DNA BARCODING OF MEDICINAL PLANTS AND ITS IMPLICATION IN THE CONSERVATION AND TRADE

M. Sc. Thesis

(2018)

Submitted to

CENTRAL DEPARTMENT OF BIOTECHNOLOGY
TRIBHUVAN UNIVERSITY
Kirtipur, Kathmandu, Nepal

A dissertation submitted as the partial fulfillment of the requirement for M.Sc. degree in Biotechnology

Pradip Kumar Chaudhary
Exam Roll No.: BT 209/071
T. U. Registration No: 5-2-33-158-2010

DNA BARCODING OF MEDICINAL PLANTS AND ITS IMPLICATION IN THE CONSERVATION AND TRADE

A dissertation submitted as the partial fulfillment of the requirement for M.Sc. in Biotechnology

(2018)

Submitted to

CENTRAL DEPARTMENT OF BIOTECHNOLOGY
TRIBHUVAN UNIVERSITY
Kirtipur, Kathmandu, Nepal

Submitted by

Pradip Kumar Chaudhary
Exam Roll No.: BT 209/071
T. U. Registration No: 5-2-33-158-2010

Supervisors

Prof. Tribikaram Bhattarai, Ph.D.
Central Department of Biotechnology,
Tribhuvan University
Kirtipur, Nepal

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

ACKNOWLEDGMENT

Foremost, I would like to express my sincere gratitude to my respected supervisors **Prof. Dr. Tribikram Bhattarai**, Central Department of Biotechnology, T.U. and **Dr. Ram Chandra Poudel**, Senior Scientific Officer, Molecular Biotechnology Unit, NAST for their continuous support motivation, encouragement, constructive remarks and dynamic supervision throughout the study. I could not have imagined having a better supervisors and mentor for my Master thesis study.

My sincere thanks also goes to **Prof. Dr. Jibraj Pokharel**, Vice-chancellor of NAST, **Prof. Dr. Budhi Ratna Khadge**, Secretary of NAST, **Mrs. Jay Shree Sijapati**, Head of Faculty of Science and **Dr. Jyoti Mahrajan**, Coordinator of Molecular Biotechnology Laboratory of NAST for entire financial support in my laboratory work and allowing me to carry out my thesis work at Molecular Biotechnology Laboratory under **Master Thesis Support Program (2074/2075)**. I would like to acknowledge **Dr. Deegendra Khadka** and **Mr. Mitesh Shrestha** for suggestions and help. My gratitude goes to entire NAST family and to my laboratory colleagues Mrs. Nirma Basnet, Mrs. Gauri Thapa, and Mrs. Rachana Regmi for their cooperation in my thesis work.

My Sincere acknowledgement is to thank **Prof. Dr. Krishana Das Manandhar**, Head of Department, Central Department of Biotechnology, T.U., Prof. Dr. Ganga Prasad Kharel, Prof.Dr.Mohan Kharel, Prof. Dr. Rajani Malla, Prof. Dr. Tilak Ram Shrestha, Dr. Suresh Subedi, Mrs. Jarina Joshi, Mr. Bal Hari Poudel, Mrs. Smitha Shrestha, and entire teaching faculty members, staffs of Central Department of Biotechnology for their support, encouragement and valuable suggestions.

I would also like to express my gratitude to my entire dear colleagues **6th batch CDBT, T.U.**, seniors, juniors of Central Department of Biotechnology for all those days spent together. You are all amazing people I would never forget your constant support and friendship.

I would also like to thank **University Grant Commission (UGC)**, Sanothimi, Bhaktapur for financial support and awarding **Master's thesis support (2073/2074)** to carry out the field work and preparing the thesis. I would also thankful to **"Dr. L. Sreerama Scholarship 2073"** provided me during **3rd semester for the partial support.**

Finally, my deepest gratitude to my **parents, brother, sisters, friends** and **relatives** for their motivation, patience, encouragement, supports and continuous trusted to my ability to successful completion of this research.

LIST OF ABBREVIATIONS

µl	Microlitre
AEC	Agro Enterprise Centre
AFLP	Amplified Fragment Length Polymorphism
BLAST	Basic Local Alignment Search Tool
BOLD	Barcode of Life Data System
BP	Base Pairs
CBOL	Consortium for the Barcode of Life
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CO1	Cytochrome Oxidase 1
CpDNA	Chloroplast DNA
CTAB	Hexadecyl (Cetyl) Trimethyl Ammonium Bromide
D/W	Double Distilled Water
DNA	Deoxyribonucleic Acid
DNTPs	Deoxynucleotide phosphates
DoF	Department of Forests
DPR	Department of Plant Resources
EDTA	Disodium Ethylene Diamine Tetra Acetate
ETBr	Ethidium Bromide
FNCCI	Federation of Nepalese Chambers of Commerce and Industry
GIZ	Gesellschaft für Internationale Zusammenarbeit
GoN	Government of Nepal
HTS	High Throughput Sequencing
iBOL	International Barcode of life
ICI	Canadian International Consortium Initiative
ITS	Internal Transcribed Spacer
*ITS*2	Internal Transcribed Spacer region 2
IUCN	International Union for Conservation of Nature
K2P	Kimura 2-parameter
Kb	Kilobase
MAPs	Medicinal and Aromatic Plants
matK	maturase K
UNEP	United Nations Environment Programme
MEGA	Molecular Evolutionary Genetic Analysis
ml	Mililitre
MoFSC	Ministry of Forests and Soil Conservation
NAST	Nepal Academy of Science and Technology
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NJ	Neighbour Joining
NTFP	Non-timber forest products
NTFPs	Non-Timber Forest Products
PCR	Polymerase Chain Reaction
RAPD	Random Amplified Polymorphic DNA
rbcL	Ribulose 1, 5-bisphosphate carboxylase/oxygenase Large subunit
RFLP	Restriction Fragment Length Polymorphism

TAE	Tris-Acetate-EDTA
TE	Tris-EDTA
TU	Tribhuvan University
UGC	University Grant Commission
WHO	World Health Organization

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

DNA barcoding is an integrative molecular taxonomic tool that uses short gene sequences taken from a standardized portion of the genome for species identification. It is entering a new phase of applications to address questions relating to taxonomy, forensics, ecology, evolution, and conservation of wildlife. Despite being small in size Nepal ranks 25th position in terms of biodiversity richness. A substantial proportion of plant diversity comprises high value species including Medicinal and Aromatic plants, (MAPs) which serve as sources of food, fodder, fiber, medicine, essential oils, paper, clothing, dye and other useful products for the local peoples' livelihood increments.

Collection and trade of these plants is a major source of rural income that in turn has generated substantial amount of revenue for the government as well. Globally the demand of Nepalese herbs has increased exponentially resulting their unsustainable collection, adulteration, overexploitation and illegal trade. Intervention of DNA barcoding technique presents tremendous potentialities to overcome existing challenges in the sustainable utilization and conservation of biodiversity of Nepal.

Mardi-Himal of Central Nepal is a single Mountain ridge that ascends from low altitude subtropical forest to high altitude alpine vegetation in single vertical transect. A total of 89 high value plants predominantly in the Nepal Governments' protection list, CITES appendices, prioritized for farming, research and development and conservation were collected from the study area. Herbarium specimens and silica preserved samples were prepared. Plant DNA was extracted from silica preserved leaves using widely used CTAB protocol and Qiagen kits. Both chloroplast (*rbcL*) and nuclear markers (*ITS*) were used for PCR optimization. The amplification success rate for *rbcL* and *ITS* markers were found to be 85.39% (76) and 83.15% (74) respectively. The sequencing efficiency of *rbcL* and *ITS* makers were found to be 73% (65) and 65% (58) respectively. Combining both markers, sequencing efficiency achieved 80% (72) of the samples. Multi-locus combination *rbcL*+*ITS2* exhibited highest discrimination at species level i.e. 61.11% by similarity method, 87% by distance method and 72% by tree method. Single locus *ITS2* has exhibited highest discrimination at species level except in tree method i.e. 72% by similarity method, 87% by distance method and 52% by tree method. Moreover, *rbcL* marker exhibited lowest discrimination by similarity method (30%) but relatively highest by distance method (88%) and tree method (62%). *ITS* marker exhibited lowest discrimination by distance (52%) and tree method (52%) but slightly high in similarity method (65%).

Combining all three methods (BLAST, Distance and Tree method), prioritized species such as *Aconitum spicatum, Berberis angulosa, B. asiatica, B. erthoyclada, Bergenia ciliata, B. purpurascens, Dactylorhiza hatagirea, Delphinium brunonianum, D. himalayai, Engelhardia spicata, Mahonia nepalensis, Michelia doltsopa, M. kisopa, Magnolia campbellii, Nardostachys grandiflora, Neopicrorhiza scrophulariiflora, Paris polyphylla, Rheum australe, Rubia manjith, Swertia chirayita, S. teres, Taxus wallichiana, Zanthoxylum armatum* and *Z. acanthopodium* were correctly identified at species level**.**

The present utility of molecular tools, possibly the first of its kind in Nepal has demonstrated the potentialities of DNA barcoding and formed a strong foundation for the establishment of "DNA barcode library of MAPs in Nepal". Findings of this study have several significant implications on species identification, cryptic species assessments;

identify adulteration, authentication and value addition of herbs and herbal products, trade regulations and sustainable utilization as well as long term conservation of the biodiversity.

Keywords: DNA Barcoding, High value plants, *rbcL*, *ITS*, *ITS2*, DNA Barcode Library, Conservation, and Trade.

CHAPTER 1. INTRODUCTION

1.1 Background

1.1.1 DNA Barcoding

DNA barcoding is an integrative molecular taxonomic tool that uses short gene sequences taken from a standardized portion of the genome to identify species, and address questions regarding the ecology and evolution of natural systems phenomena. For many users its application is to identify an unknown sample by correctly matching a specific genetic marker of reference sequence library. However, DNA barcodes can also be applied as tools for addressing fundamental questions in ecology, evolution, and conservation biology, such as: how species are evolved and how are species assembled in communities (Kress et al., 2015). The concept of a 'DNA barcode' first proposed just over a decade ago for a quick and reliable species-level identifications across all forms of life, including animals, plants, fungi, and microorganisms. A short DNA sequence of 600 base pairs (bp) in the mitochondrial gene encoding *cytochrome c oxidase subunit 1* (CO1) has been accepted as a practical and, standardized, species-level DNA barcode for many groups of animals (Hebert & Barrett, 2005). The DNA barcode loci now most commonly used for plants and fungi are a combination of plastid *rbcL, matK*, and *trnH-psbA* with nuclear internal transcribed spacer (*ITS*) for adequate species discrimination (Hollingsworth et al., 2011; China Plant BOL Group et al., 2011)

The use of DNA barcodes for the identification of new species or closely related species is entering new tool to clarify species boundaries and to quantify the species diversity (Costion et al., 2011). In many cases, these markers serve as the initiation point for the discovery of new taxa. DNA barcodes are valuable genetic tool to reveal cryptic species previously unrecognized through the analysis of standard morphological variation (Kress et al., 2015) of animals (Smith et al., 2008) and plants (Kress et al., 2009). DNA barcodes are also utilized to identify the new species in a poorly known flora (Kress et al., 2015).

The DNA barcode data is helpful to understand and reconstruction of an evolutionary relation within targeted communities. The multi-locus plant DNA barcode could robustly reconstruct evolutionary relations among phylogenetically disparate communities members (Kress et al., 2009; Erickson et al., 2014). A major challenge of reconstructing a phylogenetic tree using DNA barcode sequence data is to capture the proper evolutionary relations among both highly divergent and closely related species (Smith et al., 2008).

Combination of an evolutionary and ecological data in DNA barcoding helps to understand the role of ecological, biogeographical, and evolutionary processes in the distribution of biodiversity (Bares et al., 2004). The universal DNA barcode data for plants will help to determine the broad spatial comparisons (Erickson et al., 2014). DNA barcodes represent the unique opportunity to understand trophic interactions among organisms, especially in habitats which are difficult to access (Jones et al., 2011). DNA barcodes can provide universal markers across species in a community or a region by which means genetic distance (phylogenetic diversity) can be estimated within and across ecological communities depending on geographical scales (Chen et al., 2010).

When DNA barcode libraries becomes populated with species across globe, then, comparative measures of phylogenetic diversity will also become standard metrics for

conservation assessment (Kress et al., 2015). In addition to conservation assessment, DNA barcodes are now also being used for reliable identification and detection of illegally traded and often endangered species (Lahaye et al., 2008). The forensic use of DNA barcodes for identification of endangered species and commercially useful plants and animals is being expanded by local, state and national governments. DNA barcodes are providing useful data as evidence in criminal cases and investigations of natural and manmade disasters (Kress et al., 2015). For examples, a library of *CO1* markers for birds is now routinely used to identify avian species involved in airplane strikes (Dove et al., 2008). DNA barcoding can have a wide range of application such as to support ownership or intellectual property rights; to reveal cryptic species; in forensic to link biological sample to crime scenes; to support food safety and authenticity of labeling by confirming identity or purity and in ecological and environmental genomic studies (Li et al., 2015).

1.1.2 High value plants

Nepal is rich in biodiversity. It harbours 35 forest types and, 118 ecosystems that represents 2% of the World's flowering plants i.e. an estimates of 7,000 species of higher plants (Bhuju et al., 2007; MoFSC, 2009). Nepal is ranked 31th and 10th positions in terms of biodiversity richness in the world and Asia respectively (MoAD, 2017). The high value plants fall into two major groups: herbs and spices; and Non-Timber Forest Products (NTFPs) including Medicinal and Aromatic Plants (MAPs). Small quantities of these plants contribute to high economic benefits (Adhikari, 2001; Hamilton, 2013). Most of Nepal's high value MAPs grow in the forests and grasslands of the mountains in the northern part of the country. These high value plants are exported from far-western and Mid-western development regions of Nepal in greater quantity. The lower value MAPs grow below 2,000 m of altitude. In recent years, the market price of most of these plants, especially of high value plants, has increased significantly, leading to overharvesting of the resource (UNEP, 2012).

High value plants, those are also referred as Non-timber forest products (NTFP) are very important in the Himalaya region, which serve as sources of food, nutrition, fodder, fiber, medicine, dye and other useful products, and are important also for the local peoples' livelihood increments. The collection and marketing of NTFP is a major source of rural income, which in turn is also an important source of revenue to the government (Ghimire et al., 2008). In the mountains regions of Nepal, some studies reported 10-100% of households are engaged annually in commercial collection of medicinal plants and other NTFPs. Furthermore, some rural areas are earning up to 50% of the family income from the trade of high value plants (Chhetry 1999; Edwards, 1996).

The total contribution to national agriculture output of MAPs is small, but their value per weight is highest among the traded plants. The pharmaceutical cash crops have a high potential for remote communities that practice subsistence agriculture and have limited access to regional economy (Dubey et al., 2004; Chauhan, 2010). Medicinal plants represented by the number of species used, are perhaps one of the most significant ways in which humans directly recall the benefits provided by biodiversity (Farnsworth & Soejarto, 1991; Hamilton, 2004). More than 85% of herbal medicines used in traditional health care systems are derived from medicinal plants (Farnsworth, 1988) those ensure livelihoods of millions of people (Phondani et al., 2014). The altitudinal variation, different habitat types, and varying microclimatic conditions in the Himalayan region

favours creation of the good environment for the growth and development of MAPs (Kunwar et al., 2013). Indigenous knowledge of herbal medicine is important for poor rural communities, as these herbal medicines are used for curing variety of ailments through the traditional health care system (Phondani et al., 2014). Indigenous people's knowledge of medicinal plants and therapies of various local communities has been transmitted orally for centuries. Modern therapeutic technologies are historically based on indigenous knowledge, which become a new tool for searching a novel drugs or derivative drugs (Kunwar et al., 2013). Few plants species are used for the treatment of a specific disease, while several other have multiple uses. The plants were mainly used as stomachic, anti-allergic, anti- neuralgia, vermifuge, narcotic, laxative, anti-jaundice, emollient, hypnotic, diuretic, digestive, demulcent, carminative, astringent, aphrodisiac, anti-spasmodic, anti-emetic, anti-diabetic, anthelmintic, and many others (Alam et al., 2011). A variety of medicinally important chemical compounds have been extracted and identified from the plants including terpenes and triterpenoids, sterols and steroids, phenolics, flavonoids, gums and resins, quinones, anthocyanidines, saponins, antioxidants and fatty acids (Hameed et al., 2011).

A large number of NTFP species including medicinal plants found in the Himalayan region have been reported to be endemic to this region. 25 species of medicinal plants (1.5% of total species) have been reported to be endemic to the Nepal Himalaya (Shrestha & Joshi 1996; Ghimire 2006). According to an estimate there are over 2000 species of plants in Nepal which are known to be highly useful; about 1600-1900 species are commonly used for medicinal purposes (Shrestha et al. 2000; Baral & Kurmi, 2006; Ghimire 2008). Most of these plant species have different useful properties along with their medicinal values. About 600 species are reported to be utilized for food and food additives while about 500 species are utilized as various products such as dyestuffs, tannins, essential oils, lipids, fibers, gums/resins, etc. These products have different useful properties such as coloring agents, detergents, adhesives, lubricants, cosmetics, ornaments, deodorants, ropes, papers, binding materials, containers, clothing, etc. In addition, a number of species have used for social/religious and environmental purposes and still many species are utilized as fuel and animal food (Chaudhary, 1998; Rajbhandari, 2001; Manandhar, 2002). Distribution pattern of life forms of medicinal plant species in the Nepal Himalaya shows that 45.0% of the total naturally growing species (1714) are long-lived herbaceous perennials. The second largest life form categories of medicinal plants are shrubs (16.6%), followed by annual/biennial herbs (15.6%), trees (13.6%) woody and herbaceous climbers (8.7%) (Ghimire, 2008).

Most of NTFP including MAPs species are traded through illegal channels. Subedi (2006) estimated a total of 161 plant-based NTFP species, which are harvested for commercial purpose in Nepal. Another study of Bhattarai & Ghimire, (2006), listed a total of 143 species as commercial medicinal plants (including cultivated and exotic species). It has been estimated that the forestry sector in Nepal contributes about 15% of the national GDP, of which about 5% is contributed by NTFP. The amount of MAPs exported through legal channels from Nepal increased from around 3,400 tonnes in 1989/90 to about 11,500 tonnes in 1993/4 (Karki et al., 2003). The Department of Forests (2006) estimated that in the fiscal year 2005/06, approximately 33,000 tonnes of MAPs were exported from Nepal. In 2008, Nepal has exported MAPs worth USD 3 million which increased in 2009, to 9.8 million (GIZ, 2011). A large part of trade in Nepali MAPs are exported in China and India for production of Chinese and Ayurvedic medicine

respectively (UNEP, 2012). Among traded NTFPs, 95% of the total trade is covered by only few species. The most common NTFPs that are traded on a large scale included Pine resin (khoto), Sal seed, Kutch, Ritha, Timur, Dalchini and Tejpat, Sabai grass or Babiyo, Lokta, Satawari or Kurilo, Chirayito, Jatamansi, Padamchal and Sugandhkokila. Similarly, major 10 NTFPs traded in large quantity included Ritha, Timur, Pakhanveda, Kaulo bark, Pawan bark, Jhyayoo, Amala, Tejpat, Chiraito and Majitho (DoF/MoFSC, 2008).

