



# DNA Barcoding of Endemic Plants of Nepal

M.Sc. Thesis  
(2024)

## Submitted To:

**Central Department of Biotechnology**

Tribhuvan University

Kirtipur, Kathmandu, Nepal

## Submitted By:

Bibek Prasad Poudel

Roll no: BT 605/075

TU Registration no: 5-2-241-289-2012

## Supervisors

Prof. Krishna Das Manandhar  
Ph.D.  
(Head of Department)  
Central Department of  
Biotechnology  
Tribhuvan University, Kirtipur,  
Nepal

Ram Chandra Poudel, Ph.D.  
Senior Scientific Officer  
Nepal Academy of Science and  
Technology  
(NAST)  
Khumaltar, Lalitpur,  
Nepal



**Tribhuvan University**  
**Central Department of Biotechnology**  
**Kirtipur, Kathmandu, Nepal**

Date.....

**RECOMMENDATION**

This is to certify that the research work entitled “DNA barcoding of endemic plants of Nepal” has been carried out by **Mr. Bibek Prasad Poudel** under our supervision. This thesis work was performed for the partial fulfillment of the Master of Science in Biotechnology under the course code BT 605/075. The result presented here is his original findings. We, hereby, recommend this thesis for final evaluation.

.....

**(Supervisor)**

Prof. Krishna Das Manandhar,  
Ph.D. (Head of Department)

Central Department of  
Biotechnology

Tribhuvan University, Kirtipur,  
Nepal

.....

**(Supervisor)**

Ram Chandra Poudel, Ph.D.  
Senior Scientific Officer  
Nepal Academy of Science and  
Technology (NAST)  
Khumaltar, Lalitpur,  
Nepal



**Tribhuvan University**  
**Central Department of Biotechnology**  
**Kirtipur, Kathmandu, Nepal**

Date.....

**CERTIFICATE OF EVALUATION**

This is to certify that this thesis entitled **“DNA barcoding of the endemic plants of Nepal”** presented to evaluation committee by Mr. Bibek Prasad Poudel is found satisfactory for the partial fulfillment of Master of Science in Biotechnology.

**Signature**  
**Head of Department**

**Signature**  
**External Examiner**

**Signature**  
**Internal Examiner**

**Signature**  
**Supervisor(s)**

## ACKNOWLEDGEMENT

This study entitled “**DNA barcoding of endemic plants of Nepal**” has been conducted to satisfy the partial requirements for the degree of Master of Biotechnology of Tribhuvan University.

First and foremost, I would like to extend my immense gratitude to Prof. Dr. Krishna Das Manandhar and Dr. Ram Chandra Poudel. Their valuable supervision and guidance have been the major boost in completing this study. I am highly indebted and very thankful for their continuous support and constructive suggestions that have enabled this research project to achieve its present form.

Furthermore, I would like to express my deepest appreciation to Mrs. Neesa Rana, Chief of the science faculty; Dr. Jyoti Maharjan, Unit chief of Molecular Biotechnology unit; Dr. Deegendra Khadka, Senior Scientific officer; and the entire NAST family for providing a good laboratory environment with the required facilities and valuable supervision and guidance in completing this study. I cannot express enough thanks to them for their continuous support and supervision, which helped me complete this study in its present form. Similarly, profound gratitude goes to Mrs. Archana Chataut, Shreesti Shrestha, and Mr. Md Siddiqui, the Research Assistant of NAST, for their constant faith, supervision, guidance, and support.

My sincere gratitude to Dr. Bhaskar Adhikari, Royal Botanic Garden Edinburgh, Dr. Yadav Uprety, Central Department of Botany - TU, Dr. Mitra Pathak, Department of Plant Resources, Mr. Deepak Lamichhani and Dr. Gaurav Parmar, National Botanical Garden, Gowawari Lalitpur for providing silica gel materials of endemic plant specimens, I would also like to express my gratitude to my parents. And finally, I would like to thank my senior, Suraj Chand and friend Srijana Budathoki for their kind cooperation and encouragement, which helped me to complete this thesis.

Last but not the least, I would like to thank NAST for providing the reagents/consumables, resource materials, supportive environment, and laboratory facilities, until the completion of the project.

## LIST OF ABBREVIATION

%	Percent
°	Degree
BLAST	Basic Local Alignment Search Tool
BOLD	Barcode of Life System
Bp	Base Pair
C	Celsius
CBOL	Consortium for the Barcode of life
CpDNA	Chloroplast DNA
CTAB	Cetyltrimethylammonium bromide
DB	Diabetic Mellitus
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetra acetic acid
ITS	Internal Transcribed Spacer
ETBr	Ethidium Bromide
matK	Maturase k
MEGA	Molecular Evolutionary Genetic Analysis
Mg/ml	Milligrams per milliliter
ml	Milliliter
min	Minutes
NAST	Nepal Academy of Science and Technology
NCBI	National Center for Biotechnology Information
PCR	Polymerase Chain Reaction
RbcL Gene	ribulose bisphosphate carboxylase Gene
Sec	Seconds
Spp	Species
UV	Ultraviolet
μl	Microliter
μm	Micrometer

# TABLE OF CONTENTS

<b>RECOMMENDATION</b> .....	<b>ii</b>
<b>CERTIFICATE OF EVALUATION</b> .....	<b>iii</b>
<b>ACKNOWLEDGEMENT</b> .....	<b>iv</b>
<b>LIST OF ABBREVIATION</b> .....	<b>v</b>
<b>TABLE OF CONTENTS</b> .....	<b>vi</b>
<b>LIST OF FIGURES</b> .....	<b>x</b>
<b>LIST OF TABLES</b> .....	<b>xi</b>
<b>ABSTRACT</b> .....	<b>xii</b>
<b>CHAPTER 1</b> .....	<b>1</b>
<b>INTRODUCTION</b> .....	<b>1</b>
1.1 Background .....	1
1.2 Endemic plants for community .....	1
1.3 Morphology, phenotypic identification and problems .....	2
1.4 DNA barcoding .....	2
1.5 DNA barcoding limitations .....	4
1.6 Markers for DNA barcoding .....	5
1.7 Plants of Nepal .....	6
1.8 Endemic plants in Nepal .....	6
1.9 Contribution of DNA barcoding in conservation .....	7
1.9.1 Community level .....	8
1.9.2 Scientific sectors .....	8
1.10 Rationale .....	8
1.11 Research objectives .....	9
1.11.1 General objectives .....	9
1.11.2 Specific objectives .....	9
1.12 Research hypothesis .....	9
<b>CHAPTER 2</b> .....	<b>10</b>
<b>LITERATURE REVIEW</b> .....	<b>10</b>
2.1 Molecular markers for endemic plant identification .....	10
2.2 Errors in DNA barcoding .....	11
2.3 Plastid genome for marker selection .....	11
2.3.1 Chloroplast markers .....	12

2.3.1.1 matK (maturase K gene) .....	12
2.3.1.2 RbcL.....	12
2.3.1.3 TrnH-psbA .....	13
2.3.1.4 TrnIF-trnIC .....	14
2.4 Nuclear markers.....	15
2.5 Review on technique and instrumentation of concern .....	16
2.5.1 Herbarium and sample preparation .....	16
2.5.2 DNA extraction.....	16
2.5.3 DNA sequencing and PCR principles.....	17
2.5.4 Sequence analysis .....	18
2.5.4.1 Bioedit and Sequencer .....	19
2.5.4.2 Blast.....	19
<b>CHAPTER 3.....</b>	<b>20</b>
<b>MATERIALS AND METHODS .....</b>	<b>20</b>
3.1 Sample collection .....	21
3.2 DNA extraction.....	21
3.2.1 DNeasy Plant Mini Kit DNA extraction .....	21
3.3 Marker and primer selection .....	22
3.4 Polymerase chain reaction (PCR) amplification and PCR success rate .....	22
3.5 Agarose gel electrophoresis.....	24
3.6 Purification of the PCR products.....	25
3.7 DNA sequencing.....	25
3.7.1 Cycle sequencing.....	25
3.8 Purification of the synthesized DNA fragments.....	26
3.9 Sequencing: Capillary electrophoresis.....	27
3.10 Sequence editing and alignment .....	27
3.11 NCBI BLAST.....	27
3.12 Nucleotide composition.....	28
3.13 Genetic distance calculation .....	28
3.14 Phylogenetic tree construction and distance estimation.....	29
<b>CHAPTER 4.....</b>	<b>30</b>
<b>RESULTS .....</b>	<b>30</b>
4.1 Sampling.....	30
4.2 Sampling and preservation .....	32

4.3 DNA extraction.....	32
4.4 Amplification of markers .....	33
4.5 PCR clean up.....	35
4.6 PCR and Sequencing efficiency .....	35
4.7 Length of markers .....	36
4.8 Nucleotide composition.....	41
4.9 Sequence divergence .....	47
4.10 Comparing Sphagnum.....	60
4.11 Effectiveness of markers across different families .....	61
4.10 Phylogenetic inference: Cladogram.....	62
4.11 Phylogeny.....	66
<b>CHAPTER 5.....</b>	<b>68</b>
<b>DISCUSSION.....</b>	<b>68</b>
5.1 Identification challenges.....	68
5.2 DNA extraction and PCR .....	68
5.3 Sequences analysis and interpretation.....	69
5.4 Variations in lengths .....	69
5.5 Utilities of marker .....	70
5.5.1 RbcL.....	70
5.5.2 ITS.....	70
5.5.3 matK.....	71
5.5.4 trnH-psbA.....	71
5.5.5 TrnIF-trnIC .....	71
5.6 Phylogeny.....	72
5.7 Effectiveness of marker in different families.....	72
5.7.1 Acanthaceae.....	72
5.7.2 Apocynaceae .....	72
5.7.3 Asteraceae .....	73
5.7.4 Begoniaceae.....	73
5.7.5 Berberidaceae .....	73
5.7.6 Campanulaceae.....	73
5.7.7 Caryophyllaceae.....	73
5.7.8 Ericaceae .....	73
5.7.9 Fabaceae .....	74

5.7.10 Hypericaceae.....	74
5.7.11 Lamiaceae .....	74
5.7.12 Onagraceae .....	74
5.7.13 Orobanchaceae .....	74
5.7.14 Papaveraceae.....	75
5.7.15 Primulaceae .....	75
5.7.16 Ranunculaceae .....	75
5.7.17 Rosaceae .....	75
5.7.18 Rubiaceae.....	75
5.7.19 Saxifragaceae .....	76
5.7.20 Sphagnaceae .....	76
<b>CHAPTER 6.....</b>	<b>77</b>
<b>CONCLUSION .....</b>	<b>77</b>
<b>REFERENCES .....</b>	<b>78</b>

## LIST OF FIGURES

Figure 1 plastid markers Schematic diagram of complete plastome of Podocarpus Latifolius showing rbcL, matK, trnL, and trnH-psbA markers used in this study (Saina et al., 2019). .....	14
Figure 2 A diagram depicting the ribosome coding nuclear DNA in various groups, illustrating the ITS region and the specific location where a primer binds to amplify distinct regions of nuclear DNA (Cheng et al., 2016b). .....	16
Figure 3 An overall study design in flowchart.....	20
Figure 4 Bargraph showing number of plants from different families used in this study .....	32
Figure 5 (a) shows the positive amplification of rbcL region and (b) shows positive amplification of trnH-psbA region along with 100bp ladder .....	33
Figure 6 (c) shows positive amplification of matK region, (d) shows positive amplification of ITS region and (e) shows the positive amplification of trnIF-trnIC region along with primer .....	34
Figure 7 Bargraph showing PCR and sequencing efficiency of different markers.....	35
Figure 8 phylogenetic tree on basis of rbcL region using Neighbor-joining method.....	62
Figure 9 phylogenetic tree on basis of ITS region using Neighbor-joining method.....	63
Figure 10 phylogenetic tree on basis of matK region using Neighbor-joining method... ..	64
Figure 11 phylogenetic tree on basis of trnH-psbA region using Neighbor-joining method .....	65
Figure 12 phylogenetic tree on basis of trnIF-trnIC region using Neighbor-joining method .....	66

## LIST OF TABLES

Table 1 characteristic of plant barcodes.....	5
Table 2 Selection of markers and primers .....	22
Table 3 PCR cycle of rbcL marker.....	23
Table 4 PCR cycle of ITS marker .....	23
Table 5 PCR cycle of matK marker .....	23
Table 6 PCR cycle of trnH-psbA marker .....	24
Table 7 PCR cycle of trnL-trnF marker .....	24
Table 8 Conditions for sequence reaction .....	26
Table 9 List of plants .....	30
Table 10 PCR and sequencing efficiency of different markers .....	35
Table 11 Length of rbcL markers.....	36
Table 12 length of ITS markers .....	37
Table 13 length of trnH-psbA markers.....	38
Table 14 length of matK markers.....	40
Table 15 length of trnIF-trnIC markers. ....	41
Table 16 Nucleotide composition of rbcL markers.....	42
Table 17 Nucleotide composition of ITS region.....	43
Table 18 Nucleotide composition of matK region.....	45
Table 19 Nucleotide composition of trnH-psbA region.....	46
Table 20 Nucleotide composition of trnIF-trnIC .....	47
Table 21 Sequence divergence observed in rbcL marker, between endemic plants of Nepal and their congenics.....	48
Table 22 Sequence divergence observed in rbcL marker, between congeneric endemic plants of examined during this study.....	50
Table 23 Sequence divergence observed in ITS marker, between endemic plants of Nepal and their congenics .....	51
Table 24 Sequence divergence observed in ITS marker, between congeneric endemic plants of examined during this study .....	53
Table 25 Sequence divergence observed in matK marker, between endemic plants of Nepal and their congenics.....	54
Table 26 Sequence divergence observed in matK marker, between congeneric endemic plants of examined during this study.....	55
Table 27 Sequence divergence observed in trnH-psbA marker, between endemic plants of Nepal and their congenics.....	56
Table 28 Sequence divergence observed in trnH-psbA marker, between congeneric endemic plants of examined during this study.....	58
Table 29 Sequence divergence observed in trnIF-trnIC marker, between endemic plants of Nepal and their congenics.....	59
Table 30 Sequence divergence observed in trnIF-trnIC marker, between congeneric endemic plants of examined during this study.....	60
Table 31 variation in Sphagnum sequences.....	61

## ABSTRACT

The challenge with morphology-based plant identification underscores the significance of molecular-based methods. DNA barcoding is a technique that can be used to rapid, precise and automatable identification of species using the short standard genetic region as species tags. The introductory idea is straight forward a small quantity of DNA is extracted, amplified and then sequenced. The genetic region to be sequenced is chosen so that it is closely identical among the individuals of same species but different between species, and thus sequence can serve as identification tag for species. Due to the presence of large non-coding fragments and the degeneracy of the genetic code, DNA provides more information than protein.

Endemic plants are plants that occur naturally and exclusively to certain geographical area. We barcoded 34 flowering endemic plants and one bryophyte out of 293 endemic flowering plants and 30 endemic bryophytes of Nepal. 9 species with uncertain endemic distribution were also barcoded. With the objective of contributing to gene-based identification of endemic plants and supporting morphology-based categorization, five distinct markers (ITS, matK, rbcL, trnH-psbA, trnI-trnC) were specifically chosen and examined to ascertain their potential in aiding the identification of endemic plants collected from Nepal.

The effectiveness of matK, ITS, rbcL, trnH-psbA, and trnI-trnC was assessed across all individual species under study. The P-distance, measuring genetic divergence, between Nepal's endemic plants and those from neighboring countries, was computed. Among the markers investigated, ITS emerged as the most efficient in identifying Nepal's endemic plants. Notably, none of the individual markers used alone could accurately identify all the endemic plants. When combined together use of ITS, matK, rbcL, trnH-psbA, and trnI-trnC has greater species discrimination. Therefore, employing two or more markers is recommended to effectively identify these endemic plants.

**Keywords:** Barcoding, Endemic plants, PCR, Sequencing, Phylogeny

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Endemic plants are plants that occur naturally and exclusively to certain geographical area, and this phenomenon is called endemism. Endemic species are highly adapted to particular geographical area (Coelho et al., 2020). Endemism is relative – plants can be endemic to certain continent, country, part of country. Usually, endemic plants are found on particular area that are isolated or has unusual climate/environment (Ferreira & Boldrini, 2011). Mountainous terrain, geographical barrier, insularity and climate instability are some of factors thought to generate endemism (Harrison & Noss, 2017). Two types of endemic plants have been long recognized: neo endemic – plant that are recently diverged and that are endemic due to lack of migration or dispersion; and paleo endemic – old plant species that may be widespread before but now are restricted to a particular area (Mishler et al., 2014).

Endemic plants are usually more vulnerable to natural changes and anthropogenic threats. One or more than one set of characteristics found in most of endemic plants which make them vulnerable are: small population, restricted distribution, specific habitat, excessive collection by human, short reproduction capacity. The higher number of characters a plant display, the higher is risk of extinction (Coelho et al., 2020).

### 1.2 Endemic plants for community

Plants have been used for therapeutic purpose for thousands of years, and their constituents have contributed for number of natural based drugs used in western medicine. About one quarter of drugs commonly used drugs is obtained or derived from plants (Sanna et al., 2020). The endemic plants have antimicrobial, anticancer, antimalarial, anthelmintic, antidiabetic, mosquitocidal, antioxidant and anti-inflammatory properties (Ayalew et al., 2022). Endemic plants play a critical role in maintaining local ecosystems, providing habitat and food sources for animals, and contributing to soil health and water quality. Some endemic plants also hold cultural and spiritual significances and ornamental purposes. Some of endemic plants are also used as food and also for agricultural purpose.

### **1.3 Morphology, phenotypic identification and problems**

Morphology is basic and foundation approach for sample collection at site and primary description of plant species. At the same time, there is constant need of experts for an accurate identification of the species as well as the determination of the origin-based relationship. The lack of experienced experts often doesn't allow this to be satisfied and demands supplementary technique such as DNA barcoding (Shneyer & Rodionov, 2019). Morphological identification is also problematic in identification of sibling species – species that are morphologically identical but reproductively isolated from each other (Pereira et al., 2008). Morphological identification may not be accurate when the size of the sample is small and at times when biological sample is deteriorated due to improper storage.

After the development of technologies, first protein analysis based molecular methods were employed for the characterization of species. Although analysis of protein (protein sequencing, immunological reactions, isoenzyme analysis, protein electrophoresis, etc.) have advantages, there were few factors that limited use of protein: rapid degradation of protein in unfavorable condition, cross reaction between protein and differential expression of protein in some tissue (Skarpeid et al., 1998). With the advent of DNA based technologies, the limitation of protein based methods have been circumvented (Pereira et al., 2008).

### **1.4 DNA barcoding**

The built-in constrains of morphology based identification and the decreasing availability of taxonomic expertise propel investigator to find new approach to label the unknown samples with correct identity (Hebert et al., 2007). DNA barcoding is a technique that can be used to rapid, precise and automatable identification of species using the short standard genetic region as species tags (Hebert & Gregory, 2005). As 11-digit universal product codes (UPCs) recognize retail product, standardized DNA sequences can identify the species in the same way (Neigel et al., 2007). The introductory idea is straight forward. A small quantity of DNA is extracted, amplified and then sequenced. The genetic region to be sequenced is chosen so that it is closely identical among the individuals of same species but different between species, and thus sequence can serve as identification tag for species (Nielsen & Matz, 2006). Due to the presence of large non-coding fragments and the degeneracy of the genetic code, DNA provides more information than protein (Pereira et al., 2008).

DNA barcoding is the most appealing complement to morphological identification of the plant species (Wilkinson et al., 2017). Regarded as big science program earlier,

DNA barcoding is also thought of as the renaissance of taxonomy (X. Li et al., 2015). DNA barcoding can be applied in ecological forensics, identifying the species in shortage of taxonomist and assisting to identify the species in some groups, identifying traded materials, food inspection (Hollingsworth et al., 2011a). DNA barcoding is already successfully used in discovery of new species, biodiversity assessments, detection of medicinal plants in mixtures and ecological investigations (Gao et al., 2017).

DNA barcoding, identification of species/taxon using standardized DNA region, entered its golden age in 2003 and is now far more developed due to international initiative. Initially, DNA barcode was first used in 1993 and now it is used by taxonomist, biologist, forensic, biotechnologists and others. Professor Paul Hebert of University of Guelph in 2003 proposed the use of 'DNA barcode' for the identification of taxon (Teuchen et al., 2014). DNA barcoding sensu stricto is the identification of species using a single standardized DNA region which fits the definition of CBOL (Consortium for Barcoding of Life). DNA barcoding sensu lato is the identification of species using any DNA region that has not been recognized as DNA barcode (eg: single-nucleotide polymorphism (SNPS) or with insertions/deletions). The identification of population within species either using a diagnostic marker or difference of marker frequencies among population should not be considered as assignment test and not DNA barcoding (Valentini et al., 2009). Identification of unknown specimen by DNA barcoding involves comparing between unknown specimen to one or more sequences that have already been correctly identified by other means (Neigel et al., 2007).

DNA barcode incorporates a short-standardized sequence of DNA (400-800 bp) that in theory should be easy to generate and characterize for all species of earth. DNA barcoding has enabled users to easily identify the known species and eased the discovery and identification of unknown species. DNA sequence of yet to be identified sample is compared with large online data base as the standard. Successful universal amplification being the primary quality, the qualities DNA barcode should have are (i) contain significant divergence and genetic variability at species-level, (ii) contain conserved flanking sites required for developing universal primers and (iii) have a short sequence to ease DNA extraction and DNA amplification (Kress & Erickson, 2008). Though difficult to achieve amplification, an ideal marker should be short enough to be amplified from degraded DNA. The above-mentioned features provide DNA barcoding with rapid, accurate and automatable species identification (Taberlet et al., 2007). DNA barcoding involves choosing of minimum standard loci which can be efficiently sequenced and can resolve large and diverse specimen to provide comparable data for effortless

interpretation and better resolution as per the principles of barcoding standardization, minimalism and scalability (Hollingsworth et al., 2011a).

### **1.5 DNA barcoding limitations**

While DNA barcoding is meeting the primary goal of species identification and able to contribute to complex taxonomy, application of DNA barcoding is often hindered by the lack of consensus on proper plant marker and low success rate of specimen identification (Armenise et al., 2012). The venture to find single barcode for plant has been thought to be a key driving factor towards plants' DNA barcoding. Unlike other organism groups, plants' DNA barcoding is not easy as it does not simply rely on only DNA marker/locus (Taylor & Harris, 2012). The fact that no single gene can serve as a perfect barcode has been vulnerable spot in DNA barcoding. Expert suggest the use of barcode should tilt more towards finding out intraspecific and interspecific variation instead of marking sequence matches as perfect and imperfect (Nielsen & Matz, 2006).

Besides the use of DNA barcode for identification and assigning an unknown sample to already known species, DNA barcoding is also being used for discovery of new species delineating previously unsampled species as distinct, which is a major criticism and taken as crude approach to proclaim new species on basis of a single gene. DNA barcoding is expected to be used alongside anatomical, morphological and other informative features instead of alternative (Moritz & Cicero, 2004). Using DNA barcoding as alternative of normal taxonomy will be a bad idea and will cause much of meaningful details unused (Will et al., 2005). One of central issue of DNA barcoding is how accurately a single barcode can identify a species? The more there is overlap between the genetic divergence between sister species and genetic variation within species, the less efficient a barcode becomes (Meyer & Paulay, 2005).

Besides human error, identification of species can be misleading for at least three reasons. First, random mutation may cause the sequence to be similar to another species. Second, data base may represent true species. Third, the query may not be similar to database of true species because of random coalescences of lineages in a common ancestral species (Nielsen & Matz, 2006). Instead of being reason for decline in alpha taxonomy, DNA barcoding can aid in taxonomic investigation of cryptic species and clarify problems of synonymy that requires high taxonomic effort (Hebert & Gregory, 2005).

## 1.6 Markers for DNA barcoding

In DNA barcoding of animals, mitochondrial cytochrome oxidase subunit 1 (COI) gene is commonly used because COI is significantly conserved within species but varies highly between species. Furthermore, COI has high mutation rate which enables us to distinguish closely related species that may be morphologically identical (Ratnasingham & Hebert, 2007). COI region can be used in some groups of algae, but these sequences are highly invariant in land plants and hence, are unsuitable as DNA barcodes for all plants (Chase et al., 2005).

Several authors have variable success rates while using different plastid-DNA regions (e.g., *rpoC1*, *rpoB*, *atpF-atpH*, *psbK-psbI*, and *trnH-psbA*). So, there is also difficulty in standardizing which plastid-DNA region will function as reliable DNA barcodes for plant (Vasconcelos et al., 2021). Due to different levels of variability among standard plant barcodes, the use of various commonly used barcodes (e.g. *rbcL*, *trnH-psbA*, ITS, and *matK*) provide resolution at various phylogenetic levels, which eases the study of community phylogenies (Hollingsworth et al., 2016).

**Table 1:** characteristic of plant barcodes

Markers	Genomic source	Type	basepairs
nrITS	Nuclear	Transcribed spacer and 5.8s gene	407-1630
nrITS2	Plastid	Transcriber spacer	157-670
matK	Plastid	Protein coding	862-910
rbcL	Plastid	Protein coding	654-654
trnH-psbA	Plastid	Inter-genic spacer	226-934
rpoB	Plastid	Protein coding	536-590
rpoC1	Plastid	Protein coding	610-622
trnL-F	Plastid	Intron and Inter-genic spacer	201-2114
trnL(p6)	Plastid	Intron	51-135

Three plastid markers (*matK*, *rbcL*, and *trnH-psbA*) and one nuclear gene (ITS) region remain the standard barcode chosen for most of investigation in plants after many broad screenings of genetic region in plant genome. Ribosomal-DNA and plant-plastid barcodes normally have lower discriminatory power and there usually is graded continuum of intraspecific and interspecific distances, with barcodes commonly shared among related species (Hollingsworth et al., 2016). According to evaluation of data across laboratories by CBOL plant working group (2009) among seven plastid loci (*matK*, *ropB*, *rbcL*, *rpoC1*, *trnH-psbA*, *atpF-atpH* and *psbK-psbI*), two-locus combination of *matK* and *rbcL* region is recommended as plant barcode. Ease of recovery of *rbcL* region and discriminatory power of *matK* makes them good

pair. Combining more plastid barcodes provided higher resolution, but with diminishing return (Fazekas et al., 2008). Using two markers as plastid barcodes was more effective in distinguishing between plant species than using only one marker, but none of the other combinations of two or more markers provided a significantly better resolution of species than using both *rbcl* and *matK* markers together (Hollingsworth et al., 2011b).

In phylogenetic studies *rbcl* gene is popularly used and over 50000 sequences are available in gene bank. Though easy to amplify, *rbcl* gene evolves slowly and has low divergence, hence, not suitable for species level identification and cannot be considered as sole DNA barcode for plant (X. Li et al., 2015). *matK* gene has high evolutionary rate, high interspecific diversion and low transition/transversion rate but *matK* gene is difficult to amplify universally with currently available primer sets (Selvaraj et al., 2008a). The *trnL* gene has high level of variation between species and is commonly used in phylogenetic and molecular studies (de Groot et al., 2011a).

## **1.7 Plants of Nepal**

Nestled between the Tibetan plateau and the Gangetic plains, the Nepalese Himalayas provide a wide range of microclimates due to their substantial variations in altitude over a relatively small geographic area. While Nepal occupies less than 0.1% of the Earth's total land area, it stands out as a country with a remarkably high biodiversity. Current assessments of species diversity suggest the presence of 1,001 types of algae, 2,182 varieties of fungi, 850 kinds of lichens, 1,213 types of bryophytes, 550 species of pteridophytes, and 41 gymnosperm species (Rajbhandary, 2020).

## **1.8 Endemic plants in Nepal**

Nepal is diverse in topography, ranging from high mountains in north to tropical forests in south. This diverse topography supports various life-forms, ecosystems and genetic diversity. Varied geography and climate support the environment required for evolution of endemic plants that are not found in other parts of world. Topographic complexity aids in species richness because it may provide micro refugia in times of environmental changes and harsh weather environment (Tiwari et al., 2019a).

An enumeration of the flowering plants of Nepal (volumes 1-3) was published by Department of Plant Resources, National Herbarium and Plant Laboratories from 1978 to 1982AD. Later, Rare, Endemic and Endangered Plants of Nepal was published as compilation of endemic plants in Nepal (Shrestha et al., 1996). This book

became the source for further preparing the list of endemic plants of Nepal. Department of Plant Resources have published the Plant Fact Sheet 2021, providing information on a total figure of endemic plants. Recently, 324 species of endemic flowering plants, 30 species of endemic bryophytes, 48 species of endemic lichens, 12 species of species algae and 131 species of endemic fungi were reported in Nepal (Rajbhandary, 2020). The most recent list revealed that Nepal is home to 293 endemic flowering plant species, distributed across 129 genera and 45 families. This accounts for approximately 5% of Nepal's overall flowering plant species. In contrast, their 2017 findings had previously indicated 312 endemic flowering plant species in Nepal (Rajbhandari et al., 2021).

Regular taxonomic investigation carried in Nepal and other neighboring countries have shown that plant considered to be endemic to Nepal is also being discovered in other countries (Tiwari et al., 2019a). For eg: *Begonia pachtharensis* which was considered as endemic plant of Nepal was found in India (Pradhan et al., 2018). Similarly, *Saussurea ramchaudharyi* was reported as endemic plant of Nepal in northwestern region (Rana et al., 2018).

### **1.9 Contribution of DNA barcoding in conservation**

Endemic species are often restricted in narrow geographical habitat, highly specified environment and often are small population and are considered rare. Hence, they demand special conservational effort (Kruckeber & Rabinowitz, 1985). However, the documentation with regards to distribution, ecology and conversation status of endemic plants is scant in context of Nepal. Recent new studies about these plants have produced little new information due to specimen collection location is not properly identified or is lost/degraded (Tiwari et al., 2019a).

Due to historical reduction of biodiversity, it is necessary to devise and apply methods of biodiversity conservation (Margules & Pressey, 2000). Conservational genetics aims to avoid the extinction of species due to reduction of genetic diversity, while, preserving current genetic diversity and diversification process that are taking at species level (Bobo-Pinilla et al., 2021). Conservational genetics is one of scientific field that is required for the planning, management and sustaining biodiversity. Population genetic data is essential for the applied and conceptual planning for conservation of biodiversity and utilization (Holderegger et al., 2019).

Biodiversity hotspot all around globe are in risk, despite national and international efforts. Identification of priority species and coordinated conservational approach should be taken to preserve biodiversity utilizing available tools including in situ, ex situ and in vitro methods (Reed et al., 2011). DNA barcoding facilitates easy

identification, rapid classification and eases policy makers to determine appropriate methods and scale required for conservation (Krishna Krishnamurthy & Francis, 2012).

