Confirmed by Morphological, Cytological, and Molecular Analyses. Syst. Bot. 40, 413 – 425. https://doi.org/10.1600/036364415X688772

Ghorbani, A., Gravendeel, B., Selliah, S., Zarré, S., de Boer, H., 2017. DNA barcoding of tuberous Orchidoideae: a resource for identification of orchids used in Salep. Mol. Ecol. Resour. 17, 342 – 352. https://doi.org/10.1111/1755-0998.12615

Ginibun, F.C., Saad, M.R.M., Hong, T.L., Othman, R.Y., Khalid, N., Bhassu, S., 2010. Chloroplast DNA barcoding of *Spathoglottis* species for genetic conservation. ISHS Acta Hortic., I Int. Orchid Symp. 878, 453 – 460.

Gravendeel, B., Chase, M.W., De Vogel, E.E., Roos, M.C., Mes, T.H.M., Bachmann, K.,

2001. Molecular phylogeny of *Coelogyne* (Epidendroideae; Orchidaceae) based on plastid RFLPS, matK and nuclear ribosomal ITS sequences: Evidence for polyphyly. Am. J. Bot. 88, 1915 – 1927. https://doi.org/10.2307/3558367

Gregory, T.R., 2005. DNA barcoding does not compete with taxonomy. Nature 434, 1067. https://doi.org/10.1038/4341067b

Guo, Y.Y., Huang, L.Q., Liu, Z.J., Wang, X.Q., 2016. Promise and challenge of DNA barcoding in Venus slipper (*Paphiopedilum*). PLoS One 11, 1 – 13. https://doi.org/10.1371/journal.pone.0146880

Hartvig, I., Czako, M., Kjaer, E.D., Nielsen, L.R., Theilade, I., 2015. The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). PLoS One 10, 1 – 24. https://doi.org/10.1371/journal.pone.0138231

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313 – 321. https://doi.org/10.1098/rspb.2002.2218

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. PLoS Biol. 2(10), 1657 – 1663. https://doi.org/10.1371/journal.pbio.0020312

Hegde, S.N., 1996. Orchid wealth of India. Arunachal For. News 14, 6 – 19.

IUCN, 2017. The IUCN Red List of Threatened Species. URL http://www.iucnredlist.org/

Jalal, J.S., Rawat, G.S., Kumar, P., 2008. Abundance and habitat types of orchids in Gori valley, Eastern Uttarakhand. J. Orchid Soc. India 22, 63 – 67.

Jeanson, M.L., Labat, J.N., Little, D.P., 2011. DNA barcoding: A new tool for palm taxonomists? Ann. Bot. 108, 1445 – 1451. https://doi.org/10.1093/aob/mcr158

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111 – 120.

Kress, W.J., Erickson, D.L., 2007. A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. PLoS One 2, 1 – 10. https://doi.org/10.1371/journal.pone.0000508

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. 102, 8369 – 8374. https://doi.org/10.1073/pnas.0503123102

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870 – 1874. https://doi.org/10.1093/molbev/msw054

Lahaye, R., Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G., Savolainen, V., 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci. 105, 2923 – 2928. https://doi.org/10.1073/pnas.0709936105

Li, F.W., Kuo, L.Y., Rothfels, C.J., Ebihara, A., Chiou, W.L., Windham, M.D., Pryer, K.M., 2011. rbcl and matk earn two thumbs up as the core DNA barcode for ferns. PLoS One 6, 1 – 7. https://doi.org/10.1371/journal.pone.0026597

Marcial-quino, J., Mendoza-espinoza, J.A., Sierra-palacios, E., 2015. DNA Barcoding : An Alternative for the Identification of the Medicinal Plants Employed in Mexico. J. Plant Sci. 10, 116 – 127. https://doi.org/10.3923/jps.2015.116.127

Miwa, H., Odrzykoski, I.J., Matsui, A., Hasegawa, M., Akiyama, H., Jia, Y., Sabirov, R., Takahashi, H., Boufford, D.E., Murakami, N., 2009. Adaptive evolution of rbcL in *Conocephalum* (Hepaticae, bryophytes). Gene 441, 169 – 175. https://doi.org/10.1016/j.gene.2008.11.020

Muellner, A.N., Schaefer, H., Lahaye, R., 2011. Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). Mol. Ecol. Resour. 11, 450 – 460.