An increasing trend of harvesting has turned into greater pressure for long time on selected species since most medicinal plants and NTFPs in trade are harvested from wild population and very few are cultivated (Chaudhary et al., 2016; Luitel et al., 2014). Major conservation issues are over-harvesting, both at premature and unsustainable harvesting due to illegal trade pressure (which is often undeclared), habitat destruction, livestock grazing, forest fire, etc. It is widely believed that the harvesting of NTFPs is no longer sustainable in many areas (Chaudhary et al., 2016). There is an immediate attention regarding sustainable harvesting practices and encourage or boosting cultivation to conserve naturally occurring NTFPs and MAPs. The cultivation of high value plants including MAPs on community and private land is one of the significant methods for conserving biodiversity and generating income. There is an increasing interest to cultivate number of species in several parts of Nepal (UNEP, 2012). In 2006, 36 species were identified as having a good potential for commercial cultivation and market promotion (AEC/FNCCI, 2006). Herbs and NTFPs Coordination Committee (HNCC) of GoN has compiled a list of 30 species of MAPs and other NTFPs for research and development (DPR, 2012).

1.2 Developments on DNA Barcoding

DNA barcodes have provided a new biological tool for organismal biologists to increase their understanding of the natural world. Over the last decade, four plant DNA barcode markers, *rbcL, matK, trnH-psbA,* and *ITS* have been developed, tested, and used to address basic questions in systematics, ecology, evolutionary biology, and conservation including community assembly, species interaction networks, taxonomic discovery, and assessing priority areas for environmental protection. Forensic investigators have also applied these plant DNA barcodes in the regulatory areas of traffic in endangered species and monitoring commercial products, such as foods and herbal supplements. Major challenges ahead will focus on building the global plant DNA barcode library and adopting genomic sequencing technologies for a more efficient and cost effective workflow in applying these genetic identification markers to additional fields of biological and commercial endeavors.

One of the exciting modifications of DNA barcoding is appropriately called "metabarcoding" or eDNA; using sub-region of standardized markers to overcoming the problem of degraded DNA. Some researchers are also advocating a focus on chloroplast genome sequencing as "Super-barcodes" to replace the locus based approach.

1.3 Rationale of the study

Nepal is rich in biodiversity due to its complex variation in geomorphology and phytogeography (topology, climate and altitudinal). Despite having such abundance, relatively less is known at the species diversity at the molecular level. In some instance, even the phenotype determines the novelty of the species despite being identical at the genotype. While in other cases, despite having a distinct genotype, the medicinal plant

may show a similar phenotype (Eg. Species *Astilbe rivularis* and *Aruncus dioicus*; *Neopicrorhiza scrophulariiflora* and *Veronica* spp; *Rhododendron* spp are very similar and difficult to identify morphologically). In both of these cases, there is a chance of high degree of error in identification and the subsequent classification. In addition, trade of Non timber forest products including medicinal and aromatic plants is one of the major economic activities in Nepal. Adulteration may decrease the benefits that could be obtained from Medicinal plants. Value addition is another important issue to get high profits from medicinal plants. Illegal collection and over harvesting need to be stopped for sustainable conservation. For this reason we initially need sound taxonomy of each species and thereafter we should enrich our DNA reference library which can help us to overcome current challenges in this sector. It is imperative that the molecular study of the plants be carried out through the utilization of the DNA barcodes so as to determine the level of diversity as well as identification of new species with accuracies. Despite the limited resource settings of Nepal, the current usage of DNA barcodes in identification of the high valued plants should help in not just its molecular identification and establishment of barcode library but also its conservation necessity. Furthermore, the current research project aims to authenticate the high value plants including MAPs of central region using chloroplast (*rbcL*) and nuclear region (*ITS*-region).

1.4 Research Questions

This research aims to answer following questions:

- Does application of DNA barcoding helps in identification of High value plants including Medicinal and aromatic plants of Central Nepal at molecular level?
- How significant is the level of distinction for the intra and interspecific variablilities?
- Can the data obtained be utilized for the value addition of MAPs on trade and their sustainable conservation and effective managements?

1.5 Research Objectives

1.5.1 General objectives

i. DNA Barcoding of medicinal plants from Mardi-Himal area of Central Nepal and its implications to conservation and trade.

1.5.2 Specific objectives

i. Access the PCR and Sequencing efficiencies of the universal *rbcL* and *ITS* DNA barcode regions for medicinal plants of Mardi-Himal area.
ii. Study species discrimination efficiencies of *rbcL* and *ITS* barcodes on medicinal plants of Mardi-Himal area.
iii. Analyse the intra and interspecific divergence among sampled species.
iv. Create DNA barcode reference library for studied samples of medicinal and aromatic plants.

1.6 Scope of the Study

Nepal being rich in biodiversity has been subjected to unsustainable and overexploitation of natural sources. However, based on the outcome of this research, we can increase economic benefits through value addition of high value plants. In addition, DNA reference library that could be a helpful tool to check illegal collection

and, adulteration, new species identification, provide world standard information on our MAPs and protect our high value plants from biopiracy etc. This, in turn, would help the nation in realizing its sustainable conservation; identify adulteration and authentication of herbal products. Furthermore, biotechnological applications would also enhance our technical capabilities to implement modern molecular tools in Nepal in the sustainable utilization of many more economically potential plants those could improve the livelihood of rural communities.

This research could play a major role in the standardization of medicinal plants in trade and increase the benefits through value addition. The research would be a stepping stone towards understanding the diversity of medicinal plants of Mardi-Himal area and can become a seed of a tree whose far reaching branches shall engross fields including but not limited to that of sustainable conservation.

CHAPTER 2. LITERATURE REVIEW

2.1 Flora biodiversity in Nepal

Nepal is a landlocked country which lies along the Himalayan region and situated between China and India covering a total area of 147181 sq. km. Nepal is located between the latitude of 26° 22' and 30° 27' N and the longitude of 80° 40' and 88° 12' E. The average length of the country is 885 km from east to west and width varies 145 km to 241 km from north to south. Nepal is divided into three distinct topographic zones; 86% of the total land area is covered by hills, high mountains and remaining 14% is of flat land of Terai. Based on wide altitudinal variation (60-8848m), the climate is broadly classified into cold Arctic/Nival (above 3000m), cold temperate (2000-3000m), warm temperate (1500-2000m), subtropical (1000-1500m), and tropical (below 1000m). According to the physiological region, Nepal is divided into 7 regions including Terai, Siwaliks, Mahabharat lekh, Midhills, Himalayas, Inner Himalayas and Tibetan marginal mountain range. Himalayan region has a unique assemblage of flora and fauna. Most of this mountainous region lies in Nepal. Nepal has huge resources of natural beauty and a rich cultural heritage. Nepal has also diverse nature of soils and variable number of climatic regions, which are suitable for the growth of variety of plant species ("nepal; country profiles; fao," n.d.; Siwakoti, 2007; Paudel et al., 2012). Although Nepal covers less than 0.1% of the earth's land area but it comprises 3.2% plant species diversity at global scale. Current estimates of species number indicate that there are 465 species of lichens (2.3% of global diversity), 1,822 species of fungi (2.6%), 1,001 species of algae (2.5%), 1,150 species of bryophytes (8.2%), 534 species of pteridophytes (5.1%), 26 species of gymnosperms (5.1%), 6,973 species of angiosperms (3.2%) as shown in Table 2.1 and (GoN/MoFSc, 2014). Taxonomic research is being undertaken in Nepal to update the number of taxa (species and subspecies levels mainly) with focus on some selected groups (MoFSC 2009).

Table 2.1 Status of plant species diversity of Nepal

Group	Number of known species	Percent of known species in the World
Lichens	465	2.3
Fungi	1,822	2.6
Algae	1,001	2.5
Bryophytes	1,150	2.5
Pteridophytes	534	5.1
Gymnosperms	26	5.1
Angiosperms	6,973	3.2
Flora (total)	**11,971**	**3.2**

Source: GoN/MoFSC (2014)

2.2 Medicinal Plants

Medicinal plants are second most valuable bio-resources of Nepal after water resources (Rawal et al., 2009). Nepal's bioresources in the international trade are mostly wild in

origin. The 188 biomaterials commonly found in international trade consists of various plant parts such as roots and tubers (39 plants), barks (seven plants), leaves (26 plants), flowers (14 plants), fruits and seeds (61 plants), whole plants (12 plants), gums and resins (eight plants), and miscellaneous other plants (11) (Bhuju et al., 2007). Many highly demanded and globally important medicinal plants such as S*wertia spp., Paris polyphylla, Neo-picrorhiza scrophulariiflora, Podophyllum hexandrum, Taxus wallichiana, Nardostachys grandiflora, Podocarpus* sp. etc. are harbored in various geoclimatic region of Nepal (Press et al., 2000; Bhattarai et al., 2011). The compilation of Medicinal and Aromatic plant database of Nepal was published in 2000 (Shrestha et al. 2000), which listed 1624 species of medicinal plants. Ghimire (2008) reported 1950 species of medicinal plants in Nepal as shown in Table 2.2. Out of which 1906 species are represented by vascular a groups (angiosperms, gymnosperms and pteridophytes) that comprises 1614 native, 192 introduced and/or cultivated, 100 naturalized taxa. Most recently, Rokaya et al., (2012) reviewed and updated the list of MAPs of Nepal to 2331 species of medicinal plants of Nepal. Different life forms of medicinal plants are used in traditional health care system (Figure 2.1). Distribution patterns of MAPs of Nepal are summarized in Figure 2.2.

Table 2.2 Diversity of Medicinal plants of Nepal (Only vascular groups are considered)

Plants groups	Families	Genera	Species
Angiosperms	169	897	1810
Gymnosperms	7	11	19
Pteridophytes	24	41	77
Total	200	949	1906

(Source: Ghimire, 2008)

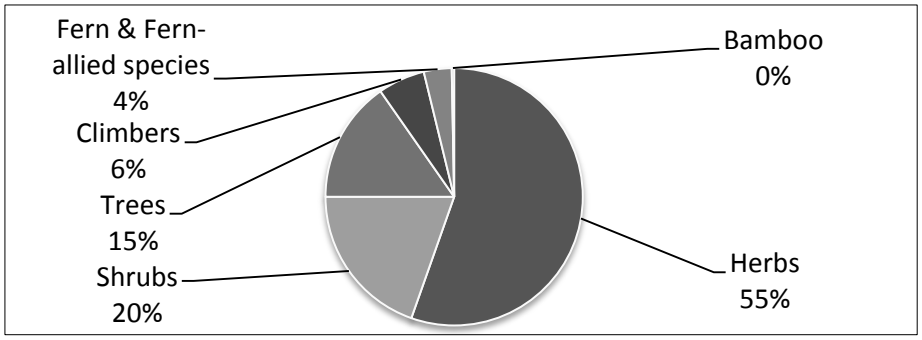


Figure 2.1 Graphical representation of different life form of plants can be used in Medicinal purposes (Source: Rokaya et al., 2012).

Medicinal plants have been harvested from ancient's time for their multipurpose uses. The indigenous people are well experienced with the properties and uses of plants of their surroundings. Until the middle of the 19th century, plants were the main therapeutic agents used by humans. According to recent estimates by the World Health Organization, more than 3.5 billion people in the developing world still rely on plants as components of their primary health care (www. fao.org) and more than 80% of the world's population relies on traditional herbal medicine for their primary healthy care needs (Inglis, 1994). About 60% of the world population and 60-90% of the population of developing countries depend on traditional medicine (Shrestha & Dhillion, 2003). Many rural peoples possess key information on traditional knowledge of medicinal plants.

Such knowledge are transferred from one generation to another generation (Shrestha & Dhillion, 2003). Traditional medicine in Nepal is used by majority of the rural population. Different types of medicinal practices viz. Ayurveda, Traditional Chinese Medicine (TCM), Unani, Tibetan Amchi and folk medical system are prevalent in Nepal (Sheng-Ji, 2001). Traditional medicine in Nepal contributed to those practices based on beliefs that were in existence often for hundreds to thousands of years before the development and spread of modern medicine, and which are still in use today. Many rural areas of Nepal practice, traditional medicine system where knowledge and practice were passed down entirely via oral tradition on a lineage mode of transmission and personal experience (Bhattarai, 1988). Medicinal plants play vital roles in the Nepalese livelihood and the use of medicinal plants is frequent in several Nepalese regions (Gaire & Subedi, 2011). These valuable herbal traditions found in developing countries have always been considered an important component of the cultural heritage and foundation for the development of modern drugs. However, unfortunately the influence of traditional herbal medicine for the creation and development of new drugs that supply modern markets with allopathic treatments is eventually leading to the replacement of traditional medicine (Sheng-Ji, 2001). Majority of the valuable plants grow in wild conditions as natural components of vegetation of particular region. The necessary plant materials (roots, leaves, bark, etc) are collected and sold by the local people to the traders, industries and exporters (Rawal et al., 2009). Nepal is a natural storehouse of medicinal plants. Each year thousands of tons of raw material are exported, mostly to India, but also to other Asian, European and American countries. The government of Nepal (GoN) has prioritized thirty medicinal plants for sustainable use, research and development (Table 2.4). Conservation programmes are designed for livelihood improvement and poverty alleviation (Uprety et al., 2010) which can have greater implications on the protection of endangered and threatened plants. GoN protected high value plants viz. endangered and threatened plants are listed in Table 2.3

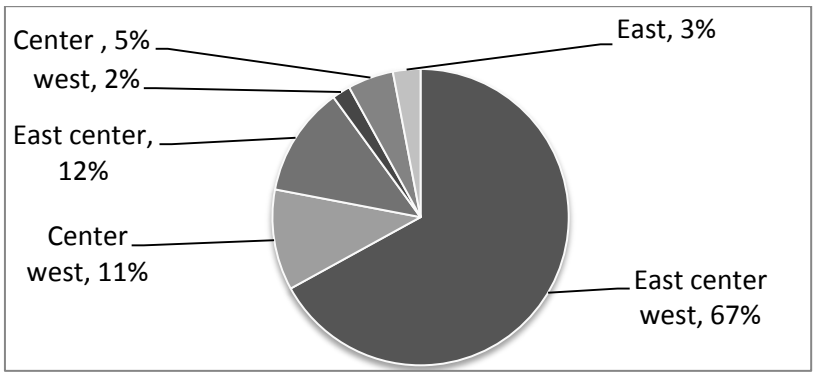


Figure 2.2 Graphical representation of medicinal plant in different region of Nepal (Source: Acharya et al., 2009).

In this scenario, handling of Nepalese medicinal biodiversity should involve an integrated approach comprising of well documentation, their sustainable utilization and conservation (Rawal et al., 2009). Conventionally, macroscopic (morphological) and microscopic characters including anatomy, cytology and chemical profiling techniques such as thin layer chromatography (TLC) and high performance thin layer chromatography (HPTLC), gas chromatography (GC), high performance liquid chromatography (HPLC), etc are being used for characterizing genotypes and chemotypes (Joshi et al. 2004).

Table 2.3 Plant species protected of Nepal

Scientific Name	Common Name	Local Name	IUCN Red List Status	CITES Appendix Status
A. Species banned for collection, use, sale, distribution, transportation, transportation and export				
Angiosperms				
Juglans regia	Walnut	Okhar		
Dactylorhiza hatagirea		Panchaule		
Neopicrorhiza scrophulariiflora	Gentian	Kutki		II
B. Species banned for export expect for processed with permission of Department of Forests				
Fungi and Angiosperms				
Cinnamomum glaucescens		Sugandakokila		
Ophiocordyceps sinensis		Yarsagumba		
Rauvolfia serpentina	Serpentine	Sarpagandha	Vulnerable	II
Valeriana jatamansi	Spike nard	Jatamansi		
Valeriana wallichii	Valerian	Sugandhwal		
Gymnosperms				
Abies spectabilis	Fir	Talis Patra	Near threatened	
Taxus contorta	Himalayan yew			
Taxus wallichiana	Himalayan yew			
Taxus mairei				
Lichens				
Parmelia spp.	Lichen	Jhyau		
C. Species banned for harvest, transportation, and export for commercial purpose				
Angiosperms				
Acacia catechu	Cutch tree	Khayer		
Bombax malabaricum	Silk cotton tree	Simal		
Dalbergia latifolia	Rose wood	Satisal	Vulnerable	
Michelia champaca	Magnolia	Champ		
Pterocarpus marsupium	Indian kino tree	Bijaysal		
Shorea robusta	Common sal	Sal		

Source: GoN/MoFSC, (2014); Poudel et al., (2012) identified 3 species of Taxus.