### **1.9.1 Community level**

To increase people's appreciation of nature and encourage their involvement in the management and protection of endemic plants, it's important to provide information on the existence, prevalence, distribution, benefits, and threats of these plants. Scientific research can explore the potential benefits of such knowledge across various sectors. In Nepal, the conservation of forests and endemic plants is challenging due to extensive exploitation for urbanization and commercial purposes, despite efforts to preserve species around communities (Tiwari et al., 2019b). DNA barcoding is a promising approach to identify endemic plants, prevent their extinction, and promote their sustainable conservation (Kress et al., 2015).

### **1.9.2 Scientific sectors**

Endemic plants have unique characteristics due to their genetic and morphological variations, that uniqueness attracts many botanist, biotechnologist, and molecular biologists towards understanding more about them. Distribution of endemic plants can be indicator of climate change (Mkala et al., 2023). Hence, scientist may study the how climate impacts various ecological changes. Various endemic plants have medicinal properties that can be used to cure illness. By studying and preserving these endemic plants, we may develop potential drug which is of pharmaceutical and economic advantages (Volenzo & Odiyo, 2020). DNA barcoding requires various tools and has various applications, it would be an opportunity for new biologist to learn not just about endemic plants but other plants as well.

### **1.10 Rationale**

The primary objective of DNA barcoding of endemic plants in Nepal is to create a repository of genetic information that can aid in identifying and classifying these plants accurately and to test morphology-based identification in these species with restricted distribution ranges. This genetic information will supplement the existing identification methods that rely on morphology, ecological data, and molecular data. Often endemic plants are rare and at more risk of extinction. The extant endemic plants should be studied thoroughly in order to conserve them and use them sustainably (Mouga & Mouga, 2021). After development of DNA barcodes of endemic plants, forensic identification can be used to restrict illegal trade of endangered endemic species as suggested by (J. Liu et al., 2018).

In Nepal, it is necessary to conduct a detailed genetic study of endemic plants, as the current information on this group is insufficient given the country's abundant botanical diversity. Additionally, there have been no previous wide-ranging molecular phylogeny and taxonomic studies conducted on the endemic plant species of Nepal.

### **1.11 Research objectives**

The principle objective of this study is to characterize selective endemic species of Nepal through DNA sequencing techniques.

#### **1.11.1 General objectives**

1. DNA barcoding of various endemic plants in Nepal.

#### **1.11.2 Specific objectives**

1. Study of PCR and sequencing efficiency of ITS, rbcL, matK, trnH-psbA, trnL-trnF markers in identifying Nepalese endemic plants.
2. Studying species level identification power of markers used in Nepalese species.
3. Create DNA reference library of endemic plants of Nepal.

### **1.12 Research hypothesis**

Hypotheses of this research are set as following.

The application of DNA barcoding would be useful to develop easy procedure to differentiate different endemic plants.

Alternatively, the resolving power of selected markers could not be enough for taxonomic identification of endemic plants of Nepal.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Molecular markers for endemic plant identification

Expert identification relies on in-depth examination of the morphology of endemic plants belonging to specific families, genera, and species. DNA barcoding serves not as a substitute for phenotype-based identification but rather as an additional scientific tool to confirm such identifications. However, expert assessment falls short when a species lacks distinctive characteristics and key features for naming and categorizing it within the scientific classification. This is precisely where DNA barcoding becomes essential (Neigel et al., 2007).

Rather than depending solely on one method, a more holistic approach for identifying and classifying plant species involves amalgamating various aspects such as morphology (including reproductive structures, leaves, phyllotaxy, and wood structure), anatomy, biochemistry, and DNA characteristics. By employing this comprehensive approach, it becomes possible to accurately identify a species while simultaneously gaining a deeper understanding of its ecological and evolutionary history (Neale & Wheeler, 2019).

After testing various loci from the chloroplast and nuclear genomes using computer simulations and laboratory experiments, it was determined that no single locus could be used as a universal DNA barcode for all plant species. It was soon recognized that multiple loci must be used to create an effective DNA barcode for plants (Hollingsworth et al., 2011b). The Consortium for the Barcode of Life (CBOL), Plant Working Group (2009) suggested combination of *rbcl* and *matK*. However, the discriminating power of these two markers in closely related species were low. So, addition of nuclear internal transcribed spacer (ITS) was proposed to be added along with *rbcl* and *matK* to improve discriminatory power (Raja et al., 2015).

During the Fourth International Barcode of Life Conference held in Adelaide, Australia (<http://www.dnabarcodes2011.org/>), attendees discussed whether a three-locus barcode (*matK* + *rbcl* + *psbA-trnH*) or a two-locus barcode would be better. Ultimately, the two-locus barcode was chosen to avoid higher sequencing costs, so the combination of *rbcl* and *matK* became the preferred and core barcode option (Hollingsworth et al., 2009). The *trnI-F* region is found as an efficient tool in identification of gametophytes and has significant potential to explore fern gametophytes' ecology (Chen et al., 2013).

## **2.2 Errors in DNA barcoding**

DNA barcoding may lead to misinterpretation of sequence data, resulting in two types of errors. False positives occur when genetically distinct individuals within a species are mistakenly identified as a cryptic species. False negatives occur when an individual is not recognized as a unique species due to a lack of sequence variation. These errors can lead to incorrect conclusions about the usefulness of DNA barcoding for species identification and taxonomy (Wiemers & Fiedler, 2007).

Due to the impracticality of collecting a representative sample from every species on earth, global efforts would be required which would encounter many political and practical difficulties. As a result, it is essential to rely on previously collected materials, which often have degraded DNA. This is why plastid and mitochondrial DNA have been considered the best regions to sequence, as these genomes are highly repetitive and more likely to remain intact in herbaria and museums (Chase et al., 2005).

## **2.3 Plastid genome for marker selection**

Plastid genome is small circular DNA molecule found in chloroplast of plant cells. Chloroplast DNA (cpDNA) ranges 120-170 kb and has 100-130 genes responsible for various metabolic activities, including photosynthesis and carbon fixation (Rogalski et al., 2015). Most of DNA barcoding in plant targets chloroplast genome sequence which proves the importance of chloroplast genome. They are maternally inherited, have a slow evolutionary rate, and are highly conserved across plant species (Li et al., 2020). Because plastomes are highly conserved in plants, it is easier to design primers for wide range of taxa that can be amplified in PCR (Ravi et al., 2008).

Plastid markers have a low mutation rate, which is advantageous compared to nuclear markers, because plastids do not undergo recombination and have uniparental inheritance. As a result, they are useful for studying plant populations that have recently undergone speciation and for investigating the evolutionary history of plants over extended time periods (Shaw et al., 2007a). Another important feature of chloroplast genome is high copy number which makes them easier to perform PCR and sequencing relative to low-copy nuclear genes (Hollingsworth et al., 2016). The DNA sequences of chloroplasts can be used as a flexible technique for identifying and categorizing plants, as well as determining their genetic relationships. Chloroplast DNA has been used to understand the evolutionary distance between plants that are both closely and distantly related. No single marker has been discovered that can distinguish between all plant species (Nock et al., 2011).

Although plastid markers have many benefits, they may not always provide an accurate representation of the evolutionary history of a plant group. One possible reason is the occurrence of hybridization, where genetic material is exchanged between two or more plant lineages through hybridization events. As plastid markers are inherited maternally, they may not fully represent the genetic diversity of a hybrid lineage, leading to erroneous phylogenetic conclusions (Raubeson & Jansen, 1992). Similarly, horizontal gene transfer can lead to insertion and incorporation of foreign gene, which may complicate the interpretation of chloroplast marker data (Smith & Lee, 2010).

### **2.3.1 Chloroplast markers**

#### **2.3.1.1 matK (maturase K gene)**

The matK gene is found within the chloroplast and spans approximately 1500 base pairs in length. It is located within the trnK intron and encodes for a maturase-like protein, which plays a role in splicing Group II introns. The two exons of the trnK gene that surround the matK have been lost over time, but the matK gene remains intact (Selvaraj et al., 2008a). The region matK is regarded as a valuable indicator for plant families because of its rapid evolution, and appropriate size. Additionally, it has a low ratio of transition to transversion (X. Li et al., 2015). Furthermore, a large proportion of nucleic acid variation is observed at the first and second codon positions of the gene. Additionally, the gene exhibits a low ratio of transition to transversion mutations and contains mutationally conserved regions. Due to these features, the matK gene has been utilized to infer relationships among families and species in plants (Selvaraj et al., 2008b).

matK as molecular marker is widely used in angiosperms but in non-angiosperm groups its utility is not quite clear. While matK as marker is still explored in bryophytes (mosses, liverworts and hornworts), other markers are found to be more reliable in ferns (Schuettpelz & Kathleen, 2007).

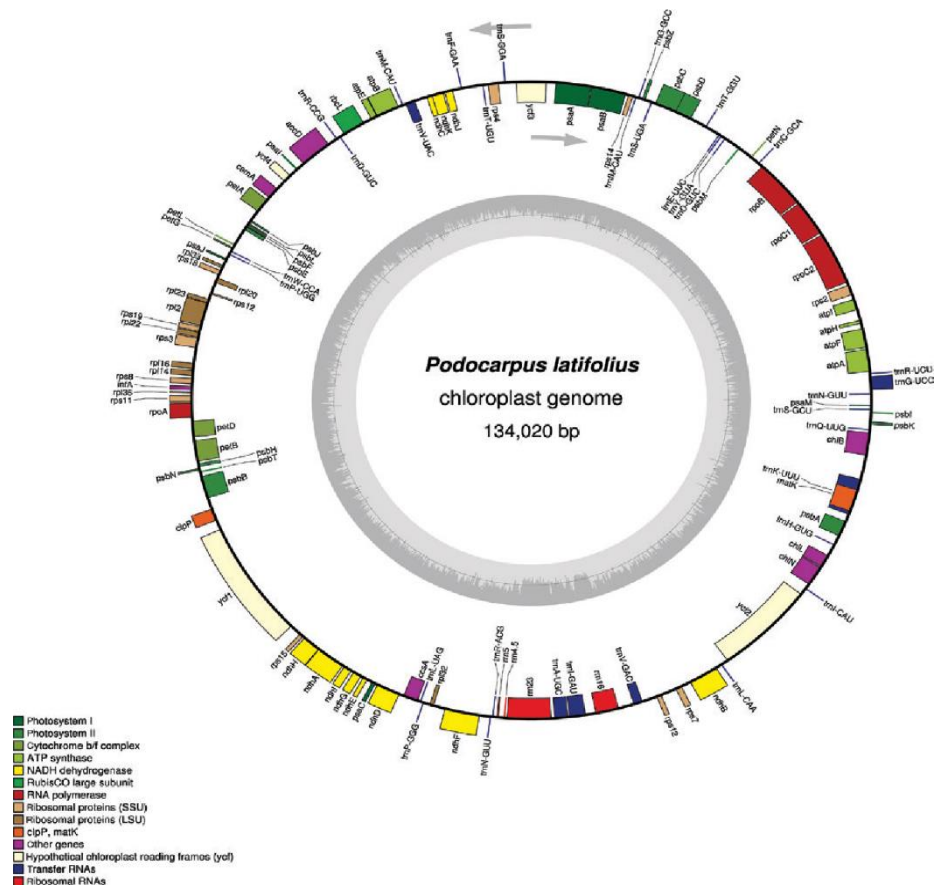
#### **2.3.1.2 RbcL**

The rbcL gene encodes for large subunit of ribulose-biphosphate carboxylase, which is responsible for photosynthesis. Rubisco is involved in photosynthetic carbon dioxide fixation. So, during evolution of photosynthetic plants this gene has been conserved (Käss & Wink, 1995). With length of about 1200 bp this gene is appropriate choice for phylogenetic study at higher taxonomic level (Hasebe et al., 1994).

The *rbcl* gene offers several advantages as a DNA barcode marker, such as its ease of amplification and sequencing in most land plants, as well as its reputation as a benchmark locus in phylogenetic studies due to its ability to provide a trustworthy classification of a taxon into a particular plant family or genus (Kress & Erickson, 2007a). However, due to slow evolving nature, low divergence and low discriminatory ability of this gene, it is not suitable for species level identification. Despite this its limitation, *rbcl* gene is still best candidate for DNA barcoding and *rbcl* with combination of nuclear marker or various plastid marker can accurately identify a plant (X. Li et al., 2015).

### **2.3.1.3 TrnH-psbA**

Currently, *trnH-psbA* is among the most extensively utilized barcode for plastids. The existence of coding sequences that are strongly conserved on both ends enables us to design universal primers (Shaw et al., 2005), with a pair of primer that can amplify almost all angiosperms (Shaw et al., 2007b). The intergenic region of the *trnH-psbA* gene that does not code for any functional molecule shows the highest level of variation in DNA sequence and has a high frequency of insertions or deletions (Kress & Erickson, 2007a). These characters make *trnH-psbA* great marker for plant DNA barcoding. Studies have shown that *trnH-psbA* can identify almost all plants in groups of Hydrocotyle, Dendrobium and Pteridophytes (X. Li et al., 2015). The *psbA-trnH* spacer region has significant challenges with alignment due to complicated evolution of *trnH-psbA* and high rates of insertion and deletions (Chang et al., 2006). Even within closely related taxa, there are substantial variations in length, making it difficult to identify shared sequences at greater taxonomic distances. Furthermore, the length of the *psbA-trnH* spacer varies significantly across different plant groups, with some being less than 300 bp while others are over 1,000 bp due to the inclusion of copies of pseudogenes (Chase et al., 2007). One major issue with using *trnH-psbA* as a standard barcode is the regular inversions that occur in certain plant groups. This can cause significant overestimations of genetic differences and incorrect classification in the phylogenetic tree. Another challenge arises when sequencing longer *trnH-psbA* regions due to the termination of sequencing reads caused by repeated single nucleotides. To obtain high-quality, bi-directional sequences, taxon-specific internal sequencing primers must be designed (X. Li et al., 2015).



**Figure 1:** Plastid markers Schematic diagram of complete plastome of *Podocarpus Latifolius* showing *rbcl*, *matK*, *trnL*, and *trnH-psbA* markers used in this study (Saina et al., 2019).

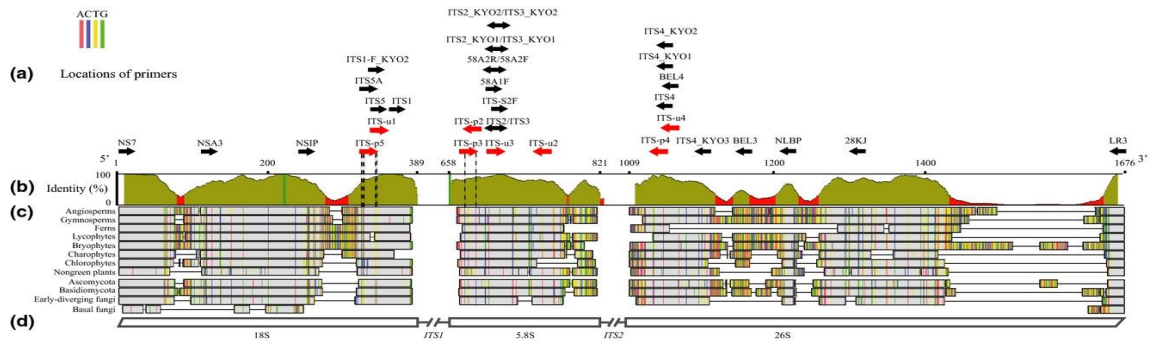
### 2.3.1.4 TrnIF-trnIC

The *trnL-F* region consists of two non-coding regions of the chloroplast genome, the *trnL* intron and the *trnL-F* intergenic spacer (pirie & vargas, 2016). Even though the non-coding region of *trnL* gene is not most variable, it is still the one of most widely used marker for identification phylogeny reconstruction of closely related species (Taberlet et al., 2007). Universal primers for *trnL* region were designed around 15 years ago (Taberlet et al., 1991) and widely used for the phylogenetic study of closely related species and genera. According to Kress et al., 2005 *trnIF* marker was found to be most universal marker, which was successful in amplifying 93% of plant across different species. P6 loop of *trnL* intron can be highly useful for barcoding degraded plant materials (Hollingsworth et al., 2011b) and *trnL* was found to predict pollen better than other markers (Omelchenko et al., 2022).

## 2.4 Nuclear markers

The Internal Transcribed Spacer (ITS) is a segment of nuclear DNA found between the 18s, 5.8s, and 6s ribosomal RNA cistron. This region is highly variable between different species and is therefore a valuable tool in phylogenetic studies aimed at identifying them. In particular, the ITS region has been extensively researched and shown to be more effective than plastid regions in distinguishing between plant species at lower taxonomic levels. As a result, the China plant BOL Group recently recommended the use of ITS as the primary plant barcode due to its superior discriminatory power when compared to plastid barcodes (X. Li et al., 2015).

Although ITS has a high differential power, it is not very useful due to the presence of incomplete lineage and divergent paralogous and pseudogenes, which may cause issues during amplification and sequencing (Gao et al., 2017). Due to its significant size, the ITS region has attracted attention in barcoding, particularly because of its heterogeneity and variability (Gernandt et al., 2001). There is also the risk of endophytic fungal DNA contamination. Difficulty in amplifying ITS DNA was also reported by few reports (Hollingsworth et al., 2011c). Focusing only on the ITS2 region can solve the issue of amplifying and sequencing the entire ITS region, but it also has some disadvantages. For certain plants, ITS2 may not offer sufficient resolution within the species (X. Li et al., 2015). ITS2 also enables for easy post PCR and sequencing procedures. Furthermore, The ITS2 region shows greater conservation of length compared to the entire ITS region (Hollingsworth et al., 2011c). Until now molecular data alone may not be enough to fully elucidate the interrelationships among seed plants, but the use of few combinations of molecular markers can provide approximate phylogenetic information (Bowe et al., 2000). The ITS (Internal Transcribed Spacer) has gained significance as a genetic marker due to several reasons, including its ability to have a high copy number of ribosomal RNA genes, a highly variable nature, low evolutionary pressure, non-functional sequences, and being inherited from both parents. Additionally, ITS is relatively easy to amplify and it has been used to study phylogeny at species and sub species level (Neale & Wheeler, 2019).



**Figure 2:** A diagram depicting the ribosome coding nuclear DNA in various groups, illustrating the ITS region and the specific location where a primer binds to amplify distinct regions of nuclear DNA (Cheng et al., 2016b).

## 2.5 Review on technique and instrumentation of concern

### 2.5.1 Herbarium and sample preparation

It is generally believed that quickly dehydrating plant samples for herbarium collections is preferable to slow drying, as it helps to preserve DNA quality and minimize damage over time. However, the DNA quality of herbarium samples collected and preserved using traditional methods is typically not satisfactory for use in further sequencing studies. Moreover, the quality of DNA from herbarium samples is generally poorer than that of fresh tissue, which is likely due to the specific collection and preservation strategies employed (Ribeiro & Lovato, 2007). The success of DNA extraction from herbarium specimens is determined by a combination of factors related to the specimen itself, such as the method of specimen preparation, any exposure to toxins or freezing, and the age of the sample, as well as factors related to the taxon, including the texture of the leaf, the type of tissue, and the biochemistry of the leaf (Särkinen et al., 2012).

It has been demonstrated that the silica gel method of drying DNA samples is the most efficient in terms of producing high yields and high-quality DNA. Furthermore, the success rate of PCR amplification of chloroplast regions from specimens preserved with silica gel is significantly higher (at 100%) compared to those prepared using air-based or alcohol-based drying methods (Särkinen et al., 2012; Staats et al., 2011).

### 2.5.2 DNA extraction

Due to their unique and diverse biochemical compositions, plants cannot rely on a single method for DNA extraction (Ribeiro & Lovato, 2007). Herbarium specimens of dried plants are a valuable source of information about their genetics. However, the

extraction and post-processing of DNA from these specimens have proven to be major challenges due to the highly degraded nature of the DNA in herbarium samples (Särkinen et al., 2012). A customary protocol for Microscale solid phase extraction involves three primary stages: firstly, DNA present in the liquid is immobilized onto the solid phase during the binding step. Secondly, unbound molecules are removed by washing with a separate buffer during the washing step. Finally, the purified DNA is recovered from the solid support and returned to the liquid using a low-salt buffer during the elution step (Carvalho et al., 2018).

### **2.5.3 DNA sequencing and PCR principles**

The Sequences-generating PCR method duplicates DNA using the same fundamental components utilized in natural DNA replication. This process is characterized by three phases: the initial cycles where the primers search for the DNA template and bind to their complementary sequences, the middle cycles where DNA copies are created at an exponential rate, and the final cycles where the amplification process is completed. In order to ensure the most successful outcome, it is important to use reagents at their optimal concentrations and adjust the concentration of components such as Mg ions and Taq polymerase as necessary. This can be accomplished by altering the number of cycles and manipulating the annealing temperature in response to any issues that arise during the amplification process. For example, to get rid of non-specific bands MgCl<sub>2</sub> can be reduced, annealing temperature can be increased and number of PCR cycle can be reduced (McPherson & Møller, 2000). The use of a combination of PCR with fluorescence-based analysis has emerged as a dependable technique for modern DNA sequencing projects (McBride et al., 1989). PCR incubation cycle is a method used to automate the preparation of a DNA template for sequencing. This process involves denaturing the template at 94° C, followed by hybridization with deoxyribonucleotide primers (at 52° C to 55° C) and then extending the hybrid through the action of DNA polymerase. The PCR protocol has a significant impact on the errors that may arise during amplification in the sequencing data, and as such, it requires careful attention and optimization efforts to minimize these errors (Wilkinson et al., 2017). Studies indicate that the addition of dimethylsulphoxide (DMSO) and 1-10% bovine serum albumin to the PCR mixture is required to prevent the non-specific amplification of the ITS region from nuclear DNA (de Miranda et al., 2010).

The sequencing method employed in this study is "cycle sequencing", which enhances the accuracy of DNA sequencing and enables the use of minimal amounts of DNA as the starting material. This technique involves incorporating ddNTPS terminators during thermal cycling (PCR) along with linear amplification using a

single primer in a reaction. The parameters of the cyclic PCR can be adjusted to promote efficient incorporation of ddNTPS (McPherson & Møller, 2000). The fundamental approach used in DNA sequencing, known as the Sanger sequencing or chain termination method, is considered the most reliable and widely accepted method. It involves the use of 2',3'-dideoxyribonucleotide triphosphates (ddNTPs), which impede the action of DNA polymerase I. Since ddNTPs lack a 3'-hydroxyl group, they prevent further elongation of the DNA strand in the 3' direction, leading to termination of the strand.

In PCR, a primer, template, fluorescent-labeled ddNTPs and dNTPs are combined with DNA polymerase to extend the hybridized primer, resulting in DNA fragments with the same 5' terminals but different 3'-fluoro-labeled dideoxyribonucleotide residues (ddATPs, ddGTPs, ddCTPs or ddTTPs). The completed sequence reaction generates fragments that don't require manual gel electrophoresis, but are instead directly loaded into a single lane capillary gel in an electrophoretic chamber for separation based on molecular size (fragment length). Thin-coated capillaries with non-polymerized gel matrices and a laser detection system are used for capillary gel electrophoretic separation. Polymer movement in and out of the capillaries during sample loading and running detects the fragments. The separated fragments are optically analyzed using a fluorescence-based DNA sequence analyzer. This method links fluorescent moieties to dideoxynucleotide triphosphates (ddNTPs) for the synthesis of all four reactions in a single vial, making each terminated DNA fragment fluorescence-tagged according to its respective base. After exciting the fluorescent moieties of the DNA fragments with a laser beam at the end of the gel column/capillary where the mixed DNA fragments terminating with different bases are run according to fragment length, the information is displayed and stored in a computer system as four different colors corresponding to the four different bases (França et al., 2002; Sanger et al., 1977).

Managing and storing sequence data have been significant challenges in DNA barcoding. The selection of appropriate methods and coherent assembly and mapping algorithms is crucial in effectively managing the vast amount of data generated. This decision is key to ensuring the accuracy of downstream steps in the analysis of DNA barcodes (Stranneheim & Lundeberg, 2012).

#### **2.5.4 Sequence analysis**

After the completion of DNA sequencing, a dataset is generated from which meaningful information must be extracted to draw conclusions about the study. Computational methods with statistical models and mathematical frameworks play a

critical role in analyzing biological data and converting it into meaningful information. These methods are essential for the processing of biological data (Yoon, 2009). To analyze a DNA sequence, the first step is to determine the nitrogen bases present. Once this is known, other aspects of the sequence can be examined. Techniques such as pairwise and multiple sequence alignment are used to identify similarities and differences between sequences. Conserved and variable sites are examined, as well as differences in sequence length. Similarity search tools are used to find related sequences and pairwise distances are calculated. Finally, the sequences are related to each other to observe the topology of the phylogenetic tree (Stranneheim & Lundeberg, 2012; Wilkinson et al., 2017).

Some of tools used in sequence analysis are given below:

#### **2.5.4.1 Bioedit and Sequencer**

The most commonly used program for sequence analysis is Bioedit 7.2.5. This program includes a sequence alignment feature, which was used in the study. Bioedit allows the alignment file to be saved in multiple formats, such as Fasta, mega, and genebank, for ease of use (hall, n.d.).

Sequencer vs 4.1.4 software allows for the import of sequences in various formats and displays the chromatogram. It is typically used to generate consensus sequences. This program displays chromatograms in multiple columns, making it easy to visualize and edit the data (Nishimura, 2000).

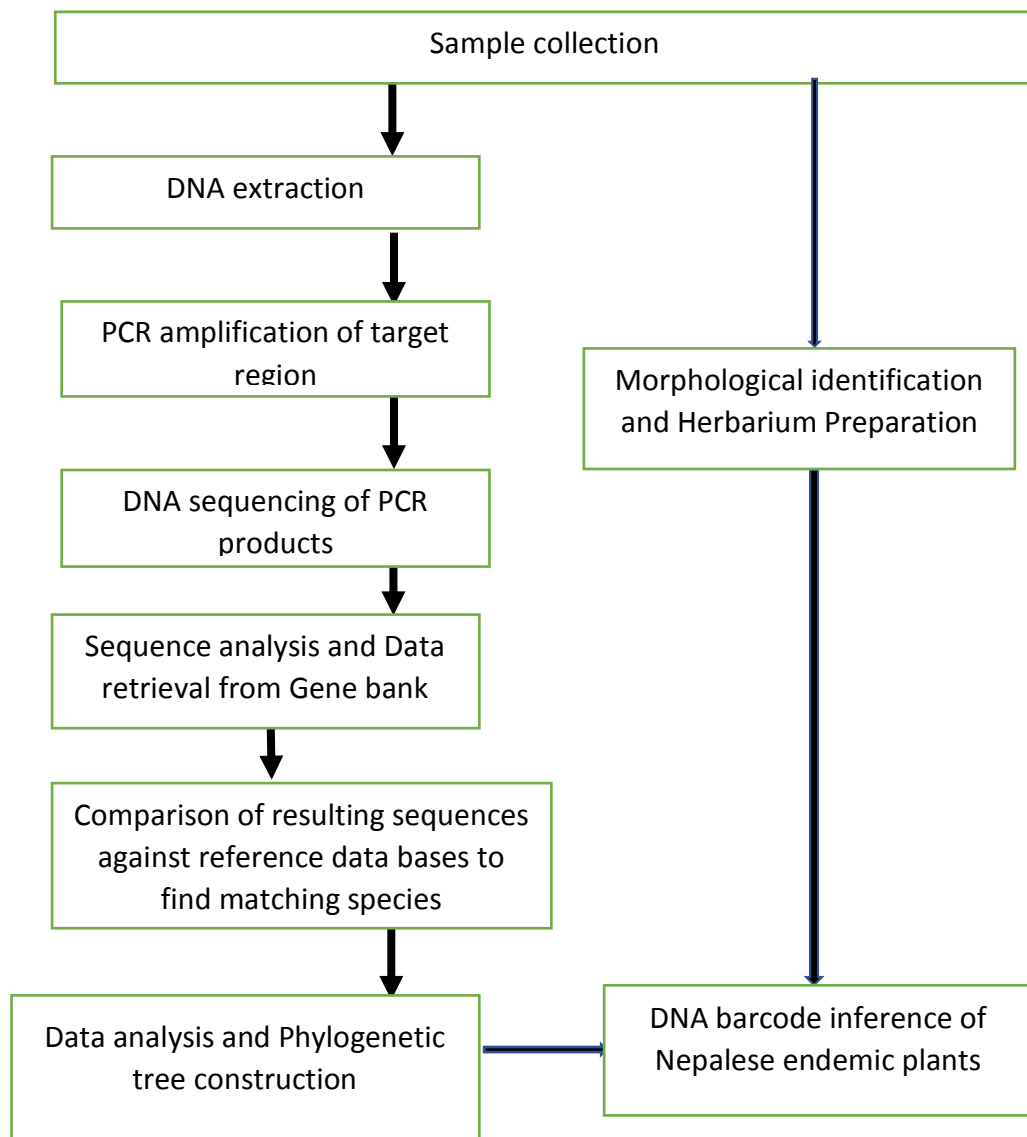
#### **2.5.4.2 Blast**

The Nucleotide-BLAST function of the Basic Local Alignment Search Tool (BLAST) is utilized to conduct similarity searches on all sequences that were retrieved. The highest score, query coverage, and percentage identity were noted and compared with other methods, such as distance-based and tree-based analysis, to identify matches using BLAST.

## CHAPTER 3

### MATERIALS AND METHODS

The research project commenced by gathering endemic plants from diverse regions of Nepal. All the specimens and information gathered were stored at the Molecular Biotechnology Unit located in NAST, Lalitpur, Nepal.



**Figure 3:** An overall study design in flowchart.

### **3.1 Sample collection**

During the process of collecting samples, a twig measuring 20-30 cm, complete with leaves, flower and fruit if available, was obtained. A small number of fresh and clean leaves were then removed from the same twig and placed in a bag or pouch made of lens paper, which was subsequently inserted into a silica gel box for preservation. The remaining part of the twig was flattened and inserted into a folded newspaper, then transported to the laboratory in a press for drying. Upon arrival at the laboratory, the silica gel in the box containing the leaf samples was replaced with fresh silica gel crystals and the samples were stored for DNA extraction. To dry the herbarium specimens, newspapers were regularly changed and replaced at least three times a week or dried under drying mechanism designed using drying fan and other tools.

### **3.2 DNA extraction**

To extract DNA from the leaves that had been preserved in silica gel, DNeasy Plant Mini Kit DNA extraction method was employed, taking into account the available resources, time limitations, and characteristics of the sample. The process of DNA extraction began by grinding the material, either by manual means using liquid nitrogen or by utilizing the tissue-lyser (MP Bio FastPrep-24) technique. Once the sample had been grounded, commercial reagent provided in the supplied kit was utilized to rupture the cell wall/membrane and nuclear membrane, thereby releasing both chloroplast and nuclear DNA. To reduce RNA contamination, the DNA mixture that had been released was treated with RNase. Lastly, the DNA was eluted using an elution buffer.