Parveen, I., 2012. Evaluation of the Recommended Loci for DNA Barcoding of Indian Orchids. University of Delhi, Ph.D. Thesis.

Parveen, I., Singh, H.K., Raghuvanshi, S., Pradhan, U.C., Babbar, S.B., 2012. DNA barcoding of endangered Indian *Paphiopedilum* species. Mol. Ecol. Resour. 12, 82 – 90. https://doi.org/10.1111/j.1755-0998.2011.03071.x

Robins, J.H., Hingston, M., Matisoo-Smith, E., Ross, H.A., 2007. Identifying Rattus species using mitochondrial DNA. Mol. Ecol. Notes 7, 717 – 729.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74, 5463 – 5467.

Saunders, G.W., 2008. A DNA barcode examination of the red algal family Dumontiaceae in Canadian waters reveals substantial cryptic species diversity. 1. The foliose Dilsea – Neodilsea complex and Weeksia. Botany 86, 773 – 789.

Schindel, D.E., Miller, S.E., 2005. DNA barcoding a useful tool for taxonomists. Nature 435, 17. https://doi.org/10.1038/435017a

Seena, S., Pascoal, C., Marvanová, L., Cássio, F., 2010. DNA barcoding of fungi: A case study using ITS sequences for identifying aquatic hyphomycete species. Fungal Divers. 44, 77 – 87. https://doi.org/10.1007/s13225-010-0056-y

Singh, B.P., Srivastava, U., 2004. Plant Genetic Resources in Indian Perspective: Theory and Practices. Directorate of Information and Publications of Agriculture, Indian Council of Agricultural Research, 1 – 715.

Singh, H., Parveen, I., Raghuvanshi, S., Babbar, S.B., 2012. The loci recommended as universal barcodes for plants on the basis of floristic studies may not work with congeneric species as exemplified by DNA barcoding of *Dendrobium* species. BMC Res. Notes 5, 1 – 11. https://doi.org/10.1186/1756-0500-5-42

Soliva, M., Kocyan, A., Widmer, A., 2001. Molecular phylogenetics of the sexually deceptive orchid genus *Ophrys* (Orchidaceae) based on nuclear and chloroplast DNA sequences. Mol. Phylogenet. Evol. 20, 78 – 88. https://doi.org/10.1006/mpev.2001.0953

Van den Berg, C., Higgins, W., Dressler, R., Whitten, W., Soto-Arenas, M., Culham, A., Chase, M., 2000. A phylogenetic analysis of Laeliinae (Orchidaceae) based on sequence data from internal transcribed spacers (ITS) of nuclear ribosomal DNA. Lindleyana 15, 96 – 114.

Varma, A., Padh, H., Shrivastava, N., 2007. Plant genomic DNA isolation: An art or a science. Biotechnol. J. 2, 386 – 392. https://doi.org/10.1002/biot.200600195

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia’s fish species. Philos. Trans. R. Soc. B Biol. Sci. 360, 1847 – 1857.

Xiang, X.G., Hu, H., Wang, W., Jin, X.H., 2011. DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): A test of DNA barcode candidates. Mol. Ecol. Resour. 11, 1012 – 1021. https://doi.org/10.1111/j.1755-0998.2011.03044.x

Xiao, J.H., Wang, N.X., Li, Y.W., Murphy, R.W., Wan, D.G., Niu, L.M., Hu, H.Y., Fu, Y.G., Huang, D.W., 2010. Molecular approaches to identify cryptic species and polymorphic species within a complex community of fig wasps. PLoS One 5, 1 – 7. https://doi.org/10.1371/journal.pone.0015067

Yao, H., Song, J.Y., Ma, X.Y., Liu, C., Li, Y., Xu, H.X., Han, J.P., Duan, L.S., Chen, S.L., 2009. Identification of *Dendrobium* Species by a Candidate DNA Barcode Sequence: The Chloroplast psbA-trnH Intergenic Region. Planta Med. 75, 667 – 669. https://doi.org/10.1055/s-0029-1185385

Yesson, C., Bárcenas, R.T., Hernández, H.M., de la Luz Ruiz-Maqueda, M., Prado, A., Rodríguez, V.M., Hawkins, J.A., 2011. DNA barcodes for Mexican Cactaceae, plants under pressure from wild collecting. Mol. Ecol. Resour. 11, 775 – 783. https://doi.org/10.1111/j.1755-0998.2011.03009.x