Table 2.4 Medicinal plants prioritized for research and development

Scientific Name	Family	Nepali Name
Aconitum heterophyllum Wall.	Ranunculaceae	Atis
Aconitum spicatum (Bruhl) Stapf	Ranunculaceae	Bish
Acorus calamus Linn.	Araceae	Bojho
Asparagus racemosus Wild.	Liliaceae	Satabari
Azadirachta indica A. Juss.	Meliaceae	Neem
Bergenia ciliata (Haw.) Sternb.	Saxifragaceae	Pashanvheda
Cinnamomum glaucescens (Nees) Hand.-Mazz.	Lauraceae	Sugandhakokila
Cinnamomum tamala (Buch-Ham.) Nees & Eberm.	Lauraceae	Tejpat
Ophiocordyceps sinensis (Berk.) Sacc	Clavicipitaceae	Yarshagumba
Dactylorhiza hatagirea (D.Don) Soo	Orchidaceae	Panchaule

Scientific Name	Family	Nepali Name
Dioscorea deltoidea Wall.	Dioscoreaceae	Vhyakur
Gaultheria fragrantissima Wall.	Ericaceae	Dhasingare
Juglans regia Linn.	Juglandaceae	Okhar
Lichens		Jhyau
Morchella spp.	Pezizaceae	Khoya/Guchi chyau
Nardostachys grandiflora DC.	Valerianaceae	Jatamansi
Neopicrorhiza scrophulariiflora (Pennell) Hong	Scrophulariaceae	Kutki
Phyllanthus emblica Linn.	Euphorbiaceae	Amala
Piper longum Linn.	Piperaceae	Pipala
Podophyllum hexandrum Royle6	Podocarpaceae	Laghupatra
Rauvolfia serpentina (L.) Benth. ex Kurz	Apocynaceae	Sarpgandha
Rheum australe D.Don	Polygonaceae	Padmachal
Rheum moorcroftianum Royle		
Rubia manjith Roxb. ex Fleming	Rubiaceae	Majitho
Sapindus mukorossi Gaertn.	Sapindaceae	Ritha
Swertia chirayita (Roxb. ex Fleming) Karsten	Gentianaceae	Chirayito
Tagetes minuta Linn.	Asteraceae	Jungli Saypatri
Taxus wallichiana Zucc.	Taxaceae	Lautha Salla
Tinospora sinensis (Lour.) Merr.	Menispermaceae	Jurjo
Valeriana jatamansi Jones	Valerianaceae	Sugandhawal
Zanthoxylum armatum DC.	Rutaceae	Timur

Source: DPR, (2012)

2.3 Genetic Diversity

Genetic diversity is the principle source of biodiversity where a total number of genetic characters contribute to variation within species. In other words, it is an estimation of the quantities of variation present within a population of a given species. Genetic diversity among individuals reflects the presence of different alleles in the gene pool, and hence that represents different genotypes in the populations (Laikre et al., 2010). Genetic diversity is the variations of genes and within populations or individuals of living organisms at the level of species including alleles, chromosomes, genotypes, individuals, populations, metapopulations, subspecies etc. Genes are principle unit of heredity that is passed from an organism to its offspring. These are composed of nucleic acids and located along an organism's chromosomes. A species consists of one or more populations and individuals of populations may be different from one another at various stages. High genetic variation within a species can support the species to adopt changing extreme environmental conditions (Primack et al., 2013; Chaudhary et al., 2016).

The study of genetic diversity is the process by which variation among individuals or groups of individuals is analyzed by a specific method or a combination of methods. Diverse of data sets have been used by researchers to measure the genetic diversity of plants (Mohammadi & Prasanna, 2003). Genetic diversity may also be estimated using morphological and biochemical markers (Mondini et al., 2009). Now, DNA-based marker data enables more reliable differentiation of genotypes. Genetic distance-similarity between two genotypes, populations, or individuals may be calculated by various statistical methods depending on the data set (Mohammadi & Prasanna, 2003).

The study of genetic diversity is needed for the development of appropriate protocol in conservation biology as well as many other applied fields. Genetic diversity is supposed to be critical for the evolutionary insights of species. Many research programs were investigated to the population structure that contribute to the evolutionary insights into the demographic pattern of diverse populations (Milligan et al., 1994). The study of the genetic diversity and population structure also assists plant breeding in the selection of parents for crossing, providing a more rational basis for expanding the gene pool, and for identifying materials that harbor genes of value for plant improvement. Variability and genetic diversity are crucial factors in evolution and also in applied sciences because they determine the responses of given organism, for examples; environmental stress, natural selection, and susceptibility of different diseases. A number of sources that affects the gene pool are: mutation, recombination, transposable genes, transfer of gene from other sources and etc (Brown, 2002). Furthermore, knowledge of population structure of genetic resources is essential for development of policies for appropriate conservation of genetic diversity. Genetic Information has essential factors in the development of conservation biology regarding on the evolutionary properties and population genetics of small populations (Milligan et al., 1994; Poczai et al., 2012). Molecular phylogenetics and genetic diversity analysis can allow clarifying the taxonomic identity and evolutionary relationships of the wild varieties of plant species. These methods can also enable to prevent misidentification. Interest in genetic diversity has also increasing in applied field of biology such as agriculture, where the technique of multiline represents of varietal mixes within a single field has long been known to increase crop yields (Poczai et al., 2012).

Genetic diversity studies of wild species are lacking in Nepal. This indicates that more people researches needs to be carried out in future on biodiversity sector at genetic level. It is considered that isolation of mountain peaks facilitates speciation and promotes genetic diversity. Brown oak (*Quercus semecarpifolia*), one of the most widely distributed tree species should have high genetic attributes capable of coping with severe climatic change (Singh et al. 2011). However, a substantial genetic diversity reflects among flora [and fauna] and is evident in terms of morpho-geographical variations. Nepal has rich diversity in terms of cereals, grain legumes, vegetables, fruits, timbers and medicinal plants etc. Molecular techniques (Isozyme, RAPD and Microsatellite, ISSR, DNA barcoding) are being used to characterize the selected species of crops, and almost rare in medicinal and high value wild plants. In Nepal, some of the Governmental organizations viz. Biological resources laboratory, Molecular Biotechnology unit of Nepal Academy of Science and Technology (NAST), Biotechnology laboratory at Nepal Agricultural Research Council (NARC), Central Department of Biotechnology, Tribhuvan University and Central Department of Botany, Tribhuvan University are engaged on investigating characterization and evaluation for disease/insect resistance, drought, biotic and abiotic traits, genetic diversity assessment of protected and other economic crops and DNA Barcoding of medicinal plants.

2.4 Common Methods Applied to Estimate of Genetic Diversity

2.4.1 Morphological identification/Phenotypic markers

Figure 2.3 the main components of typical leaf and flowers (Source: Remagnino et al., 2017).

The recognition of species of plants and animals by their appearance (morphology) and behavior goes back to humanity's own origins and certainly even before that (Atran, 1990). The morphological similarities among individuals of a species based on hereditary characters has always been acknowledged by taxonomists (Richards 2010; Wilkins 2009, 2010) Plant taxonomists use several characters which are relatively easy to see and characterize visually. These characters are predominantly external parts of the plant such as the sizes, shapes, colors, patterns and textures of stem, leaves, flowers, fruits and seeds as shown in Figure 2.3. Other, non-morphological characteristics have also been traditionally used to help characterize species, e.g. habitat, phenology and easily verifiable genetic features, such as the production of progeny which strongly resemble the parents (Remagnino et al., 2017). However, several limitations are there in the morphological identification and classifying life's diversity. In addition the morphological variation and their close related traits are difficult to distinguish and also complex process and subtle that most taxonomists specialize in a single group of closely related organisms. As a result, a multitude of taxonomic experts may be required to identify specimens from a single biodiversity survey. Searching of appropriate experts and distributing specimens can be a time-consuming and expensive process. The non-specialist who needs to identify specimens now, but is far from the few taxonomic centers in the world, is confronted with a nearly impossible task (Remagnino et al., 2017). But, now modern taxonomic work includes analysis of a host of other traits, including genes, isoenzymes, physiology, behavior, population biology, ecology and geography of the species.

Morphometric techniques can employ to enhanced transformations or manipulations of traditional morphological characters for the goal of identification (Remagnino et al., 2017). Morphometric, the study of shape, has been applied to plants and their organs for many years. The shapes of leaves, petals and whole plants are of high significance to plant science, as they can allow easy discrimination between different species. The increasing trends in biodiversity studies and the availability of digital images contribute to make morphometric analysis robust and reliable. A robust automated species identification system would allow people with only limited botanical training and expertise to carry out valuable field work. The computational, morphometric and image

processing methods that have been used in recent decades to analyze images of plants, introducing readers to relevant botanical concepts along the way (Cope et al., 2012).

2.4.2 Biochemical based identification/ Allozymes (Isozyme)

Medicinal or high value plants contain different constituents of secondary metabolites; depend on the planting material, environment and farming conditions, and age at harvest, storage, and processing. Chemical profiling of herbal products using chemical analytical methods such as chromatography, spectroscopy, or hyphenated chromatography-mass spectrometry (Lolita et al., 2017) allow to identify the quantitative molecular description of the whole extract of plant secondary metabolites (MW <1,000 Da) and product quality (Yongyu et al., 2011).. Nuclear magnetic resonance (NMR) spectroscopy can yield considerable information in an untargeted analysis of a plant extract (Lolita et al., 2017). Chemometrics is a family of techniques that applies statistics to voluminous chemical data, such as spectroscopic signals from a collection of samples, with the objective of gaining insights into the characteristics of the samples through graphical representation or pattern- recognition (Wold, 1995). Chemometric analysis is tool for the characterization of spectroscopic data from whole plant extracts to differentiate plants based on species, origin, processing treatment, age, and other quality parameters (Kim et al., 2011).

Isozyme refers to multiple form of an enzyme, with similar or identical catalytic activities, occurring within same organism (Markert & Moller, 1959). Isozymes provides marker system for a wide range of investigations in the genetics, biochemistry and developmental biology of plants (Scandalios, 1969). Isozyme consists of multiple forms of two or more polypeptide subunit encoded by different genes. Amino acid changes alter isoelectric points of polypeptides and permit their separation by electrophoresis and, as consequence, their use as genetic markers. When variant proteins are produced by different alleles at the same locus called allozymes. The advantages of allozymes over morphologic markers reside in their general codominance combined with rare epistatic interactions which make them particularly suitable for elucidating the mechanism of variation at single defined loci. On the other hand, they are subject to alteration by environment and development and limited to coding regions of soluble proteins (Cloutier & Landry, 1994). Isozymes is helpful for accessing the genetic diversity of population studies (Harry, 1986), taxonomy, genetic variability, identification of Crosses, Hybrids, and Genetic Lines (Micales et al., 1992), phylogenetic relationships and in plant breeding (Staub et al., 1997).

2.4.3 Molecular Techniques

The assessment of genetic diversity within and between populations or species is routinely performed at the molecular level using various laboratory-based techniques such as DNA analysis, which measures the level of variation directly (Mondini et al., 2009). DNA-based species identification is a powerful tool to both fill the gap and build on the taxonomic base is already in place (50% of the species ordinarily encountered by most people on earth are already described and can be identified through often at high cost with light to intensive applications of classical methods). The uses of DNA barcoding in taxonomic studies and the identification process, using short DNA sequence in a single gene can contain more than enough information to identify 10 million or more species. For example, a 600 nucleotide segment of a protein coding gene contains 200

nucleotides that are in the third position within a codon. At these sites, substitutions are (usually) selectively neutral and mutations accumulate randomly. Even if a group of organisms was completely biased to either adenosine or thymine (or alternatively, to either guanosine or cytosine) at the third nucleotide positions there would still be 2^{200}, or 10^{60}, possible sequences based on third position nucleotides alone (Stoeckle et al., 2003).

Molecular tools have been widely used for species differentiation and identification throughout the past decades. The large scales application of molecular data is clearly bound to revolutionize taxonomy, but the validity and practicalities of molecular approaches to taxonomy have been subject to a variety of criticisms. A major distinction should be made between species identification, generally associated with the idea of 'molecular barcodes', and species circumstances and delineation, broadly referred to as 'DNA taxonomy' (Vogler & Monaghan, 2007).

DNA sequence analysis of a single gene sequence to allow species identification is termed "DNA barcoding". The DNA Barcoding is an analogy with the Uniform Product Code barcodes on manufactured goods, which may represent to a "Species" of product (as well as individual item). The availability of broad range primers for amplification of a 645 bp fragment of *cytochrome c oxidase subunit I* (*COI*) from diverse invertebrate and vertebrate phyla establishes this gene sequence as a particularly promising target for species identification in animals (Folmer et al., 1994), as does the moderate change rate in this sequence and the large numbers of copies per cell owing to *COI* being a mitochondrial gene. Proof of principle for DNA barcoding is now being provided and probed by comparison of *COI* sequences among closely related species and across diverse phyla in the animal kingdom (Hebert et al., 2003)

DNA barcoding enables discrimination between closely related species. Comparison of *COI* sequences from 13,000 pairs of congeneric species showed a mean divergence of 11.3%, corresponding to approximately 50 diagnostic substitutions per 500 bp of the *COI* gene (Hebert et al. 2003). Furthermore, *COI* sequence variation within species is generally quite low, less than 2%, and has not been an impediment to species discrimination, including among many assemblages of closely related organisms (Hebert et al. 2003).

Regarding the role of DNA barcoding in diagnosing new species, it will be a helpful in addition to the existing tools, but it is not visualized as replacing them. In many groups, alpha taxonomy requires data from morphology, behavior, ecology, natural history, and geographic variation. These data will certainly be enhanced by complementary information on associated DNA sequences. The aim of DNA barcoding is to analyze the smallest sequence that will provide the required information, thus reducing cost and facilitating automation. It allows assisting a small portion of a *COI* may enable identification of most animal species (and a sequence of some other gene is very likely to work as well for plants, fungi, etc).

2.4.3.1 Molecular Markers

The concept of genetic markers is not new; in the nineteenth century, Gregor Mendel employed phenotypic-based genetic markers in his experiments. Later, phenotypic based genetic markers for *Drosophila melanogaster* led to founding theory of genetic linkage, occurring when particular genetic loci or alleles for genes are inherited jointly.

The limitations of phenotypic-based genetic markers led to the development of DNA-based markers i.e. Molecular markers. Molecular markers can be defined as a genomic locus, detected using probe or specific primers (Barcaccia et al., 2000; Mondini et al., 2009). Molecular markers may or may not correlate with phenotypic expression of a genomic trait. An ideal molecular marker should possess following features;

- Polymorphic and evenly distributed throughout the genome.
- Provides adequate resolution of genetic differences.
- Generate multiple, independent and reliable markers.
- Simple, quick, inexpensive.
- Need small quantity of tissues or DNA samples.
- Linked to distinct phenotypes.
- Require no prior information about the genome of an organism.

Nevertheless, no molecular markers possess all the listed advantages. The different methods of molecular assessment differ from each other with respect to important features such as genomic abundance, level of polymorphism detected, locus specificity, reproducibility, technical requirements and cost (Mondini et al., 2009; (Primmer, 2009).

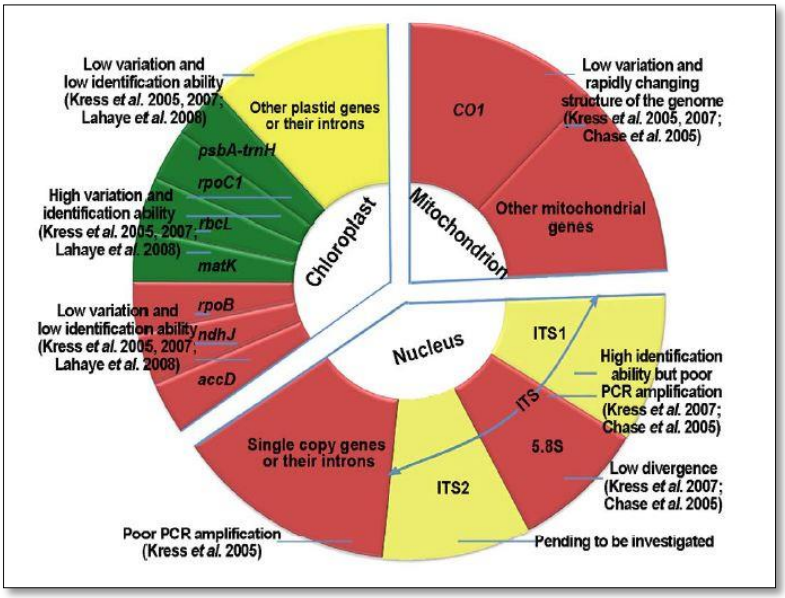


Figure 2.4 Genes from three genomes in plants that are potential candidate barcodes marker (Source: Chen et al., 2010)**.**

DNA sequencing techniques (Maxam & Gilbert, 1977; Sanger et al., 1977) were followed by the advent of molecular markers such as restriction fragment length polymorphism {RFLP; (Botstein et al., 1980)}, polymerase chain reaction {PCR; (Mullis et al., 1986}; and a derivative there of, termed random amplified polymorphic DNA {RAPD; (Williams et al, 1990}. These DNA markers along with isozyme and morphologic markers allowed extensive genome mapping in several species (Paterson et al., 1991).

Genomic authentication is based on biomarker development. Biomarkers are molecular phenotypes in the form of a pattern, like a fingerprint that is used to recognize an individual or a group of individuals in different taxonomic categories, such as a species. The development of biomarkers entails a discovery phase followed by validation and then by verification. Biomarkers are of different kinds and different approaches and

methods. Biomarkers are increasingly developed as new and efficient methods and techniques are being invented (Baum, 2013).

In animal CO1 region is established as universal DNA barcode marker, but in plants several mitochondrial, chloroplast and nuclear potential barcode markers were under studies (Figure 2.4). CBOL Plant group (2009) recommends chloroplast region *rbcL* and *matK* as core DNA barcode markers in Plant groups.

2.4.3.1.1 Why Employ Molecular Genetic Markers to access Diversity

In the molecular era, standard approaches to estimating association and phylogeny necessarily entailed comparisons of phenotypic data from morphology, physiology, behavior, or other organismal characteristics amenable to observation. Molecular ecologists and evolutionists also employ the comparative method, but the comparisons now include direct or indirect genotypic information from nucleic acid and protein sequences (Avise, 2004).

a. Molecular Data are Generic

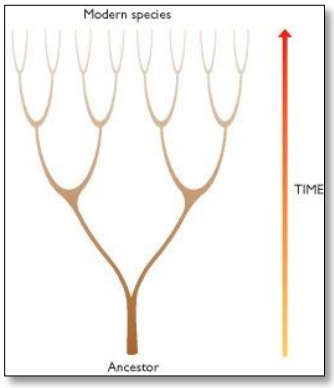


Figure 2.5 Showing evolutionary tree of life (Source: Brown, 2002).