#### **3.2.1 DNeasy Plant Mini Kit DNA extraction**

A sample weighing less than 100mg was disrupted using liquid nitrogen in a sterile mortar and pestle. The resulting material was crushed in 200 ul of AP1 buffer to make a suspension, which was then poured into a 2 ml microcentrifuge tube. Next, 200 ul of AP1 buffer was added, followed by the addition of 4 ul of RNase A. The mixture was vortexed and then incubated at 65°C for 20 minutes. After this step, Buffer P3 (130 ul) was added and mixed well, followed by a 5-minute incubation on ice. The lysate was then centrifuged at 14,000 rpm for 8 minutes. The supernatant (excluding the pellet) was pipetted into a QIAshredder spin column that was placed in a 2 ml collection tube and centrifuged at 14,000 rpm for 2 minutes. The resulting flow-through was transferred to a new tube, without disturbing any pellet that might have been present. About 1.5 volumes of Buffer AW1 were added and mixed by

pipetting, and then 650 ul of the mixture was transferred to a DNeasy Mini spin column that was placed in a 2 ml collection tube and centrifuged at 8,000 rpm for 1 minute. This step was repeated with the remaining volume of the mixture. The spin column was then placed into a new 2 ml collection tube, and 500 ul of Buffer AW2 was added. The spin column was centrifuged at 14,000 rpm for 2 minutes. Finally, the spin column was transferred to a new 1.5 ml microcentrifuge tube, and 100 ul of Buffer AE was added for elution. The spin column with elution buffer was then incubated for 10 minutes at room temperature and centrifuged at 8,000 rpm for 1 minute. This elution step was repeated for better recovery of DNA.

### 3.3 Marker and primer selection

Following the global literatures on molecular characterization nuclear region ITS and chloroplast region *rbcl*, *matK*, *trnH-psbA*, *trnL-trnF* were selected.

**Table 2:** Selection of markers and primers

Regions	Direction	Primer sequence 5'-3'	Reference
<b>ITS</b>	P5-forward	CCTTATCAYTTAGAGGAAGGAG	(Cheng et al., 2016a)
	U4-reverse	RGTTTCTTTTCCTCCGCTTA	(Cheng et al., 2016a)
<b>matK</b>	F' matK 472F	CCRTCATCTGGAAATCTTGGTT	(Fatima et al., 2019)
	R' 1248R	GCTRTRATAATGAGAAAGATT TCTGC	(Fatima et al., 2019)
<b>rbcl</b>	<i>rbcl</i> a-f	ATGTCACCACAAACAGAGACTAAAGC	(Kress & Erickson, 2007b)
	<i>rbcl</i> a-rev	GTA AAAATCAAGTCCACCRCG	(Vere et al., 2015)
<b>trnH-psbA</b>	<i>psbA</i> f	GTTATGCATGAACGTAATGCTC	(Bhagya Chandrasekara et al., 2021)
	<i>trnH</i> R	CGCGCATGGTGGATT CACAATC	(Bhagya Chandrasekara et al., 2021)
<b>trnL-trnF</b>	<i>trnL</i>	CGAAATCGGT AGACGCTACG	(F. Li et al., 2023)
	<i>trnF</i>	ATTTGAACTGGTGACAGAG	(F. Li et al., 2023)

### 3.4 Polymerase chain reaction (PCR) amplification and PCR success rate

Following the selection of suitable markers and a thorough review of the PCR conditions, the PCR reaction mixture was prepared to amplify *rbcl*, *matK*, *trnL-trnF*, *trnH-psbA*, and ITS. All essential adjustments were made to optimize the PCR process, including modifying the number of cycles, alternating annealing temperatures, and components to ensure successful amplification. The PCR

reactions were conducted in PCR tubes that were prepared on ice and subsequently incubated in a PCR machine for approximately three hours on average.

The PCR reaction mixture was created by combining the Master mix (PCR buffer, dNTPs, Taq polymerase, and Magnesium chloride) to achieve a final concentration of 1X. The forward and reverse primers were added at a concentration of 0.4 pmoles each. If necessary, Magnesium chloride was included at a concentration of 1mM, and Taq polymerase at a concentration of 0.25 U/10  $\mu$ l. The total volume of the reaction mixture was 10  $\mu$ l, which was made up with nuclease-free water. For amplification of ITS region bovine serum albumin was added to make final concentration of 0.1  $\mu$ g/ $\mu$ l and 4% DMSO was added.

After adding template to the PCR mixture, PCR tubes were vortexed for 5 seconds and short-spined and then incubated in thermocycler (PCR machine) for amplification. Thermal cycler (Applied Biosystem, USA) was used for the PCR amplification. PCR condition were set according to primers used.

**Table 3:** PCR cycle of rbcL marker

Steps	Initial denaturation	Denaturation	Annealing	Extension	Final extension	Hold
Cycles	1	35			1	1
Temperature	94°C	94°C	52°C	72°C	72°C	4°C
Time	2 min	40 sec	45 sec	1 min	7 min	$\infty$

**Table 4:** PCR cycle of ITS marker

Steps	Initial denaturation	Denaturation	Annealing	Extension	Final extension	Hold
Cycles	1	35			1	1
Temperature	94°C	94°C	55°C	72°C	72°C	4°C
Time	2 min	50 sec	55 sec	1:1 min	7 min	$\infty$

**Table 5:** PCR cycle of matK marker

Steps	Initial denaturation	Denaturation	Annealing	Extension	Final extension	Hold
Cycles	1	35			1	1
Temperature	94°C	94°C	52°C	72°C	72°C	4°C
Time	2 min	40 sec	45 sec	55 sec	7 min	$\infty$

**Table 6:** PCR cycle of trnH-psbA marker

Steps	Initial denaturation	Denaturation	Annealing	Extension	Final extension	Hold
Cycles	1	35			1	1
Temperature	94°C	94°C	50°C	72°C	72°C	4°C
Time	4 min	45 sec	40 sec	1 min	7 min	∞

**Table 7:** PCR cycle of trnL-trnF marker

Steps	Initial denaturation	Denaturation	Annealing	Extension	Final extension	Hold
Cycles	1	35			1	1
Temperature	94°C	94°C	52°C	72°C	72°C	4°C
Time	2 min	40 sec	45 sec	55 sec	7 min	∞

### 3.5 Agarose gel electrophoresis

To verify the success of PCR amplification, the method of agarose gel electrophoresis was employed. Following electrophoresis, the agarose gel containing the target DNA-EtBr complex was exposed to UV light and compared to a standard DNA ladder. This process confirmed the presence of the desired band.

1% agarose gel was prepared in a conical flask using freshly prepared 1x TAE buffer. The agarose granules were fully dissolved by heating the flask in an oven at medium heat for 2 minutes. After slight cooling, EtBr (ethidium bromide) was added to the agarose gel, which was still hot but bearable, to achieve a final concentration of 0.5 ug/ml. The gel, after thorough mixing with EtBr, was poured into a gel casting tray with a comb and allowed to solidify for approximately 20-30 minutes. Carefully removing the comb, the gel was placed in an electrophoretic chamber containing freshly prepared 1x TAE buffer. The buffer was added until the gel completely sank into it.

For loading onto the gel, the PCR product (2.5-3ul) was mixed with loading dye to achieve a 1x concentration of the dye. This mixture was meticulously loaded into the wells of the solidified gel, taking care to avoid air bubbles. Additionally, a standard DNA marker (Generuler Plus 100bp ladder) was loaded to serve as a reference for band sizes.

The electrophoresis was initiated at 100 volts and run for 30 minutes. Subsequently, the gel was removed and visualized under UV light using a UV-transilluminator and a Syngene gel documentation instrument. Throughout the process, gloves were worn

for self-protection and to prevent potential cross-contamination hazards associated with EtBr.

### **3.6 Purification of the PCR products**

To ensure that only the desired amplicon was used for subsequent processing of the target PCR product, a purification step was performed using the EXOSap-IT™ kit (Product No. 75001/75002; Applied Biosystems, USA). The purification followed the protocol provided with the EXOSap. In this process, 5µl of the PCR product was mixed well with 2µl of EXOSap mix by pipetting. The mixture was then incubated at 37°C for 10 minutes, followed by incubation at 80°C for an additional 10 minutes. Both incubation steps were carried out using a PCR machine.

To confirm the presence of the purified product and assess the difference in purity before and after purification, 1µl of each product was loaded onto a 1% agarose gel, and electrophoresis was performed. The purified product's presence was verified, and the analysis included a comparison of purity levels. The remaining purified product was used for DNA sequencing.

### **3.7 DNA sequencing**

#### **3.7.1 Cycle sequencing**

The DNA sequencing process involves incubating the amplified DNA (to be sequenced) with Ready Reaction mix and sequencing buffer using only a single primer at a time, either forward or reverse. High-quality PCR-purified product, displaying a single band after gel visualization under UV light, was used for DNA sequencing. The sequencing reactions were set up on a 96-well plate following the guidelines provided by the BigDye™ Terminator v3.1 Cycle Sequencing Kit.

The primers, Ready Reaction mix (containing fluorescent-labeled ddNTPs, dNTPs, and DNA polymerase), PCR buffer, PCR products, and BSA were thawed completely and kept on ice after a brief spin. BSA was added to enhance the reaction's effectiveness. The Ready Reaction mix, being light-sensitive, was handled with utmost care to minimize exposure to light, such as covering it with aluminum foil and turning off chamber lights.

Separate sequencing reaction mixtures were prepared for each respective marker and direction (forward or reverse), as specified in Table below. Once the mixtures were prepared and transferred to PCR strips. Then PCR strips were properly capped were briefly vortexed for 2-3 seconds and then spun in a swinging bucket centrifuge

(ST 8R, Thermo Scientific, USA). Subsequently, the plate was immediately placed in a PCR machine for incubation under the appropriate sequencing reaction conditions.

**Table 8:** Conditions for sequence reaction

Steps	Initial denaturation	Denaturation	Annealing	Extension	Hold
Cycles	1	40			1
Temperature	96°C	96°C	50°C	60°C	4°C
Time	60sec	100 sec	5 sec	4 min	Infinite

### 3.8 Purification of the synthesized DNA fragments

In order to prevent any potential disruption caused by previously used buffers, primers, and other reagents, the DNA fragments that were produced during the PCR incubation were purified prior to the electrophoresis process for sequence reading.

A mixture containing 3 microliters of 3M Sodium acetate and 25 microliters of NFW per reaction was prepared. Then, 27 microliters of this mixture were added to each reaction. For each reaction tube 63 microliters of absolute ethanol were added. The PCR strips were properly capped, and the reagents were mixed softly using hand by gently turning PCR strips up and down. The capped strips were then incubated at room temperature for 20 minutes and centrifuged at 4200 rpm for 25 minutes at room temperature. The liquid on top (supernatant) was removed by turning the plate upside down on tissue paper. To eliminate any remaining liquid, the plate was spun upside down at 300 RPM for 2 minutes. Next, 200 microliters of freshly prepared 70% ethanol were added to each reaction in the plate.

Again, sequence fragment and ethanol were softly mixed by turning plates up and down. The capped strips were spun at 4200 g for 15 minutes at room temperature. The ethanol was removed by inverting the PCR strips over layers of UV-sterile tissue paper after opening the caps. Again, plate was spun upside down for 2 minutes at 300 RPM. This washing process was carried out with great caution and repeated to ensure complete removal of ethanol residue. The sequence fragment was then dried in thermocycler at 40°C for 10 minutes. Finally, 10 microliters of Hi-Di formamide were added to each reaction in the plate and spun. The plate was then subjected to denaturation at 95°C for 2 minutes using a thermocycler machine, followed by immediate cooling on ice (snap-chilling).

### **3.9 Sequencing: Capillary electrophoresis**

The plate containing the purified sample dissolved in Hi-Di was placed into the sequencer, secured, and connected to the 3500 Genetic analyzer from Applied Biosystems, USA. The Applied Biosystems Gene analyzer ABI 3500XL housed by molecular biotechnology unit of NAST operates based on Sanger's dideoxy chain termination technique. The analyzer provided data on the allocation of bases and chromatography patterns for markers including ITS, matK, rbcL, trnH-psbA, and trnL-trnF, which were further evaluated to generate information.

### **3.10 Sequence editing and alignment**

The sequence files were imported into Sequencher vs 4.1.4 to verify the quality of the sequences by examining the sequence peaks. Ambiguous peaks were taken as it is for single-stranded sequences. For the double-stranded sequences (comprising both forward and reverse sequences), consensus sequences (contigs) were generated.

To create the contigs, the sequences obtained through the Genetic Analyzer (based on forward and reverse primers) were imported into a Sequencher vs 4.1.4 project, and overlapping fragments were assembled to form contigs. The assembly parameters used were a minimum match percentage of 65-85 and a minimum overlap percentage of 20. The resulting sorted sequence was exported as a consensus sequence for further analysis.

In the process of preparing the reliable and valid sequence, the assignment of bases, deletion of bases, deletion of gaps, assignment of gaps, and replacement of nitrogen bases were performed, based on the base calls of nitrogen bases corresponding to the forward and reverse sequences. The generated consensus sequences were then imported into Bioedit vs 7.2.5 for multiple sequence alignment. The primer regions (both forward and reverse) were excluded from the multiple sequence alignment by comparing with the primer sequences, and the resulting alignment was saved as a fasta file for further analysis. Additionally, individual sequences for all the markers (ITS, matK, rbcL, trnH-psbA, trnL-trnF) were saved for further analysis.

### **3.11 NCBI BLAST**

The process of similarity search was conducted using NCBI-nucleotide BLAST. This method involves utilizing the most prominent match and percentage similarity to establish an identity score in a comparison between the query and reference sequences.

The resultant consensus sequence was then submitted to NCBI nucleotide-BLAST to compare the query samples with available reference specimens. By considering the highest score, identity, and E-value, the samples were examined for sequence-based species identification in relation to phenotypic-based sample labeling.

Endemic species identified morphologically were matched against informative DNA sequences, and the matches were analyzed. The corresponding barcode sequences were subjected to 'NCBI nucleotide BLAST' to assess potential discrepancies, query coverage, matching scores, and identical outcomes. Furthermore, the species identification was cross-referenced with the BLAST method (percentage identity), distance method (interspecific value exceeding intraspecific value), and tree method (branch bootstrap value surpassing 50).

After performing BLAST search the sequence of highly related plant of same species from neighbor country was downloaded to find the sequence divergence of our endemic plant.

### **3.12 Nucleotide composition**

The MEG11 tool was utilized to acquire the nucleotide composition of marker sequences for each sample's respective markers. The known percentages of A, T, G, and C were then examined to determine the average G+C content within the markers of endemic plants' genes.

### **3.13 Genetic distance calculation**

The outcome of performing multiple sequence alignment was observed with the aim of identifying variations. The sequences were retraced to verify whether the detected base variation was indeed genuine or if it resulted from sequencing errors, appearing as false positives. The regions with variability were marked as hotspots, and distinct variable sections were identified for each gene, including matK, ITS, rbcL, trnH-psbA, and trnIF-trnIC. Sequences from diverse sections of both chloroplast and nuclear markers were saved in Fasta format, aligned using CLUSTALW within MEGA11, and subsequently exported in the mega format. These files were then imported into MEGA11 for further processing, where parameters such as the count of constant sites (C), variable sites (V), parsimony sites (P), and singleton sites (S) were estimated. P-distance between interspecific species were calculated using distance tool of MEGA11.

### **3.14 Phylogenetic tree construction and distance estimation**

The sequences, aligned in an acceptable fasta format, were imported into MEGA11. The analysis was continued using the Analyze option. The Phylogeny application employed the Neighbor-Joining method with a modified parameter of 10000 bootstrap replicates to construct the tree. The resulting tree was visualized in various formats provided by MEGA11, with one format saved for later interpretation and analysis.

Phylogenetic trees were generated for the markers ITS, matK, trnH-psbA, rbcL, trnL-trnF using the neighbor-joining (NJ) approach in the MEGA11 software. The p-distance was utilized as a measure of genetic divergence. To assess the reliability of the tree's structure, bootstrap testing involving 1000 replicates was conducted, estimating confidence levels for the consensus tree's topology. Node support was categorized as weak (50-74%) or strong (75-100%) based on the bootstrap scale as suggested by (Manning et al., 2014). Successful species resolution required that all individual branches exhibit at least 50% bootstrap value. Furthermore, successful species discrimination was indicated when individuals from the same species formed a monophyletic group in the phylogenetic tree, according to (Ford et al., 2009).

## CHAPTER 4

### RESULTS

#### 4.1 Sampling

A total of 44 voucher samples of plants were collected, among which 9 vouchers represented potential endemic plants with uncertain information about their endemic status in Nepal, while the remaining samples comprised known endemic plants. Herbarium specimens were made for species that were collected for the study and healthy leaves were stored within lens cleaning pouches made of Whatman paper and then preserved in silica gel for DNA study.

**Table 9:** List of plants

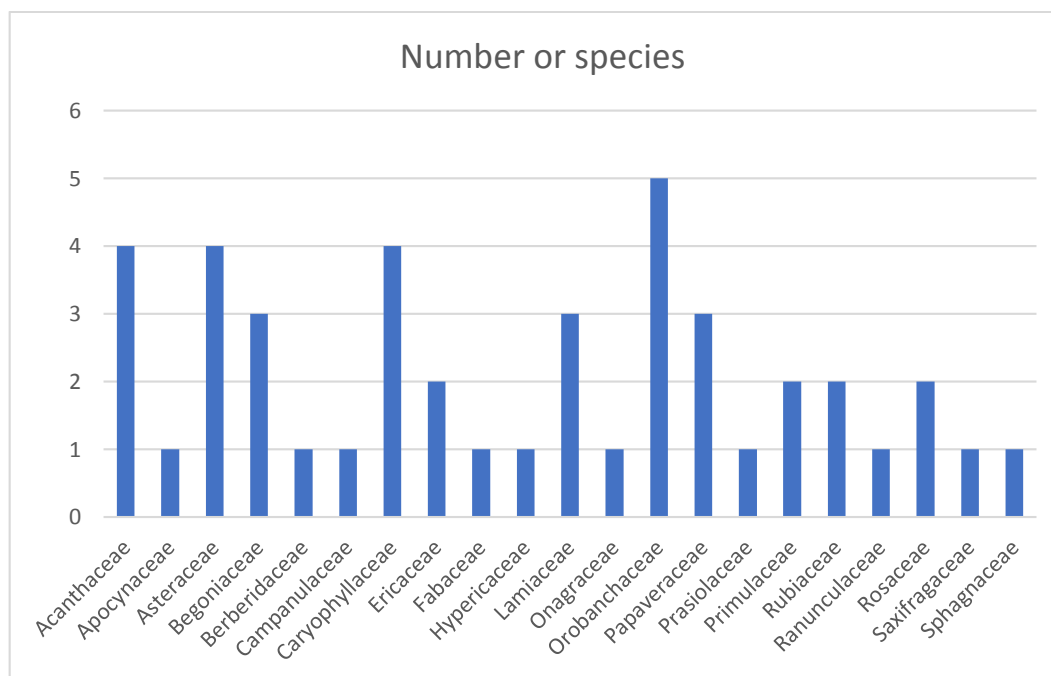
SN	Names of species	Lab code	Family	Location
1	<i>Justicia tukuchensis</i> V.A.W.Graham	M24	Acanthaceae	Mustang
2	<i>Strobilanthes saccata</i> J.R.I.Wood	6S	Acanthaceae	Kaski
3	<i>Thunbergia kasajuana</i> Bh. Adhikari & J. R. I. Wood	S6	Acanthaceae	Bhaktapur
4	<i>Thunbergia nepalensis</i> Bh.Adhikari & J.R.I.Wood	6T	Acanthaceae	Doti
5	<i>Ceropegia meleagris</i> H.Huber	C1	Apocynaceae	Mustang
6	<i>Artemisia mustangensis</i> Yonek.	M08	Asteraceae	Mustang
7	<i>Artemisia nepalica</i> Yonek.	M13	Asteraceae	Mustang
8	<i>Cicerbita nepalensis</i> Kitam.	M22	Asteraceae	Mustang
9	<i>Begonia flagellaris</i> H.Hara	3B	Begoniaceae	Gul Bhanjang,
10	<i>Begonia nuwakotensis</i> S. Rajbh	S9	Begoniaceae	Nuwakot
11	<i>Begonia tligera</i> S. Rajbh	S7	Begoniaceae	Kaski
12	<i>Berberis mucrifolia</i> Ahrendt	M10	Berberidaceae	Mustang
13	<i>Codonopsis grey-wilsonii</i> J.M.H.Shaw	C4	Campanulaceae	Humla
14	<i>Eremogone mukerjeeana</i> (Majumdar) Rabeler & W.L.Wagner	M03	Caryophyllaceae	Mustang
15	<i>Silene helleboriflora</i> Exell & Bocquet	4S	Caryophyllaceae	Mulmuley Khola, Rukum
16	<i>Silene vautierae</i> Bocquet	8S	Caryophyllaceae	Mustang
17	<i>Rhododendron cowanianum</i> Davidian	1R	Ericaceae	Langtang lateral valley, Rasuwa
18	<i>Rhododendron lowndesii</i> Davidian	M01	Ericaceae	Manang
19	<i>Oxytropis nepalensis</i> Vassilcz.	M18	Fabaceae	Mustang

SN	Names of species	Lab code	Family	Location
20	<i>Hypericum cordifolium</i> Buch.-Ham. ex D. Don	S8	Hypericaceae	Kathmandu
21	<i>Isodon namikawanus</i> Murata	M21	Lamiaceae	Mugu
22	<i>Salvia</i> sp.	3I	Lamiaceae	Manang
23	<i>Salvia transhimalaica</i> Yonek.	M20	Lamiaceae	Mustang
24	<i>Epilobium staintonii</i> P.H.Raven	M19	Onagraceae	Mustang
25	<i>Pedicularis anserantha</i> T.Yamaz.	10P	Orobanchaceae	Dolpa
26	<i>Corydalis calycina</i> Lidén	C3	Papaveraceae	Jumla
27	<i>Corydalis megacalyx</i> Ludlow & Stearn	C2	Papaveraceae	Rukum
28	<i>Corydalis terracina</i> Lidén	4c	Papaveraceae	Kaski
29	<i>Primula sharmae</i> H.R.Fletcher	M02	Primulaceae	Mustang
30	<i>Primula wigramiana</i> W.W.Sm.	5p	Primulaceae	Sindhupalchok
31	<i>Clematis phlebantha</i> L.H.J.Williams	5c	Ranunculaceae	Dolpa
32	<i>Griffitharia sharmae</i> (M.F.Watson, V.Manandhar & Rushforth) Rushforth	9S	Rosaceae	Rasuwa
33	<i>Prunus jajarkotensis</i> H.Hara	MP	Rosaceae	Jajarkot
34	<i>Saxifraga excellens</i> Harry Sm.	7S	Saxifragaceae	Kaski
35	<i>Sphagnum nepalense</i> H. Suzuki,	2M	Sphagnaceae	Maipokhari, Ilam
	<b>Species with uncertain endemic distributions</b>			
36	<i>Galium</i> sp.	1G	Rubiaceae	Kaski
37	<i>Galium</i> sp.	2G	Rubiaceae	Kaski
38	<i>Pedicularis</i> SP.1 ( <i>Pedicularis annapurnensis</i> T.Yamaz.)	M04	Orobanchaceae	
39	<i>Pedicularis</i> SP.2 ( <i>Pedicularis anserantha</i> T.Yamaz.)	M05	Orobanchaceae	
40	<i>Pedicularis</i> SP.3 ( <i>Pedicularis breviscaposa</i> T.Yamaz.)	M06	Orobanchaceae	
41	<i>Pedicularis</i> SP.4 ( <i>Pedicularis chamissonoides</i> T.Yamaz.)	M07	Orobanchaceae	
42	S01 ( <i>Prasiola formosana</i> Okada)	S01	Prasiolaceae	Kavre
43	<i>Silene</i> sp.	M17	Caryophyllaceae	Mustang
44	<i>Taraxacum staintonii</i> Soest	M09	Asteraceae	

Among the studied 10 species (plants whose lab code initials is 'M') were obtained from the Silica Gel preserved collection of Royal Botanical Garden, Edinburgh UK,

while others were received under collaborative research activities between institutes and experts working in endemic plants of Nepal.

We barcoded a total of 44 plants representing 19 different families. Within these, five individuals from the Orobanchaceae Family were barcoded. Additionally, ten plants from ten different families, two plants each from four different families, three plants each from three different families, and four plants each from three different families were also barcoded.



**Figure 4:** Bargraph showing number of plants from different famlies used in this study

#### 4.2 Sampling and preservation

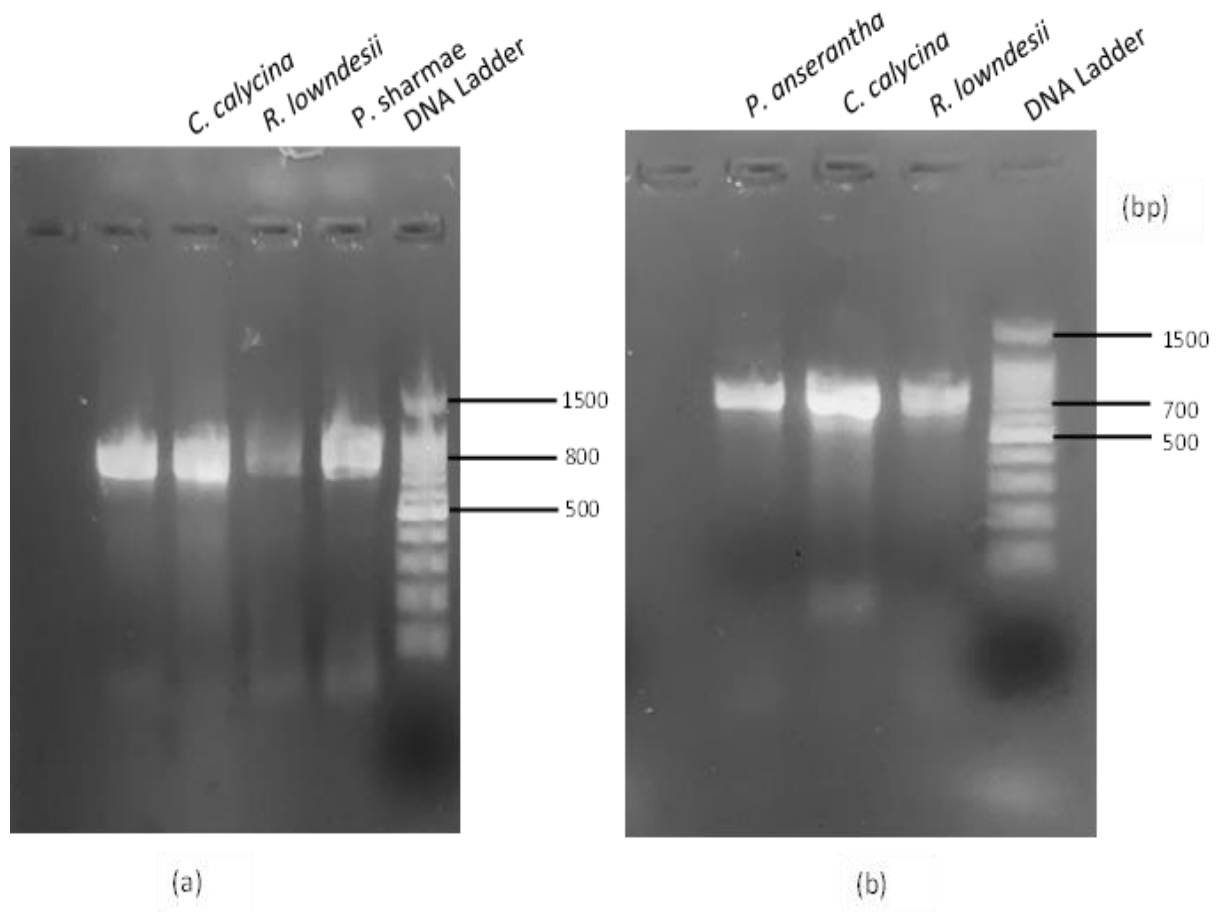
Silica gel preservation was selected and proven to be an effective choice for preserving chemicals. Unfortunately, the *Parsiola* sample was inadequately dried and not stored with silica gel, leading to difficulties in extracting DNA and conducting PCR amplification.

#### 4.3 DNA extraction

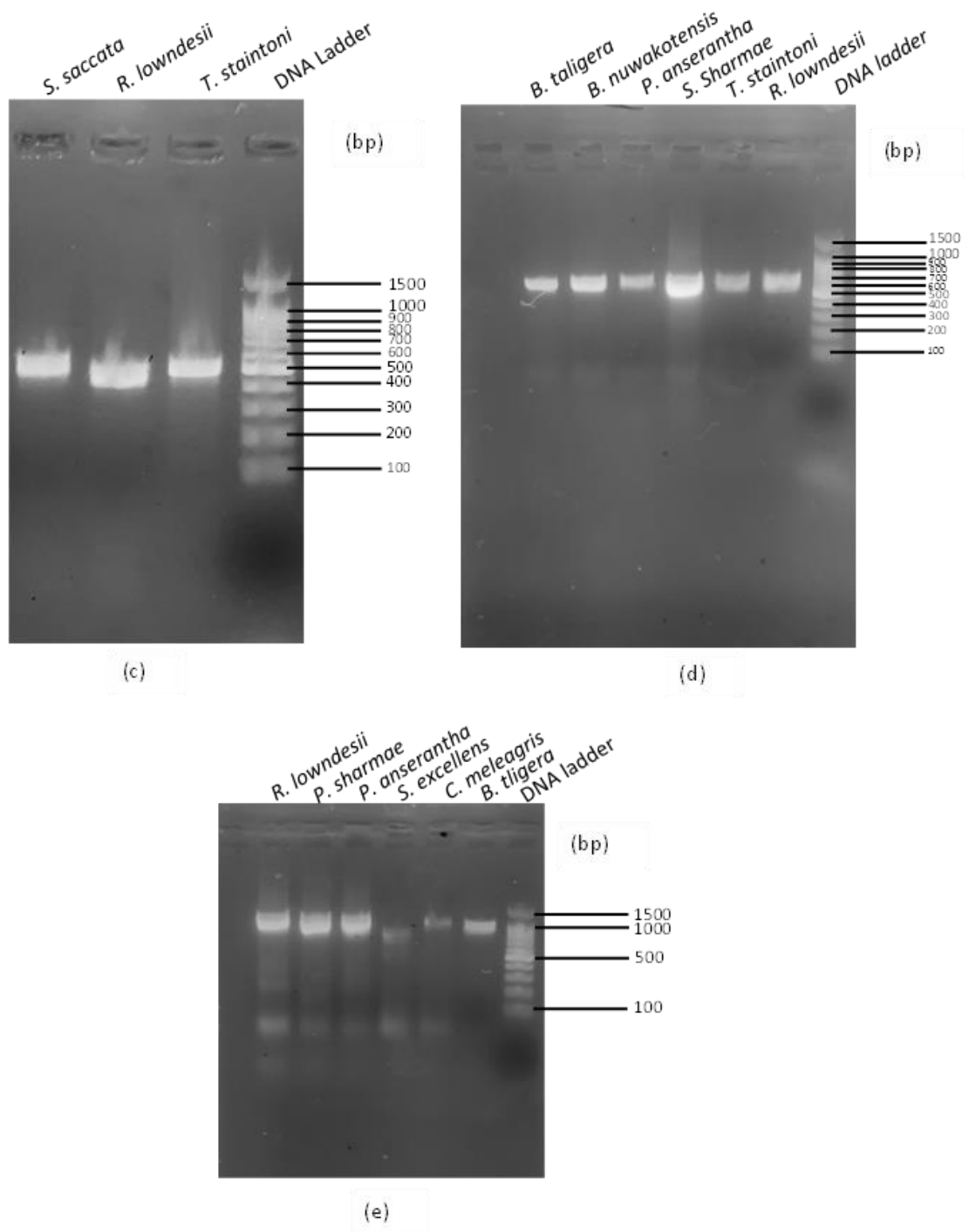
DNA of 44 endemic plants of Nepal was extracted using DNeasy plant mini kit and agarose gel electrophoresis method was performed to visualize the DNA band under UV illumination.