Molecular data provide genetic information. Because phylogeny is "the core of heredity," only genetic traits are genealogically informative. Molecular assays not only provide details features of DNA (or sometimes, their protein products) but also variable character whose particular genetic bases and modes transmission can be precisely specified viz. an ancestral species is at the bottom of the 'trunk' of the tree. As time passes, new species evolve from earlier ones so the tree repeatedly branches until we reach the present time, when there are many species descended from the ancestor as shown in Figure 2.5. In Plants, phenotypic or development plasticity has been recognized as informative source of phenotypic variation (Avise, 2004). Molecular data provide information on many different characters in a DNA sequences, for example, every nucleotide position is a character with four character states, A, C, G and T. Molecular characters states are unambiguous, easily recognizable and cannot be confused with another. Some morphological characters, such as those based on the shape of a structure, can be less easy to distinguish because of overlaps between different character states. Molecular data are easily converted to numerical form and hence are amenable to mathematical and statistical analysis (Brown, 2002).

b. Molecular Methods Open the Entire Biological World for Genetic Scrutiny

Most genetic studies were confined to a small handful of species that could be reared and crossed in the laboratory or garden; bacteria such as *Escherichia coli* and their

phages, Mendel's pea plants (*Pisum sativum*), corn (*Zea mays*), fruit flies (*Drosophila* species) and mice (*Mus musculus*). From hereditary pattern across generations, the genetic bases of particular morphological or physiological traits in these species have deduced. However, such analyses could hardly be expected to capture full richness of diversity among the multitudinous genes within these study organisms, much less to embrace the broader sense of genetic across the Earth's biota. In contrast, molecular assays, can provide direct physical evidence on essentially any DNA sequences or protein, and they can be useful to the genetics of any or all creatures (Avise, 2004).

c. Molecular Methods Access an Unlimited Pool of Genetic Diversity

Each genome contains storehouse of information, not only encoding the ribonucleic acids and proteins that are working machinery of cellular life, but also retaining within its nucleotide sequence a detailed historical record of phylogenetic links to others form of life (Avise, 2004). Molecular genetic methods are accessing the conservation and use of plant genetic resources. Molecular techniques have been applied in the analysis of specific genes, understanding gene action, generate genetic maps and assist in the development of gene transfer technologies. These techniques have also helpful for studying the phylogeny and species evolution, understanding of the distribution and extent of genetic variation within and between species (Mondini et al., 2009).

d. Molecular data differentiate the Homology from Analogy

The challenges of phylogenetic are to distinguish the component of biological similarity that is due to descent from a common ancestor (homology) from that due to evolutionary convergence from different ancestors (analogy). Molecular character states, such as duplications, deletions, or rearrangement of DNA are rare or unique events likely to be single (monophyletic) evolutionary origin. These data provides unique features for exploring or distinguishing the phylogenetic history (Avise, 2004).

2.4.3.2 Molecular Identification of Commercialized Medicinal Plants

There are an estimated 300000 plant species in the world (IUCN, 2012) but relatively few of these can be identified based on traditional identification method (Hebert et al., 2003; Chase & Fay, 2009)**.** Accurate classification and identification of this large number of species remains a significant challenge even for specialist taxonomists. The emergence of DNA barcoding has had a positive impact on biodiversity classification and identification (Gregory, 2005). DNA barcoding has attracted much attention of taxonomists. Global DNA barcoding was initially regarded as big 'science' programme (Gregory, 2005). However, the *cytochrome c oxidase 1* (*CO1*) sequence, which has been developed as a universal barcode in animals, does not discriminate most plants because of much slower mutation rate (Kress et al., 2005). Although many studies have searching for a universal plant barcode, none of the available loci work across all species (Chase & Fay, 2009; Chen et al., 2010). The Consortium for the Barcode of Life-Plant Working Group (CBOL) recommended the two-locus combination of *matK*+ *rbcL* as the best plant barcode with a discriminatory efficiency of only 72% (CBOL Plant Working Group, 2009). Taxonomists have suggested that a multi-locus method may be necessary to discriminate plant species (Hebert et al., 2004; Kress & Erickson, 2007). The concept of 'super-barcode' is derived from the comparison between plastid genome sequences from a target group of taxa is presented as an effective option that might be widely applicable to plant identification studies. Specific barcode may provide new perspectives

in the search for rapid and accurate methods for species discrimination, especially for closely related species (Li et al., 2015). There are various techniques which are used for the molecular identification of medicinal plants. These techniques are described below:

2.4.3.2.1 Traditional Based Identification

High value or Medicinal plants are extremely popular all over the world as botanical supplements, herbal medicine and sources of lead compounds for pharmaceutical development. For the protection of consumers and illegal trades, authentication of medicinal plants is a critical issue. Ideally, authentication should occur from the harvesting of the plant material to the final product. Unfortunately, there is no single or superior method to assure 100% authentication of entire process, but the goal can be achieved through the variety of different methodologies. Macroscopic and microscopic examinations can be used as rapid and inexpensive identification techniques. Chemical analysis is by far the best method for the detection of contaminants and can be an excellent method for plant identification. Each of these methodologies has limitations and more analytical methods are needed to assist in the authentication process. Molecular techniques is superior than previously mentioned techniques that can be very useful for authentication of high value or medicinal plants (Techen et al., 2004).

To understand the application and limitations of molecular techniques for the authentication of source materials of plants, it is important to first understand Polymerase Chain Reaction (PCR) of particular loci and factor influence on it, as a number of possible identification are based on this technology. PCR is an *in vitro* method to amplify a specific portion of an organism's DNA. Under special conditions, DNA fragments with nucleotide sequence up to 40,000 base pairs in length can be amplified (Soltis & Soltis, 1998). Standardization of authentication protocol for plants using PCR based technology is a difficult due to testing of multiple locations as well as some technical problems such as several types of thermal DNA polymerase supplied from many suppliers using different buffer formulations, preparing PCR reactions of ingredients, quality of DNA template of plant materials (Penner et al., 1993; Techen et al., 2004). Nowadays standardization is easier due to the availability of commercial kits or commercial polymerase mixtures (containing DNA polymerase, buffer, $MgCl_2$ and dNTPs).

Molecular plant systematics determines the evolutionary relationship of one species to another. One of the most reliable methods is by comparing the DNA sequences of the same genetic region of each species to another. Often a nuclear and chloroplast region is used for this comparison. There are number of methodologies that can be used to characterize the plants without prior knowledge of DNA sequences, e.g. Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLPs), Random Amplified Polymorphic DNA (RAPDs), Inter-Simple Sequence Repeat (ISSRs) Simple Sequence Repeats (SSRs) and others technique.

1. Restriction Fragment Length Polymorphism (RFLP)

RFLP is the hybridization-based molecular marker. RFLP is useful genetic markers for various genetic and plant breeding applications including molecular systematics (Dowling et al., 1990). The variation among individual or organisms in the length of particular restriction fragment depends on difference of the genome sequences of individuals. The individual genome sequences may differ due to the resulting point

mutation, insertion/deletion, translocation, inversion and duplication. Some of the differences in DNA sequences at the restriction sites can result in the gain, loss, or relocation of a restriction site. These genotypic changes can all be recognized by the altered mobility of restriction fragments on agarose gel electrophoresis (Botstein et al, 1980; Semagn et al., 2006)

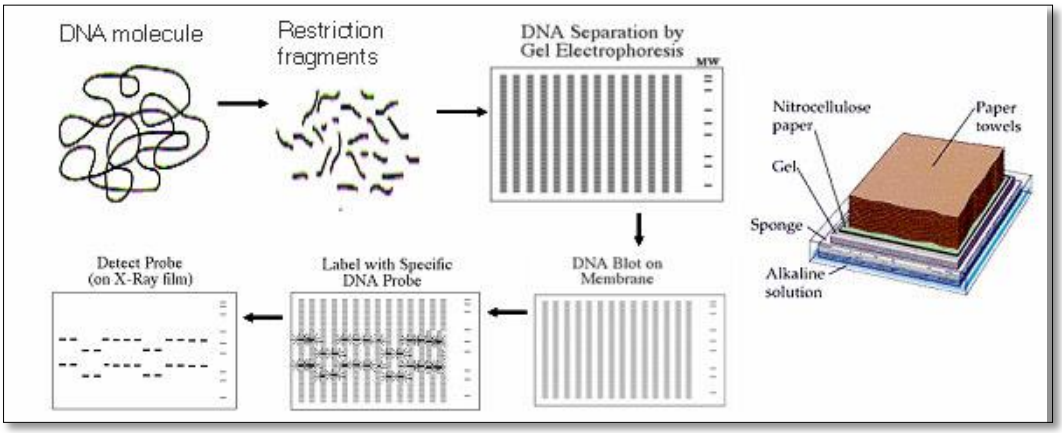


Figure 2.6 Showing an overview of the different steps of RFLP (Semagn et al., 2006).

RFLP analysis can be applied to rare and endangered or dried samples such as herbarium specimens (Mizukami et al., 1993). Some of the limitations of RFLP are; it requires the presence of high quantity and quality of DNA, the level of polymorphism is low, and few loci are detected per assay, it is time consuming, laborious, and expensive and it usually requires radioactively labeled probes (Figure 2.6) (Semagn et al., 2006).

2. Amplified Fragment Length Polymorphism (AFLP)

AFLP analysis is a very reliable and robust technology. AFLP techniques based on the combination of RFLP and PCR-based technology (Techen et al., 2004; Mondini et al., 2009). AFLP analysis generates a multiple banding pattern, but the PCR protocol is more stable compared to RAPD analysis. In AFLPs, the DNA is first digested with restriction enzymes, which cut the DNA at specific sequence sites as shown in Figure 2.7. Most AFLP fragments represent to the unique positions on the genome and hence can be considered as markers in genetic and physical mapping. This technique can be used to distinguish closely related individuals at the subspecies level (Althoff et al., 2007) and can also map genes. The AFLP polymorphism depend upon the mutations of the restriction site, mutations of sequences flanking the restriction site, and complementary to the extension of the selective primers, insertions, duplications or deletions inside amplification fragments. These mutations can cause the appearance/disappearance of a fragment or the modification (increase or decrease) of an amplified-restricted fragment (Mondini et al., 2009).

The advantages of AFLP analysis are to exploring the plant genetic diversity (Mba & Tohme, 2005), and determine genetic variability within and natural populations AFLP markers have been extensively used for phylogenetic analysis and determining the genetic diversity for conservation of endangered plant species (Li et al., 2008). Some limitations regarding AFLP analyses are laborious, nuclear and chloroplast sequences sometimes fail to reveal variability when plant species are closely related (Després et al., 2003); unacceptable as an across the board method for the authentication of medicinal

plants, even though the results are very reproducible between laboratories (Techen et al., 2004)

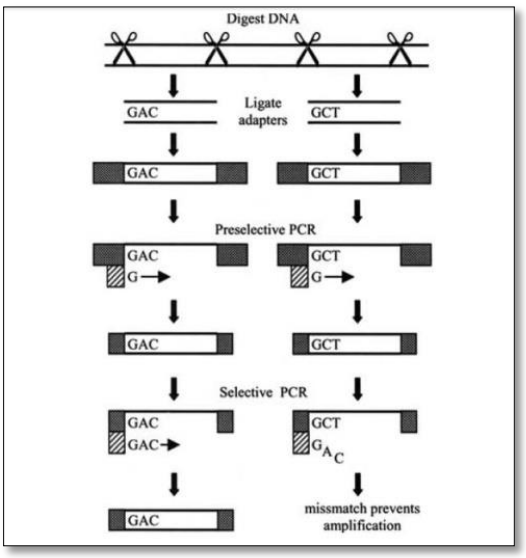


Figure 2.7 Showing diagrammatic representation of the AFLP principle (Techen et al., 2004).

3. Random Amplified Polymorphic DNA (RAPD)

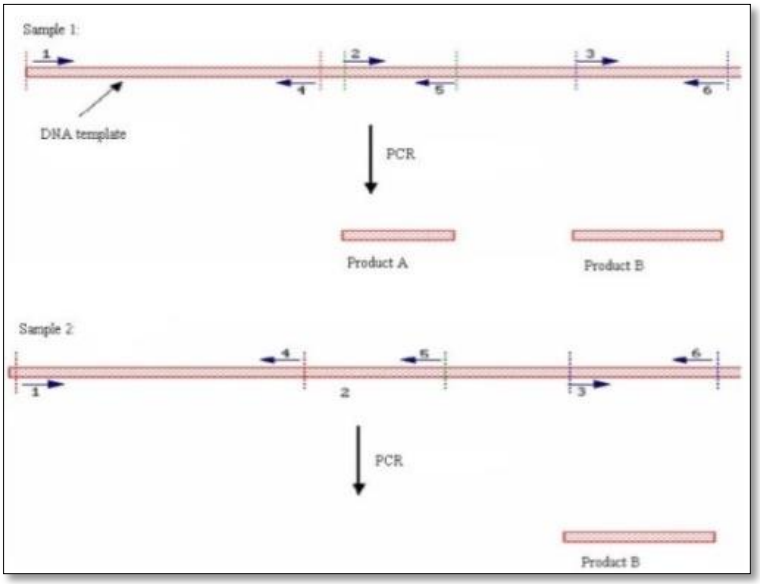


Figure 2.8 Schematic diagram of principle of RAPD analysis (Semagn et al., 2006).

RAPDs were the first PCR-based molecular markers to be employed in genetic variation analyses (Welsh & McClelland, 1990; Williams et al, 1990). RAPD is based on the amplification of genomic DNA with single primers of arbitrary nucleotide sequence (short primers; decamers). These primers detect polymorphisms in the absence of specific nucleotide sequence information and the polymorphisms makes the RAPD as genetic markers and can be used to construct genetic maps.

In RAPD analysis, series of different primers are used that can generate genetic fingerprint (different DNA fragments) of given plant or species specific bands, through random amplification of genomic DNA. The use of short primers is necessary to increase the probability that, although the sequences are random, they are able to find

homologous sequences suitable for annealing. DNA polymorphisms are then produced by "rearrangements or deletions at or between oligonucleotide primer binding sites in the genome" as shown in Figure 2.8 (Williams et al, 1990).

The RAPD markers have been used for detecting genomic variations within and between varieties. RAPD has also been used for estimation of genetic diversity in various endangered plant species (Zheng et al., 2007). The major drawback of this technique is that the profiling is dependent on reaction conditions and PCR parameters which can vary between laboratories. A slight variation in temperature can cause significant alterations in the banding pattern. Due to above limitations, which make it unacceptable as an authentication method of medicinal plants (Techen et al., 2004).

4. Microsatellites

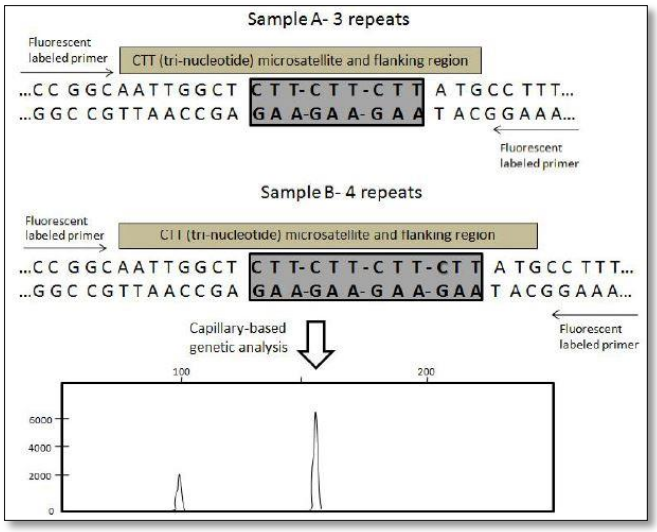


Figure 2.9 Diagrammatic representation of CTT (tri-nucleotide) microsatellite and flanking region and the detection method. Arrows indicate positions of PCR primers. Two length variants are shown (A and B) (Arif et al., 2010).

Microsatellites or Simple Sequence Repeats (SSR) are polymorphic loci present in DNA, contains multiple copies of repeated sequences. The repeated sequences consist of sequences of repetitions, comprising basic short motifs generally between 2 and 6 base pairs length (Mondini et al., 2009). The basic principle of microsatellite is as shown in Figure 2.9 (Arif et al., 2010). Microsatellites are highly popular genetic markers as they possess: co-dominant inheritance, high abundance, enormous extent of allelic diversity, ease of assessing SSR size variation through PCR with pairs of flanking primers and high reproducibility. Microsatellites have proved to be versatile molecular markers, particularly for population analysis. With the abundance of PCR technology, primers that flank microsatellite loci are simple and quick to use. Microsatellites developed for particular species can often be applied to closely related species, but the percentage of loci that successfully amplify may decrease with increasing genetic distance (Jarne & Lagoda, 1996). Microsatellite technique has been also used to establish conservation strategy of endangered plants (Arif et al., 2010).

Major molecular markers based on assessment of variability generated by microsatellites sequences are: STMSs (Sequence Tagged Microsatellite Site), SSLPs (Simple Sequence Length Polymorphism), SNPs (Single Nucleotide Polymorphisms),

SCARs (Sequence Characterized Amplified Region) and CAPS (Cleaved Amplified Polymorphic Sequences) (Mondini et al., 2009).

2.4.3.2.2 Single-Locus DNA Barcodes (DNA barcoding based)

Traditional barcodes have been widely studied but still have some limitations as discussed previously. Some of these widely used single locus barcodes are described below.

1. Ribulose 1, 5-bisphosphate carboxylase/oxygenase (*rbcL)*

Ribulose 1,5-bisphosphate carboxylase/oxygenase (rubisco) is the first enzyme of C3 cycle in plants. It is the most abundant and most important protein on the planet and central to the global carbon cycle (Chase, et al., 1993). The *rbcL* gene is located on chloroplast genome as a single copy gene and has an enormous phylogenetic utility. The *rbcL* gene is ~1428 bp long and is universal to all plants (except in some parasites). It is very convenient to study, easy to align and its secondary structure is known and present in many copies with less insertions and deletions. The *rbcL* gene encodes the large subunit of rubisco, while the small subunit is encoded by *rbcS* gene in nucleus. The *rbcL* gene was one of the first plant genes to be sequenced (Zurawski & Clegg, 1984) and is still among the most frequently sequenced segments of plant DNA. This gene has been used widely in systematic studies of land plants, angiosperms in particular (Chase, et al., 1993).

rbcL is easy to amplify, sequence, and align in most land plants and provides a useful backbone to the barcode dataset, despite it having only modest discriminatory power. Two-marker plastid barcodes gave better discrimination than single marker barcodes, but no other 2-marker or multi-marker plastid barcode gave appreciably greater species resolution than the *rbcL+matK* combination. The *rbcL* barcode consists of a 599 bp region at the 5'end of the gene, located at bp 1–599 (including primer sites) in the complete *Arabidopsis thaliana* plastid genome sequence (Hollingsworth et al., 2011). About 500 *rbcL* sequences were used to address phylogenetic relationships within angiosperms and secondarily among extant seed plants (Chase, et al., 1993).

However, *rbcL* sequences evolve slowly and this locus has by far the lowest divergence of plastid genes in flowering plants (Kress, et al., 2005). Consequently, it is not suitable at the species level due to its modest discriminatory power (Fazekas et al., 2008; Lahaye et al., 2008; CBOL Plant Working Group, 2009; Chen et al., 2010). The length of the gene can also be problematic as double-stranded sequencing of the entire gene sequence may require four primers. Despite these limitations, *rbcL* is still suggested as one of the best potential candidate plant barcodes based on the straightforward recovery of the gene sequence (CBOL Plant Working Group, 2009; Hollingsworth et al., 2011). Although *rbcL* by itself does not meet the desired attributes of a barcoding locus, it is accepted that *rbcL* in combination with various plastid or nuclear loci can make accurate identifications (Newmaster et al., 2006; Chase et al., 2007; Kress & Erickson, 2007; CBOL Plant Working Group, 2009; Hollingsworth et al., 2009).

2. Internal Transcribed Spacer (*ITS*)

The internal transcribed spacers of nuclear ribosomal DNA (nr*ITS*) is one of the most commonly used DNA markers in plant phylogenetic and DNA barcoding analyses, and it has been recommended as a core plant DNA Barcode (Cheng et al., 2016). Among these DNA regions, *ITS* (or a part of it, *ITS2*) is one of the most widely used DNA fragments in

plant molecular systematic at genus and species level showing high levels of interspecific and intraspecific divergence (Alvarez & Wendel, 2003; Yuan et al., 2015). *ITS* was first proposed as a barcode for flowering plants (Kress et al., 2005). The greater discriminatory power of *ITS* over plastid regions at low taxonomic level has been widely studied to it also being suggested as a plant barcode (Stoeckle, 2003; Kress et al., 2005), but lost popularity for some time due to concerns about the incomplete concerted evolution of multiple copies, different alleles from paternal and maternal parents, DNA contamination of different species (e.g. through symbiosis) and some technical problems. It was demonstrated that these imperfections did not cause large problems, and it was reproposed as a core barcode for seed plants (Hollingsworth et al., 2011; China Plant BOL Group et al., 2011). Methodological studies indicate that it is more effective or even necessary to use a combination of barcodes from both the biparentally inherited nuclear genome and the uniparentally inherited plastid genome for accurate identification of species, and *ITS* is so far the most promising candidate from the nuclear genome (Chase & Fay 2009; Fazekas et al. 2009; Roy et al. 2010).

Presenting a different point of view, the China Plant BOL Group, recently argued that when direct sequencing was possible, the *ITS* region should be incorporated into the core barcodes because *ITS* has higher discriminatory power than plastid barcodes (CBOL Plant Working Group, 2011). Despite showing the highest discriminatory power of existing candidate barcodes and greatest popularity in plant systematics, amplification and sequencing of *ITS* sometimes suffers from nonspecificity and low PCR and sequencing (Hollingsworth et al., 2011) due to primer related problems. The most popular *ITS* primer pairs, for example *ITS1*+*ITS4* (White et al., 1990), was originally designed for fungi. As fungi in many cases are symbiotic with plants in natural ecosystems, it is easy to obtain non-target amplicons for *ITS* from fungi when amplifying plant *ITS* fragments. The other major problems is that existing primers lack satisfactory universality for many plant groups, resulting low PCR and sequencing success rates.

To improve the taxonomic coverage of plants *ITS* primers while reducing coamplification and fungal contaminations, Recently, Cheng et al., (2016) designed new universal and plant specific whole *ITS* primers pairs (Figure 2.10) which shows higher universality and specificity than existing primer pairs. *ITS2* is recommended as a standard DNA barcode to identify medicinal plants. *ITS2* is considered as one of the candidate DNA barcodes because of its valuable characteristics, including the availability of conserved regions for designing universal primers, the ease of its amplification, and enough variability to distinguish even closely related species (Chen et al., 2010). Song et al., (2012) recently showed that the *ITS2* intra-genomic distances were markedly smaller than those of the intraspecific or interspecific variants in a wide range of plant families. However, Cheng et al., (2016) has reported *ITS* as the sole choice owing to the availability of universal primers. The existence of conservative 5.8s and 28s region between *ITS*1 and *ITS2* (Figure 2.11) makes the use of *ITS*1 and *ITS2* or entire *ITS* region very flexible and powerful even for highly degraded DNA templates.

Although the internal problems of incomplete concerted evolution of multiples copies of *ITS*, different alleles due to hybridization or polyploidization, and contamination of other plant cannot be thus be eliminated. Such problems can be solved by cloning and more recently using NGS technologies (Cheng et al., 2016).

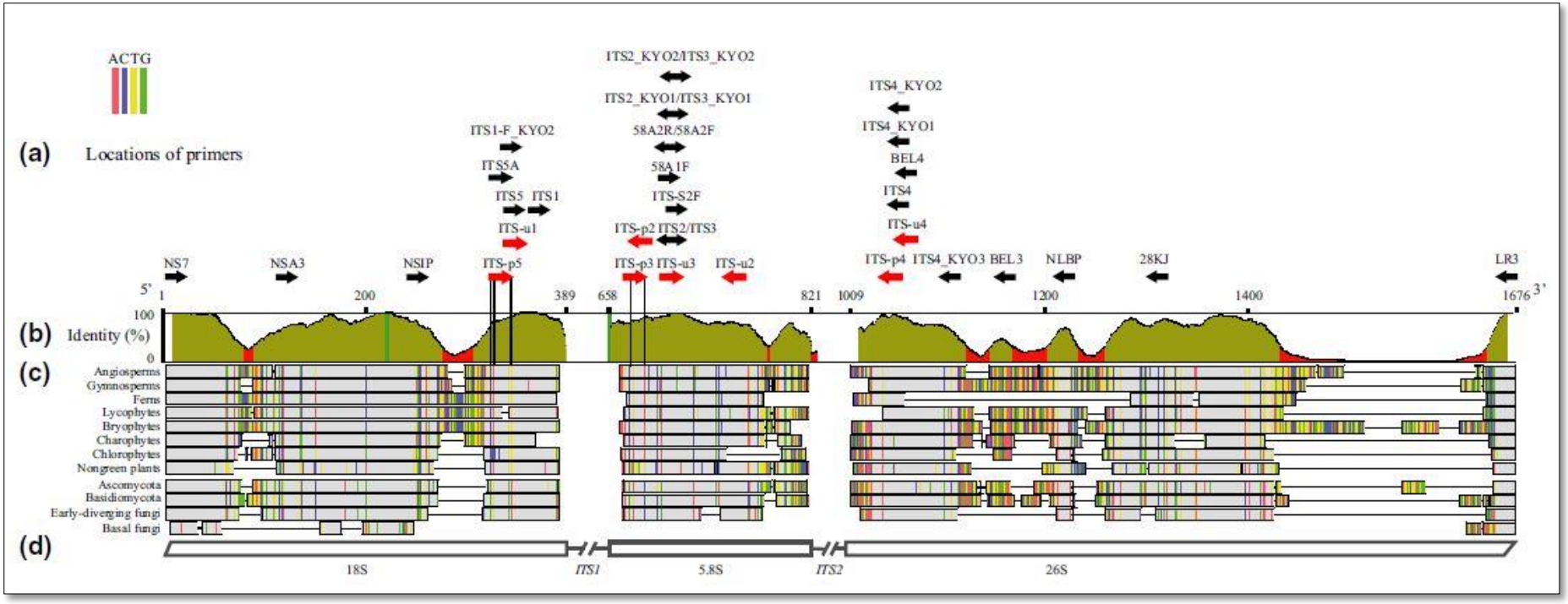


Figure 2.10 schematic diagram of whole *ITS* region. (a) Locations of forward (right-pointing arrows) and reverse (left-pointing arrows) primers of newly designed (red arrows) and common-used primers (black arrows). (b) Nucleotide identity in plants and fungi. (c) Conservative nucleotide regions in each plant or fungus group with variable positions highlighted. (d) Schematic diagram of the nuclear ribosomal RNA genes based on sequences from *Arabidopsis thaliana* (GQ380689 and X52320) (Cheng et al., 2016).

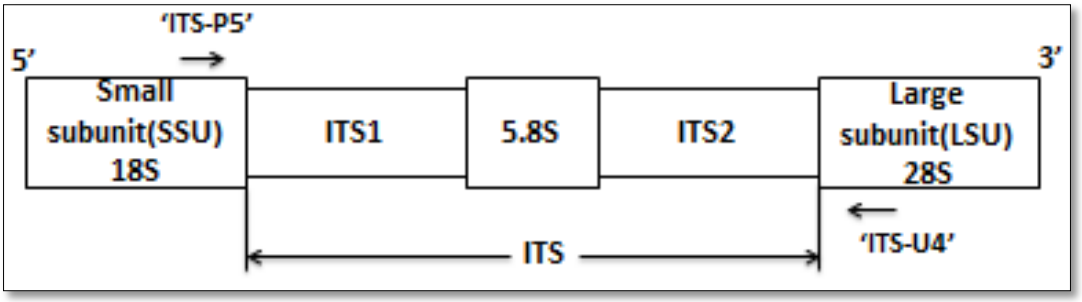


Figure 2.11 showing schematic diagram of the organization of the *ITS* region of the 18S-5.8S-28S nuclear DNA repeat. Arrows indicate approximate positions of primers for sequencing (redraw adapted from Cheng et al., 2016).

3. Maturase K (*matK)*

The *matK* (maturase) gene previously called *orfK,* is approximately 1500 base pairs (bp), located within the the intron of the chloroplast gene *trnK* (lysine tRNA), and encodes a maturase involved in splicing type II introns from RNA transcripts. Studies have shown the usefulness of this gene in estimating intergeneric or interspecific relationships among flowering plants. The *matK* gene is known to have relatively high rates of substitution compared with other genes used in grass systematics, possesses high proportions of transversion mutations, and the 3' end of its coding region has proven quite useful for constructing phylogenies. Sequences from noncoding regions of the chloroplast genome is often used in systematics because such regions tend to evolve relatively rapidly (Liang & Hilu, 1996; Palmer et al., 1988).

The *matK* barcode region consists of the 841 bp region at the center of gene, located between bp 205-1046 (including primer sites) in the complete *Arabidopsis thaliana* plastid genome sequence (Hollingsworth et al., 2011). *matK* is one of the most rapidly evolving coding sections of the plastid genome, high evolutionary rate, suitable length and obvious interspecific divergence as well as low transition/transversion rate (Liang & Hilu, 1996; Li et al., 2015) thus, has recommended as a locus for DNA barcoding by consortium for the Barcode of life (CBOL) plant Working group. Unfortunately, *matK* can difficult to amplify using existing primer sets – particularly in non-angiosperms. Amplification and sequencing of the matK barcoding regions is difficult due to high sequence variability in the primer binding sites (Hollingsworth et al., 2011). The CBOL Plant Working group (2009) revealed nearly 90% success rate in amplifying angiosperm DNA using single primer pair. However, the success rate was limited in gymnosperms (83%) and much worse in cryptogams (10%) even with multiple primer pairs. Different primer pairs were required in different taxonomic groups (Chase et al., 2007; Hollingsworth, 2008). Currently, there are three popular primer *matK* pairs available (Figure 2.12) to amplify approximately the same region of the gene; 390F and 1326R (Sun et al., 2001); XF and 5R (ford et al., 2009) and 1R_KIM and 3F_KIM (Hollingsworth et al., 2009). Kress et al., (2009) used these three primer pairs to amplify DNA barcodes from 296 shrub and tree species. These primer combinations showed amplification success in 85% and sequencing success in 69% of the species, proving that reliable amplification is possible across a range of plants, using several primer combinations. However, using more than one pair can be time consuming as well as costly and is often complex for large-scale projects. Heckenhauer et al., (2016) designed a set of cocktail universal primers that can be multiplexed in one PCR to amplify *matK* successfully in

angiosperms and expedite high-throughput, rapid, automated, and cost-effective species identification. This cocktail primer pairs enable 100% efficient PCR amplification and sequencing of the *matK* barcode region.

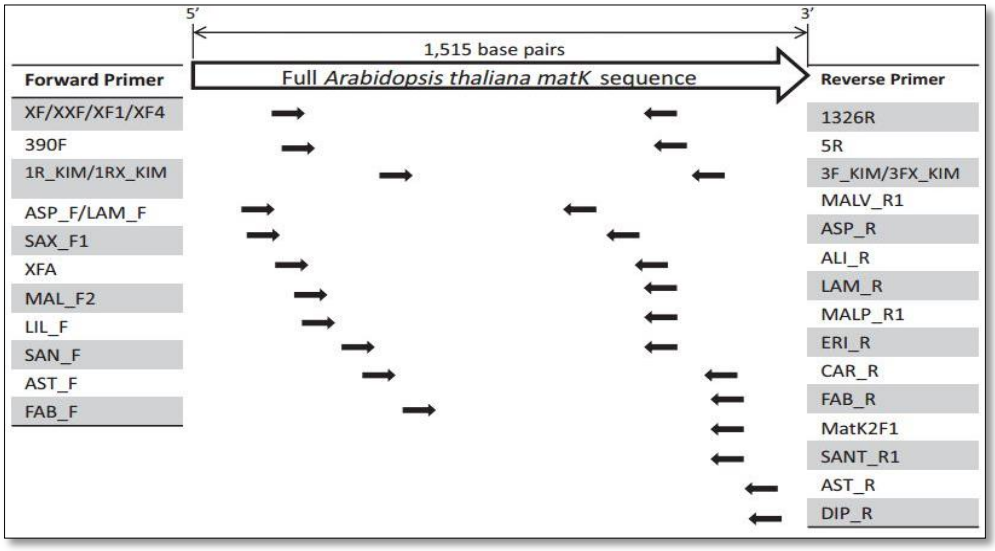


Figure 2.12 Showing map and orientation of primers with *matK* of A. thaliana as template (Source: Dunning & Savolainen, 2010)

A further challenge is the differentiation rates in different taxonomic groups. *matK* can discriminate more than 90% of species in the Orchidaceae (Kress & Erickson, 2007) but less than 49% in nutmeg family (Newmaster et al., 2008). Fazekas et al., (2008) attempted the identification of 90 species from 32 genera using the *matK* barcode but only achieved a success rate 56%. These finding demonstrate that the matK barcode alone is not a suitable universal barcode.

4. *trnH–psbA*

The *psbA-trnH* intergenic region is among the most variable regions in the angiosperm chloroplast genome. It is a popular tool for plant population genetics and species level phylogenetics (Štorchová & Olson, 2007), and is currently the most widely used plastid barcode (Xiwen Li et al., 2015). The presence of highly conserved coding sequences on both sides make the design of universal primers feasible (Shaw et al., 2005), with a single primer pair likely to amplify nearly all angiosperms (Shaw et al., 2007).

The *psbA-trnH* intergenic region contains two parts that differ in their evolutionary conservation: 1) the psbA 3'UTR, which is important for posttranscriptional regulation of psbA gene expression, and thus is subject to selection pressure, and 2) the psbA-trnH non-transcribed intergenic spacer, which appears to lack function because of its high variability across angiosperms as shown in the Figure 2.13. The psbA gene encodes the D1 reaction center protein of photosystem II (He et al. 1998). The role of chloroplast psbA UTRs in the regulation of gene expression has been investigated intensively for more than twenty years (Zurawski et al. 1982).The *psbA-trnH* spacer offers a unique opportunity for investigation of comparative patterns in evolution of chloroplast 3' UTR stem-loop structures. *psbA-trnH* spacer contains 200-500 bp length in most species, nucleotide variation in this region is very high. This could influence the potential for phylogenetic interference across highly divergent taxa (Štorchová & Olson, 2007). Štorchová & Olson (2007) suggested that 3' stem- loop region has frequent patterns

consistent with the independent evolution of a large inversion and different mutation hot spots in *psbA-trnH* spacer. This information is further useful for functional investigation and phylogenetic analysis.

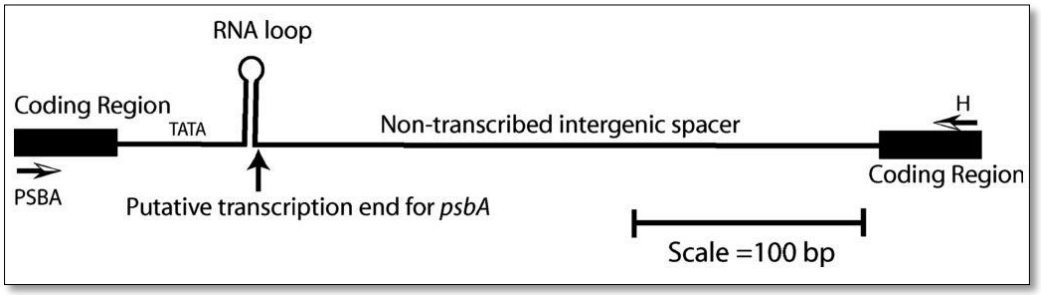


Figure 2.13 Schematic diagram represents the *psbA-trnH* intergenic region; PSBA and H region is universal primers for this region (source: Štorchová & Olson, 2007).

Alignment of the *trnH-psbA* spacer can be highly ambiguous because of its complicated molecular evolution, considerable length variation (Chang et al., 2006), and high rates of insertion/deletion in larger families of angiosperms (Chase et al., 2007). Furthermore, due to the presence of duplicated loci and a pseudogene, the trnH-psbA sequence is much longer [>1000 base pairs (bp)] in some conifers and monocots (Chase et al., 2007; Hollingsworth et al., 2009) while it is exceedingly short, less than 300 bp, in other groups (Kress et al., 2005) and shorter than 100 bp in bryophytes (Stech & Quandt, 2014). One of the key problems associated with the use of *trnH-psbA* as a standard barcode is the frequent inversions in some plant lineages, which may lead to large overestimates of genetic divergence and to incorrect phylogenetic assignment (Whitlock, Hale, & Groff, 2010). Additionally, because of the premature termination of sequencing reads caused by mononucleotide repeats, longer *trnH-psbA* regions can be difficult to retrieve without taxon specific internal sequencing primers designed to obtain high quality bi-directional sequences (Devey et al., 2009). Shorter *trnH-psbA* spacers may not have adequate sequence variation for species discrimination, such as in the genera Solidago (Kress et al., 2005). As a consequence, Kress et al. (2005) and Chase et al. (2007), respectively, proposed that *trnH-psbA* can be used in two-locus or three-locus barcode combination to provide adequate resolution.

5. Other widely used plastid barcodes

At present, DNA barcoding technology relies heavily on chloroplast loci because of their relatively low evolutionary rate compared with nuclear loci (Li et al., 2015). Beyond the candidate barcodes described above, there are many other widely used plastid barcoding markers, such as *ndhj*, accD, *rpoB, rpoC1, atpF-atpH, psbK-psbI, ycf5* and *trnL-F* (Figure 2.14). The *atpF-atpH* and *psbK-psbI* intergenic spacers were proposed as plant barcoding regions at the second international Barcode of Life Conference. These two markers have not been widely used in plant systematic and phylogeographic studies and as a result there is a paucity of data on their performance. In the study by the CBOL Plant Working Group, *psbK-psbI* showed high levels of discriminatory power, but lower sequence quality and universality, whereas *atpF-atpH* showed relatively modest discriminatory power, intermediate sequence quality and universality (Hollingsworth et al., 2011).