#### 4.4 Amplification of markers

The effectiveness of the PCR amplification process was witnessed by conducting agarose gel electrophoresis. The DNA fragments that were amplified were visualized within the gel using UV light. The appearance of clear bands under the UV light indicated positive outcomes. To confirm the presence of the desired DNA segments, a negative control (reaction mixture without a template) and a positive standard DNA (DNA ladder or marker) were employed for comparison of DNA length.



**Figure 5:** (a) shows the positive amplification of *rbcl* region and (b) shows positive amplification of *trnH-psbA* region along with 100bp ladder



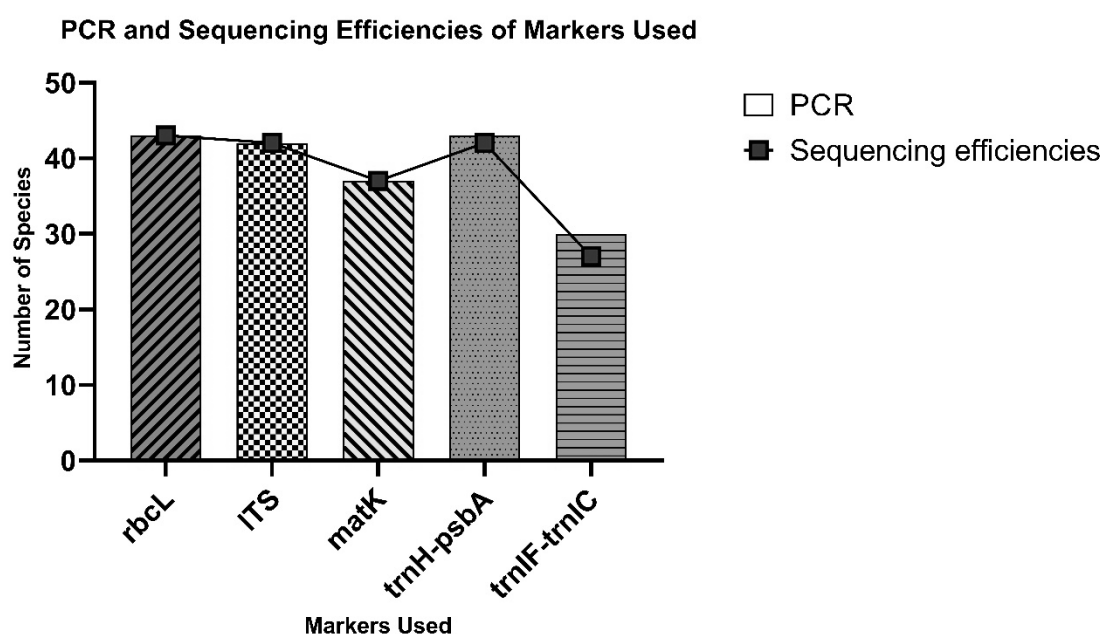
**Figure 6:** (c) shows positive amplification of *matK* region, (d) shows positive amplification of ITS region and (e) shows the positive amplification of *trnIF-trnIC* region along with primer

## 4.5 PCR clean up

The PCR products that were effectively amplified underwent purification using the SAP-EXO purification kit to prevent any chemical disruptions during sequencing. Following this, the samples were placed onto a 1% agarose gel to enhance the visibility of DNA bands and to make a tentative comparison of the DNA band intensity per microliter of each sample.

## 4.6 PCR and Sequencing efficiency

The graph below displays the PCR and sequencing effectiveness of five distinct markers. The PCR success rates for *rbcl* and *trnH-psbA* were the highest at 97.7%, making them the most straightforward to amplify for target products. For ITS, the PCR efficiency stood at 95%, while *matK* had an efficiency of 84%. In contrast, amplifying *trnIF-trnIC* proved to be the most challenging, with an efficiency of just 68%. All PCR products of *rbcl*, *matK* and ITS were successfully sequenced, while 97.6% and 90% sequencing efficiency was found in *trnH-psbA* and *trnIF-trnIC*.



**Figure 7:** Bargraph showing PCR and sequencing efficiency of different markers

**Table 10:** PCR and sequencing efficiency of different markers

Markers	PCR	Sequencing
<i>rbcl</i>	43	43
ITS	42	42
<i>matK</i>	37	37
<i>trnH-psbA</i>	43	42
<i>trnIF-trnIC</i>	30	27

## 4.7 Length of markers

The analysis involved examining the differences in marker lengths for each sample, and the outcomes were presented in the provided tables. Some sequences were notably shorter in length compared to their corresponding forward or reverse counterparts. These shortened lengths were attributed to sequencing errors or interruptions during the sequencing process. The *rbcl* markers exhibited a consistent length in the majority of cases (553 bp) across 38 out of 43 samples. On the other hand, the *trnH-psbA* markers displayed significant variation in length, ranging from 207 bp to 781 bp.

**Table 11:** Length of *rbcl* markers

Name of species	Lab code	Without Primer(bp)	Remarks
<i>Artemesia mustangensis</i>	M08	553	F/R
<i>Artemisia nepalica</i>	M13	553	F/R
<i>Begonia flagellans</i>	3B	553	F/R
<i>Begonia nuwakotensis</i>	S9	553	F/R
<i>Begonia taligera</i>	S7	553	F/R
<i>Berberis mucrifolia</i>	M10	528	F/R
<i>Ceropegia neleagris</i>	C1	553	F/R
<i>Cicerbita nepalensis</i>	M22	546	F/R
<i>Clematis phlebantha</i>	5C	553	F/R
<i>Codonopsis grey-wilsoni</i>	C4	553	F/R
<i>Corydalis calycina</i>	C3	553	F/R
<i>Corydalis megacalyx</i>	C2	553	F/R
<i>Corydalis terracina</i>	4C	553	F/R
<i>Epilobium staintonii</i>	M19	553	F/R
<i>Eremogone mukerjeeana</i>	M03	553	F/R
<i>Galium sp. 1</i>	1G	553	F/R
<i>Glaium sp. 2</i>	2G	553	F/R
<i>Griffitharia sharmae</i>	9S	553	F/R
<i>Isodon namikawanus</i>	M21	553	F/R
<i>Justicia tukuchensis</i>	M24	553	F/R
<i>Oxytropis nepalensis</i>	M18	553	F/R
<i>Pedicularis annapurnensis</i>	M04	553	F/R
<i>Pedicularis anserantha</i>	10P	553	F/R
<i>Pedicularis breviscaposa</i>	M06	552	F/R
<i>Pedicularis chamissonoides</i>	M07	556	F/R
<i>Pedicularis sp. 2</i>	M05	553	F/R
<i>Prasiola formosana</i>	S01	553	F/R
<i>Primula sharmae</i>	M02	553	F/R
<i>Primula wigramiana</i>	5S	553	F/R
<i>Prunus jajarkotium</i>	MP	553	F/R
<i>Rhodendron lowendesii</i>	M01	553	F/R

Name of species	Lab code	Without Primer(bp)	Remarks
<i>Rhododendron cowanianum</i>	1R	553	F/R
<i>Salvia spp.</i>	3L	553	F/R
<i>Salvia transhimalaica</i>	M20	553	F/R
<i>Saxifraga excellens</i>	7S	553	F/R
<i>Silene hellebriflora</i>	4S	553	F/R
<i>Silene spp.</i>	M17	554	F/R
<i>Silene vautierae</i>	8S	553	F/R
<i>Sphagnum nepalense</i>	2M	553	F/R
<i>Strobilanthes saccata</i>	6S	553	F/R
<i>Taraxacum staintonii</i>	M09	553	F/R
<i>Thunbergia kasajuana</i>	S6	553	F/R
<i>Thunbergia nepalensis</i>	6T	553	F/R

The length of rbcL markers for all samples were 553 bp except for two *Pedicularis* species., One *Silene sp.*, *Berberis mucrifolia* and *Cicerbita nepalensis*.

**Table 12:** length of ITS markers

Name of species	Lab code	Without primer(bp)	Remarks
<i>Artemisia nepalica</i>	M13	734	F\R
<i>Artemisia mustangesis</i>	M08	733	F\R
<i>Begonia flagellaris</i>	3B	800	F\R
<i>Begonia nuwakotensis</i>	S9	799	F\R
<i>Begonia taligera</i>	S7	798	F\R
<i>Berberis mucrifolia</i>	M10	701	F\R
<i>Ceropegia meleagris</i>	C1	715	F\R
<i>Cicerbita nepalensis</i>	M22	743	F\R
<i>Clematis phlebantha</i>	5C	645	F\R
<i>Codonopsis grey-wilsonii</i>	C4	701	F\R
<i>Corydalis calycina</i>	C3	640	F\R
<i>Corydalis megacalyx</i>	C2	647	F\R
<i>Corydalis terracina</i>	4C	633	F\R
<i>Epilobium staintonii</i>	M19	710	F\R
<i>Eremogone mukerjeeana</i>	M03	720	F\R
<i>Galium sp.1</i>	1G	715	F\R
<i>Galium sp.2</i>	2G	735	F\R
<i>Griffitharia sharmae</i>	9S	721	F\R
<i>Hypericum cordifolium</i>	S8	729	F\R
<i>Isodon namikawanus</i>	M21	660	F\R
<i>Jajarkotium nepalensis</i>	MP	697	F\R
<i>Justicia tukuchensis</i>	M24	728	F\R
<i>Oxytropis nepalensis</i>	M18	696	F\R
<i>Pedicularis annapurnensis</i>	M04	717	F\R
<i>Pedicularis anserantha</i>	10P	717	F\R
<i>Pedicularis breviscaposa</i>	M06	704	F\R
<i>Pedicularis chamissonoides</i>	M07	696	F

Name of species	Lab code	Without primer(bp)	Remarks
<i>Pedicularis sp.2</i>	M05	700	F\R
<i>Primula sharmae</i>	M02	728	F\R
<i>Primula wigramiana</i>	5S	730	F\R
<i>Rhodendron cowanianum</i>	1R	741	F\R
<i>Rhodendron lowndesii</i>	M01	739	F\R
<i>Salvia sp.</i>	3L	712	F\R
<i>Salvia transhimalaica</i>	M20	689	F
<i>Saxifraga excellens</i>	7S	772	F\R
<i>Silene helleboriflora</i>	4S	717	F\R
<i>Silene sp.</i>	M17	718	F\R
<i>Silene vautierae</i>	8S	717	F\R
<i>Sphagnum nepalense</i>	2M	789	F\R
<i>Taraxaccum staintoni</i>	M09	744	F\R
<i>Thunbergia kasajuana</i>	S6	712	F\R
<i>Thunbergia nepalensis</i>	6T	707	F\R

The length of ITS markers for *Corydalis terracina*, *clematis phlebantha*, *Corydalis megacalyx* and *Corydalis calycina* were less than 650bp, *Corydalis terracina* being shortest with length of 633bp. *Jajarkotium nepalensis*, *Isodon namikawanus*, *Oxytropis nepalensis*, *Pedicularis chamissonoides* and *Salvia transhimalaica* had length between 650 to 700 bp. *Saxifraga excellens* and three *Begonia* species had length more than 750bp. *Begonia flagellaris* was longest with 800 bp length, while other *Begonia* species were just one or two bp shorter. Remaining plants were of 700bp to 750bp length.

**Table 13:** length of trnH-psbA markers

Name of species	Lab code	Without primer(bp)	Remarks
<i>Artemisia cf. nepalica</i>	M13	471	F/R
<i>Artemisia mustangesis</i>	M08	465	F/R
<i>Begonia flagellaris</i>	3B	451	F/R
<i>Begonia nuwakotensis</i>	S9	781	F/R
<i>Begonia taligera</i>	S7	781	F/R
<i>Berberis mucrifolia</i>	M10	533	F/R
<i>Ceropegia meleagris</i>	C1	423	F/R
<i>Cicerbita nepalensis</i>	M22	475	F/R
<i>Clematis phlebantha</i>	5C	413	F/R
<i>Codonopsis grey-wilsonii</i>	C4	439	F/R
<i>Corydalis calycina</i>	C3	561	F/R
<i>Corydalis megacalyx</i>	C2	470	F/R
<i>Corydalis terracina</i>	4C	503	F/R
<i>Epilobium staintonii</i>	M19	661	F/R
<i>Eremogone mukerjeeana</i>	M03	352	F/R
<i>Galium sp.2</i>	2G	317	F/R

Name of species	Lab code	Without primer(bp)	Remarks
<i>Galium sp.1</i>	1G	335	F/R
<i>Griffitharia sharmae</i>	9S	350	F/R
<i>Hypericum cordifolium</i>	S8	721	F/R
<i>Isodon namikawanus</i>	M21	660	F/R
<i>jajarkotium nepalensis</i>	MP	366	F/R
<i>Justicia tukuchensis</i>	M24	372	F/R
<i>Oxytropis nepalensis</i>	M18	413	F/R
<i>Pedicularis annapurnensis</i>	M04	625	F/R
<i>Pedicularis anserantha</i>	10P	625	F/R
<i>Pedicularis breviscaposa</i>	M06	588	F/R
<i>Pedicularis chamissonoides</i>	M07	521	F/R
<i>Pedicularis sp.2</i>	M05	598	F/R
<i>Primula sharmae</i>	M02	345	F/R
<i>Primula wigramiana</i>	5S	507	F/R
<i>Rhodendron cowanianum</i>	1R	431	F/R
<i>Rhodendron lowndesii</i>	M01	431	F/R
<i>Salvia sp.</i>	3L	387	F/R
<i>Salvia sp.</i>	M20	388	F/R
<i>Saxifraga excellens</i>	7S	207	F/R
<i>Silene helleboriflora</i>	4S	381	F/R
<i>Silene sp.</i>	M17	374	F/R
<i>Silene vautierae</i>	8S	400	F/R
<i>Sphagnum nepalense</i>	2M	270	F/R
<i>Strobilanthes saccata</i>	6S	518	F/R
<i>Taraxaccum staintoni</i>	M09	478	F/R
<i>Thunbergia kasajuana</i>	S6	520	F/R
<i>Thunbergia nepalensis</i>	6T	538	F/R

Length of trnH-psbA was found to be most variable. *Saxifraga excellens* was only 207bp long while, *Begonia nuwakotensis* and *Begonia taligera* were 781bp long. *Sphagnum nepalense* and *Saxifraga excellens* were less than 300bp long. 11 of our studied plants were of 300 to 400bp length. Similarly, 13 of studied plants were of 400 to 500bp length and 10 plants had length between 500 to 600bp. *Pedicularis anserantha*, *Pedicularis annapurnensis*, and *Isodon namikawanus* had length between 600 to 700bp. *Begonia taligera*, *Hypericum cordifolium* and *Begonia nuwakotensis* were of 700 to 800bp.

**Table 14:** length of matK markers

Name of speies	Lab code	Without primer(bp)	Remarks
<i>Artemisia mustangensis</i>	M08	747	F/R
<i>Artemisia nepalica</i>	M13	747	F/R
<i>Begonia flagellaris</i>	3B	747	F/R
<i>Begonia nuwakotensis</i>	S9	741	F/R
<i>Begonia taligera</i>	S7	741	F/R
<i>Berberis mucrifolia</i>	M10	744	F/R
<i>Ceropegia meleagris</i>	C1	765	F/R
<i>Cicerbita nepalensis</i>	M22	747	F/R
<i>Clematis phlebantha</i>	5C	744	F/R
<i>Codonopsis grey-wilsonii</i>	C4	741	F/R
<i>Corydalis calycina</i>	C3	744	F/R
<i>Corydalis megacalys</i>	C2	744	F/R
<i>Corydalis terracina</i>	4C	744	F/R
<i>Epilobium staintonii</i>	M19	753	F/R
<i>Eremogone mukerjeeana</i>	M03	747	F/R
<i>Galium sp.2</i>	2G	744	F/R
<i>Galium sp.1</i>	1G	744	F/R
<i>Griffitharia sharmae</i>	9S	741	F/R
<i>Isodon namikawanus</i>	M21	732	F/R
<i>Jajarkotium nepalensis</i>	MP	741	F/R
<i>Justicia tukuchensis</i>	M24	747	F/R
<i>Oxytropis nepalensis</i>	M18	741	F/R
<i>Pedicularis annapurnensis</i>	M04	747	F/R
<i>Pedicularis anserantha</i>	10P	747	F/R
<i>Pedicularis breviscaposa</i>	M06	745	F/R
<i>Pedicularis chamissonoides</i>	M07	751	F/R
<i>Pedicularis sp.2</i>	M05	747	F/R
<i>Salvia sp.</i>	3L	747	F/R
<i>Salvia transhimalaica</i>	M20	747	F/R
<i>Saxifraga excellens</i>	7S	749	F/R
<i>Silene helleboriflora</i>	4S	741	F/R
<i>Silene sp.</i>	M17	741	F/R
<i>Silene vautierae</i>	8S	741	F/R
<i>Strobilanthes saccata</i>	6S	747	F/R
<i>Taraxaccum staintoni</i>	M09	747	F/R
<i>Thunbergia kasajuana</i>	S6	753	F/R
<i>Thunbergia nepalensis</i>	6T	753	F/R

Then length of matK markers were similar. *Isodon namikawanus* was found to be shortest with 732 bp. 9 of the plant had length of 741bp, 7 plants had 744 bp length and 12 plants had 747bp length. *Thunbergia kasajuana* and *Epilobium* were of 753bp length. *Saxifraga excellens* was 749 bp and *Pedicularis chamissonoides* was 751 bp long. *Ceropegia meleagris* was longest with 765bp length.

**Table 15:** length of trnIF-trnIC markers.

Name of species	Lab code	Without primer(bp)	Remarks
<i>Artemisia mustangensis</i>	M08	882	F/R
<i>Begonia nuwakotensis</i>	S9	1035	F/R
<i>Begonia taligera</i>	S7	1040	F/R
<i>Ceropegia meleagris</i>	C1	924	F/R
<i>Cicerbita nepalensis</i>	M22	881	F/R
<i>Clematis phlebantha</i>	5C	734	F/R
<i>Corydalis calycina</i>	C3	952	F/R
<i>Corydalis megacalyx</i>	C2	955	F/R
<i>Epilobium staintonii</i>	M19	912	F/R
<i>Galium sp.1</i>	1G	965	F/R
<i>Galium sp.2</i>	2G	945	F/R
<i>Griffitharia sharmae</i>	9S	1013	F/R
<i>Hypericum cordifolium</i>	S8	649	F/R
<i>Isodon namikawanus</i>	M21	874	F/R
<i>Justicia tukuchensis</i>	M24	898	F/R
<i>Oxytropis nepalensis</i>	M18	1033	F/R
<i>Pedicularis anserantha</i>	10P	920	F/R
<i>Pedicularis breviscaposa</i>	M06	924	F/R
<i>Pedicularis chamissonoides</i>	M07	926	F/R
<i>Primula sharmae</i>	M02	983	F/R
<i>Rhodendron cowanianum</i>	1R	966	F
<i>Rhodendron lowndesii</i>	M01	983	F/R
<i>Saxifraga excellens</i>	7S	819	F/R
<i>Silene vautierae</i>	8S	1066	F/R
<i>Sphagnum nepalense</i>	2M	739	F/R
<i>Strobilanthes saccata</i>	6S	884	F/R
<i>Thunbergia kasajuana</i>	S6	896	F/R
<i>Thunbergia nepalensis</i>	6T	899	F/R

There was huge variation of length in trnIF-trnIC marker when compared to other markers. The length of *Hypericum cordifolium* was shortest (649bp), While *Begonia taligera* was longest (1040bp).

#### 4.8 Nucleotide composition

The nucleotide composition offered insights into the proportional representation of A, T, G, and C, as a percentage of the overall sequence length.

RbcL region

The average composition of A was 29.2%, T was 27.7%, G was 20.4% and C was 22.7%.

**Table 16:** Nucleotide composition of *rbcl* markers.

<b>Name of species</b>	<b>T(U)</b>	<b>C</b>	<b>A</b>	<b>G</b>	<b>Total</b>
<i>Artemisia nepalica</i>	27.5	23.0	30.4	19.0	552
<i>Artemisia mustangensis</i>	27.5	23.1	30.4	19.0	553
<i>Begonia flagellaris</i>	26.8	22.8	30.1	20.3	552
<i>Begonia nuwakotensis</i>	26.8	23.1	30.2	19.9	553
<i>Begonia taligera</i>	26.8	23.0	30.7	19.5	553
<i>Berberis mucrifolia</i>	26.8	21.6	29.5	22.0	518
<i>Ceropegia meleagris</i>	28.0	22.8	27.1	22.1	553
<i>Cicerbita nepalensis</i>	27.2	22.1	30.7	20.0	515
<i>Clematis phlebantha</i>	27.5	22.8	29.5	20.3	553
<i>Codonopsis grey-wilsonii</i>	28.0	22.6	28.6	20.8	553
<i>Corydalis calycina</i>	27.5	23.0	28.9	20.6	553
<i>Corydalis megacalys</i>	27.9	22.6	28.6	20.8	552
<i>Corydalis terracina</i>	27.7	23.0	28.4	21.0	553
<i>Epilobium staintonii</i>	28.0	24.2	28.0	19.8	546
<i>Eremogone mukerjeeana</i>	27.5	22.6	28.8	21.0	552
<i>Galium sp.1</i>	28.0	23.3	28.8	19.9	553
<i>Galium sp.2</i>	28.0	23.3	28.9	19.7	553
<i>Griffitharia sharmae</i>	26.9	23.0	30.2	19.9	553
<i>Isodon namikawanus</i>	28.2	22.4	29.5	19.9	553
<i>Jajarkotium nepalensis</i>	26.5	23.2	29.8	20.5	551
<i>Justicia tukuchensis</i>	27.5	23.1	28.2	21.2	553
<i>Oxytropis nepalensis</i>	27.1	23.5	29.5	19.9	553
<i>Pedicularis annapurnensis</i>	28.4	22.0	28.2	21.5	550
<i>Pedicularis anserantha</i>	28.4	22.2	28.2	21.2	553
<i>Pedicularis breviscaposa</i>	28.4	22.5	28.1	21.0	552
<i>Pedicularis chamissonoides</i>	28.8	22.2	27.5	21.5	531
<i>Pedicularis sp.2</i>	28.6	22.3	28.2	20.9	546
<i>Prasiola formosana</i>	31.3	21.2	32.0	15.6	553
<i>Primula sharmae</i>	26.6	23.5	29.5	20.4	553
<i>Primula wigramiana</i>	26.4	23.1	29.8	20.6	553
<i>Rhodendron cowanianum</i>	27.8	23.0	28.8	20.4	553
<i>Rhodendron lowndesii</i>	27.9	22.9	28.7	20.5	551

Name of species	T(U)	C	A	G	Total
<i>Salvia sp.</i>	28.1	22.4	29.1	20.4	549
<i>Saxifraga excellens</i>	27.1	22.4	28.9	21.5	553
<i>Silene helleboriflora</i>	27.3	22.4	29.5	20.8	553
<i>Silene sp.</i>	27.5	22.6	29.3	20.5	552
<i>Silene vautierae</i>	27.3	22.4	29.5	20.8	553
<i>Sphagnum nepalense</i>	28.6	21.5	30.2	19.7	553
<i>Strobilanthes saccata</i>	27.5	23.4	27.7	21.4	552
<i>Taraxaccum staintoni</i>	27.5	22.4	31.4	18.7	545
<i>Thunbergia kasajuana</i>	27.8	22.5	28.1	21.6	551
<i>Thunbergia nepalensis</i>	27.8	22.2	28.4	21.5	553
Average	27.7	22.7	29.2	20.4	537.3

ITS region

The average composition of A was 22.4%, T was 20.5%, G was 28.7% and C was 29.1%.

**Table 17:** Nucleotide composition of ITS region.

Name of species	T(U)	C	A	G	Total
<i>Artemisia nepalica</i>	23.1	27.0	22.6	27.4	727
<i>Artemisia mustangensis</i>	22.8	27.3	22.4	27.6	729
<i>Begonia flagellaris</i>	23.2	24.8	23.0	29.0	797
<i>Begonia nuwakotensis</i>	20.7	29.2	21.5	28.6	797
<i>Begonia taligera</i>	20.5	29.2	21.8	28.5	785
<i>Berberis mucrifolia</i>	24.7	24.8	24.3	26.2	701
<i>Ceropegia meleagris</i>	25.1	22.9	25.3	26.7	708
<i>Cicerbita nepalensis</i>	23.6	26.4	23.1	26.9	732
<i>Clematis phlebantha</i>	16.1	31.6	24.4	27.9	639
<i>Codonopsis grey-wilsonii</i>	19.1	29.5	23.2	28.2	698
<i>Corydalis calycina</i>	13.7	35.1	21.3	29.9	635
<i>Corydalis megacalyx</i>	13.9	34.5	21.2	30.3	631
<i>Corydalis terracina</i>	14.0	35.0	21.6	29.4	629
<i>Epilobium staintonii</i>	20.5	29.5	23.4	26.7	709
<i>Eremogone mukerjeeana</i>	23.9	26.6	22.7	26.8	704
<i>Galium sp.1</i>	18.1	30.8	23.2	28.0	708
<i>Galium sp.2</i>	19.4	30.0	22.3	28.2	726

<b>Name of species</b>	<b>T(U)</b>	<b>C</b>	<b>A</b>	<b>G</b>	<b>Total</b>
<i>Griffitharia sharmae</i>	21.1	27.0	23.2	28.8	660
<i>Hypericum cordifolium</i>	20.4	28.8	24.0	26.9	726
<i>Isodon namikawanus</i>	18.8	31.5	20.0	29.7	654
<i>Jajarkotium nepalensis</i>	21.0	27.9	21.2	29.9	692
<i>Justicia tukuchensis</i>	16.3	32.5	19.0	32.2	720
<i>Oxytropis nepalensis</i>	23.4	26.0	21.6	29.0	696
<i>Pedicularis annapurnensis</i>	19.0	31.0	19.6	30.3	715
<i>Pedicularis anserantha</i>	19.0	31.2	19.5	30.3	717
<i>Pedicularis breviscaposa</i>	20.3	30.8	19.9	29.0	704
<i>Pedicularis chamissonoides</i>	19.2	29.9	20.8	30.1	678
<i>Pedicularis sp.2</i>	20.3	30.7	19.4	29.6	700
<i>Primula sharmae</i>	22.9	27.7	21.7	27.7	725
<i>Primula wigramiana</i>	24.5	26.0	22.7	26.8	723
<i>Rhodendron cowanianum</i>	24.9	26.3	22.4	26.4	738
<i>Rhodendron lowndesii</i>	25.1	26.2	22.4	26.4	738
<i>Salvia sp.</i>	18.8	31.5	19.6	30.2	709
<i>Salvia transhimalaica</i>	18.6	31.3	19.7	30.4	661
<i>Saxifraga excellens</i>	24.0	25.5	22.8	27.7	772
<i>Silene helleboriflora</i>	20.1	28.6	22.3	28.9	716
<i>Silene sp.</i>	20.3	29.1	21.6	29.0	714
<i>Silene vautierae</i>	21.3	27.8	21.3	29.7	715
<i>Sphagnum nepalense</i>	18.7	32.6	16.4	32.3	798
<i>Taraxaccum staintoni</i>	24.9	24.9	23.7	26.6	723
<i>Thunbergia kasajuana</i>	16.2	33.7	19.4	30.7	710
<i>Thunbergia nepalensis</i>	16.3	34.0	19.0	30.7	700
Average	20.5	29.1	21.7	28.7	694.4

matK region

The average composition of A was 29.0%, T was 36.7%, C was 17.6% and G was 16.7%.

**Table 18:** Nucleotide composition of matK region.

<b>Name of species</b>	<b>T(U)</b>	<b>C</b>	<b>A</b>	<b>G</b>	<b>Total</b>
<i>Artemisia mustangensis</i>	36.4	18.2	28.5	16.9	747
<i>Artemisia nepalica</i>	36.4	17.9	28.5	17.1	747
<i>Begonia flagellaris</i>	37.8	16.2	29.0	17.0	741
<i>Begonia nuwakotensis</i>	38.1	16.1	28.9	16.9	740
<i>Begonia taligera</i>	38.1	15.9	29.1	16.9	740
<i>Berberis mucrifolia</i>	35.8	18.3	29.4	16.5	744
<i>Ceropegia meleagris</i>	35.0	19.4	28.6	17.0	761
<i>Cicerbita nepalensis</i>	36.6	18.1	28.1	17.2	745
<i>Clematis phlebantha</i>	38.2	16.8	30.2	14.8	744
<i>Codonopsis grey-wilsonii</i>	35.5	18.6	28.3	17.5	741
<i>Corydalis calycina</i>	34.7	18.8	29.7	16.7	743
<i>Corydalis megacalyx</i>	35.1	18.6	28.9	17.4	743
<i>Corydalis terracina</i>	34.5	19.1	28.9	17.5	743
<i>Epilobium staintonii</i>	36.7	17.7	29.1	16.5	752
<i>Eremogone mukerjeeana</i>	38.8	15.7	30.3	15.3	740
<i>Galium sp. 1</i>	37.8	16.4	28.8	17.0	743
<i>Galium sp. 2</i>	37.2	16.8	29.0	17.0	742
<i>Griffitharia sharmae</i>	37.5	17.3	29.0	16.1	738
<i>Isodon namikawanus</i>	37.6	17.5	28.3	16.7	732
<i>Jajarkotium nepalensis</i>	37.5	17.9	28.8	15.8	739
<i>Justicia tukuchensis</i>	37.2	15.8	29.0	17.9	747
<i>Oxytropis nepalensis</i>	38.6	16.6	29.4	15.4	741
<i>Pedicularis annapurnensis</i>	35.9	17.8	29.5	16.8	746
<i>Pedicularis anserantha</i>	35.8	17.7	29.6	16.9	746
<i>Pedicularis breviscaposa</i>	35.7	17.7	29.9	16.7	742
<i>Pedicularis chamissonoides</i>	35.7	18.0	29.5	16.8	750
<i>Pedicularis sp.2</i>	36.0	17.3	30.0	16.7	747
<i>Salvia sp.</i>	35.9	18.4	28.6	17.2	746
<i>Salvia transhimalaica</i>	36.4	18.1	28.2	17.3	747
<i>Saxifraga excellens</i>	35.1	18.9	29.9	16.1	735
<i>Silene helleboriflora</i>	38.1	17.0	28.7	16.2	741
<i>Silene sp.</i>	37.8	17.1	29.0	16.2	739
<i>Silene vautierae</i>	37.8	17.1	28.9	16.2	741
<i>Strobilanthes saccata</i>	36.5	17.5	28.6	17.3	747
<i>Taraxaccum staintoni</i>	36.7	18.1	27.4	17.8	747
<i>Thunbergia kasajuana</i>	36.7	17.8	28.4	17.1	753
<i>Thunbergia nepalensis</i>	36.4	17.8	28.8	17.0	751
<b>Average</b>	<b>36.7</b>	<b>17.6</b>	<b>29.0</b>	<b>16.7</b>	<b>724.5</b>

The average composition of A was 32.3%, T was 38.6%, G was 16.3% and C was 12.7%.