The *trnL* intron and the intergenic spacer between *trnL* and *trnF* have been widely used in plant systematics and phylogeography since the early 1990s. This frequent use is

attributable to the early publication of a robust set of primers that allow routine recovery (Taberlet et al., 1991). The regions are generally simple to sequence, although mononucleotide repeats (Hollingsworth et al., 2011) can impact on sequencing reads in some taxa. Some studies have noted that other regions of the plastid genome may be more variable and informative for plant phylogenetic studies (Shaw et al., 2007), but a major strength of the trnL intron for species identification is the presence of a small stem-loop structure within the intron, the P6 loop (Pierre Taberlet et al., 2007). P6 has conserved priming sites flanking a variable loop of ca 10–143 bp. This very short 'mini-barcode' has proved very useful to ecologists studying highly degraded DNAs and using next generation sequencing technologies to assess the diversity of complex environmental samples (Valentini et al., 2009). This '*trnL* approach' of ecological barcoding has developed somewhat in parallel to the major international barcoding consortia of the International Barcode of Life Project (iBOL) and the Consortium for the Barcode of Life (CBOL).

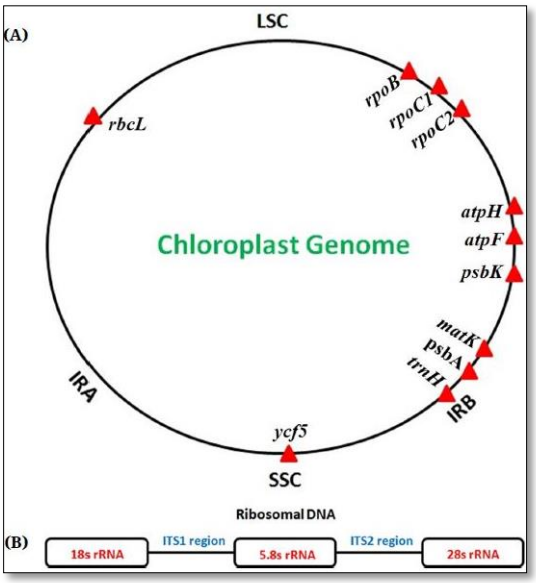


Figure 2.14 Schematic representation of Plastid (A) and nuclear (B) markers commonly used in Plant DNA barcoding (Saddhe & Kumar, 2017).

Dong et al., (2015) reported *ycf1* were most variable loci that were better than existing plastid candidate barcodes and can serve as a barcode of land plants. Molecular evolution of cp-genome sequences shows both lineage-specific and nonrandom spatial pattern of substitution (Dong et al., 2012; Li et al., 2015), for example (Dong et al., 2012) demonstrated that the region of *ycf1* located in the invert repeats (IRb) region is conservative while two regions located in the small single copy (SSC) region are extremely variable. Such substitution patterns in chloroplast genomes indicate complex processes of mutation that are asymmetric, and lack independence between sequence positions. Thus, the patterns of substitution are not well described by currently used substitution models, particularly with respect to deeper phylogenetic divergences (Lockhart & Steel, 2005). Chloroplast sequence evolution can be inconsistent across lineages, and phylogenetic incongruence between different chloroplast gene loci is possible (Lockhart & Steel, 2005; Dong et al., 2012). It is difficult, and even unlikely, to identify such regions in all taxa. Therefore, it can be problematic to find an ideal universal barcode applicable at various taxonomic levels.

2.4.3.2.3 Multi-Locus DNA Barcodes

Despite extensive efforts to identify a universal plant barcode comparable to *CO1* in animals, the task has proved difficult due to the lack of adequate variation within single loci (Kress et al., 2005; Newmaster et al., 2006; Chase et al., 2007; Kress & Erickson, 2007; Fazekas et al., 2008; Lahaye et al., 2008). Many researchers have suggested that a multi-locus method will be required to obtain adequate species discrimination (Hebert et al., 2004; Kress & Erickson, 2007; Erickson et al., 2008; Lahaye et al., 2008; CBOL Plant Working Group, 2009; Chase & Fay, 2009). Various combinations of plastid loci have been proposed including *rbcL+trnH-psbA* (Kress & Erickson, 2007), *matK+rpoC1+rpoB* or rpoC1+matK+trnH-psbA (Chase et al., 2007), *matK+atpF-atpH+psbK-psbI or matK+atpF-atpH+ trnH-psbA* (Pennisi, 2007). These combined barcodes exhibit higher species discrimination than single-locus approaches. Different research groups have tested different combinations using different taxa while attempting to select a universal barcode, however universal agreement is yet to be reached. Fazekas et al. (2008) compared these barcode combinations using the same large-scale taxonomic samples, but none could identify more than 70% of tested species.

Table 2.5 List of most commonly used DNA barcodes markers used in the identification of various plant groups with reference cited.

Plant Division	DNA Barcode	References
Algae	*COI, rbcL, tufA,* and *ITS*	(Saunders & McDevit, 2012; Hadi et al., 2016)
Bryophytes	*rbcL, trnH-psbA, ITS, nad5, trnL-trnF, trnSrps4/rps4, atpB-rbcL, rpoC1,* and *trnG*	(Liu et al., 2010; Stech & Quandt, 2010)
Pteridophytes	rbcL, matK, trnH-psbA, trnL-trnF, and *ITS2*	(Ebihara et al., 2010; Li et al., 2011; Gu et al., 2013)
Gymnosperms	*rbcL, matK* and *ITS*	(Sass et al., 2007; Li et al., 2011)
Angiosperms	*rbcL , matK, trnH-psbA, ITS, trnL-trnF, rpoB, rpoC1, YCF5 , atpF-atpH,* and *psbK-psbI*	(Chase et al., 2007; Pennisi, 2007; CBOL Plant Working Group, 2009; Chen et al., 2010; Dong et al., 2015;)

The CBOL Plant Working Group recently recommended *matK+rbcL* as the universal barcode combination due to the straightforward recovery of the *rbcL* region and the discriminatory power of the *matK* sequence (CBOL Plant Working Group, 2009) Although the choice of *rbcL+matK* offered slightly higher identification efficiency than other combinations, the *rbcL+matK* barcode still failed to meet the original goal of a universal DNA barcode. Firstly, the combination of *rbcL+matK* cannot avoid the low PCR efficiency of matK and secondly, the success of *rbcL+matK* in discriminating plants is typically lower than that of *CO1* in animals. At the Fourth International Barcode of Life Conference (http://www.dnabarcodes2011.org/) the option of a three locus barcode (*matK* + *rbcL* + *psbA-trnH)* versus a two locus barcode was discussed. The two-locus barcode was referred to avoid the increased costs of sequencing three loci rather than two in very large sample sets, and to prevent further delays in implementing a standard barcode for land plants. The barcode combination *rbcL* + *matK* were the preferred choice. There are various barcode marker used to identify various plant groups Table 2.5. Zuo et al., (2011) analyzed 12 genomic regions of 95 ginseng samples, representing

all the species in the genus. They demonstrated that the combination of *psbA-trnH* and *ITS* was sufficient for the identification of all species and species clusters in the genus. Chen et al., (2010) analyzed >6600 plant samples belonging to 4800 species from 753 distinct genera using the genomic regions *psbA-trnH, matK, rbcL, rpoC1, ycf5*, *ITS*, and *ITS2*. Their data suggested that the *psbA-trnH* and *ITS2* represents the most suitable region for DNA barcoding applications.

2.4.4 Sequencing Techniques

DNA sequencing technique was advent in early 1977s by Maxam & Gilbert (1977); Sanger et al., (1977). Frederick Sanger (1988) devoted his scientific life to the determination of primary sequences of the nucleotide bases A, T, G and C present in the target of molecule of DNA. Earlier, there were two types of DNA sequencing techniques such as chemical cleavage protocol and chain terminator protocol (Maxam & Gilbert 1977; Sanger et al., 1977). Frederick Sanger, (1988) Dye terminator sequencing is the standard method in automated sequencing analysis. The dye-terminator sequencing method, along with automated high-throughput DNA sequence analyzers, is now being used for the vast majority of sequencing work. The basic technique related with dye terminator sequencing and phylogenetic analysis as shown in. Dye-terminator sequencing utilizes labeling of the chain terminator ddNTPs, which allows sequencing in a single reaction, rather than four reactions as in the previously used labeled-primer method. In dye-terminator sequencing, the four dideoxynucleotide chain terminators are labeled with fluorescent dyes, each with a different wavelength of fluorescence emission. The main advantages of this technique are its robustness, automation and high accuracy (>98%). On the other hand, the limitations of this technique include dye effects due to differences in the incorporation of the dye-labeled chain terminators into the DNA fragment. Such incorporation of dye can result in unequal peak heights and shapes in the electronic DNA sequence trace chromatogram after capillary electrophoresis. Another drawback is its inability to handle long sequences; however, it can reliably sequence up to approximately 1000 bp long DNA fragments in a single reaction. The advent of new generation sequencers with solid state chemistry has significantly overcome these problems.

DNA sequencing technologies have undergone impressive improvements with the recent emergence of next-generation sequencing (NGS) platforms (Glenn, 2011). These new platforms can provide billions of sequence reads in a single experiment, which corresponds to an improvement of at least five orders of magnitude when compared to traditional Sanger sequencing using capillary electrophoresis. Such a dramatic leap in sequencing capacity has the potential to revolutionize many areas of scientific inquiry. However, the impact of the new technologies on biodiversity research is more difficult to assess. Another, major challenge will be to develop new bioinformatic pipelines especially designed for exploiting such massive amounts of sequence data in the most efficient way for DNA-based species identification (Taberlet et al., 2012).

2.4.5 BLAST Method of Similarity Searches

The Basic Local Alignment Search Tool (BLAST) finds identities of local similarity between query sequences and reference sequences and maximizes the segment pair scores. The program compares nucleotide or protein sequences to sequences databases and calculates the statistical significant of matches of the query (Altschul et al, 1990, 1997).

For local pairwise alignments, the best approach to defining statistical significance is to estimate an expect value (E value), which is closely related to a probability value (p value) (Pevsner, 2017). MegaBLAST is currently the default module program NCBI-BLASTN, which is a local nucleotide database search tool from the NCBI BLAST software distribution (Chen et al., 2015). MegaBLAST is an optimized program for the rapid alignment of very large DNA queries (Zhang et al., 2000). MegaBLAST can specify the percent identify threshold viz. 99%, 90% or 80% identity with corresponding match and mismatch scores. For examples, 95% to 99% identity, a match score of +1 and mismatch of -3 is applied and 85% or 90% identity the mismatch score is instead set to -2. Non-affine gapping parameters are used: the gap opening penalty is 0 (MegaBLAST has alignments with more gaps but with the benefit of enhanced speed), and the gap extension penalty is based on the selected match and mismatch scores. Discontiguous MegaBLAST is a related algorithm at NCBI that is designed to align more distantly related genomic sequences. It is useful for comparing relatively divergent sequences (Pevsner, 2017).

2.4.6 Sequence Alignment

Sequence alignment is the first step in comparative sequence analysis, identifying homologous nucleotide (or amino acid) positions among collection of different sequences and widely used tool in molecular biology (Kumar et al., 2004). Aligned sequences are usually interpreted as sharing their homologous region. If sequence has no corresponding residue which mean of an insertion or deletion event occured, the position is displayed as '–' or another symbol and is called a 'gap'(Katoh & Toh, 2008). Pairwise alignment is the process of comparing two sequences to achieve maximal levels of identity. Pairwise alignment is to assess the degree of similarity and the possibility of homology between two molecules. Pairwise alignment is useful for identifying unknown samples, mutation during evolution, evolutionary relationship (Pevsner, 2017). Multiple sequence alignment (MSA) is also widely used tool in various types of comparative studies of biological sequences. MSA is used in phylogenetic reconstruction, conserved region detection, structure prediction and proteins and other analysis (Katoh & Toh, 2008). Algorithms "CLUSTAL W" is used for MSA that improves sensitivity and progress through sequence weighting, position specific gap penalties and weight matrix choice (Thompson et al., 1994).

2.4.7 Phylogenetic Study

Before the advent of DNA sequencing technologies, phylogenetic trees were used almost exclusively to describe relationships among species in systematics and taxonomy. Today, phylogenies are used in almost every branch of biology (Yang & Rannala, 2012). Besides representing the relationships among species on the tree of life, phylogenies are used to describe relationships between paralogues in a gene family (Zhu et al., 2000), histories of populations (Edwards, 2009), to interpret modern and ancient individual genomes (Li & Durbin, 2011); and to reconstruct ancestral genomes (Zhu et al., 2000).

A phylogenetic tree or an evolutionary tree is a graph that shows the evolutionary relationships among various biological species based on their genetic distances (Mahapatro et al., 2012). A phylogenetic tree estimates the relationships among taxa (or sequences) and their hypothetical common ancestors. Phylogenetic trees are built from molecular data: DNA or protein sequences (Nei & Kumar, 2000). The purpose of most

molecular phylogenetic trees was to estimate the relationships among the species represented by those sequences, but today the purposes have expanded to include understanding the relationships among the sequences themselves without regard to the host species, inferring the functions of genes that have not been studied experimentally (Hall et al., 2009). Genetic structuring of plant populations is strongly influenced by both common ancestry and current pattern of interpopulation genetic exchange. The interaction of these two forces is particularly confounding and hence interesting in plants. This complexity of plant genetic structures is due to part to a diversity of reproductive ecologies affecting genetic exchange and the fact that reproductive barriers are often weak between otherwise morphologically well-defined species.

Table 2.6 List of methods for inferring phylogenetic trees (Horiike, 2016)

Method	Group	Algorithm	Software
UPGMA	distance-matrix	Clustering for the shortest evolutionary distance	MEGA 7
Neighbor-joining	distance-matrix	Clustering for minimum total branch length	PHYLIP, Clustal X, MEGA 7
Maximum parsimony	character-based	Searching tree with minimum total number of character-state changes	PHYLIP, MEGA 7
Maximum likelihood	character-based	Searching tree with maximum likelihood	PHYLIP, PhyML, RAxML, FastTree, MEGA 7, TOPALi v2
Bayesian	character-based	Searching tree with maximum posterior probability	MrBayes, TOPALi v2

Since the discovery of molecular clock (constancy of molecular evolution rate), the revolutionary phylogenetic studies have been used to clarify the evolutionary branching patterns up to the current species revolutionarily changed. In contrast to the conventional palaeontological studies that is mainly based on morphological differences between species, molecular evolutionary (sometimes RNA or protein) of the homologous, genes among the different species, as the basis to determine the evolutionary relations. On the basis of the constancy of the base substitution (mutation) rate during the evolution, the ‘evolutionary distances’ between the species are calculated from these DNA sequence differences, and from them, by using relevant methods, we can reconstruct the evolutionary phylogenetic (Ren et al., 1995). In population genetics, the development of 'coalescent theory'(The process of joining ancestral lineages when the genealogical relationships of a random sample of sequences from a modern population are traced back) (Kingman, 1982)and widespread of availability of gene sequences for multiple individuals from the same species have promoted the development of genealogy-based inference methods, which have revolutionized modern computational populations genetics (Yang & Rannala, 2012).

A phylogenetic is a tree containing nodes that are connected by branches. Each branch represents the persistence of a genetic lineage through time, and each node represents the birth of a new lineage as shown in Figure 2.15. If the tree represents the relationship among a group of species, then the nodes represent speciation events. Phylogeny reconstruction methods are either distance-based or character-based. In distance matrix methods, the distance between every pair of sequences is calculated, and the resulting

distance matrix is used for tree reconstruction. For instance, neighbour joining (Saitou & Nei, 1987) applies a cluster algorithm to the distance matrix to arrive at a fully resolved phylogeny. Character-based methods include maximum parsimony, maximum likelihood and Bayesian inference methods. These approaches simultaneously compare all sequences in the alignment, considering one character (a site in the alignment) at a time to calculate a score for each tree. The 'tree score' is the minimum number of changes for maximum parsimony, the log-likelihood value for maximum likelihood and the posterior probability for Bayesian inference. In theory, the tree with the best score should be identified by comparing all possible trees (Yang & Rannala, 2012).

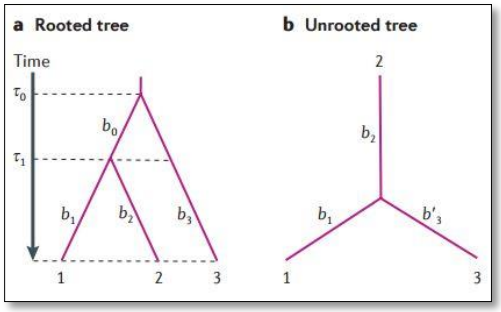


Figure 2.15 Showing the rooted and unrooted tree of two speciation events that occurred at time τ0 and τ1. The branch lengths (b0, b1, b2 and b3) are typically expressed in unITS of expected number of substitutions per site and measure the amount of evolution along the branches (Yang & Rannala, 2012).

Building a phylogenetic tree requires four distinct steps: (Step 1) identify and acquire a set of homologous DNA or protein sequences, (Step 2) align those sequences, (Step 3) estimate a tree from the aligned sequences, and (Step 4) present that tree in such a way as to clearly convey the relevant information to others (Hall, 2013). Various computational methods have been developed so far to reconstruct the phylogenetic tree from the sequence differences. Some of the commonly used methods for the phylogeny reconstruction are summarized in the Table 2.6.

2.4.7.1 Distance Matrix Method

Pairwise sequence distances are based on the Markov chain model of nucleotide substitution and most commonly used models as shown in Figure 2.16. The JC69 model (Jukes et al., 1969) supposes an equal rate of substitution between any two nucleotides, whereas the K80 model (Kimura, 1980) supposes different rates for transitions and transversions. Both models estimate equal frequencies of the four nucleotides. There are three methods for evaluating the distances such as least square method, minimum evolution and neighbor joining. The least square calculated pairwise sequence method that (Cavalli-Sforza & Edwards, 1967) reduces the measures of the difference between the calculated distances (d_{ij}) in the distance matrix and the expected distances ($\hat{d}_{ij}$) on the tree (i.e. the sum of branch lengths on the tree linking the two species i and j):

$$Q = \sum_{i=1}^{s} \sum_{i=1}^{s} (\hat{d}_{ij} - d_{ij})^2$$

This is the same least squares method used in statistics for fitting a straight line y=a+bx to a scatter plot. Optimizing branch lengths (or $d\hat{}_{ij}$) leads to the score Q for the given tree, and the tree with the smallest score is the least squares estimate of the true tree.