**Table 19:** Nucleotide composition of trnH-psbA region.

<b>Name of species</b>	<b>T(U)</b>	<b>C</b>	<b>A</b>	<b>G</b>	<b>Total</b>
Artemisia mustangensis	41.4	12.3	30.6	15.7	464
Artemisia nepalica	41.3	12.3	30.6	15.7	470
Begonia flagellaris	43.3	12.8	27.5	16.3	436
Begonia nuwakotensis	45.4	11.7	29.3	13.5	392
Begonia taligera	43.5	12.8	28.0	15.7	446
Berberis mucrifolia	41.0	15.6	28.2	15.2	532
Ceropegia meleagris	28.9	11.9	44.9	14.3	419
Cicerbita nepalensis	39.1	11.2	32.4	17.4	466
Clematis phlebantha	28.4	21.2	36.7	13.7	401
Codonopsis grey-wilsonii	37.6	16.2	27.8	18.5	439
Corydalis calycina	35.2	16.4	28.7	19.7	543
Corydalis megacalys	37.5	14.9	29.6	17.9	469
Corydalis terracina	38.2	14.3	29.5	17.9	502
Eremogone mukerjeeana	39.9	13.5	33.0	13.5	348
Galium sp.1	40.9	11.9	33.4	13.7	335
Galium sp.2	39.5	10.8	33.4	16.2	314
Griffitharia sharmae	40.4	9.9	31.6	18.1	342
Hypericum cordifolium	39.4	13.0	32.5	15.1	624
Isodon namikawanus	36.5	11.1	31.9	20.5	351
Justicia tukuchensis	40.7	14.6	27.5	17.3	371
Oxytropis nepalensis	35.2	9.3	37.9	17.6	409
Pedicularis annapurnensis	39.2	10.4	34.9	15.5	625
Pedicularis anserantha	39.2	10.4	35.0	15.4	625
Pedicularis breviscaposa	40.3	9.9	34.3	15.4	583
Pedicularis chamissonoides	37.5	11.7	35.4	15.4	520
Pedicularis SP.2	39.0	10.2	35.3	15.4	597
Primula sharmae	39.9	11.4	31.2	17.5	343
Primula wigramiana	42.8	11.0	30.3	15.9	502
Prunus jajarkotensis	42.7	10.7	31.4	15.2	363
Rhodendron cowanianum	40.1	13.4	29.6	16.9	426
Rhodendron lowndesii	39.2	13.6	29.3	17.8	426
Salvia sp.	34.6	12.4	37.0	16.0	387
Saxifraga excellens	32.4	14.7	30.9	22.1	204
Silene helleboriflora	40.8	12.2	33.2	13.8	377
Silene sp.	40.2	11.3	33.7	14.8	371
Silene vautierae	41.0	11.6	33.2	14.2	395
Sphagnum nepalense	28.7	19.4	32.1	19.8	268
Strobilanthes saccata	38.4	13.0	31.4	17.2	516
Taraxaccum staintoni	40.1	12.2	31.1	16.6	476
Thunbergia kasajuana	34.7	14.1	35.5	15.8	519
Thunbergia nepalensis	34.4	13.2	34.4	18.0	523
Average	38.6	12.7	32.4	16.3	431.0

TrnIF- trnIC

The average composition of A was 35.0%, T was 29.8%, G was 16.4% and C was 19.2%.

**Table 20:** Nucleotide composition of trnIF-trnIC

Name of species	T(U)	C	A	G	Total
<i>Artemisia mustangensis</i>	29.1	17.2	35.3	18.4	882
<i>Begonia nuwakotensis</i>	29.3	16.7	37.7	16.4	1033
<i>Begonia taligera</i>	28.8	17.2	37.5	16.5	1036
<i>Ceropegia meleagris</i>	28.0	19.4	34.1	18.5	924
<i>Cicerbita nepalensis</i>	30.0	17.4	34.4	18.2	880
<i>Clematis phlebantha</i>	30.0	15.6	37.2	17.2	731
<i>Corydalis calycina</i>	27.9	19.9	32.7	19.5	952
<i>Corydalis megacalyx</i>	28.3	19.6	32.7	19.4	954
<i>Epilobium staintonii</i>	30.2	16.8	33.9	19.2	912
<i>Galium sp.1</i>	31.2	17.5	33.9	17.3	964
<i>Galium sp.1</i>	31.4	17.4	34.0	17.3	944
<i>Griffitharia sharmae</i>	31.9	15.9	35.5	16.6	1011
<i>Hypericum cordifolium</i>	28.5	18.3	35.4	17.8	646
<i>Isodon namikawanus</i>	27.8	18.2	35.2	18.8	874
<i>Justicia tukuchensis</i>	29.1	20.2	33.4	17.3	896
<i>Oxytropis nepalensis</i>	32.4	15.3	36.3	16.0	1027
<i>Pedicularis anserantha</i>	29.9	17.5	35.2	17.3	912
<i>Pedicularis breviscaposa</i>	29.4	17.4	35.9	17.3	923
<i>Pedicularis chamissonoides</i>	30.1	17.5	35.0	17.4	925
<i>Primula sharmae</i>	31.5	16.4	34.5	17.7	978
<i>Rhodendron cowanianum</i>	29.3	19.2	35.1	16.4	955
<i>Rhodendron lowndesii</i>	29.9	18.0	35.8	16.4	978
<i>Saxifraga excellens</i>	28.7	15.6	37.9	17.7	818
<i>Silene vautierae</i>	33.0	16.2	35.0	15.8	994
<i>Sphagnum nepalense</i>	30.8	16.4	35.6	17.2	738
<i>Strobilanthes saccata</i>	27.7	18.2	35.6	18.4	884
<i>Thunbergia kasajuana</i>	29.4	18.4	33.7	18.4	895
<i>Thunbergia nepalensis</i>	29.4	18.1	33.9	18.6	897
Average	29.8	17.6	35.0	17.6	888.2

#### 4.9 Sequence divergence

Files were imported in bioedit vs 7.2.5 and saved in fasta format. Then, fasta format files were opened in Mega11 and aligned and pairwise distance (p value) was calculated and then transferred to percentage of sequence divergence between endemic species of Nepal and other congeneric species. Sequence divergence on studied all 5 markers between endemic species of Nepal and their congeneric species growing in the neighboring countries or the nearest country to Nepal are presented under this section.

**Table 21:** Sequence divergence observed in rbcL marker, between endemic plants of Nepal and their congenics.

<b>Replica.1 Name of species</b>	<b>Replica.2 Name and accession no.</b>	<b>Country of replica2</b>	<b>P-distance</b>	<b>Divergence</b>
<i>Artemisia mustangensis</i>	<i>Artemisia salsoloides</i> (MF694951.1)	India	0.0000000000	0.0%
<i>Artemisia nepalica</i>	<i>Artemisia salsoloides</i> (MF694951.1)	India	0.0000000000	0.0%
<i>Begonia flagellaris</i>	<i>Begonia acetosella</i> (KC840120.1)	China	0.0054347826	0.54%
<i>Begonia nuwakotensis</i>	<i>Begonia acetosella</i> (KC840120.1)	China	0.0054249548	0.54%
<i>Begonia taligera</i>	<i>Begonia acetosella</i> (KC840120.1)	China	0.0072332731	0.72%
<i>Berberis mucrifolia</i>	<i>Berberis diaphana</i> (MH457119.1)	China	0.0019417476	0.19%
<i>Ceropegia meleagris</i>	<i>Ceropegia juncea</i> (MG254904.1)	India	0.0887681159	8.8%
<i>Cicerbita nepalensis</i>	<i>Melanoseris cyanea</i> (KF196043.1)	China	0.0155339806	1.55%
<i>Clematis phlebantha</i>	<i>Clematis gracillifolia</i> (MN205185.1)	China	0.0036166365	0.36%
<i>Codonopsis grey-wilsonii</i>	<i>Leptocodon hirsutus</i> (KY070680.1)	China	0.0000000000	0.0%
<i>Corydalis calycina</i>	<i>Corydalis gortschakovii</i> (MT931857.1)	China	0.0090415913	0.9%
<i>Corydalis megacalyx</i>	<i>Corydalis gortschakovii</i> (MT931857.1)	China	0.0000000000	0.0%
<i>Corydalis terracina</i>	<i>Corydalis gortschakovii</i> (MT931857.1)	China	0.0054249548	0.54%
<i>Eremogone mukerjeeana</i>	<i>Mesostemma dichotomum</i> (MK534810.1)	China	0.0018115942	0.18%
<i>Galium sp.1</i>	<i>Galium bungee</i> (MN204937.1)	China	0.0000000000	0.0%
<i>Galium sp.2</i>	<i>Galium bungee</i> (MN204937.1)	China	0.0108499096	0.10%
<i>Griffitharia sharmae</i>	<i>Cydonia oblonga</i> (MT931933.1)	China	0.0000000000	0.0%
<i>Isodon namikawanus</i>	<i>Isodon rugosus</i> (OP245107.1)	Pakistan	0.0018083183	0.18%
<i>Justicia tukuchensis</i>	<i>Rhinacanthus nasutus</i> (GQ436493.1)	China	0.0054249548	0.54%
<i>Oxytropis nepalensis</i>	<i>Oxytropis globiflora</i> (MT931465.1)	China	0.0036166365	0.36%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis decora</i> (MN185206.1)	China	0.0036363636	0.36%
<i>Pedicularis</i>	<i>Pedicularis decora</i>	China	0.0000000000	0.0%

Replica.1 Name of species	Replica.2 Name and accession no.	Country of replica2	P-distance	Divergence
<i>antherantha</i>	(MN185206.1)			
<i>Pedicularis breviscaposa</i>	<i>Pedicularis decora</i> (MN185206.1)	China	0.0072463768	0.072%
<i>Pedicularis chamissonoides</i>	<i>Pedicularis decora</i> (MN185206.1)	China	0.0301318267	3.01%
<i>Pedicularis sp.2</i>	<i>Pedicularis decora</i> (MN185206.1)	China	0.0091575092	0.91%
<i>Primula sharmae</i>	<i>Primula kialensis</i> (JF943656.1)	China	0.0090415913	0.9%
<i>Primula wigramiana</i>	<i>Primula kialensis</i> (JF943656.1)	China	0.0000000000	0.0%
<i>Rhododendron cowanianum</i>	<i>Rhododendron intricatum</i> ( KM606576.1)	China	0.00000000	0.0%
<i>Rhododendron lowndesii</i>	<i>Rhododendron intricatum</i> ( KM606576.1)	China	0.0018148820	0.18%
<i>Salvia sp.</i>	<i>Salvia digitaloides</i> (HM590090.1)	China	0.0054249548	0.54%
<i>Salvia transhimalaica</i>	<i>Salvia digitaloides</i> (HM590090.1)	China	0.0000000000	0.0%
<i>Saxifraga excellens</i>	<i>Saxifraga pseudohirculus</i> (MF100689.1)	China	0.0018083183	0.18%
<i>Silene helleboriflora</i>	<i>Silene wallichiana</i> (MF158741.1)	China	0.0000000000	0.0%
<i>Silene sp.</i>	<i>Silene wallichiana</i> (MF158741.1)	China	0.0127272727	1.27%
<i>Silene vautierae</i>	<i>Silene wallichiana</i> (MF158741.1)	China	0.0000000000	0.0%
<i>Sphagnum nepalense</i>	<i>Sphagnum subsecundum</i> (FJ572348)	China	0.0019646365	0.19%
<i>Strobilanthes saccata</i>	<i>Strobilanthes hamiltoniana</i> (OM674587.1)	Srilanka	0.0000000000	0.0%
<i>Taraxacum staintoni</i>	<i>Taraxacum officinale</i> (LC530261.1)	Japan	0.0386029412	3.8%
<i>Thunbergia kasajuana</i>	<i>Thunbergia laurifolia</i> (LC456372.1)	Thailand	0.0000000000	0.0%
<i>Thunbergia nepalensis</i>	<i>Thunbergia laurifolia</i> (LC456372.1)	Thailand	0.0018115942	0.18%

The most significant difference was observed between *Ceropegia meleagris* from Nepal and *Ceropegia juncea* from India. While 14 Plants could not be differentiated using rbcL. The range of genetic variation within different plant species was as follows: *Rhododendron* ranged from 0.0% to 0.18%, *Begonia* from 0.54% to 0.72%, *Silene* from 0.0% to 1.27%, *Primula* from 0.0% to 0.9%, *Pedicularis* from 0.0% to

3.01%, *Galium* from 0.0% to 0.10%, *Salvia* from 0.0% to 0.9%, *Corydalis* from 0.0% to 0.9%, *Thunbergia* from 0.0% to 0.18%, and *Artemisia* exhibited no genetic divergence.

**Table 22:** Sequence divergence observed in *rbcl* marker, between congeneric endemic plants of examined during this study.

Species name	Congenerics	p-distance	Divergence
<i>Artemisia mustangensis</i>	<i>Artemisia nepalica</i>	0.0000000000	0.0%
<i>Begonia flagellaris</i>	<i>Begonia taligera</i>	0.0090579710	0.90%
<i>Begonia flagellaris</i>	<i>Begonia nuwakotensis</i>	0.0108695652	1.08%
<i>Begonia taligera</i>	<i>Begonia nuwakotensis</i>	0.0054249548	0.54%
<i>Corydalis megacalyx</i>	<i>Corydalis calycina</i>	0.00906	0.90%
<i>Corydalis terracina</i>	<i>Corydalis megacalyx</i>	0.0054347826	0.54%
<i>Corydalis terracina</i>	<i>Corydalis calycina</i>	0.00723	0.72%
<i>Galium sp.1</i>	<i>Galium sp.1</i>	0.0108499096	1.08%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis sp.2</i>	0.0092081031	0.92%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis breviscaposa</i>	0.0109289617	1.09%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis breviscaposa</i>	0.0323809524	3.23%
<i>Pedicularis anserantha</i>	<i>Pedicularis annapurnensis</i>	0.0036363636	0.36%
<i>Pedicularis anserantha</i>	<i>Pedicularis sp.2</i>	0.0091575092	0.91%
<i>Pedicularis anserantha</i>	<i>Pedicularis breviscaposa</i>	0.0072463768	0.72%
<i>Pedicularis anserantha</i>	<i>Pedicularis breviscaposa</i>	0.0284090909	2.84%
<i>Pedicularis breviscaposa</i>	<i>Pedicularis chamissonoides</i>	0.0303605313	3.03%
<i>Pedicularis sp.2</i>	<i>Pedicularis breviscaposa</i>	0.0055045872	0.55%
<i>Pedicularis sp.2</i>	<i>Pedicularis chamissonoides</i>	0.0325670498	3.25%
<i>Primula wigramiana</i>	<i>Primula sharmae</i>	0.0090415913	0.90%
<i>Rhodendron cowanianum</i>	<i>Rhodendron lowndesii</i>	0.0054249548	0.54%
<i>Salvia transhimalaica</i>	<i>Salvia sp.</i>	0.0054644809	0.54%
<i>Silene helleboriflora</i>	<i>Silene vautierae</i>	0.0000000000	0.0%
<i>Silene helleboriflora</i>	<i>Silene sp.</i>	0.0127272727	1.27%
<i>Silene vautierae</i>	<i>Silene sp.</i>	0.0127272727	1.27%
<i>Thunbergia nepalensis</i>	<i>Thunbergia kasajuana</i>	0.0018148820	0.18%

Intraspecific distance between endemic plants of Nepal was 0.000 to 0.032.

**Table 23:** Sequence divergence observed in ITS marker, between endemic plants of Nepal and their congeners

Name of species	Replica2 Name and accession no.	Country of replica 2	P-distance	Divergence
<i>Artemisia cf. nepalica</i>	<i>Artemisia japonica</i> (LC645200.1)	Japan	0.0565517241	5.65%
<i>Artemisia mustangesis</i>	<i>Artemisia japonica</i> (LC645200.1)	China	0.0000000000	0.0%
<i>Begonia flagellaris</i>	<i>Begonia aptera</i> (AJ491196.1)	Taiwan	0.1037131882	10.37%
<i>Begonia nuwakotensis</i>	<i>Begonia aptera</i> (AJ491196.1)	China	0.0382653061	3.82%
<i>Begonia taligera</i>	<i>Begonia aptera</i> (AJ491196.1)	China	0.0453367876	4.53%
<i>Berberis mucrifolia</i>	<i>Berberis circumserrata</i> (MH711369.1)	China	0.0042857143	0.42%
<i>Ceropegia meleagris</i>	<i>Ceropegia salicifolia</i> (MZ927496.1)	China	0.0588235294	5.88%
<i>Cicerbita nepalensis</i>	<i>Melanoseris cyanea</i> (KF485617.1)	China	0.0015625000	0.15%
<i>Clematis phlebantha</i>	<i>Clematis chinensis</i> (MH711588.1)	China	0.0125786164	1.25%
<i>Codonopsis grey-wilsonii</i>	<i>Pseudocodon grey-wilsonii</i> (BOP018127)	China	0.0175438596	1.75%
<i>Corydalis calycina</i>	<i>Corydalis gortschakovii</i> (MT924006.1)	China	0.0318979266	3.18%
<i>Corydalis megacalyx</i>	<i>Corydalis gortschakovii</i> (MT924006.1)	China	0.0241157556	2.41%
<i>Corydalis terracina</i>	<i>Corydalis gortschakovii</i> (MT924006.1)	China	0.0367412141	3.67%
<i>Epilobium staintonii</i>	<i>Epilobium amurense</i> (MH710698.1)	China	0.0000000000	0.0%
<i>Eremogone mukerjeeana</i>	<i>Mesostemma platyphyllum</i> (T624584.1)	Afganstan	0.0753623188	7.53%
<i>Galium sp.1</i>	<i>Galium asperifolium</i> (MG906010.1)	China	0.0174927114	1.74%
<i>Galium sp.2</i>	<i>Galium asperifolium</i> (MG906010.1)	China	0.0958702065	9.58%
<i>Hypericum cordifolium</i>	<i>Hypericum ascyron</i> (MH712615.1)	China	0.0096418733	0.96%
<i>Isodon namikawanus</i>	<i>Isodon wardii</i> (MG232763.1)	China	0.0076452599	0.76%
<i>Jajarkotium nepalensis</i>	<i>Prunus cerasoides</i> (MT811820.1)	India	0.0087847731	0.87%
<i>Justicia</i>	<i>Justicia quadrifaria</i>	China	0.0252365931	2.52%

Name of species	Replica2 Name and accession no.	Country of replica 2	P-distance	Divergence
<i>tukuchensis</i>	(KP744335.1)			
<i>Oxytropis nepalensis</i>	<i>Oxytropis chinglingensis</i> (MH808471.1)	China	0.0057887120	0.57%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis striata</i> (MH711295.1)	China	0.0225352113	2.25%
<i>Pedicularis anserantha</i>	<i>Pedicularis striata</i> (MH711295.1)	China	0.0224719101	2.24%
<i>Pedicularis breviscaposa</i>	<i>Pedicularis striata</i> (MH711295.1)	China	0.0758226037	7.58%
<i>Pedicularis sp.2</i>	<i>Pedicularis striata</i> (MH711295.1)	China	0.0787965616	7.87%
<i>Primula sharmae</i>	<i>Primula purdomii</i> (MH711136.1)	China	0.0497925311	4.97%
<i>Primula wigramiana</i>	<i>Primula purdomii</i> (MH711136.1)	China	0.0417246175	4.17%
<i>Rhodendron cowanianum</i>	<i>Rhododendron intricatum</i> (KM605999.1)	China	0.0000000000	0.0%
<i>Rhodendron lowndesii</i>	<i>Rhododendron intricatum</i> (GLM-081075)	China	0.0027972028	0.27%
<i>Salvia sp.</i>	<i>Salvia nubicola</i> (MG824236.1)	China	0.0015151515	0.15%
<i>Salvia transhimalaica</i>	<i>Salvia nubicola</i> (MG824236.1)	China	0.0423861852	4.23%
<i>Saxifraga excellens</i>	<i>Saxifraga przewalskii</i> (LN812508.1)	China	0.0181582361	1.81%
<i>Silene helleboriflora</i>	<i>Silene adenopetala</i> (MT923247.1)	China	0.0223463687	2.23%
<i>Silene sp.</i>	<i>Silene adenopetala</i> (MT923247.1)	China	0.0140252454	1.40%
<i>Silene vautierae</i>	<i>Silene adenopetala</i> (MT923247.1)	China	0.0223776224	2.23%
<i>Taraxacum staintoni</i>	<i>Taraxacum sp. 1</i> YX-2018 (MH712636.1)	China	0.0237099024	2.37%
<i>Thunbergia kasajuana</i>	<i>Thunbergia coccinea</i> (HG004805.1)	China	0.0085227273	0.85%
<i>Thunbergia nepalensis</i>	<i>Thunbergia coccinea</i> (HG004805.1)	China	0.0143884892	1.43%

The genetic distance, measured by P distance, between Nepal's native plants and those from neighboring countries using ITS ranged from 0.00 to 0.07. Within specific plant species, the divergence spanned as follows: *Rhodendron* exhibited a range of 0.0% to 0.27%, *Begonia* ranged from 3.82% to 10.37%, *Silene* showed divergence between 1.40% to 2.23%, *Primula* varied between 4.17% to 4.97%, *Pedicularis* showed divergence from 2.24% to 7.87%, *Galium* showed a range of 1.74% to 9.58%, *Salvia* exhibited divergence between 0.15% to 4.23%, *Corydalis* varied from 2.41% to

3.67%, *Thunbergia* showed divergence from 0.85% to 1.43%, and *Artemisia* exhibited a range of 0.0% to 5.65%.

**Table 24:** Sequence divergence observed in ITS marker, between congeneric endemic plants of examined during this study

Species name	Congenerics	P-distance	Divergence
<i>Artemisia mustangensis</i>	<i>Artemisia nepalica</i>	0.0527045770	5.27%
<i>Begonia flagellaris</i>	<i>Begonia taligera</i>	0.1014120668	10.14%
<i>Begonia flagellaris</i>	<i>Begonia nuwakotensis</i>	0.0986093552	9.86%
<i>Begonia taligera</i>	<i>Begonia nuwakotensis</i>	0.0217113665	2.17%
<i>Corydalis megacalyx</i>	<i>Corydalis calycina</i>	0.0417335474	4.17%
<i>Corydalis terracina</i>	<i>Corydalis megacalyx</i>	0.0388978930	3.88%
<i>Corydalis terracina</i>	<i>Corydalis calycina</i>	0.0448000000	4.48%
<i>Galium sp.1</i>	<i>Galium sp.2</i>	0.0981240981	9.81%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis sp.2</i>	0.0760401722	7.60%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis breviscaposa</i>	0.0744985673	7.44%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis chamissonoides</i>	0.0471841705	4.71%
<i>Pedicularis anserantha</i>	<i>Pedicularis annapurnensis</i>	0.0027972028	0.27%
<i>Pedicularis anserantha</i>	<i>Pedicularis sp.2</i>	0.0772532189	7.72%
<i>Pedicularis anserantha</i>	<i>Pedicularis breviscaposa</i>	0.0757142857	7.57%
<i>Pedicularis anserantha</i>	<i>Pedicularis chamissonoides</i>	0.0471841705	4.71%
<i>Pedicularis breviscaposa</i>	<i>Pedicularis chamissonoides</i>	0.0900621118	9.00%
<i>Pedicularis sp.2</i>	<i>Pedicularis breviscaposa</i>	0.0772532189	7.72%
<i>Pedicularis sp.2</i>	<i>Pedicularis chamissonoides</i>	0.0870917574	8.70%
<i>Primula wigramiana</i>	<i>Primula sharmae</i>	0.0574229692	5.74%
<i>Rhodendron cowanianum</i>	<i>Rhodendron lowndesii</i>	0.0040816327	0.40%
<i>Salvia transhimalaica</i>	<i>Salvia sp.</i>	0.0406250000	4.06%
<i>Silene helleboriflora</i>	<i>Silene vautierae</i>	0.0322128852	3.22%
<i>Silene helleboriflora</i>	<i>Silene sp.</i>	0.0252808989	2.52%
<i>Silene vautierae</i>	<i>Silene sp.</i>	0.0267229255	2.67%
<i>Thunbergia nepalensis</i>	<i>Thunbergia kasajuana</i>	0.0200286123	2.00%

P distance of ITS loci of endemic plants of Nepal were found to be between 0.2% to 10.14%.

**Table 25:** Sequence divergence observed in matK marker, between endemic plants of Nepal and their congenics.

Name of species	Replica 2 Nam and accession no.	Country of replica 2	P-distance	Divergence
<i>Artemisia mustangesis</i>	<i>Artemisia japonica</i> (LC622107.1)	Japan	0.0000000000	0.00%
<i>Artemisia nepalica</i>	<i>Artemisia japonica</i> (LC622107.1)	Japan	0.0026773762	0.26%
<i>Begonia flagellaris</i>	<i>Begonia grandis</i> (MH660172.1)	China	0.0013869626	0.13%
<i>Begonia nuwakotensis</i>	<i>Begonia grandis</i> (MH660172.1)	China	0.0027739251	0.277%
<i>Begonia taligera</i>	<i>Begonia grandis</i> (MH660172.1)	China	0.0013850416	0.13%
<i>Berberis mucrifolia</i>	<i>Berberis chitria</i> (MZ945423.1)	China	0.0040322581	0.40%
<i>Ceropegia meleagris</i>	<i>Ceropegia juncea</i> (MG254903.1)	India	0.0078947368	0.78%
<i>Cicerbita nepalensis</i>	<i>Melanoseris cyanea</i> (KF195999.1)	China	0.0013793103	0.13%
<i>Clematis phlebantha</i>	<i>Clematis apiifolia</i> (MN492773.1)	China	0.0040322581	0.40%
<i>Codonopsis grey-wilsonii</i>	<i>Pseudocodon grey-wilsonii</i> (KY070942.1)	China	0.0000000000	0.00%
<i>Corydalis calycina</i>	<i>Corydalis crithmifolia</i> (OQ926894.1)	China	0.0188425303	1.88%
<i>Corydalis megacalys</i>	<i>Corydalis crithmifolia</i> (OQ926894.1)	China	0.0121130552	1.21%
<i>Corydalis terracina</i>	<i>Corydalis crithmifolia</i> (OQ926894.1)	China	0.0121130552	1.21%
<i>Eremogone mukerjeeana</i>	<i>Stellaria ovatifolia</i> (OP595547.1)	China	0.0381471390	3.814%
<i>Galium sp.1</i>	<i>Galium pusillosetosum</i> (KX911116.1)	China	0.0013458950	0.13%
<i>Galium sp.2</i>	<i>Galium pusillosetosum</i> (KX911116.1)	China	0.0148247978	1.48%
<i>Griffitharia sharmae</i>	<i>commixta</i> (LC437819.1)	Japan	0.0027100271	0.27%
<i>Isodon namikawanus</i>	<i>Isodon rugosus</i> (MH939199.1)	India	0.0013661202	0.13%
<i>Jajarkotium nepalensis</i>	<i>Prunus campanulate</i> (OP279337.1)	India	0.0000000000	0.0%
<i>Justicia tukuchensis</i>	<i>Justicia wynaadensis</i> (MH373654.1)	China	0.0250696379	2.50%
<i>Oxytropis nepalensis</i>	<i>Oxytropis coerulea</i> (HM049544.1)	China	0.0094466937	0.94%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis decora</i> (MN317525.1)	China	0.0067024129	0.67%
<i>Pedicularis anserantha</i>	<i>Pedicularis decora</i> (MN317525.1)	China	0.0067024129	0.67%
<i>Pedicularis breviscaposa</i>	<i>Pedicularis decora</i> (MN317525.1)	China	0.0377358491	3.77%
<i>Pedicularis chamissonoides</i>	<i>Pedicularis decora</i> (MN317525.1)	China	0.0360480641	3.6%
<i>Pedicularis sp.2</i>	<i>Pedicularis decora</i> (MN317525.1)	China	0.0254350736	2.54%
<i>Salvia sp.</i>	<i>Salvia miltiorrhiza</i>	China	0.0013404826	0.13%

Name of species	Replica 2 Nam and accession no.	Country of replica 2	P-distance	Divergence
	(OM888653.1)			
<i>Salvia transhimalaica</i>	<i>Salvia miltiorrhiza</i> (OM888653.1)	China	0.0053547523	0.53%
<i>Saxifraga excellens</i>	<i>Saxifraga tsangchanensis</i> (KU524313.1)	China	0.0122783083	1.22%
<i>Silene helleboriflora</i>	<i>Silene nigrescens</i> (JF970637.1)	China	0.0027063599	0.27%
<i>Silene sp.</i>	<i>Silene nigrescens</i> (JF970637.1)	China	0.0067842605	0.67%
<i>Silene vautierae</i>	<i>Silene nigrescens</i> (JF970637.1)	China	0.0027063599	0.27%
<i>Strobilanthes saccata</i>	<i>Strobilanthes cusia</i> (KJ939223.1)	China	0.0040160643	0.40%
<i>Taraxacum staintoni</i>	<i>Taraxacum platyepidium</i> (HM989734.1)	China	0.00669	0.66%
<i>Thunbergia kasajuana</i>	<i>Thunbergia laurifolia</i> (LC456371.1)	China	0.0040816327	0.40%
<i>Thunbergia nepalensis</i>	<i>Thunbergia laurifolia</i> (LC456371.1)	China	0.0258503401	2.58%

The sequence divergence of matK with plants of foreign countries in terms of pairwise distance was found to be between 0.00 to 0.03. Within specific plant species, the divergence spanned as follows: The divergence range of *Begonia* was 0.13% to 0.277%, *Silene* ranged from 0.27% to 0.67%, *Pedicularis* varied from 0.67% to 3.77%, *Galium* exhibited divergence from 0.13% to 1.48%, *Salvia* ranged from 0.13% to 0.53%, *Corydalis* varied from 1.1% to 1.88%, *Thunbergia* ranged 0.40% to 2.58%, and *Artemisia* showed range of 0.0% to 0.26%.

**Table 26:** Sequence divergence observed in matK marker, between congeneric endemic plants of examined during this study.

Species name	Congenerics	p-distance	Divergence
<i>Artemisia mustangensis</i>	<i>Artemisia nepalica</i>	0.0026773762	0.26%
<i>Begonia flagellaris</i>	<i>Begonia taligera</i>	0.0027173913	0.27%
<i>Begonia flagellaris</i>	<i>Begonia nuwakotensis</i>	0.0040816327	0.40%
<i>Begonia taligera</i>	<i>Begonia nuwakotensis</i>	0.0013531800	0.13%
<i>Corydalis megacalys</i>	<i>Corydalis calycina</i>	0.0283018868	2.83%
<i>Corydalis terracina</i>	<i>Corydalis megacalys</i>	0.0215633423	2.15%
<i>Corydalis terracina</i>	<i>Corydalis calycina</i>	0.0229110512	2.29%
<i>Galium sp.1</i>	<i>Galium sp.2</i>	0.0161943320	1.61%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis sp.2</i>	0.0268096515	2.68%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis breviscaposa</i>	0.0377867746	3.77%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis chamissonoides</i>	0.0335570470	3.35%
<i>Pedicularis anserantha</i>	<i>Pedicularis annapurnensis</i>	0.0026845638	0.26%

Species name	Congenerics	p-distance	Divergence
<i>Pedicularis anserantha</i>	<i>Pedicularis sp.2</i>	0.0254691689	2.54%
<i>Pedicularis anserantha</i>	<i>Pedicularis breviscaposa</i>	0.0391363023	3.91%
<i>Pedicularis anserantha</i>	<i>Pedicularis chamissonoides</i>	0.0335570470	3.35%
<i>Pedicularis breviscaposa</i>	<i>Pedicularis chamissonoides</i>	0.0512820513	5.12%
<i>Pedicularis sp.2</i>	<i>Pedicularis breviscaposa</i>	0.0390835580	3.90%
<i>Pedicularis sp.2</i>	<i>Pedicularis chamissonoides</i>	0.0402144772	4.02%
<i>Primula wigramiana</i>	<i>Primula Sharmae</i>		0
<i>Rhodendron cowanianum</i>	<i>Rhodendron lowndesii</i>	0.0	0.0%
<i>Salvia transhimalaica</i>	<i>Salvia sp.</i>	0.0067024129	0.67%
<i>Silene helleboriflora</i>	<i>Silene vautierae</i>	0.0026990553	0.26%
<i>Silene helleboriflora</i>	<i>Silene sp.</i>	0.0067658999	0.67%
<i>Silene vautierae</i>	<i>Silene sp.</i>	0.0067658999	0.67%
<i>Thunbergia nepalensis</i>	<i>Thunbergia kasajuana</i>	0.0213049268	2.13%

The divergence of matK among endemic plants was found between 0.0% TO 5.12%.

**Table 27:** Sequence divergence observed in trnH-psbA marker, between endemic plants of Nepal and their congenics.

Name of species	Replica 2 Name and accession no.	Country of replica 2	P-distance	Divergence
<i>Artemisia mustangesis</i>	<i>Artemisia pubescens</i> (KX581970.1)	China	0.0000000000	0.0%
<i>Artemisia nepalica</i>	<i>Artemisia pubescens</i> (KX581970.1)	China	0.0107758621	1.07%
<i>Begonia flagellaris</i>	<i>Begonia acetosella</i> (KC840100.1)	China	0.0053619303	0.53%
<i>Begonia nuwakotensis</i>	<i>Begonia acetosella</i> (KC840100.1)	China	0.0000000000	0.0%
<i>Begonia taligera</i>	<i>Begonia acetosella</i> (KC840100.1)	China	0.0000000000	0.0%
<i>Berberis mucrifolia</i>	<i>Berberis fallax</i> (MH117004.1)	China	0.0056497175	0.56%
<i>Ceropegia meleagris</i>	<i>Ceropegia longifolia</i> (LT905019.1)	China	0.0204603581	2.04%
<i>Cicerbita nepalensis</i>	<i>Melanoseris atropurpurea</i> (HQ436158.1)	China	0.0046838407	0.46%
<i>Clematis phlebantha</i>	<i>Clematis acerifolia</i> (GU732653.1)	China	0.0140056022	1.40%
<i>Codonopsis grey-wilsonii</i>	<i>Leptocodon hirsutus</i> (KY070815.1)	China	0.0000000000	0.0%
<i>Corydalis calycina</i>	<i>Corydalis curviflora</i> (MF786000.1)	China	0.0642398287	6.42%
<i>Corydalis megacalys</i>	<i>Corydalis curviflora</i> (MF786000.1)	China	0.0428265525	4.28%
<i>Corydalis terracina</i>	<i>Corydalis curviflora</i> (MF786000.1)	China	0.0427698574	4.27%

Name of species	Replica 2 Name and accession no.	Country of replica 2	P-distance	Divergence
<i>Eremogone mukerjeeana</i>	<i>Stellaria sessiliflora</i> (LC728203.1)	Japan	0.0887573964	8.87%
<i>Galium sp.1</i>	<i>Galium asperifolium</i> (MH117097.1)	China	0.0029850746	0.29%
<i>Galium sp.2</i>	<i>Galium asperifolium</i> (MH117097.1)	China	0.0541401274	5.41%
<i>Hypericum cordifolium</i>	<i>Hypericum monogynum</i> (JN044928.1)	China	0.0107334526	1.07%
<i>Isodon namikawanus</i>	<i>Isodon sp. 5 XY-2014</i> (KF855612.1)	China	0.0838323353	8.38%
<i>Jajarkotium nepalensis</i>	<i>Prunus cerasoides</i> (OL739701.1)	China	0.0027548209	0.27%
<i>Justicia tukuchensis</i>	<i>Justicia candicans</i> (KT161342.1)	Thailand	0.0831024931	8.31%
<i>Oxytropis nepalensis</i>	<i>Oxytropis glabra</i> (KF011569.1)	China	0.0131233596	1.31%
<i>Pedicularis annapurnensi</i>	<i>Pedicularis clarkei</i> (JN045913.1)	China	0.0036297641	0.36%
<i>Pedicularis anserantha</i>	<i>Pedicularis clarkei</i> (JN045913.1)	China	0.0036297641	0.36%
<i>Pedicularis breviscapos</i>	<i>Pedicularis clarkei</i> (JN045913.1)	China	0.0450097847	4.50%
<i>Pedicularis chamissonoides</i>	<i>Pedicularis clarkei</i> (JN045913.1)	China	0.1153846154	11.53%
<i>Pedicularis sp.2</i>	<i>Pedicularis clarkei</i> (JN045913.1)	China	0.0392927308	3.92%
<i>Primula sharmae</i>	<i>Primula pumilio</i> (KT259775.1)	China	0.0208955224	2.08%
<i>Primula wigramiana</i>	<i>Primula pumilio</i> (KT259775.1)	China	0.0200000000	2.0%
<i>Rhododendron cowanianum</i>	<i>Rhododendron lepidotum</i> (KM605647.1)	China	0.0093896714	0.93%
<i>Rhododendron lowndesii</i>	<i>Rhododendron lepidotum</i> (KM605647.1)	China	0.0164319249	1.64%
<i>Salvia sp.</i>	<i>Salvia kiaometiensis</i> (KC473183.1)	China	0.0051813472	0.51%
<i>Salvia transhimalaica</i>	<i>Salvia kiaometiensis</i> (KC473183.1)	China	0.0129198966	1.29%
<i>Saxifraga excellens</i>	<i>Saxifraga tibetica</i> (KR065640.1)	China	0.0319148936	3.19%
<i>Silene helleboriflora</i>	<i>Silene firma</i> (LC622134.1)	Japan	0.0353982301	3.53%
<i>Silene sp.</i>	<i>Silene firma</i> (LC622134.1)	Japan	0.0235294118	2.35%
<i>Silene vautierae</i>	<i>Silene firma</i> (LC622134.1)	Japan	0.0059171598	0.59%
<i>Sphagnum nepalense</i>	<i>Sphagnum subsecundum</i> (FJ572571.1)	China	0.0000000000	0.0%
<i>Strobilanthes saccata</i>	<i>Strobilanthes cusia</i> (KP095477.1)	China	0.0000000000	0.0%
<i>Taraxacum</i>	<i>Taraxacum officinale</i>	Pakistan	0.0063025210	0.63%

Name of species	Replica 2 Name and accession no.	Country of replica 2	P-distance	Divergence
<i>staintoni</i>	(MG582094.1)			
<i>Thunbergia kasajuana</i>	<i>Thunbergia coccinea</i> (HG005045.1)	China	0.0077972710	0.77%
<i>Thunbergia nepalensis</i>	<i>Thunbergia coccinea</i> (HG005045.1)	China	0.0638722555	6.38%

The sequence variation in trnH-psbA between plants native to Nepal and those of foreign origin ranged from 0.00 to 0.08. Within specific plant species, the divergence showed the following ranges: *Rhodendron* exhibited a range of 0.93% to 1.64%, *Begonia* ranged from 0.00% to 0.53%, *Silene* showed divergence between 0.69% to 3.53%, *Primula* varied between 2.0% to 2.08%, *Pedicularis* showed divergence from 0.36% to 11.53%, *Galium* showed a range of 0.29% to 5.41%, *Salvia* exhibited divergence between 0.51% to 1.29%, *Corydalis* varied from 4.27% to 6.42%, *Thunbergia* showed divergence from 0.77% to 6.38%, and *Artemisia* exhibited a range of 0.0% to 1.07%.

**Table 28:** Sequence divergence observed in trnH-psbA marker, between congeneric endemic plants of examined during this study.

Species name	Congenerics	p-distance	Divergence
<i>Artemisia mustangesis</i>	<i>Artemisia nepalica</i>	0.0107758621	1.07%
<i>Begonia flagellaria</i>	<i>Begonia nuwakotensis</i>	0.0518134715	5.18%
<i>Begonia flagellaris</i>	<i>Begonia taligera</i>	0.0161290323	1.61%
<i>Begonia taligera</i>	<i>Begonia nuwakotensis</i>	0.0434782609	4.34%
<i>Corydalis Megacalyx</i>	<i>Corydalis calycina</i>	0.0413943355	4.13%
<i>Corydalis terracina</i>	<i>Corydalis megacalyx</i>	0.0191897655	1.91%
<i>Corydalis terracina</i>	<i>Corydalis calycina</i>	0.0386266094	3.86%
<i>Galium sp.1</i>	<i>Galium sp.2</i>	0.0509554140	5.09%
<i>Pedicularis annapurnensi</i>	<i>Pedicularis sp.2</i>	0.0362068966	3.62%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis breviscaposa</i>	0.0466321244	4.66%
<i>Pedicularis annapurnensis</i>	<i>Pedicularis chamissonoides</i>	0.1013645224	10.13%
<i>Pedicularis anserantha</i>	<i>Pedicularis annapurnensis</i>	0.0016000000	0.16%
<i>Pedicularis anserantha</i>	<i>Pedicularis sp.2</i>	0.0362068966	3.62%
<i>Pedicularis anserantha</i>	<i>Pedicularis breviscaposa</i>	0.0449050086	4.49%
<i>Pedicularis anserantha</i>	<i>Pedicularis chamissonoides</i>	0.1013645224	10.13%
<i>Pedicularis breviscaposa</i>	<i>Pedicularis chamissonoides</i>	0.1120000000	11.2%
<i>Pedicularis sp.2</i>	<i>Pedicularis breviscaposa</i>	0.0332167832	3.32%
<i>Pedicularis sp.2</i>	<i>Pedicularis chamissonoides</i>	0.0982318271	9.82%
<i>Primula wigramiana</i>	<i>Primula sharmae</i>	0.0331325301	3.31%
<i>Rhodendron cowanianum</i>	<i>Rhodendron lowndesii</i>	0.0235849057	2.35%
<i>Salvia sp.</i>	<i>Salvia transhimalaica</i>	0.0129533679	1.29%

Species name	Congenerics	p-distance	Divergence
<i>Silene helleboriflora</i>	<i>Silene vautierae</i>	0.0272479564	2.72%
<i>Silene helleboriflora</i>	<i>Silene sp.</i>	0.0538243626	5.38%
<i>Silene vautierae</i>	<i>Silene sp.</i>	0.0174418605	1.74%
<i>Thunbergia nepalensis</i>	<i>Thunbergia kasajuana</i>	0.0596421471	5.96%

Divergence of trnH-psbA among endemic plants of Nepal was found between 0.001 to 0.1.

**Table 29:** Sequence divergence observed in trnIF-trnIC marker, between endemic plants of Nepal and their congenics.

Species name	Replica 2 Names and accession no.	Country of Replica 2	P-distance	Divergence
<i>Rhododendron cowanianum</i>	<i>Rhododendron mucronulatum</i> (AF394251.1)	China	0.05792	5.79%
<i>Sphagnum nepalense</i>	<i>Sphagnum sp. 1 ND-2016</i> (KR067389.1)	China	0.00547	0.54%
<i>Strobilanthes saccata</i>	<i>Strobilanthes colorata</i> (AB161983.1)	Thailand	0.00581	0.58%
<i>Saxifraga excellens</i>	<i>Saxifraga heleonastes</i> (LN812680.1)	China	0.00639	0.63%
<i>Silene vautierae</i>	<i>Silene aprica</i> (FN821322.1)	Japan	0.06406	6.40%
<i>Pedicularis anserantha</i>	<i>Pedicularis tongolensis</i> (KF277748.1)	China	0.00689	0.68%
<i>Thunbergia nepalensis</i>	<i>Crossandra infundibuliformis</i> (KP744154.1)	China	0.07117	7.11%
<i>ceropegia meleagris</i>	<i>Ceropegia dolichophylla</i> (LT635679.1)	China	0.00224	0.22%
<i>Corydalis megacalys</i>	<i>Corydalis thyrsoflora</i> (OQ927040.1)	China	0.01310	1.31%
<i>Corydalis calycina</i>	<i>Corydalis thyrsoflora</i> (OQ927040.1)	China	0.02295	2.29%
<i>Rhododendron lowndesii</i>	<i>Rhododendron mucronulatum</i> (AF394251.1)	China	0.00430	0.43%
<i>Primula sharmae</i>	<i>Primula modesta</i> (EF218544.1)	Japan	0.01969	1.96%
<i>Pedicularis breviscaposa</i>	<i>Pedicularis tongolensis</i> (KF277748.1)	China	0.01736	1.73%
<i>Pedicularis chamissonoides</i>	<i>Pedicularis tongolensis</i> (KF277748.1)	China	0.04828	4.82%
<i>Artemisia mustangensis</i>	<i>Artemisia capillaris</i> (AB551235.1)	Japan	0.00341	0.34%
<i>Isodon namikawanus</i>	<i>Isodon coetsa</i> (MH557917.1)	China	0.00244	0.24%
<i>Justicia tukuchensis</i>	<i>Justicia quadrifaria</i> (KP744171.1)	China	0.02190	2.19%
<i>Thunbergia kasajuana</i>	<i>Crossandra infundibuliformis</i>	China	0.07494	7.49%

Species name	Replica 2 Names and accession no.	Country of Replica 2	P-distance	Divergence
	(KP744154.1)			
<i>Begonia taligera</i>	<i>Begonia cathayana</i> (AY091818.1)	China	0.01288	1.28%
<i>Hypericum cordifolium</i>	<i>Hypericum kamtschaticum</i> (AB219639.1)	Japan	0.03306	3.30%
<i>Begonia nuwakotensis</i>	<i>Begonia cathayana</i> (AY091818.1)	China	0.01390	1.39%

Divergence in terms of p-distance of trnIF-trnIC marker when compared to foreign plants ranged from 0.002 to 0.05. Within specific plant species, the divergence showed the following ranges: *Rhodendron* varied from 0.43% to 5.79%, *Begonia* showed divergence from 1.28% to 1.39%, *Pedicularis* ranged from 0.68% to 4.82%, *Corydalis* exhibited divergence of 1.31% to 2.29% and *Thunbergia* ranged from 7.11% to 7.49%.

**Table 30:** Sequence divergence observed in trnIF-trnIC marker, between congeneric endemic plants of examined during this study.

Plant 1	Plant 2	P distance	Divergene
<i>Begonia taligera</i>	<i>Begonia nuwakotensis</i>	0.01170	1.17%
<i>Corydalis megacalys</i>	<i>Corydalis calycina</i>	0.02944	2.94%
<i>Galium sp.1</i>	<i>Galium sp.</i>	0.02123	2.12%
<i>Pedicularis anserantha</i>	<i>Pedicularis breviscaposa</i>	0.03000	3.0%
<i>Pedicularis anserantha</i>	<i>Pedicularis chamissonoides</i>	0.05519	5.51%
<i>Pedicularis breviscaposa</i>	<i>Pedicularis chamissonoides</i>	0.05482	5.48%
<i>Rhodendron cowanianum</i>	<i>Rhodendron lowndesii</i>	0.05720	5.72%
<i>Thunbergia nepalensis</i>	<i>Thunbergia kasajuana</i>	0.00896	0.89%

Sequence divergence of trnIF-trnIC between endemic plants of Nepal was found to be between 0.008 to 0.057.

#### 4.10 Comparing Sphagnum

While comparing sequence of *Sphagnum nepalense* we obtained with the sequence obtained by (Shudan et al., 2022), we found the following pairwise distance. In case of rbcL and trnH-psbA few bases were found to be different but in case of trnIF-trnIC insertion of 3,4 basepairs were found as difference. Other than *Sphagnum nepalense*, sequence of any other Nepali endemic plant was not found in NCBI.

**Table 31:** variation in Sphagnum sequences.

Markers	Pairwise distance	Percentage distance
rbcl	0.00383	0.3%
trnH-psbA	0.01	1%
trnIF-trnIC	0.0123	1.2%

#### 4.11 Effectiveness of markers across different families

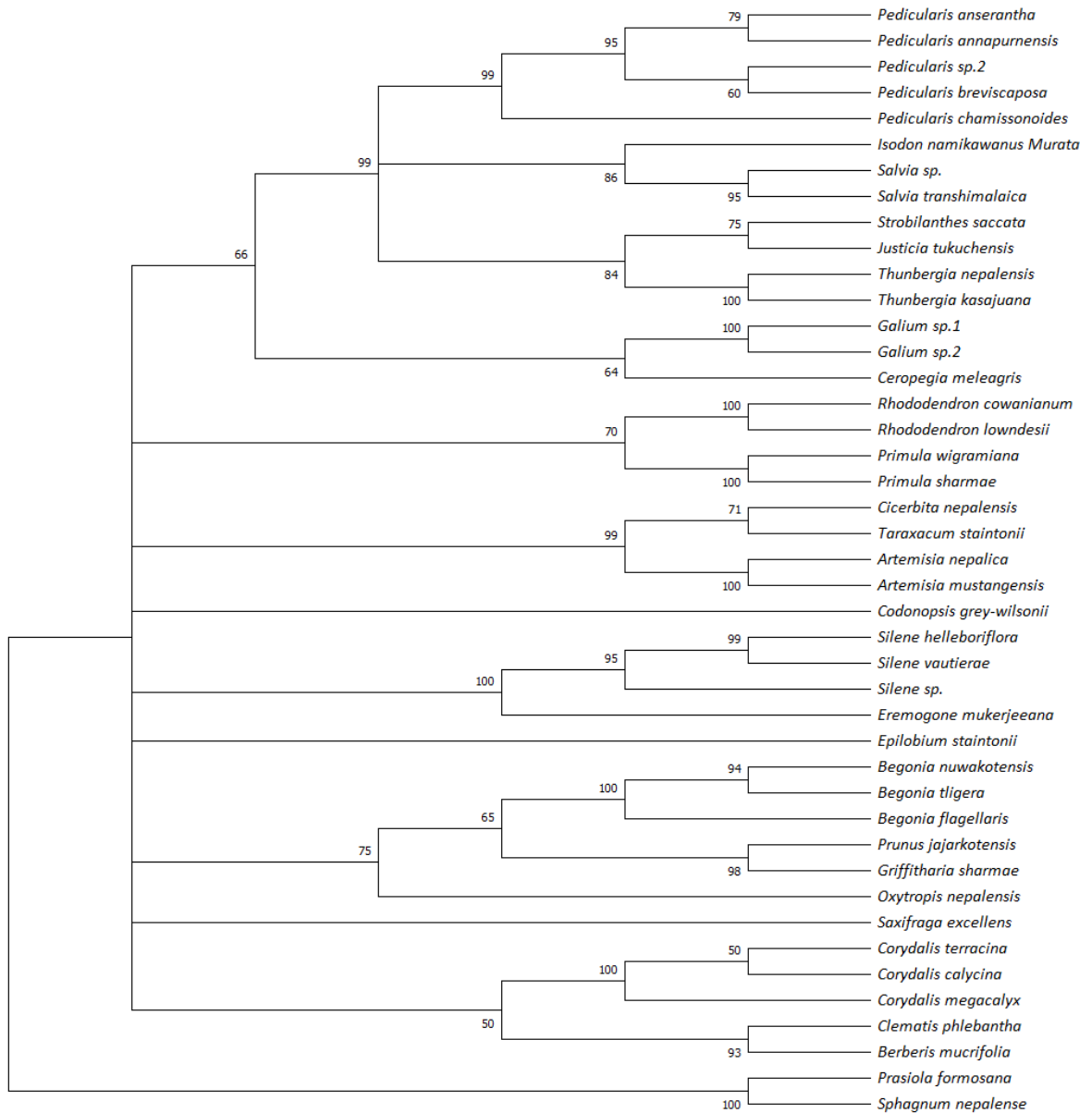
In this study, it was observed that TrnH-psbA exhibited the highest degree of divergence among the seven families examined. Following closely was the ITS marker, which displayed the highest divergence in six families. TrnIF-trnIC demonstrated the highest divergence in five families. RbcL exhibited the highest divergence within the Apocynaceae family. However, while matK showed significant overall divergence, it was not the most efficient marker in any specific family.

**Table 32:** Marker effectiveness in different families

Families	Effectiveness of markers
Acanthaceae	trnIF-trnIC > trnH-psbA/rbcl
Apocynaceae	rbcl > trnH-psbA/ITS/matK > trnIF-trnIC
Asteraceae	trnIF-trnIC/ITS > ITS/matK
Begoniaceae	ITS > matK > rbcl > trnH-psbA
Berberidaceae	trnH-psbA > ITS/matK > rbL
Campanulaceae	ITS > matK/rbcl/trnH-psbA
Caryophyllaceae	trnH-psbA > ITS > matK > rbcl
Ericaceae	Trnlf-trnlc > trnH-psbA > ITS > rbcl
Fabaceae	trnH-psbA > matK > ITS > rbcl
Hypericaceae	trnIF-trnIC > ITS > trnH-psbA
Lamiaceae	trnH-psbA > ITS/matK > rbcl
Onagraceae	
Orobanchaceae	ITS > matK > trnH-psbA
Papaveraceae	trnH-psbA > ITS > matK > trnIF-trnIC > rbcl
prasiolaceae	
Primulaceae	ITS > trnH-psbA/trnIF-trnIC > rbcl
Ranunculaceae	trnH-psbA > ITS > matK > rbcl
Rosaceae	ITS > matK
Rubiceae	ITS > trnH-psbA > matK > rbcl
Saxifragaceae	trnH-psbA > ITS > matK > rbcl
Sphagnaceae	trnIF-trnIC > rbcl > trnH-psbA

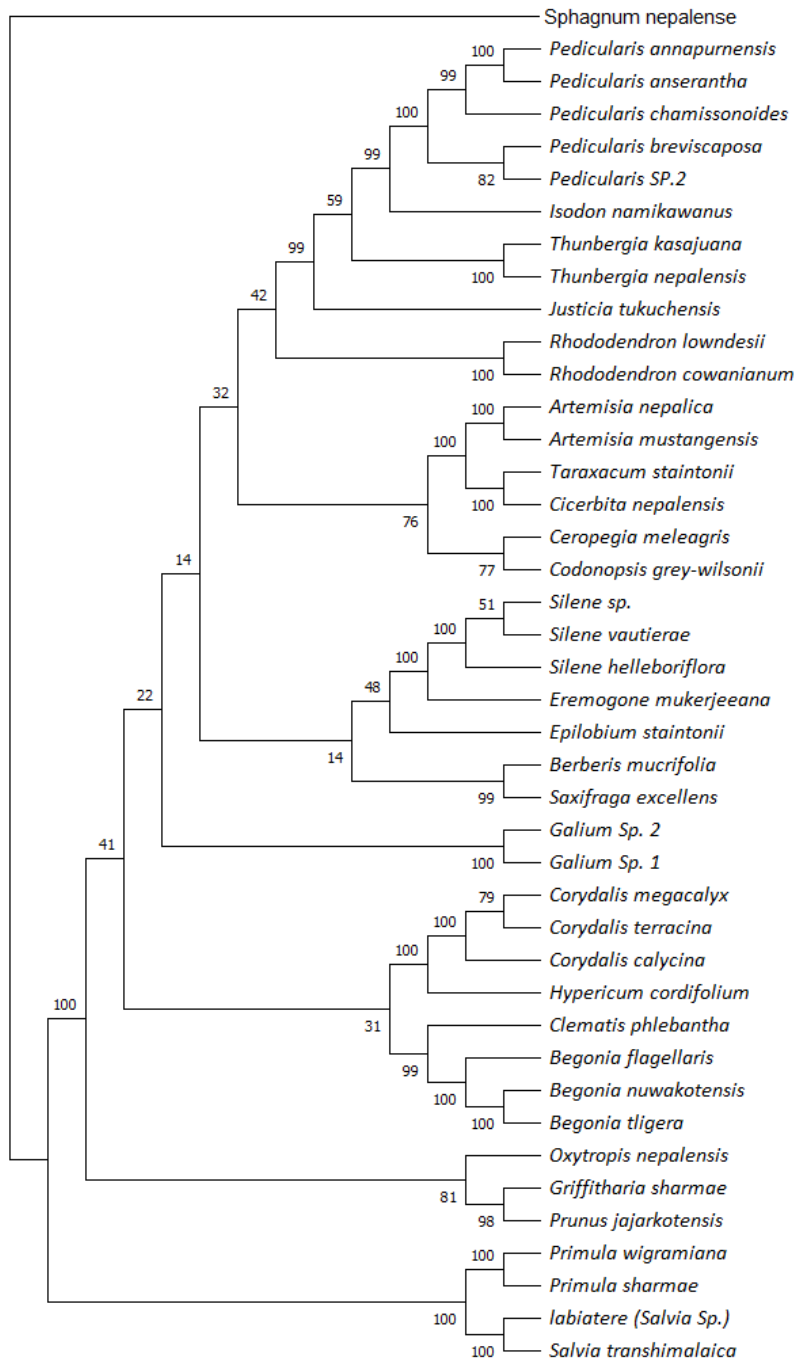
#### 4.10 Phylogenetic inference: Cladogram

Phylogenetic tree was constructed using five markers viz. ITS, matK rbcL, trnH-psbA, trnI-trnK.



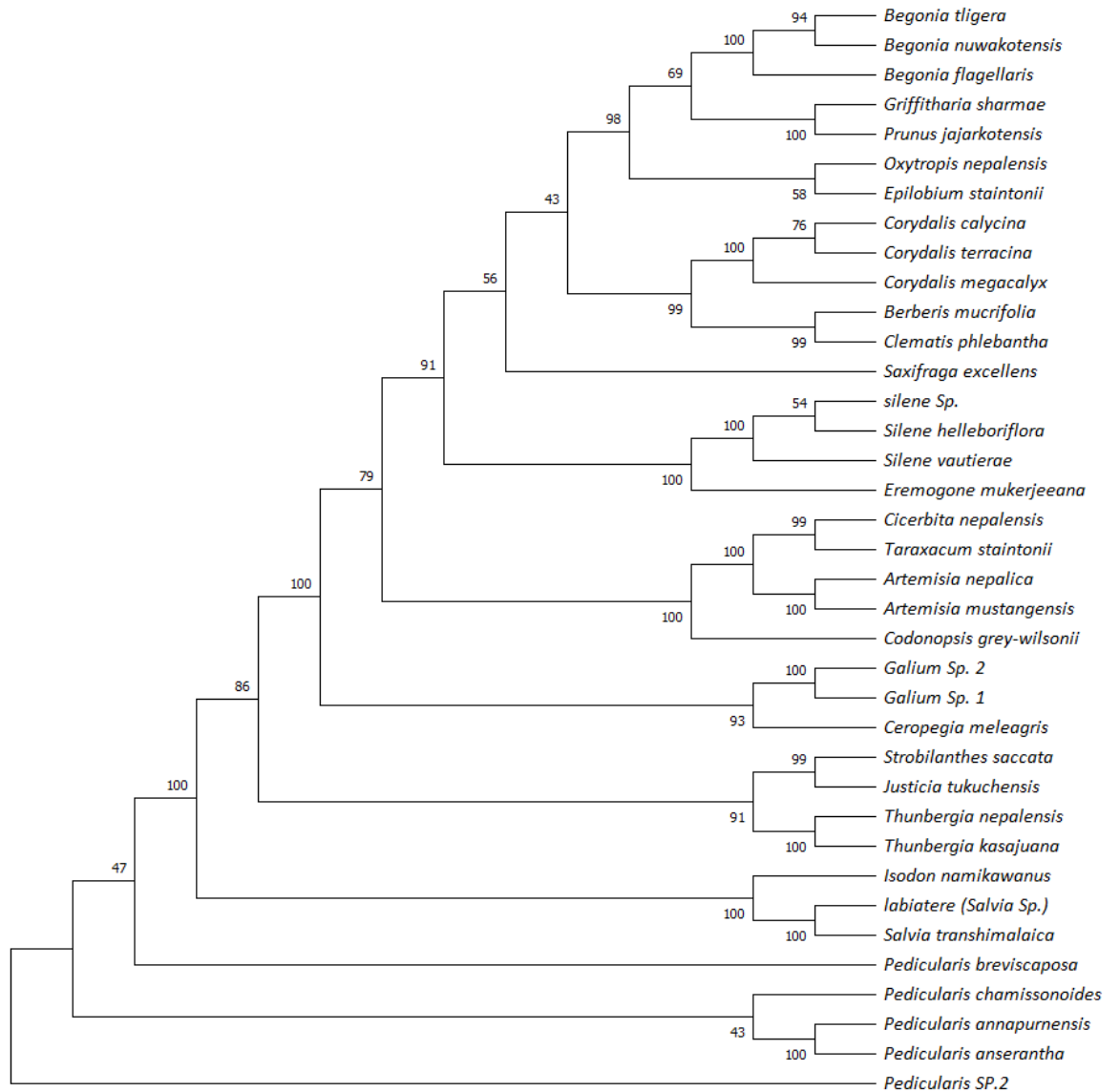
**Figure 8:** phylogenetic tree on basis of rbcL region using Neighbr-joining method.

The evolutionary analysis employed the Neighbor-joining method within MEGA11 software. The percentage of identical tree replicas where related taxas clustered together in the bootstrap examination (1000 repetitions) is displayed alongside branches. Evolutionary distances, calculated using the p-distance method and are expressed as base differences per site units.