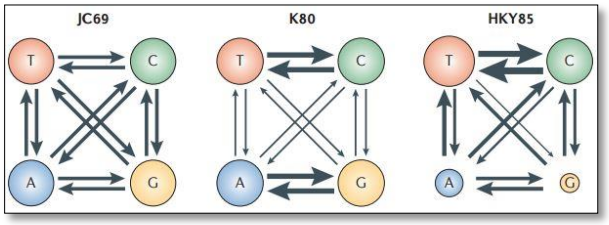


Figure 2.16 Markov models of nucleotide substitution. The thickness of the arrows indicates the substitution rates of the four nucleotides (T, C, A and G), and the sizes of the circles represent the nucleotide frequencies when the substitution process is in equilibrium. Note that both JC69 and K80 predict equal proportions of the four nucleotides (Yang & Rannala, 2012).

The minimum evolution method (Rzhetsky & Nei, 1992; Desper & Gascuel, 2002) based on the tree length (which is the sum of branch lengths) instead of Q for tree selection, even though the branch lengths can still be predicted using the least squares criterion. Under the minimum evolution criterion, shorter trees are more likely to be correct than longer trees.

The most commonly employed distance method is neighbor joining (Saitou & Nei, 1987). Neighbour joining is a heuristic approach that does not guarantee to find the perfect result, but under normal conditions has a very high probability to do so. It has a very good computational efficiency, making it well suited for large datasets. This is a cluster algorithm and starting with a star tree and selecting a pair of taxa to join together (based on the taxon distances), until a fully resolved tree is obtained. The taxa to be joined are selected in order to minimize an estimate of tree length. The two joined taxa (for example, species 1 and 2 as shown in Figure 2.17) are then represented by their ancestor (for example, node y in Figure 2.17 and the number of taxa that are joined to the root (node x in Figure 2.17) is reduced by one (Figure 2.17). The distance matrix is revised by Gascuel & Steel (2006), replacing the two original taxa with the joined taxa and based on the neighbour- joining updating formula. Neighbour-joining method is efficient, easily accessible in the program MEGA (Tamura et al., 2011).

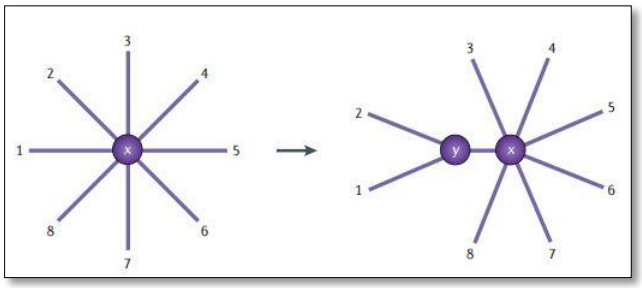


Figure 2.17 The neighbour joining algorithm. The neighbour joining algorithm is a divisive cluster algorithm. It starts from a star tree: two nodes are then joined together on this tree (in this example, nodes 1 and 2), reducing the number of nodes at the root (node x) by one. The process is repeated until a fully resolved tree is generated (Yang & Rannala, 2012).

The advantage of distance methods (especially of neighbour joining) is their computational efficiency. The cluster algorithm is fast because it does not need to compare as many trees under an optimized criterion as maximum parsimony and maximum likelihood. Neighbour joining is applicable for analyzing large data sets that have low levels of sequence divergence. Realistic substitution model is important to calculate the pairwise distances. But, distance methods can perform poorly for very divergent sequences because large distances involve large sampling errors, and most distance methods (such as Neighbour joining), do not differentiate for the high variances of large distance estimates. Distance methods are also sensitive to gaps in the sequence alignment (Bruno et al., 2000).

2.4.7.2 Maximum Parsimony

Parsimony was originally developed for use in analyzing discrete morphological characters. During the late 1970s, it was applied for molecular data as well. The maximum parsimony (MP) method of phylogenetic tree reconstruction using nucleotide sequence data conducts a site-by-site analysis. For each tree topology, it determines the minimum number of nucleotide changes (substitutions) that are required to explain the observed site pattern. The numbers of changes are summed over sites to give a parsimony score for each tree topology, and the topology having the smallest total number of changes is taken as the estimate of the phylogeny, which is known as the most parsimonious tree (Yang, 1996). Alternatively, the maximum parsimony method reduces the number of changes on a phylogenetic tree by assigning characters states to internal nodes on the tree. The character (or site) length is the minimum number of changes required for that site, whereas the tree score is the sum of character lengths overall the sites (Yang & Rannala, 2012). Parsimony appears to involve very stringent predictions concerning the process of sequence evolution, such as constancy of substitution rates between nucleotides, constancy of rates across nucleotide sites, and equal branch lengths in the tree. For practical data analysis, the requirement of equal branch lengths means similar substitution rates among lineages (the existence of an approximate molecular clock), relatively long interior branches, and also few species in the data. However, a small amount of evolution is not useful for the analysis of this method (Yang, 1996).

An algorithm for finding the minimum number of changes on a binary tree (and for reconstructing the ancestral states to achieve the minimum) was developed by Fitch, (1971) and (Hartigan, 1973). PAUP (Swofford, 2000), MEGA (Tamura et al., 2011) and TNT (Goloboff et al., 2008) are commonly used parsimony programs (Table 2.7). The use of parsimony is still common: not because it is believed to be assumption-free, but because it often produces reasonable results, simplicity, easy to understand mathematical analysis and is computationally effective (Yang & Rannala, 2012). The failure of parsimony to correct for multiple substitutions at the same site makes it suffer from a problem known as long-branch attraction (Felsenstein, 1978).

2.4.7.3 Maximum likelihood

Maximum likelihood was developed by Fisher in the 1920s as a statistical methodology for estimating unknown parameters in a model. The likelihood function explains the probability of the data given the parameters but considered as a function of the parameters with the data observed and fixed. It represents all information in the data about the parameters. The maximum likelihood estimates (MLEs) of parameters are the

parameter values that maximize the likelihood. MLEs are commonly used numerical value for using iterative optimization algorithms. The MLEs have desirable asymptotic (large-sample) properties: they are unbiased, consistent (they approach the true values) and efficient (they have the smallest variance among unbiased estimates) (Yang & Rannala, 2012).

The first algorithm for maximum likelihood analysis of DNA sequence data was developed by Felsenstein (1981). Later, the algorithm was modified for amino acid sequence data by Kishino et al., (1990). This method is useful for finding a tree which maximizes the probability of observing the data for a specific substitution model (Nei and Kumar, 2000; Hall, 2008). Phylogeny is inferred based on likelihood values and the topology with the highest likelihood is selected (Nei and Kumar, 2000). The Maximum Likelihood function calculates branch lengths by regarding every possible nucleotide/amino acid for each interior node (hypothetical ancestor) to maximize the likelihood for each observed site. The probability of a topology is the sum of the probabilities calculated for each site. For easier computational handling, the probability (likelihood) is expressed as log likelihood (Hall, 2008). The topology with the highest probability (log likelihood) is selected as the Maximum Likelihood tree (Bromham, 2008)

The method is now most commonly used due to increasing computer power, software implementations and to the development of increasingly realistic models of sequence evolution. Two optimization steps are involved in maximum likelihood tree estimation: optimization of branch lengths to calculate the tree score for each candidate tree and a search in the tree space for the maximum likelihood tree. From a statistical point of view, the tree (topology) is a model instead of a parameter, whereas branch lengths on the given tree and substitution parameters are parameters in the model. Maximum likelihood tree assumption is thus equivalent to comparing many statistical models, each with the same number of parameters. The attractive asymptotic properties of MLEs mentioned above apply to parameter estimation when the true tree is given but not to the maximum likelihood tree (Yang, 1996, 2006).

Early algorithm for finding maximum likelihood include PHYLIP (Felsenstein, 2005), MOLPHY (Adachi & Hasegawa, 1996) and PAUP* 4.0 (Swofford, 2000) . Modern implementations, such as PhyML (Guindon et al., 2003), RAxML (Stamatakis, 2006) and GARLI (Zwickl, 2006) (Table 2.7), are not only computationally much faster but are also more effective in finding trees with high likelihood scores. The recent inclusion of maximum likelihood in MEGA v.5 (Tamura et al., 2011) has made the method more accessible to biologists who are not experienced computer users.

Maximum likelihood is superior to distance or parsimony methods in case of sequence evolution understanding. The likelihood ratio test can be used to examine the fit of evolutionary models (Goldman, 1993) and to test interesting biological hypotheses, such as the molecular clock (Felsenstein, 1981) and Darwinian selection affecting protein evolution. The main drawback of maximum likelihood is that the likelihood calculation and, in particular, tree search under the likelihood criterion is computationally demanding. Another drawback is that the method has potentially poor statistical properties if the model is mis-specified (Yang & Rannala, 2012).

2.4.7.4 Bayesian methods

Bayesian inference of phylogeny reconstruction was first introduced in the late 1990s (Rannala & Yang, 1996; 1997; Li et al., 2000) Bayesian inference is similar to the Maximum Likelihood method in that both methods examine and calculate the likelihood of possible trees. However, unlike Maximum Likelihood, Bayesian methods sample trees from the tree space and do not calculate the likelihood of all possible branch lengths per tree (Bromham, 2008). Bayesian inference is a general methodology of statistical inference. This model is considered to be random variables with statistical distributions, whereas in maximum likelihood they are unknown fixed constants. Before the analysis of the data, parameters are assigned a prior distribution, which is combined with the data (or likelihood) to generate the posterior distribution. All inferences concerning the parameters are then based on the posterior distribution. In the past two decades, Bayesian inference has gained popularity thanks to advances in computational methods, especially Markov chain Monte Carlo algorithms (MCMC algorithms) (Larget & Simon, 1999; Yang & Rannala, 2012). A more recent Bayesian implementation in the program BEAST (Drummond et al., 2006) (Table 2.7) uses the so-called relaxed-clock models to infer rooted trees even though the model allows substitution rates to vary across lineages. Both likelihood and Bayesian methods use the likelihood function and thus share many statistical properties, such as consistency and efficiency (Yang & Rannala, 2012).

Table 2.7 Functionalities of a few commonly used phylogenetic programs

Name	Brief description	Link	References
Bayesian evolutionary Analysis sampling trees (BEAST)	A Bayesian MCMC program for inferring rooted trees under the clock or relaxed clock models. It can be used to analyse nucleotide and amino acid sequences, as well as morphological data. A suite of programs, such as Tracer and FigTree, are also provided to diagnose, summarize and visualize results.	http://beast.bio.ed.ac.uk	(Drummond & Rambaut, 2007)
Genetic algorithm for rapid likelihood inference (GARLI)	A program that uses genetic algorithms to search for maximum likelihood trees. It includes the GTR + Γ model and special cases and can analyses nucleotide, amino acid and codon sequences. A parallel version is also available.	http://code.google.com/p/garli	(Zwickl, 2006)
Hypothesis testing using phylogenies (HYPHY)	A maximum likelihood program for fitting models of molecular evolution. It implements a high-level language that the user can use to specify models and to set up likelihood ratio tests	http://www.hyphy.org	(Pond & Muse, 2005)
Molecular evolutionary genetic analysis (MEGA)	A Windows-based program with a full graphical user interface that can be run under Mac OSX or Linux using Windows emulators. It includes distance, parsimony and likelihood methods of phylogeny reconstruction, although its strength lies in the distance methods. It incorporates the alignment program ClustalW and can retrieve data from GenBank	http://www.megasoftware.net	(Tamura et al.,2011)
MrBayes	A Bayesian MCMC program for phylogenetic inference. It includes all of the models of nucleotide, amino acid and codon substitution developed for likelihood analysis	http://mrbayes.nt	(Huelsenbeck & Ronquist, 2001)
Phylogenetic analysis By maximum likelihood (PAML)	A collection of programs for estimating parameters and testing hypotheses using likelihood. It is mostly used for tests of positive selection, ancestral reconstruction and molecular clock dating. It is not appropriate for tree searches	http://abacus.gene.ucl.ac.uk/software	(Yang, 2007)
Phylogenetic analysis Using parsimony* and other methods (PAUP* 4.0)	PAUP* 4.0 is still a beta version (at the time of writing). It implements parsimony, distance and likelihood methods of phylogeny reconstruction	http://paup.phylosolutions.com	
PHYLIP	A package of programs for phylogenetic inference by distance, parsimony and likelihood methods	http://evolution.genetics.washington.edu/phylip.html	
PhyML	A fast program for searching for the maximum likelihood trees using nucleotide or protein sequence data	http://www.atgc-montpellier.fr/phyml/binaries.php	(Guindon et al., 2003)

Name	Brief description	Link	References
RAxML	A fast program for searching for the maximum likelihood trees under the GTR model using nucleotide or amino acid sequences. The parallel versions are particularly powerful		(Stamatakis, 2006)
Tree analysis using new technology (TNT)	A fast parsimony program intended for very large data sets	http://www.zmuc.dk/public/phylogeny/TNT	(Goloboff et al., 2008)

Note: all programs can run on Windows, Mac OSX and UNIX or Linux platforms. GTR, general time reversible; MCMC, Markov chain Monte Carlo. All the program listed http://evolution.genetics.washington.edu/phylip/software.html (Yang & Rannala, 2012).

2.5 Status of DNA barcoding of Plants

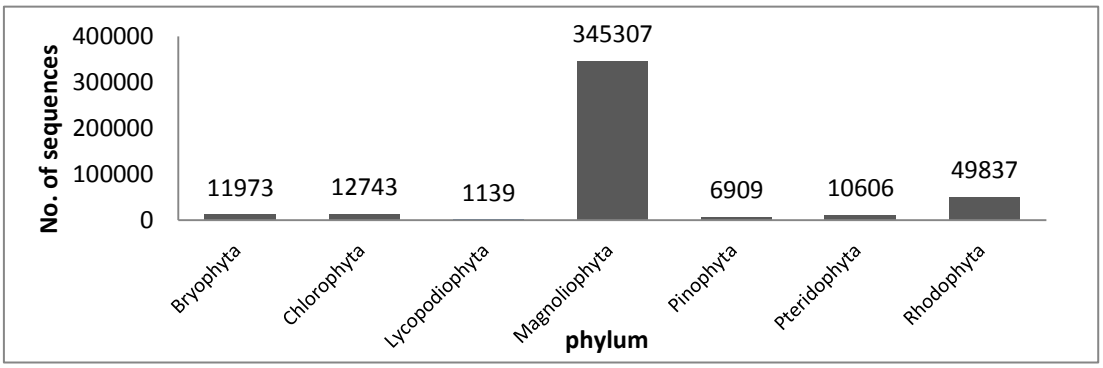


Figure 2.18 Number of sequences of each phylum of plants present in the BOLD database (Source: data accessed 9th Jun 2018, www.barcodinglife.org; Ratnasingham & Hebert, 2007).

Barcode of Life project was proposed to promote DNA barcoding as a global standard for sequence-based identification of eukaryotes. In 2004, this project was formally initiated by the establishment of the Consortium for the Barcode of Life (CBOL), The CBOL is an international initiative devoted to developing DNA barcoding as a global standard for the identification of biological species. CBOL has more than 130 Member Organizations from more than 40 countries, aims to develop a standard protocol for DNA barcoding and to construct a comprehensive DNA barcode library. Recently, the Barcode of Life project transformed to the International Barcode of Life project (iBOL; International Barcode of Life 2010). The iBOL is an international collaboration of 28 countries that aims to establish an automated identification system based on a DNA barcode library of all eukaryotes. The iBOL will also address the development of technologies, including new or improved protocols, informatics, equipment, DNA extraction methods and faster information systems. The Barcode of Life Data Systems (BOLD) is the official informatics workbench for the Barcode of Life project (Ratnasingham & Hebert, 2007), developed by the Canadian Center for DNA Barcoding (CCDB). BOLD provides a data repository, publication DNA barcodes, an identification support system, acquisition, data analysis and web services for other system developers. BOLD is freely available to any researcher via the Internet, although registration is required to create private databases and/or access restricted data. To identify unknown samples, researchers simply search for their sequenced barcode regions on the BOLD website. Importantly, the BOLD system is open to the public. Some other important websites and databases are also available for accessing resources of DNA barcoding of plants viz. http://www.ibol.org, http://biodiversitygenomics.net, https://www.ncbi.nlm.nih.gov, http://www.eol.org, http://ww.ccdb.ca, http://www.gbif.org/, http://www.ggi.si.edu/, http://www.barcodeofwildlife.org/.

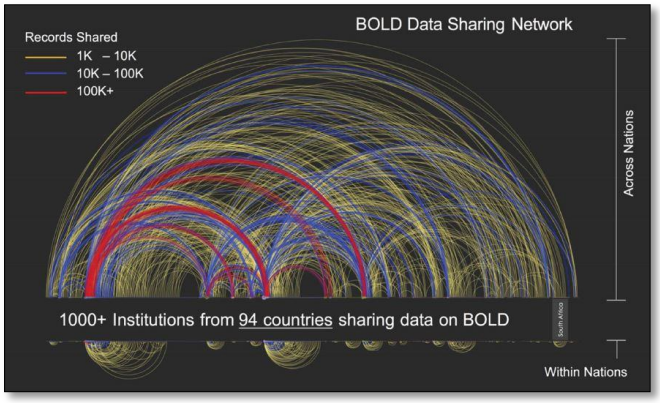


Figure 2.19 Network diagram of data shared between institutions on the Barcode of Life Data (BOLD) System workbench, where at least 1K records are shared between two institutions from August, 2015, with South Africa highlighted. Each line represents the sharing of data between two institutions (yellow, 1K to 10K records; green, 10K to 100K records; red, more than 100K records). The volume and incidence of data sharing across nations greatly exceeds sharing within nations, reflecting collaboration patterns in the International Barcode of Life community (Source: Adamowicz et al., 2017).