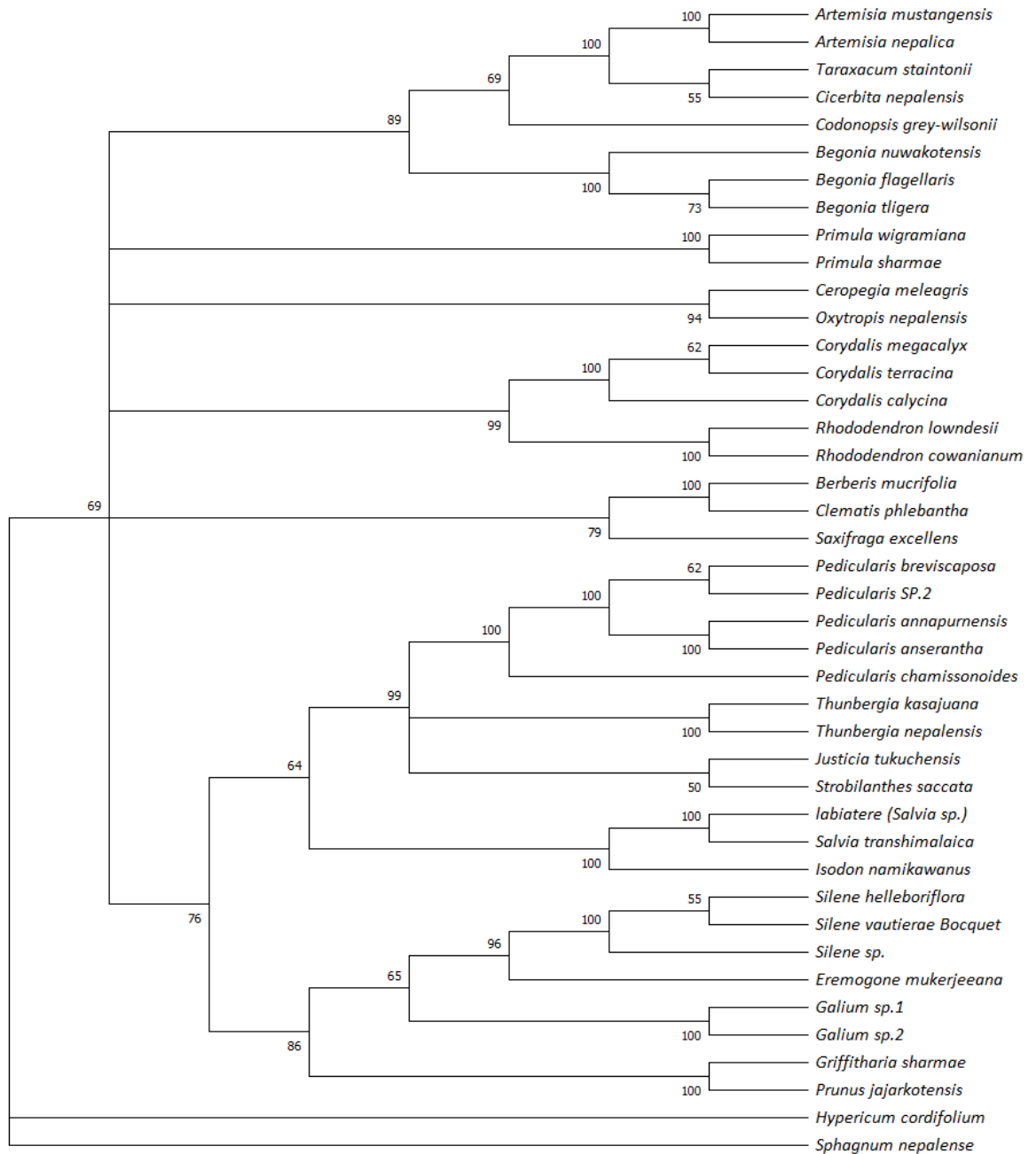
**Figure 9:** phylogenetic tree on basis of ITS region using Neighbor-joining method.

The evolutionary analysis employed the Neighbor-joining method within MEGA11 software. The percentage of identical tree replicas where related taxas clustered together in the bootstrap examination (1000 repetitions) is displayed alongside branches. Evolutionary distances, calculated using the p-distance method and are expressed as base differences per site units.



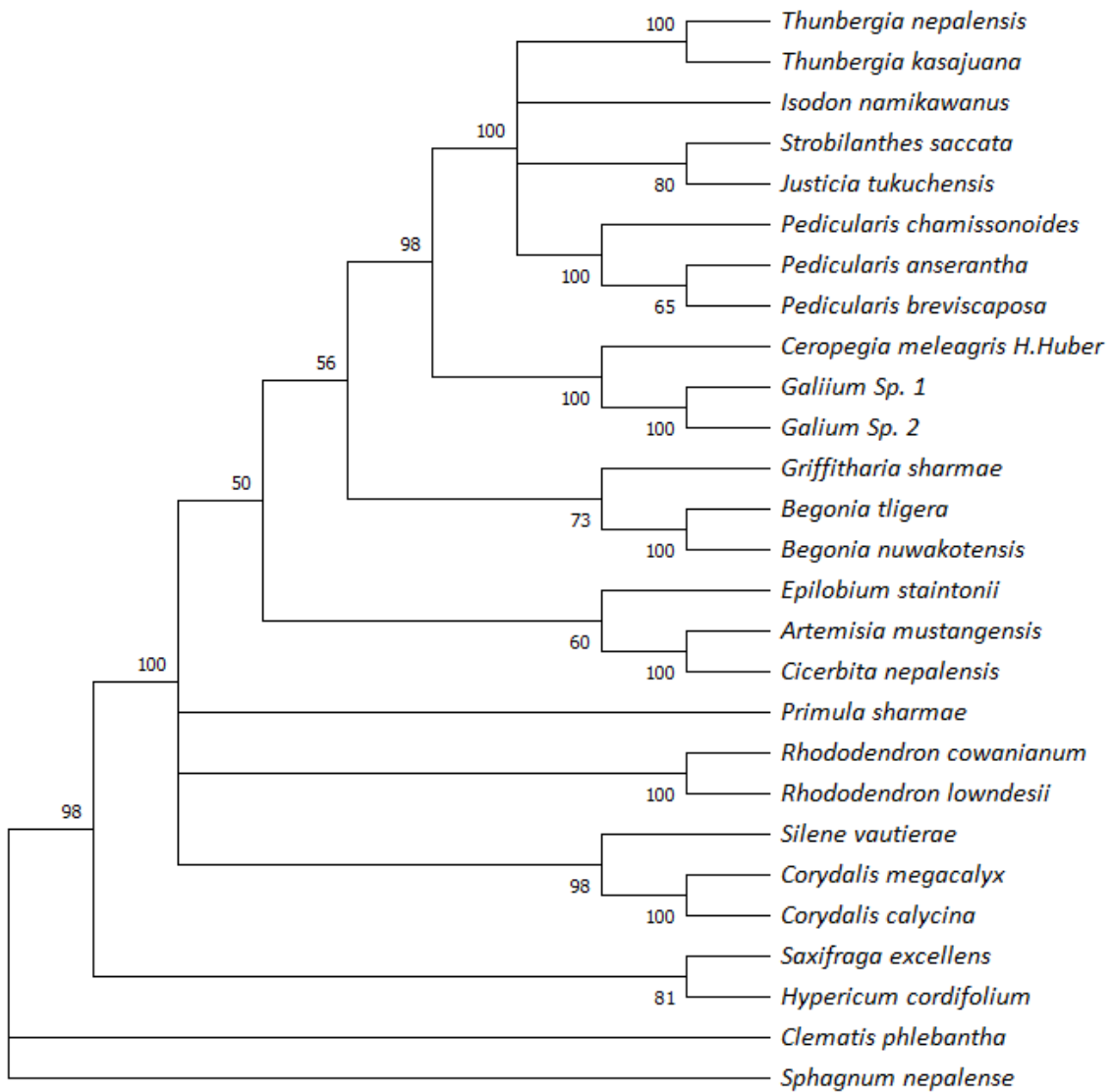
**Figure 10:** Phylogenetic tree on basis of matK region using Neighbr-joining method.

The evolutionary analysis employed the Neighbor-joining method within MEGA11 software. The percentage of identical tree replicas where related taxas clustered together in the bootstrap examination (1000 repetitions) is displayed alongside branches. Evolutionary distances, calculated using the p-distance method and are expressed as base differences per site units.



**Figure 11:** phylogenetic tree on basis of trnH-psbA region using Neighbr-joining method

The evolutionary analysis employed the Neighbor-joining method within MEGA11 software. The percentage of identical tree replicas where related taxa clustered together in the bootstrap examination (1000 repetitions) is displayed alongside branches. Evolutionary distances, calculated using the p-distance method and are expressed as base differences per site units.



**Figure 12:** phylogenetic tree on basis of trnIF-trnIC region using Neighbor-joining method

The evolutionary analysis employed the Neighbor-joining method within MEGA11 software. The percentage of identical tree replicas where related taxa clustered together in the bootstrap examination (1000 repetitions) is displayed alongside branches. Evolutionary distances, calculated using the p-distance method and are expressed as base differences per site units.

#### 4.11 Phylogeny

Species of family Acanthaceae, lamiaceae and Orobanchaceae made a clade while using rbcL, matK, trnH-psbA and trnIF-trnIC markers but while using ITS marker Two salvia species of family lamiaceae were found to form separate clade. Galium species and *Ceropogia meleagris* made a clade while using all markers except trnH-psbA. *Ceropogia meleagris* and *Oxytropis nepalensis* were seen to form a clade in

phylogenetic tree formed using trnH-psbA marker. Rhodendron species were seen to form clade with Primula species while using rbcL, but formed clade with Papaveraceae family while using trnH-psbA. Similarly, Rhodendron species were seen to be closer to different other species (Primulaceae and Acanthaceae) while using trnF-trnC and ITS marker. Species of family Begoniaceae and Rosaceae formed a clade while using matK, rbcL and trnF-trnC marker but when using ITS marker Rosaceae family made a clade with *Oxytropis nepalensis*. Species of Asteraceae made clade with *Codonopsis grey-wilsonii* when using rbcL, ITS, matK, and trnH-psbA markers. *Berbersi mucrifolia* and *Clematis phlebantha* were found to form a clade, while using rbcL, matK and trnH-psbA markers, but while using ITS marker *Berbersi mucrifolia* made clade with *Saxifraga excellens* and *Clematis phlebantha* made clade with members of Begoniaceae.

## CHAPTER 5

### DISCUSSION

#### 5.1 Identification challenges

Amalgamating various aspects such as morphology (including reproductive structures, leaves, phyllotaxy, and wood structure), anatomy and biochemistry were used to identify plants. Generally morphological based identification of species becomes challenging when congeneric species including endemic species are taxonomically discriminated based on few overlapping character states.

In the present study we have 18 species under 8 families (Acanthaceae, Asteraceae, Begoniaceae, Caryophyllaceae, Ericaceae, Lamiaceae, Papveraceae and Primulaceae) having more than one species in each genus. Morphological discrimination of these species with congeners is always a challenge. DNA barcoding, which targets specific genomic regions, revealed significant genetic differences between endemic species and their congeners. This was apparent in *Artemisia*, *Begonia*, *Thunbergia* and *Corydalis* genera while compared using ITS, MatK, TrnLF and TrnH-PsbA markers. Similar to what Hebert & Gregory, (2005) suggested, our finding highlights the utility of DNA barcoding in elucidating cryptic species diversity and refining taxonomic classifications. Furthermore, the DNA barcodes generated in our study have contributed to the establishment of a comprehensive reference database for endemic plant species in Nepal. This database serves as a valuable resource for researchers, conservationists, and policymakers involved in biodiversity conservation and management efforts. With DNA barcodes readily available for reference, the identification process becomes more accessible, precise, and efficient.

#### 5.2 DNA extraction and PCR

The DNA extraction process for endemic plants was relatively easy due to the use of a DNA extraction kit. The overall success rate for PCR amplification of these plants was 88.6%. Among the markers, *rbcl* proved to be the easiest to amplify, followed by the *trnH-psbA* marker. The addition of BSA and DMSO made amplifying the ITS marker less challenging when compared to *matK* and *trnL-trnC*. Utilizing BSA and DMSO aids in PCR amplification of samples that may have inhibitors such as phenolic compounds or have high GC content (Farell & Alexandre, 2012). This might be why we observed easier amplification of ITS marker.

The PCR optimization for *rbcl* marker was not challenging. PCR efficiency of *rbcl* to be 97.7 %. The *trnH-psbA* marker showed same PCR efficiency. Successful PCR amplification of *rbcl* and *trnH-psbA* were close as found by Kress & Erickson, 2007b. In case of *Parsiola*, which was not preserved properly, it only successfully amplified and sequenced using *rbcl* marker.

Optimizing PCR reagents for ITS was a little challenging than *rbcl* and *trnH-psbA* but not difficult as *matK* and *trnL-trnC*. In our research, the PCR efficiency of the ITS region was found to be 95%, a result that closely aligns with an in-silico investigation done by Cheng et al., 2016 where ITS-u4 exhibited a PCR success rate exceeding 97%, while ITS-p5 displayed a coverage rate of 89%.

The pursuit of discovering a single, universally effective set of primers for consistently amplifying the *matK* gene across all terrestrial plant species to enable efficient, quick, automated, and cost-effective species identification has been a challenging goal to achieve. An examination of sequence data available in GenBank/EBI reveals that even among monocots and eudicots, the majority of bases within the *matK* gene exhibit variability (Dunning & Savolainen, 2010). Eventhough, we achieved 84%. Success with *matK* amplification, PCR optimization of *matK* was huge challenge.

The study on fern shown PCR success rate of *trnL* to be as high as *rbcl* (around 95%) (Groot et al., 2011a). However, in our study PCR success of *trnL* was only around 68%. The non-specific bands, smearing and lack of full length were some of difficulty faced when optimizing *trnL* marker.

### **5.3 Sequences analysis and interpretation**

*RbcL* was easiest marker to make consensus from forward and reverse sequences. ITS and *trnL* markers had more ambivalent peaks than *trnH-psbA* and *matK*. Hence making ITS and *trnL* difficult to prepare consensus sequence. We utilized the NCBI's BLASTn tool to match consensus with available data sets. Most of them showed only genus level identification. While aligning the sequences *trnL-trnC* and *trnH-psbA* sequences were difficult to align and we had to reverse compliment the sequences for several trials to form the proper alignment.

### **5.4 Variations in lengths**

In our study all *rbcl* regions except 4 individuals had exactly 553 bp length and even these 4 plants had around 553bp, similar result was reported by Trujillo-Argueta et al. (2022). The length of ITS for Angiosperm and Bryophytes using p5/u4 primer pair is proposed to be 617 to 897bp and 855 to 1109bp (Cheng et al., 2016b). Our

findings were consistent with proposed length of Angiosperm. However, *Sphagnum* was only 789bp long, suggesting bryophytes might have short length than proposed.

The length of matK marker's amplified using primer 472 F/1248R is reported to be 710 bp (Fatima et al., 2019). In our data, we found the length of matK in our sample to be between 732bp to 765bp. Loera-Sánchez et al. (2020) suggested the length of trnH-psbA to be 268 to 614 bp. However, in our study, one sample had less than 268 bp length and 3 samples were more than 614bp long. The length of trnH-psbA was between 207bp to 625bp and the average was 431bp. TrnIF-trnIC was about 1000bp long, which was as suggested by Wilson et al. (2016).

## **5.5 Utilities of marker**

### **5.5.1 RbcL**

We analyzed a total of 40 plants with rbcL marker, and among them, 14 of our plants showed no sequence differences when compared to foreign plants. Notably, *Ceropegia meleagris* exhibited the highest sequence divergence at 8.87%. when comparing the rbcL sequences among congenics used in this study, *Silene helleboriflora* and *Silene vautierae*, there was no discernible differentiation. A similar situation was observed between *Artemisia mustangensis* and *Artemisia nepalica*. On the other hand, the most significant divergence occurred between *Pedicularis spp* and *Pedicularis chamissonoides*, with a divergence value of 3.25%. As suggested by Vere et al. (2015), we found that rbcL sequences were quite similar in closely related species.

### **5.5.2 ITS**

When comparing the ITS genes of 40 of our plant species to foreign ones, the ITS analysis successfully differentiated our native plants from foreign ones, with the exception of *Artemisia mustangensis*, *Rhodendron cowanianum* and *Epilobium*. The greatest genetic divergence was observed between *Begonia flagellaris* and *Begonia aptera* of Taiwan, reaching a value of 10.37%. ITS exhibited the highest level of genetic differentiation between these species. When focusing on our native plant species, the ITS analysis effectively differentiated all of them from each other. The most significant genetic disparity was found between *Begonia flagellaris* and *Begonia taligera*, measuring 10.14% divergence. The level of divergence ranged from 0.2% to 10.14% among these endemic plants. The internal transcribed spacer region within nuclear ribosomal DNA is considered a potential DNA barcode due to several beneficial attributes. These include the presence of conserved regions suitable for

designing universal primers, the simplicity of amplification, and adequate variability to differentiate closely related species(Yao et al., 2010).

### **5.5.3 matK**

When comparing matK markers of 36 plants with foreign plants, all plant showed some genetic divergence with foreign species except *Jajarkotium nepalensis*, *Artemisia mustangensis* and *codonopsis grey-wilsonii*. *Eremogone mukerjeeana* was found to have highest divergence of 0.038 with foreign plant. Sequence divergence was not as high as ITS region. While comparing between of our plant of Nepal matK was able to differentiate among all plants of Nepal sequenced in our research. The highest divergence was between *Pedicularis breviscaposa* and *Pedicularis chamissonoides* with divergence of 5.1%. And divergence ranged from 0.13% to 5.1%. One downside of matK is matK presents more challenges and frequently necessitates the use of a variety of primer combinations, particularly when dealing with herbarium material(Vere et al., 2015).

### **5.5.4 trnH-psbA**

Out of the 41 comparisons made, six plant showed no divergence with foreign plants while using trnH-psbA marker. When compared to a foreign plant of the same genus *Pedicularis chamissonoides* and *Pedicularis clarkei* (JN045913.1) showed highest divergence (11.53%). When examining the divergence among plants of Nepal, it became apparent that there were differences in the trnH-psbA genes of all the plants. The most significant divergence observed was 11.2% between *Pedicularis chamissonoides* and *Pedicularis breviscaposa*. Despite the complications posed by intraspecific inversions and rps19 insertions, which can increase intraspecific variation and potentially complicate the utilization of trnH-psbA, it remains a highly efficient plant barcode(Pang et al., 2012).

### **5.5.5 TrnIF-trnIC**

Among 21 comparisons, trnIF-trnIC was able to separate all the plants compared with foreign plants. The most notable divergence, reaching a value of 7.49%, was found in *Thunbergia kasajuana*. The marker trnIF-trnIC effectively distinguished all the sequenced samples of our native plants. Among native plants of same family, the highest sequence divergence, measuring 5.51%, was observed between *Pedicularis anserantha* and *Pedicularis Chamissonoides*. Some of the challenges associated with trnL-F include technical concerns regarding primer universality, sequence quality and complexity, as well as the inability to differentiate between samples(Groot et al., 2011b).

## 5.6 Phylogeny

Usually, sequences derived from individuals belonging to the same species demonstrate a strong evolutionary connection. Nonetheless, there are scenarios where these sequences do not unite to create a clearly distinguishable evolutionary cluster. Discrepancies between species classification based on physical characteristics and the genetic barcoding sequences warrant deeper exploration. Such inconsistencies could suggest the existence of concealed or cryptic species, occurrences of genetic exchange between species (introgression), or historical genetic diversity that has endured due to incomplete separation of lineages (Little et al., 2013).

Taking the example of rbcL tree, let's elaborate on its structure. In this tree, each node gives rise to two sister taxa, indicating that these taxa are of the same age, and neither of the two is considered basal or ancestral. It's important to note that the terminal taxon, which is the one found at the end of a cladogram or any type of phylogenetic tree, is not regarded as the basal species or a direct ancestor. This doesn't imply that the other species evolved after the emergence of the taxus species in the tree. Instead, it simply illustrates the existing relationships and inclinations among the species and clades. To understand this better, consider that the terminal species is positioned farther away from its ancestral group and is relatively closer to its sister group in terms of evolutionary relationships (Zachos, 2016).

## 5.7 Effectiveness of marker in different families

### 5.7.1 Acanthaceae

In silico study has suggested ITS to be the most effective marker in differentiating the plants of Acanthaceae family (Suesatpanit et al., 2017). However, in our research ITS marker of *Strobilanthes* could not be amplified and neither rbcL nor trnH-psbA could differentiate all the plants. TrnIF-trnIC was the most effective plant barcode for Acanthaceae family.

### 5.7.2 Apocynaceae

Among ITS, rbcL, matK, and trnH-psbA, Mishra et al., (2017) found rbcL, trnH-psbA to be most effective and matK to have lowest intraspecific difference in Apocynaceae family. In our study, we found rbcL had highest divergence, followed by trnH-psbA, ITS, and matK, while trnIF-trnIC had lowest divergence.

### **5.7.3 Asteraceae**

*Artemisia Mustangensis* was only discriminated by trnIF-trnIC marker. rbcL failed to discriminate both *Artemisia* samples. ITS showed the highest divergence in Asteraceae family, followed by trnH-psbA and matK. Which was similar to result found by Braukmann et al., (2017).

### **5.7.4 Begoniaceae**

ITS marker showed the highest pairwise distance, followed by rbcL and matK, while comparing plants of Begoniaceae family. Eventhough, trnH-psbA had high discrimination rate in other families, it failed to discriminate 2 out of 3 *Begonia* plants used in our study. As recommended by Ford et al., (2009) , we found that it is better to use two or more markers for *Begonia* for proper identification.

### **5.7.5 Berberidaceae**

The Hien et al., (2018) revealed that the nucleotide diversity indexes for ITS and psbA-trnH barcodes were markedly greater than those for rbcL, matK, and trnL in Berberidaceae family. In contrast, our research showed that trnH-psbA exhibited the highest diversity, followed by ITS and matK, with rbcL displaying the lowest diversity. Unfortunately, trnLF-trnLC could not be amplified in our study.

### **5.7.6 Campanulaceae**

Unlike the results reported by Wang et al., (2017), in 2017, where matK showed the highest discrimination capability among the markers ITS, rbcL, trnH-psbA, and matK for *Codonopsis*, our research found that only ITS was effective in distinguishing *Codonopsis* from foreign plants.

### **5.7.7 Caryophyllaceae**

All generated markers except rbcL were able to identify the plants of Caryophyllaceae. The highest average genetic distance within species was observed in the trnH-psbA marker, followed by ITS, matK, and rbcL, which is similar to what was found byJin et al., (2023).

### **5.7.8 Ericaceae**

As per Yan et al., (2015), rbcL was of limited practical value and ITS showed a low species resolution, matK exhibited strong species identification capabilities albeit requiring two sets of primers for 100% PCR success, with the trnH-psbA marker offering the most detailed resolution for rhododendron identification. Similarly, in

our research, *rbcl* had lowest resolution and ITS showed low resolution, *matK* could not be amplified and *trnH-psbA* showed high resolution but not higher than *trnIF-trnIC*. In our study *trnIF-trnIC* marker showed highest utility while barcoding *Rhodendron* as suggested by (Brown et al., 2006).

#### **5.7.9 Fabaceae**

Among ITS, *matK* and *rbcl*, Braukmann et al., (2017) found ITS to be most effective followed by *rbcl* in plants of Fabaceae family. However, we found *matK* was more divergent followed by ITS. Among all five markers *trnH-psbA* was found to be most divergent and *rbcl* had lowest divergence.

#### **5.7.10 Hypericaceae**

*RbcL* and *matK* sequence of *Hypericum cordifolium* were not amplified. In our study *trnIF-trnIC* showed highest divergence followed by ITS and *trnH-psbA*.Pyrka et al., (2021) suggested that *matK* could be difficult to amplify and ITS + *trnH-psbA* could be useful to identify plants of *Hypericum*, which is similar to our finding.

#### **5.7.11 Lamiaceae**

In line with the research conducted by Braukmann et al., (2017) , where they evaluated ITS, *rbcl*, and *matK*, we also found ITS to be the most effective, in Lamiaceae, with *matK* following as the second most effective marker, confirming our findings. However, among all five markers, *trnH-psbA* emerged as the most effective marker, while *rbcl* was found to be the least effective.

#### **5.7.12 Onagraceae**

While comparing *Epilobium* of family Onagraceae, we could only compare ITS marker and unfortunately, ITS showed no divergence. While *trnH-psbA* could not be amplified, we could not access the other markers of *Epilobium* from NCBI.

#### **5.7.13 Orobanchaceae**

In contrast with the findings of Wang et al., (2017), where *matK* showed the highest discrimination capability among the markers ITS, *rbcl*, *trnH-psbA*, and *matK* for *Pedicularis*, our research showed ITS to be most effective followed by *matK* and *trnH-psbA*.

#### **5.7.14 Papaveraceae**

All the markers except *rbcl* were able to discriminate the plants of Papveraceae studied in our research. *TrnH-psbA* showed the highest divergence followed by ITS, *matK*, *trnIF-trnIC* and *rbcl*. Similar to Yun & Oh, (2018), we found ITS, *matK* and *rbcl* would be enough to identify plants of papaveraceae family.

#### **5.7.15 Primulaceae**

*RbcL* showed lowest divergence, while ITS u4/p5 primer pair showed highest intraspecific divergence followed by *trnIF-trnIC* and *trnH-psbA*. Amplification of *matK* was not successful. However, H. F. Yan et al., (2015) reported *rbcl* to have the least intra-specific variation, whereas *trnH-psbA* displayed the highest intra-specific divergence, followed by ITS2 among ITS1, ITS, *rbcl*, *matK* and *trnH-psbA* in Primulaceae family.

#### **5.7.16 Ranunculaceae**

Similar to the study conducted by Braukmann et al., (2017) in which they assessed ITS, *rbcl*, and *matK*, we reached a similar conclusion that ITS was the most effective marker, followed by *matK* in Ranunculaceae family, thus validating our results. However, when considering all five markers in our study, *trnH-psbA* stood out as the most effective, whereas *rbcl* proved to be the least effective. Unfortunately, we couldn't include the *trnLF-trnLC* gene in our comparison due to our inability to access this gene from NCBI.

#### **5.7.17 Rosaceae**

In line with the research conducted by Braukmann et al., (2017) , where they evaluated ITS, *rbcl*, and *matK* in Rosaceae family. we also found ITS to be the most effective, with *matK* following as the second most effective marker, confirming our findings. *RbcL* could not discriminate the endemic plant from congenetics from nearby country, while, we could not access the *trnH-psbA* gene and *trnIF-trnIC* gene of plant of neighbor country to compare.

#### **5.7.18 Rubiceae**

ITS was found to have the highest mean divergence, followed by *trnH-psbA*, *matK* and *rbcl* in Rubiceae family, similar to work done by Guo et al., (2011). Unfortunately, we could not access the *trnIF-trnIC* sequence in NCBI.

### **5.7.19 Saxifragaceae**

All the markers successfully distinguished Saxifragaceae. In line with the research conducted by Braukmann et al., (2017), where they evaluated ITS, rbcL, and matK, we also found ITS to be the most effective, with matK following as the second most effective marker in Saxifragaceae, confirming our findings. However, among all five markers, trnH-psbA emerged as the most effective marker, while rbcL was found to be the least effective.

### **5.7.20 Sphagnaceae**

We compared available three different markers of *Sphagnum*, where trnIF-trnIC showed highest resolution followed by rbcL marker. TrnH-psbA failed to show any divergence, even though, ITS2, rbcL, trnH-psbA, and trnIF-trnIC were suggested for barcoding of bryophyta by Liu et al., (2010). Same article found amplification rate of matK to be only 20%, similarly, matK could not be amplified in our study. While ITS barcode of sphagnum from neighbor countries could not be found.

## CHAPTER 6

### CONCLUSION

Our study helped us assess if the chosen barcode markers were adequate in distinguishing Nepal's endemic plants across different taxonomic levels such as family, genus, and species. The markers utilized were indeed capable of discriminating the endemic plants of Nepal. Our study generated the barcodes of Endemic plants of Nepal for the very first time.

The *rbcl* marker was easy to amplify and sequence, but its discriminatory capability was the least effective. On the other hand, the ITS markers, while also easy to amplify and sequence, exhibited the highest discriminatory ability, aligning with existing literature. *matK* markers, known for their discriminative potential, were challenging to amplify and displayed less variation compared to the ITS markers. The *trnH-psbA* marker demonstrated decent discriminatory ability and was easily amplified through PCR. Conversely, the *trnI-trnC* markers were the most challenging to PCR amplify, yet they could effectively discriminate many of the endemic plants.

Furthermore, our findings revealed that marker efficiency varies across different plant families. Some markers displayed higher discriminatory power within certain families but lower effectiveness in others. *matK* and *trnI-trnC* markers posed challenges for a few families, suggesting the need for family-specific primers for these markers. Our study underscores the advantage of using two or more markers to identify endemic plants. Additionally, we recommend incorporating anatomical, morphological, and biochemical methods alongside DNA barcodes for plant identification.

The study's outcome highlighted the efficiency of both nuclear marker (ITS) and chloroplast markers (*rbcl*, *matK*, *trnH-psbA*, *trnI-trnC*) in identifying Nepal's endemic plants. These results can serve as a reference for identifying endemic plant species within sample mixtures. Ultimately, the analysis based on molecular markers is expected to contribute to the conservation of Nepal's endemic plants. It has provided foundational molecular-level information about these plants, directing its use in species-level identification and supporting claims regarding the regional endemism of Nepal's plants.

## REFERENCES

- Armenise, L., Simeone, M. C., Piredda, R., & Schirone, B. (2012). Validation of DNA barcoding as an efficient tool for taxon identification and detection of species diversity in Italian conifers. *European Journal of Forest Research*, *131*(5), 1337–1353. <https://doi.org/10.1007/S10342-012-0602-0>
- Ayalew, H., Tewelde, E., Abebe, B., Alebachew, Y., & Tadesse, S. (2022). Endemic medicinal plants of Ethiopia: Ethnomedicinal uses, biological activities and chemical constituents. *Journal of Ethnopharmacology*, *293*, 115307. <https://doi.org/10.1016/J.JEP.2022.115307>
- Bhagya Chandrasekara, C. H. W. M. R., Naranpanawa, D. N. U., Bandusekara, B. S., Pushpakumara, D. K. N. G., Wijesundera, D. S. A., & Bandaranayake, P. C. G. (2021). Universal barcoding regions, *rbcl*, *matK* and *trnH-psbA* do not discriminate *Cinnamomum* species in Sri Lanka. *PLoS ONE*, *16*(2). <https://doi.org/10.1371/JOURNAL.PONE.0245592>
- Bobo-Pinilla, J., Salmerón-Sánchez, E., Mota, J. F., & Peñas, J. (2021). Genetic conservation strategies of endemic plants from edaphic habitat islands: The case of *Jacobaea auricula* (Asteraceae). *Journal for Nature Conservation*, *61*, 126004. <https://doi.org/10.1016/J.JNC.2021.126004>
- Bowe, L. M., Coat, G., & DePamphilis, C. W. (2000). Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4092–4097. <https://doi.org/10.1073/PNAS.97.8.4092/ASSET/28D698EB-146A-460F-81E0-2B0AB0AB5AE0/ASSETS/GRAPHIC/PQ0502978003.JPEG>
- Braukmann, T. W. A., Kuzmina, M. L., Sills, J., Zakharov, E. V., & Hebert, P. D. N. (2017). Testing the Efficacy of DNA Barcodes for Identifying the Vascular Plants of Canada. *PLOS ONE*, *12*(1), e0169515. <https://doi.org/10.1371/JOURNAL.PONE.0169515>

- Brown, G. K., Craven, L. A., Udovicic, F., & Ladiges, P. Y. (2006). Phylogeny of *Rhododendron* section *Vireya* (Ericaceae) based on two non-coding regions of cpDNA. *Plant Systematics and Evolution*, 257(1–2), 57–93. <https://doi.org/10.1007/S00606-005-0367-8/METRICS>
- Carvalho, J., Puertas, G., Gaspar, J., Azinheiro, S., Diéguez, L., Garrido-Maestu, A., Vázquez, M., Barros-Velázquez, J., Cardoso, S., & Prado, M. (2018). Highly efficient DNA extraction and purification from olive oil on a washable and reusable miniaturized device. *Analytica Chimica Acta*, 1020, 30–40. <https://doi.org/10.1016/J.ACA.2018.02.079>
- Chang, C. C., Lin, H. C., Lin, I. P., Chow, T. Y., Chen, H. H., Chen, W. H., Cheng, C. H., Lin, C. Y., Liu, S. M., Chang, C. C., & Chaw, S. M. (2006). The Chloroplast Genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative Analysis of Evolutionary Rate with that of Grasses and Its Phylogenetic Implications. *Molecular Biology and Evolution*, 23(2), 279–291. <https://doi.org/10.1093/MOLBEV/MSJ029>
- Chase, M. W., Cowan, R. S., Hollingsworth, P. M., Van Den Berg, C., Madriñán, S., Petersen, G., Seberg, O., Jørgensen, T., Cameron, K. M., Carine, M., Pedersen, N., Hedderson, T. A. J., Conrad, F., Salazar, G. A., Richardson, J. 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(2), 295–299. <https://doi.org/10.1002/TAX.562004>
- 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. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1462), 1889–1895. <https://doi.org/10.1098/RSTB.2005.1720>
- Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016a). Barcoding the kingdom Plantae: New PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources*, 16(1), 138–149. <https://doi.org/10.1111/1755-0998.12438>

- Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016b). Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources*, *16*(1), 138–149. <https://doi.org/10.1111/1755-0998.12438>
- Coelho, N., Gonçalves, S., & Romano, A. (2020). Endemic Plant Species Conservation: Biotechnological Approaches. *Plants*, *9*(3). <https://doi.org/10.3390/PLANTS9030345>
- de Groot, G. A., During, H. J., Maas, J. W., Schneider, H., Vogel, J. C., & Erkens, R. H. J. (2011a). Use of rbcL and trnL-F as a Two-Locus DNA Barcode for Identification of NW-European Ferns: An Ecological Perspective. *PLOS ONE*, *6*(1), e16371. <https://doi.org/10.1371/JOURNAL.PONE.0016371>
- de Groot, G. A., During, H. J., Maas, J. W., Schneider, H., Vogel, J. C., & Erkens, R. H. J. (2011b). Use of rbcL and trnL-F as a Two-Locus DNA Barcode for Identification of NW-European Ferns: An Ecological Perspective. *PLOS ONE*, *6*(1), e16371. <https://doi.org/10.1371/JOURNAL.PONE.0016371>
- de Miranda, V. F. O., Martins, V. G., Furlan, A., & Bacci, M. (2010). Plant or fungal sequences? An alternative optimized PCR protocol to avoid ITS (nrDNA) misamplification. *Brazilian Archives of Biology and Technology*, *53*(1), 141–152. <https://doi.org/10.1590/S1516-89132010000100018>
- Dunning, L. T., & Savolainen, V. (2010). Broad-scale amplification of matK for DNA barcoding plants, a technical note. *Botanical Journal of the Linnean Society*, *164*, 1–9.
- Farell, E. M., & Alexandre, G. (2012). *Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates*. <https://doi.org/10.1186/1756-0500-5-257>
- Fatima, T., Srivastava, A., Somashekar, P. V., Hanur, V. S., & Rao, M. S. (2019). Development of DNA-based species identification and barcoding of three important timbers. *Bulletin of the National Research Centre* *2019 43:1*, *43*(1), 1–17. <https://doi.org/10.1186/S42269-019-0116-8>

- 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*(7). <https://doi.org/10.1371/JOURNAL.PONE.0002802>
- Ferreira, P. M. A., & Boldrini, I. I. (2011). Potential Reflection of Distinct Ecological Units in Plant Endemism Categories. *Conservation Biology*, *25*(4), 672–679. <https://doi.org/10.1111/J.1523-1739.2011.01675.X>
- Ford, C. S., Ayres, K. L., Toomey, N., Haider, N., Van, J., Stahl, A., Kelly, L. J., Wikström, N., Hollingsworth, P. M., Duff, R. J., Hoot, S. B., Cowan, R. S., Chase, M. W., & Wilkinson, M. J. (2009). *Selection of candidate coding DNA barcoding regions for use on land plants.*
- França, L. T. C., Carrilho, E., & Kist, T. B. L. (2002). A review of DNA sequencing techniques. *Quarterly Reviews of Biophysics*, *35*(2), 169–200. <https://doi.org/10.1017/S0033583502003797>
- Gao, L. M., Li, Y., Phan, L. K., Yan, L. J., Thomas, P., Phan, L. K., Möller, M., & Li, D. Z. (2017). DNA barcoding of East Asian Amentotaxus (Taxaceae): Potential new species and implications for conservation. *Journal of Systematics and Evolution*, *55*(1), 16–24. <https://doi.org/10.1111/JSE.12207/SUPPINFO>
- Gernandt, D. S., Liston, A., & Piñero, D. (2001). Variation in the nrDNA ITS of Pinus Subsection Cembroides: Implications for Molecular Systematic Studies of Pine Species Complexes. *Molecular Phylogenetics and Evolution*, *21*(3), 449–467. <https://doi.org/10.1006/MPEV.2001.1026>
- Guo, X., Simmons, M. P., But, P. P. H., Shaw, P. C., & Wang, R. J. (2011). Application of DNA barcodes in Hedyotis L. (Spermacoceae, Rubiaceae). *Journal of Systematics and Evolution*, *49*(3), 203–212. <https://doi.org/10.1111/J.1759-6831.2011.00130.X>
- hall, tom. (n.d.). BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences.*

- Harrison, S., & Noss, R. (2017). Endemism hotspots are linked to stable climatic refugia. *Annals of Botany*, 119(2), 207–214. <https://doi.org/10.1093/AOB/MCW248>
- Hasebe, M., Omori, T., Nakazawa, M., Sano, T., & Kato, M. (1994). rbcL gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns (pteridophyte/molecular systematics/phylogenetics/ribulose-bisphosphate carboxylase). *Proc. Natl. Acad. Sci. USA*, 91, 5730–5734.
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & Dewaard, J. R. (n.d.). *Biological identifications through DNA barcodes*. <https://doi.org/10.1098/rspb.2002.2218>
- Hebert, P. D. N., & Gregory, T. R. (2005). The Promise of DNA Barcoding for Taxonomy. *Systematic Biology*, 54(5), 852–859. <https://doi.org/10.1080/10635150500354886>
- Holderegger, R., Balkenhol, N., Bolliger, J., Engler, J. O., Gugerli, F., Hochkirch, A., Nowak, C., Segelbacher, G., Widmer, A., & Zochs, F. E. (2019). Conservation genetics: Linking science with practice. *Molecular Ecology*, 28(17), 3848–3856. <https://doi.org/10.1111/MEC.15202>
- Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S., Erickson, D. L., Fazekas, A. J., Graham, S. W., James, K. E., Kim, K. J., John Kress, W., Schneider, H., van AlphenStahl, J., Barrett, S. C. H., van den Berg, C., Bogarin, D., ... Little, D. P. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), 12794–12797. [https://doi.org/10.1073/PNAS.0905845106/SUPPL\\_FILE/0905845106SI.PDF](https://doi.org/10.1073/PNAS.0905845106/SUPPL_FILE/0905845106SI.PDF)
- Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011a). Choosing and Using a Plant DNA Barcode. *PLOS ONE*, 6(5), e19254. <https://doi.org/10.1371/JOURNAL.PONE.0019254>
- Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011b). Choosing and using a plant DNA barcode. *PLoS ONE*, 6(5). <https://doi.org/10.1371/JOURNAL.PONE.0019254>

- Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011c). Choosing and Using a Plant DNA Barcode. *PLOS ONE*, 6(5), e19254. <https://doi.org/10.1371/JOURNAL.PONE.0019254>
- Hollingsworth, P. M., Li, D. Z., Van Der Bank, M., & Twyford, A. D. (2016). Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1702). <https://doi.org/10.1098/RSTB.2015.0338>
- Jin, D. P., Sim, S., Park, J. W., Choi, J. E., Yoon, J., Lim, C. E., & Kim, M. H. (2023). Identification of the Plant Family Caryophyllaceae in Korea Using DNA Barcoding. *Plants*, 12(10), 2060. <https://doi.org/10.3390/PLANTS12102060/S1>
- Käss, E., & Wink, M. (1995). Molecular Phylogeny of the Papilionoideae (Family Leguminosae): RbcL Gene Sequences versus Chemical Taxonomy. *Botanica Acta*, 108(2), 149–162. <https://doi.org/10.1111/J.1438-8677.1995.TB00845.X>
- Kress, W. J., & Erickson, D. L. (2007a). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLoS ONE*, 2(6). <https://doi.org/10.1371/JOURNAL.PONE.0000508>
- Kress, W. J., & Erickson, D. L. (2007b). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLOS ONE*, 2(6), e508. <https://doi.org/10.1371/JOURNAL.PONE.0000508>
- Kress, W. J., & Erickson, D. L. (2008). DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences of the United States of America*, 105(8), 2761–2762. <https://doi.org/10.1073/PNAS.0800476105/ASSET/B56297BF-B2F7-4142-BE23-B92F4D2068EB/ASSETS/GRAPHIC/ZPQ0100897780001.JPEG>
- Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution*, 30(1), 25–35. <https://doi.org/10.1016/J.TREE.2014.10.008>

- Krishna Krishnamurthy, P., & Francis, R. A. (2012). A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation* 2012 21:8, 21(8), 1901–1919. <https://doi.org/10.1007/S10531-012-0306-2>
- Kruckeberg, A. R., & Rabinowitz, D. (1985). *Biological Aspects of Endemism in Higher Plants on JSTOR*.
- Li, C., Zheng, Y., & Huang, P. (2020). Molecular markers from the chloroplast genome of rose provide a complementary tool for variety discrimination and profiling. *Scientific Reports* 2020 10:1, 10(1), 1–15. <https://doi.org/10.1038/s41598-020-68092-1>
- Li, F., Zhao, R., Li, Y., Li, X., Wei, C., Zhang, X., He, H., Qu, S., Huang, M., & Huang, H. Q. (2023). *Phylogenetic Analysis of Impatiens in the Eastern Yunnan, China*. <https://doi.org/10.21203/rs.3.rs-2375964/v1>
- Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., & Chen, S. (2015). Plant DNA barcoding: from gene to genome. *Biological Reviews*, 90(1), 157–166. <https://doi.org/10.1111/BRV.12104>
- Little, D. P., Knopf, P., & Schulz, C. (2013). DNA Barcode Identification of Podocarpaceae-The Second Largest Conifer Family. *PLoS ONE*, 8(11), 81008. <https://doi.org/10.1371/journal.pone.0081008>
- Liu, J., Milne, R. I., Möller, M., Zhu, G. F., Ye, L. J., Luo, Y. H., Yang, J. B., Wambulwa, M. C., Wang, C. N., Li, D. Z., & Gao, L. M. (2018). Integrating a comprehensive DNA barcode reference library with a global map of yews (*Taxus* L.) for forensic identification. *Molecular Ecology Resources*, 18(5), 1115–1131. <https://doi.org/10.1111/1755-0998.12903>
- Liu, Y., Yan, H. F., Cao, T., & Ge, X. J. (2010). Evaluation of 10 plant barcodes in Bryophyta (Mosses). *Journal of Systematics and Evolution*, 48(1), 36–46. <https://doi.org/10.1111/J.1759-6831.2009.00063.X>
- Manning, J., Boatwright, J. S., Daru, B. H., Maurin, O., & Bank, M. van der. (2014). A Molecular Phylogeny and Generic Classification of Asphodelaceae subfamily

- Alooideae: A Final Resolution of the Prickly Issue of Polyphyly in the Aloooids? *Systematic Botany*, 39(1), 55–74.
- Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature* 2000 405:6783, 405(6783), 243–253. <https://doi.org/10.1038/35012251>
- McBride, L. J., Koepf, S. M., Gibbs, R. A., Salser, W., Mayrand, P. E., Hunkapiller, M. W., & Kronick, M. N. (1989). Automated DNA sequencing methods involving polymerase chain reaction. *Clinical Chemistry*, 35(11), 2196–2201. <https://doi.org/10.1093/CLINCHEM/35.11.2196>
- McPherson, M. J., & Møller, S. G. (2000). *PCR The basics*. BIOS.
- Meyer, C. P., & Paulay, G. (2005). DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLOS Biology*, 3(12), e422. <https://doi.org/10.1371/JOURNAL.PBIO.0030422>
- Mishler, B. D., Knerr, N., González-Orozco, C. E., Thornhill, A. H., Laffan, S. W., & Miller, J. T. (2014). Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. *Nature Communications* 2014 5:1, 5(1), 1–10. <https://doi.org/10.1038/ncomms5473>
- Mishra, P., Kumar, A., Sivaraman, G., Shukla, A. K., Kaliamoorthy, R., Slater, A., & Velusamy, S. (2017). Character-based DNA barcoding for authentication and conservation of IUCN Red listed threatened species of genus *Decalepis* (Apocynaceae). *Scientific Reports* 2017 7:1, 7(1), 1–12. <https://doi.org/10.1038/s41598-017-14887-8>
- Mkala, E. M., Mwanza, V., Nzei, J., Oluoch, W. A., Ngarega, B. K., Wanga, V. O., Oulo, M. A., Munyao, F., Kilingo, F. M., Rono, P., Waswa, E. N., Mutinda, E. S., Ochieng, C. O., Mwachala, G., Hu, G. W., Wang, Q. F., Katunge, J. K., & Victoire, C. I. (2023). Predicting the potential impacts of climate change on the endangered endemic annonaceae species in east africa. *Heliyon*, 9(6), e17405. <https://doi.org/10.1016/J.HELIYON.2023.E17405>
- Moritz, C., & Cicero, C. (2004). DNA Barcoding: Promise and Pitfalls. *PLOS Biology*, 2(10), e354. <https://doi.org/10.1371/JOURNAL.PBIO.0020354>

- Mouga, T., & Mouga, T. (2021). Conserving Endemic Plant Species in Oceanic Island's Protected Areas. *Protected Area Management - Recent Advances*. <https://doi.org/10.5772/INTECHOPEN.100571>
- Neale, D. B., & Wheeler, N. C. (2019). The conifers: Genomes, variation and evolution. In *The Conifers: Genomes, Variation and Evolution*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-46807-5/COVER>
- Neigel, J., Domingo, A., & Stake, J. (2007). DNA barcoding as a tool for coral reef conservation. *Coral Reefs*, 26(3), 487–499. <https://doi.org/10.1007/S00338-007-0248-4/METRICS>
- Nielsen, R., & Matz, M. (2006). Statistical Approaches for DNA Barcoding. *Systematic Biology*, 55(1), 162–169. <https://doi.org/10.1080/10635150500431239>
- Nishimura, D. (2000). Editor's Page. In *Biotech Software & Internet Report* (Vol. 1, Issue 5). <https://doi.org/10.1089/152791600750034695>
- Nock, C. J., Waters, D. L. E., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., & Henry, R. J. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*, 9(3), 328–333. <https://doi.org/10.1111/J.1467-7652.2010.00558.X>
- Omelchenko, D. O., Krinitsina, A. A., Kasianov, A. S., Speranskaya, A. S., Chesnokova, O. V., Polevova, S. V., & Severova, E. E. (2022). Assessment of ITS1, ITS2, 5'-ETS, and trnL-F DNA Barcodes for Metabarcoding of Poaceae Pollen. *Diversity*, 14(3), 191. <https://doi.org/10.3390/D14030191/S1>
- Pang, X., Liu, C., Shi, L., Liu, R., Liang, D., Li, H., Cherny, S. S., & Chen, S. (2012). Utility of the trnH-psbA Intergenic Spacer Region and Its Combinations as Plant DNA Barcodes: A Meta-Analysis. *PLoS ONE*, 7(11). <https://doi.org/10.1371/JOURNAL.PONE.0048833>
- Pereira, F., Carneiro, J., & Amorim, A. (2008). Identification of species with DNA-based technology: Current progress and challenges. *Recent Patents on DNA and Gene Sequences*, 2(3), 187–200. <https://doi.org/10.2174/187221508786241738>

- pirie, michael D., & vargas, M. paula balcazar. (2016). *Ancient paralogy in the cpDNA trnL-F Region in Annonaceae: Implications for Plant Molecular Systematics.*
- Pradhan, A., Rai, D., & Barik, Saroj. k. (2018). *Begonia panchtharensis (Begoniaceae), a New Record to India from Sikkim, Eastern Himalaya.*
- Pyrka, I., Stefanaki, A., & Vlachonasios, K. E. (2021). DNA Barcoding of St. John's wort (*Hypericum* spp.) Growing Wild in North-Eastern Greece. *Planta Medica*, 87(7), 528–537. <https://doi.org/10.1055/A-1379-3249>
- Raja, K., Balachandran, S., Mohanasundaram, S., & Ramalingam, S. (2015). DNA barcoding: a genomic-based tool for authentication of phytomedicinals and its products. *Botanics: Targets and Therapy*, 5, 77–84. <https://doi.org/10.2147/BTAT.S61121>
- Rajbhandari, K. R., Sanjeev Kumar, J., Mohan Dev, K., Subash, B., Datt, G., & Chhetri, R. (2021). Endemic Flowering Plants of Nepal: Status and Distribution. *Integrating Biological Resources for Prosperity*, Rajbhandari, K. R., Sanjeev Kumar, J., Mohan Dev, K., Subash, B., Datt, G., Chhetri, R. (2021). *Endemic Flowering Plants of Nepal: Status and Distribution. Integrating Biological Resources for Prosperity*, 23–72., 23–72.
- Rajbhandary, S. (2020). An Overview of Plant Diversity in Nepal. *Plant Diversity in Nepal.*
- Rana, H. K., Sun, H., Paudel, A., & Ghimire, S. K. (2018). *Saussurea ramchaudharyi* (Asteraceae), a new species from Nepal. *Phytotaxa*, 340(3), 271–276. <https://doi.org/10.11646/PHYTOTAXA.340.3.7>
- Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System: Barcoding. *Molecular Ecology Notes*, 7(3), 355–364. <https://doi.org/10.1111/J.1471-8286.2007.01678.X>
- Raubeson, L. A., & Jansen, R. K. (1992). Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science (New York, N.Y.)*, 255(5052), 1697–1699. <https://doi.org/10.1126/SCIENCE.255.5052.1697>

- Ravi, V., Khurana, J. P., Tyagi, A. K., & Khurana, P. (2008). An update on chloroplast genomes. *Plant Systematics and Evolution*, 271(1–2), 101–122. <https://doi.org/10.1007/S00606-007-0608-0>
- Reed, B. M., Sarasan, V., Kane, M., Bunn, E., & Pence, V. C. (2011). Biodiversity conservation and conservation biotechnology tools. *In Vitro Cellular and Developmental Biology - Plant*, 47(1), 1–4. <https://doi.org/10.1007/S11627-010-9337-0>
- Ribeiro, R. A., & Lovato, M. B. (2007). Comparative analysis of different DNA extraction protocols in fresh and herbarium specimens of the genus *Dalbergia*. *Genetics and Molecular Research : GMR*.
- Rogalski, M., Vieira, L. D. N., Fraga, H. P., & Guerra, M. P. (2015). Plastid genomics in horticultural species: Importance and applications for plant population genetics, evolution, and biotechnology. *Frontiers in Plant Science*, 6(JULY), 586. <https://doi.org/10.3389/FPLS.2015.00586/BIBTEX>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463. <https://doi.org/10.1073/PNAS.74.12.5463>
- Sanna, C., Maxia, A., Fenu, G., & Loi, M. C. (2020). So Uncommon and so Singular, but Underexplored: An Updated Overview on Ethnobotanical Uses, Biological Properties and Phytoconstituents of Sardinian Endemic Plants. *Plants 2020, Vol. 9, Page 958, 9(8)*, 958. <https://doi.org/10.3390/PLANTS9080958>
- Särkinen, T., Staats, M., Richardson, J. E., Cowan, R. S., & Bakker, F. T. (2012). How to Open the Treasure Chest? Optimising DNA Extraction from Herbarium Specimens. *PLOS ONE*, 7(8), e43808. <https://doi.org/10.1371/JOURNAL.PONE.0043808>
- Schuettpelz, E., & Kathleen, M. P. (2007). *fern phylogeny from 400 leptosporangiate species and three plastid genes*.

- Selvaraj, D., Sarma, R. K., & Sathishkumar, R. (2008a). Phylogenetic analysis of chloroplast matK gene from Zingiberaceae for plant DNA barcoding. *Bioinformation*, 3(1), 24. <https://doi.org/10.6026/97320630003024>
- Selvaraj, D., Sarma, R. K., & Sathishkumar, R. (2008b). Phylogenetic analysis of chloroplast matK gene from Zingiberaceae for plant DNA barcoding. *Bioinformation*, 3(1), 24. <https://doi.org/10.6026/97320630003024>
- Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., Siripun, K. C., Winder, C. T., Schilling, E. E., & Small, R. L. (2005). The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, 92(1), 142–166. <https://doi.org/10.3732/AJB.92.1.142>
- Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007a). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, 94(3), 275–288. <https://doi.org/10.3732/AJB.94.3.275>
- Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007b). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. *American Journal of Botany*, 94(3), 275–288. <https://doi.org/10.3732/AJB.94.3.275>
- Shneyer, V. S., & Rodionov, A. V. (2019). Plant DNA Barcodes. *Biology Bulletin Reviews* 2019 9:4, 9(4), 295–300. <https://doi.org/10.1134/S207908641904008X>
- Shudan, M., Magar<sup>1</sup>, T., Maharjan<sup>1</sup>, S., Pathak<sup>1</sup>, J., Kandel, D. R., & Rijal<sup>1</sup>, G. (2022). Some Barcoding DNA Sequence Analysis of *Sphagnum nepalense* H. Suzuki, a Bryophyte Species Endemic to East Nepal. *Journal of Plant Resources*, 20(1). <https://doi.org/10.3126/bdpr.v20i01.56577>
- Skarpeid, H. J., Kvaal, K., & Hildrum, K. I. (1998). Identification of animal species in ground meat mixtures by multivariate analysis of isoelectric focusing protein profiles. *ELECTROPHORESIS*, 19(18), 3103–3109. <https://doi.org/10.1002/ELPS.1150191811>

- Smith, D. R., & Lee, R. W. (2010). Low nucleotide diversity for the expanded organelle and nuclear genomes of *Volvox carteri* supports the mutational-hazard hypothesis. *Molecular Biology and Evolution*, *27*(10), 2244–2256. <https://doi.org/10.1093/MOLBEV/MSQ110>
- Staats, M., Cuenca, A., Richardson, J. E., Ginkel, R. V. van, Petersen, G., Seberg, O., & Bakker, F. T. (2011). DNA Damage in Plant Herbarium Tissue. *PLOS ONE*, *6*(12), e28448. <https://doi.org/10.1371/JOURNAL.PONE.0028448>
- Stranneheim, H., & Lundeberg, J. (2012). Stepping stones in DNA sequencing. *Biotechnology Journal*, *7*(9), 1063. <https://doi.org/10.1002/BIOT.201200153>
- Suesatpanit, T., Osathanunkul, K., Madesis, P., & Osathanunkul, M. (2017). Should DNA sequence be incorporated with other taxonomical data for routine identifying of plant species? *BMC Complementary and Alternative Medicine*, *17*(1). <https://doi.org/10.1186/S12906-017-1937-3>
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermet, T., Corthier, G., Brochmann, C., & Willerslev, E. (2007). Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3), e14–e14. <https://doi.org/10.1093/NAR/GKL938>
- Taberlet, P., Gielly, L., Pautou, G., & Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, *17*(5), 1105–1109. <https://doi.org/10.1007/BF00037152>
- Taylor, H. R., & Harris, W. E. (2012). An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, *12*(3), 377–388. <https://doi.org/10.1111/J.1755-0998.2012.03119.X>
- Techen, N., Parveen, I., Pan, Z., & Khan, I. A. (2014). DNA barcoding of medicinal plant material for identification. *Current Opinion in Biotechnology*, *25*, 103–110. <https://doi.org/10.1016/J.COPBIO.2013.09.010>
- Thi Thu Hien, L., Nhat Linh, N., Le Bich Hang, P., Phuong Mai, N., Hong Hanh, H., Thi Thu Hue, H., Van Huan, H., Quoc Viet, H., & Giay, C. (2018). Developing DNA Barcodes for Species Identification of *Berberis* and *Dysosma* Genera in Vietnam.

*INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY Dysosma Genera in Vietnam. Int. J. Agric. Biol, 20, 18. <https://doi.org/10.17957/IJAB/15.0608>*

Tiwari, A., Uprety, Y., & Rana, S. K. (2019a). Plant endemism in the Nepal Himalayas and phytogeographical implications. *Plant Diversity, 41*(3), 174–182. <https://doi.org/10.1016/J.PLD.2019.04.004>

Tiwari, A., Uprety, Y., & Rana, S. K. (2019b). Plant endemism in the Nepal Himalayas and phytogeographical implications. *Plant Diversity, 41*(3), 174–182. <https://doi.org/10.1016/J.PLD.2019.04.004>

Trujillo-Argueta, S., Del Castillo, R. F., & Velasco-Murguía, A. (2022). Testing the effectiveness of rbcLa DNA-barcoding for species discrimination in tropical montane cloud forest vascular plants (Oaxaca, Mexico) using BLAST, genetic distance, and tree-based methods. *PeerJ, 10*. <https://doi.org/10.7717/PEERJ.13771/SUPP-2>

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology and Evolution, 24*(2), 110–117. <https://doi.org/10.1016/j.tree.2008.09.011>

Vasconcelos, S., Nunes, G. L., Dias, M. C., Lorena, J., Oliveira, R. R. M., Lima, T. G. L., Pires, E. S., Valadares, R. B. S., Alves, R., Watanabe, M. T. C., Zappi, D. C., Hiura, A. L., Pastore, M., Vasconcelos, L. V., Mota, N. F. O., Viana, P. L., Gil, A. S. B., Simões, A. O., Imperatriz-Fonseca, V. L., ... Oliveira, G. (2021). Unraveling the plant diversity of the Amazonian canga through DNA barcoding. *Ecology and Evolution, 11*(19), 13348–13362. <https://doi.org/10.1002/ECE3.8057>

Vere, N. De, Rich, T. C. G., Trinder, S. A., & Long, C. (2015). Dna barcoding for plants. *Methods in Molecular Biology, 1245*, 101–118. [https://doi.org/10.1007/978-1-4939-1966-6\\_8](https://doi.org/10.1007/978-1-4939-1966-6_8)

Volenzo, T., & Odiyo, J. (2020). Integrating endemic medicinal plants into the global value chains: the ecological degradation challenges and opportunities. *Heliyon, 6*(9), e04970. <https://doi.org/10.1016/J.HELIYON.2020.E04970>

- Wang, D. Y., Wang, Q., Wang, Y. L., Xiang, X. G., Huang, L. Q., & Jin, X. H. (2017). Evaluation of DNA barcodes in *Codonopsis* (Campanulaceae) and in some large angiosperm plant genera. *PLoS ONE*, *12*(2). <https://doi.org/10.1371/JOURNAL.PONE.0170286>
- Wei Chen, C., Moan Huang, Y., Yaung Kuo, L., Dat Nguyen, Q., Truong Luu, H., Rey Callado, J., Farrar, D. R., & Liang Chiou, W. (n.d.). *trnL-F is a powerful marker for DNA identification of field vittarioid gametophytes (Pteridaceae)*. <https://doi.org/10.1093/aob/mct004>
- Wiemers, M., & Fiedler, K. (2007). Does the DNA barcoding gap exist? - A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, *4*(1), 1–16. <https://doi.org/10.1186/1742-9994-4-8/FIGURES/6>
- Wilkinson, M. J., Szabo, C., Ford, C. S., Yarom, Y., Croxford, A. E., Camp, A., & Gooding, P. (2017). Replacing Sanger with Next Generation Sequencing to improve coverage and quality of reference DNA barcodes for plants. *Scientific Reports*, *7*. <https://doi.org/10.1038/SREP46040>
- Will, K. W., Mishler, B. D., & Wheeler, Q. D. (2005). The Perils of DNA Barcoding and the Need for Integrative Taxonomy. *Systematic Biology*, *54*(5), 844–851. <https://doi.org/10.1080/10635150500354878>
- Yan, H. F., Liu, Y. J., Xie, X. F., Zhang, C. Y., Hu, C. M., Hao, G., & Ge, X. J. (2015). DNA Barcoding Evaluation and Its Taxonomic Implications in the Species-Rich Genus *Primula* L. in China. *PLOS ONE*, *10*(4), e0122903. <https://doi.org/10.1371/JOURNAL.PONE.0122903>
- Yan, L. J., Liu, J., Möller, M., Zhang, L., Zhang, X. M., Li, D. Z., & Gao, L. M. (2015). DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya–Hengduan Mountains. *Molecular Ecology Resources*, *15*(4), 932–944. <https://doi.org/10.1111/1755-0998.12353>
- Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., & Chen, S. (2010). Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. *PLoS ONE*, *5*(10). <https://doi.org/10.1371/JOURNAL.PONE.0013102>

- Yoon, B.-J. (2009). Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics*, 10(6), 402. <https://doi.org/10.2174/138920209789177575>
- Yun, N., & Oh, S. H. (2018). Phylogenetic relationships of *Coreanomecon* (Papaveraceae: Papaveroideae), an endemic genus in Korea, using DNA sequences. *Korean Journal of Plant Taxonomy*, 48(4), 289–300. <https://doi.org/10.11110/KJPT.2018.48.4.289>
- Zachos, F. E. (2016). Tree thinking and species delimitation: Guidelines for taxonomy and phylogenetic terminology. *Mammalian Biology*, 81(2), 185–188. <https://doi.org/10.1016/J.MAMBIO.2015.10.002>