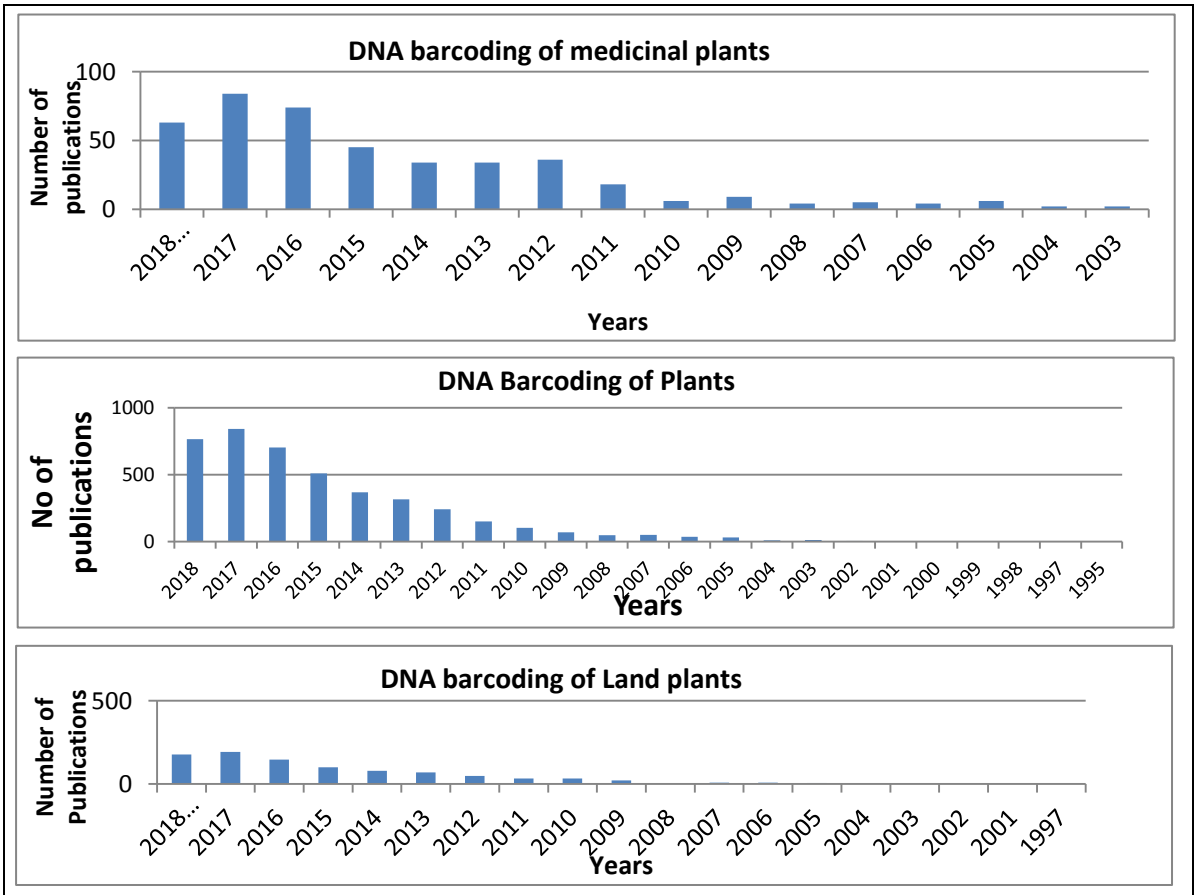


Figure 2.20 Number of publications of research articles and related articles till may 2018 based on keywords search on DNA Barcoding of Plants, land plants and medicinal plants (Database: https://www.sciencedirect.com/).

Plant barcodes are extra flexible for analysis phylogenetic signal, that allowing to higher-level systematics and understanding of macroevolutionary trends of diversification.

Moreover, combined regional-scale phylogenies data enable to research into phylogenetic community structure, presenting patterns of species coexistence and community assembly (Adamowicz et al., 2017). The DNA barcoding of plants commonly uses multiple DNA barcode markers, generally 2-4 plastid genes (Hollingsworth., 2016), while many researchers uses additional or complementary markers to increase the efficiency of species level-resolution.

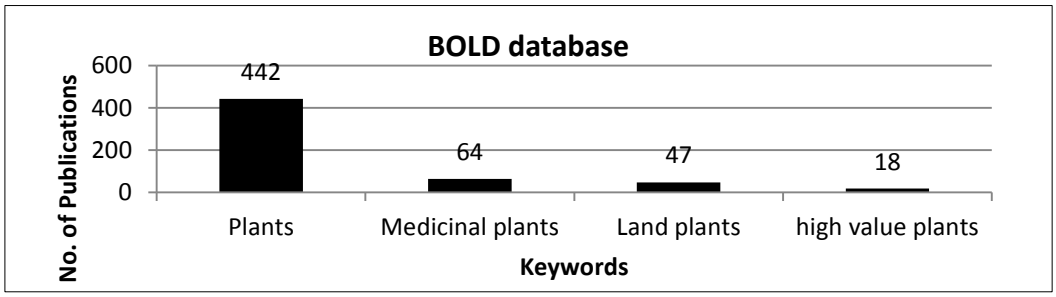


Figure 2.21 List of published article in BOLD database on different keywords searches (www.barcodinglife.org

In 2016, available barcodes represents 363,584 sequences from 50,039 species. The criteria of DNA barcoding (Srivastava et al., 2016), i.e. minimum sequence length of 500 bp and more than three organisms per single species have been convincing by 13,761 species. In January 2009, international Barcode of life (iBOL) under development by Canadian International Consortium Initiative (ICI), started with the target to collect barcodes for 5 million species in first 5 years. The researchers from 25 countries contributed this work (Sarwat & Yamdagni, 2016). Recently, 9th Jun 2018, total available barcode of plants in BOLD database represents 438,514 sequences of 67,064 species as shown in Figure 2.18 {(www.barcodinglife.org), Ratnasingham & Hebert, 2007}. Total 1000+ institutions from 94 countries were sharing data on BOLD as shown in Figure 2.19.

The survey of published research articles and its related papers on DNA barcoding of plants are increasing in trends. Based on Science direct web portal (https://www.sciencedirect.com/) keywords search for "DNA barcoding of plants" revealed a total of 4292 published research papers, followed by 938 on "DNA barcoding of land plant", and "426 on DNA barcoding of medicinal plants". These papers are published in different thematic area as well as different publication title as shown in Figure 2.20 and Figure 2.22. However, in BOLD database only 442 research articles or related articles have been published on keywords "DNA barcoding of plants" followed by "medicinal plants" (64), "land plants" (47) and "high value plants" (18) as shown in Figure 2.21. In BOLD database only highly specific areas or thematic area related articles have been submitted, that's why the number of article are less in comparison to Science direct web portal.

The ultimate goal of DNA barcoding is exploring all groups of living organism to make DNA reference library available for allowing to understand, conservation, and utilize the world's biodiversity (de Vere et al., 2015). Now most Researchers group focus on barcoding of mixed or environmental samples (Hollingsworth., 2016), market samples. The importance of barcoding and its related techniques is applied to conservation, validation commercial products such as medicinal products and to wildlife forensics (Adamowicz et al., 2017).

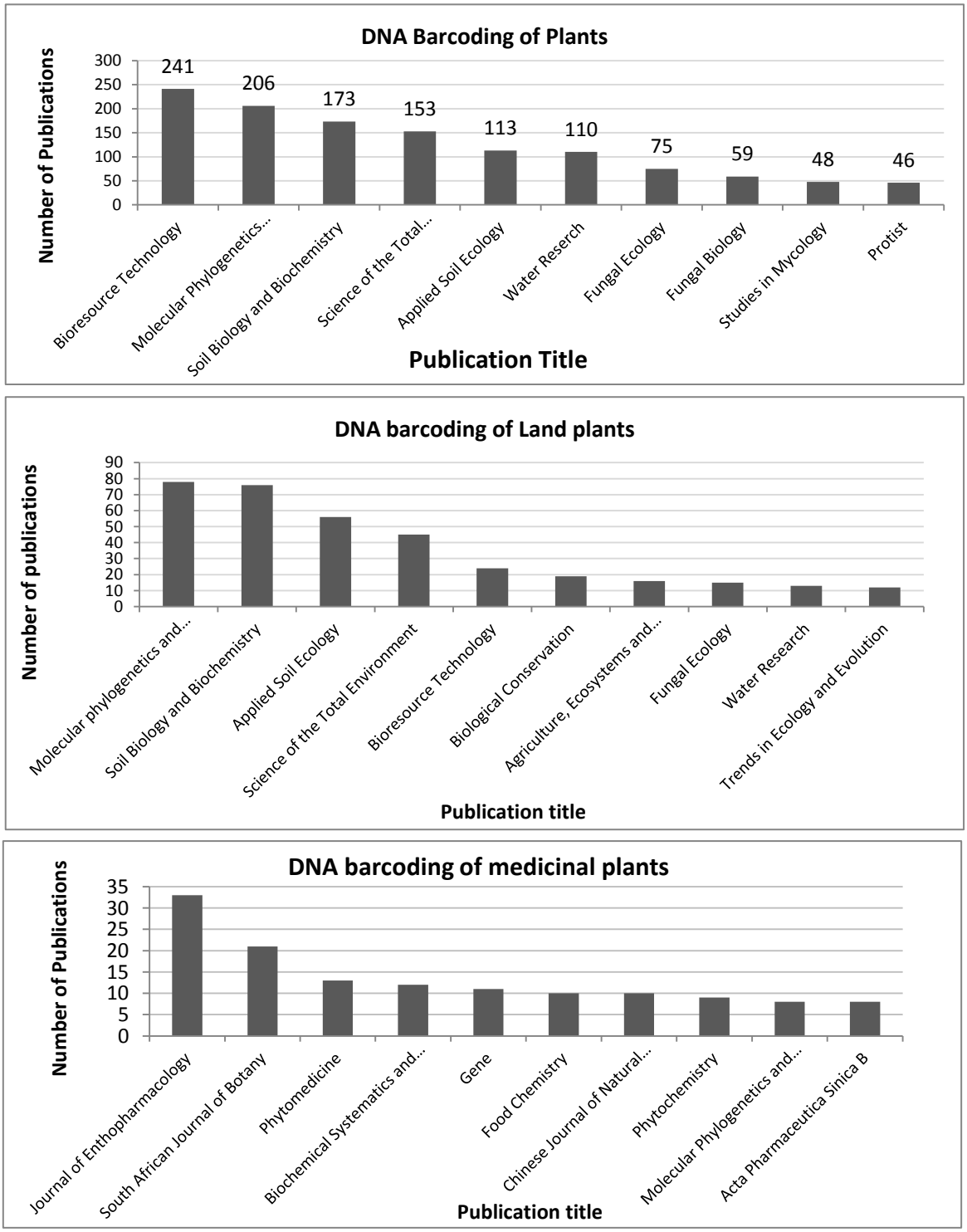


Figure 2.22 Number of research article or related articles in different publication title based on keywords DNA barcoding of plant, land plants and medicinal plants (Database: https://www.sciencedirect.com/).

Medicinal plants

A wide spread use of herbal medicine for the boosting of the health and treatment of diseases can be tracked back to prehistoric time throughout the many cultures and regions (Li et al., 2011). Even today, 70-80% of World population uses traditional

medicine for healthcare and therapeutic purposes (WHO, 2013). According to the World Health Organization (WHO), total international seasoning, the drug market is calculated as US$62 billion and is anticipated to grow to the extent of US$5 trillion by the year 2050. The demand of the medicinal plants at industrial level is higher due to global growth within the herbal industries. World herbal market is estimated 60 million yearly (Srivastava et al., 2016). However, adulteration is the major problems in the herbals and herbal products sold in the market. Therefore authentication and standardization is the essential factor to minimize the unfair trade.

The DNA barcoding project has a major objective to standardize medicinal herbs and formulation through the development of DNA reference library. Very few ethnopharmacological studies have attempted to evaluate their bio-efficacy (Uprety et al., 2010). Since, there is a complete lack of phyto-therapeutic evidence for many species. Phytochemical and pharmacological studies should be recommended for validation of properties attributed to these species (Shrestha & Dhillion, 2003). The improvement of quality control, standardization, scientific way of production and analysis of business products is necessary. The standardized mass production of herbal products tested scientifically would not only maintain the efficacy of the herbal but also offer competition to other medicines. DNA Barcoding would have key role to overcome such pertaining challenges.

China is presently leading the efforts on DNA barcoding of medicinal plants and has developed database of barcode. DNA barcode of some of the medicinal plants are given in the Table 2.8. However, majority of them are not the native species of Nepal. In the given list barcoded medicinal plants *rbcL* and *ITS2* barcode markers are the leading markers. In Nepal, few studies have carried out DNA barcoding of few Nepalese medicinal plants (*Taxus wallichiana, T. contorta, T. mairei, Swertia chirayita, Podophyllum hexandrum, Neopicrorhiza scrophulariiflora, Cordyceps sinensis, Citrus aurantifolia, Citrus spp.* etc) (Poudel et al., 2012; Neupane et al.,2017; Lamichane et al., 2014 unpublished thesis; Tamang et al., 2014 unpublished thesis, Gyanwali et al., 2014 unpublished thesis). Only some limited National and private organization of Nepal initiated DNA barcoding of high value plant and NTFPs including MAPs

Table 2.8 List of DNA barcode of some high value medicinal plants (Source: Srivastava et al., 2016)

S.N.	Plant Species	Family	DNA Barcode Region
1	*Acanthopanacis cortex*	Araliaceae	*ITS2*
2	*Aconitum* spp.	Ranunculaceae	*psbA-trnH*
3	*Acori tatarinowii rhizoma*	Araceae	*ITS2*
4	*Andrographis paniculata*	Acanthaceae	*rbcL, rpoC1, trnH-psbA*
5	*Angelica* spp.	Apiaceae	*matK, rbcL, ITS, ITS2, psbA-trnH*
6	*Arisaematis rhizoma*	Araceae	*matK, rbcL*
7	*Astragalus* spp.	Fabaceae	*matK, rbcL, ITS*
8	*Boerhavia* spp.	Nyctaginaceae	*ITS, ITS2*
9	*Brugmansia , Datura*	Solanaceae	*ITS2*
10	*Bupleuri radix*	Apiaceae	*ITS2*
11	*Butea superb*	Fabaceae	*matK*
12	*Cassia species*	Fabaceae	*rbcL*
13	*Centella asiatica*	Apiaceae	*ITS2, rpoC1, trnH-psbA*
14	*Citrus* spp.	Rutaceae	*Matk*
15	*Cleome* spp.	Cleomaceae	*matK, rbcL, ITS1*
16	*Clinacanthus nutans*	Acanthaceae	*ITS2, rpoC1, trnH-psbA*
17	*Cosmos caudatus*	Asteraceae	*ITS2, rpoC1, trnH-psbA*
18	*Cymbidium* spp.	Orchidaceae	*ITS2*
19	*Cynanchum auriculatum*	Apocynaceae	*trnL-F*

S.N.	Plant Species	Family	DNA Barcode Region
20	*Cynanchum wilfordii*	Asclepiadaceae	*trnL-F*
21	*Dalbergiae odoriferae*	Fabaceae	*ITS2*
22	*Dendrobium* spp.	Orchidaceae	*psbA-trnH*
23	*Dipsacus* spp.	Caprifoliaceae	*ITS*
24	*Gentiana* spp.	Gentianaceae	*matK + ITS*
25	*Ginseng genus*	Araliaceae	*matK, rbcL, ITS, psbA-trnH, rpoB, rpoC1, ITS2*
26	*Ginseng radix*	Araliaceae	*ITS2*
27	*Hedyotis diffusa*	Rubiaceae	*ITS*
28	*Hypericum spp.*	Hypericaceae	*ITS*
29	*Illicium*	Schisandraceae	*ITS, trnH-psbA*
71	*Inulae flos*	Compositae	*ITS2*
30	*Isatis indigotica*	Cruciferae	*ITS2*
31	*Justicia gendarussa*	Acanthaceae	*ITS2, rpoC1, trnH-psbA*
32	*Lonicera* spp.	Caprifoliaceae	*matK, rbcL, ITS, psbA-trnH,trnL-F*
33	*Lonicerae japonicae Flos*	Caprifoliaceae	*ITS2*
34	*Meconopsis* spp.	Papaveraceae	*ITS*
35	*Mentha aquatica L.*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
36	*Mentha spicata L.*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
37	*Mucuna collettii*	Fabaceae	*matK*
38	*Murraya koenigii*	Rutaceae	*ITS2, rpoC1, trnH-psbA*
40	*Ochradenus spp.*	Resedaceae	*ITS, rpoB, rpoC1*
39	*Ocimum basilicum L.*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
41	*Ocimum gratissimum*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
42	*Ocimum tenuifl orum L.*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
43	*Origanum heracleoticum L.*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
44	*Origanum majorana L.*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
45	*Orthosiphon stamineus*	Lamiaceae	*ITS2, rpoC1, trnH-psbA*
46	*Paris spp.*	Melanthiaceae	*ITS2*
47	*Persicaria odorata*	Polygonaceae	*ITS2, rpoC1, trnH-psbA*
48	*Phyllanthus niruri*	Phyllanthaceae	*ITS2, rpoC1, trnH-psbA*
49	*Phyllanthus* spp.	Phyllanthaceae	*psbA-trnH*
50	*Pinelliae ternata*	Araceae	*matK, rbcL*
51	*Piper betel*	Piperaceae	*ITS2, rpoC1, trnH-psbA*
52	*Piper sarmentosum*	Piperaceae	*ITS2, rpoC1, trnH-psbA*
53	*Plectranthus asirensis*	Lamiaceae	*rps16, rpoB*
54	*Polygonum multifl orum*	Polygonaceae	*trnL-F*
55	*Pueraria candollei*	Fabaceae	*matK*
56	*Radix Astragali*	Fabaceae	*ITS*
57	*Radix Rubi Parvifolii*	Gentianaceae	*ITS2*
58	Rehmannia spp.	Scrophulariaceae	*ITS*
59	*Rhodiola*	Crassulaceae	*ITS*
60	*Rhododendron* spp.	Ericaceae	*matK, rbcL,ITS, ITS2, psbA-trnH*
61	*Rhubarb*	Polygonaceae	*matK*
62	*Rosmarinus officinalis L.*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
63	*Rubus* spp.	Rosaceae	*ITS, psbA-trnH, trnL-F*
64	*Ruta* spp.	Rutaceae	*ITS, rpoB, rpoC1*
65	*Sabia* spp.	Sabiaceae	*matK, rbcL, psbA-trnH*
66	*Salvia divinorum*	Lamiaceae	*trnL*
67	*Salvia officinalis*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
68	*Salvia rutilans*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
69	*Salvia sclarea*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
70	*Salvia uliginosa*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
72	*Sambucus chinensis*	Acanthaceae	*ITS2, rpoC1, trnH-psbA*
73	*Scutellaria baicalensis*	Lamiaceae	*psbA-trnH*
74	*Scutellaria* spp.	Lamiaceae	*matK, rbcL, psbA-trnH*
75	*Senna* spp.	Fabaceae	*psbA-trnH*
76	*Smilax* spp.	Smilacaceae	*psbA-trnH*
77	*Solanum* spp.	Solanaceae	*matK, rbcL, ITS, psbA-trnH, trnL-F*
78	*Thymus vulgaris*	Lamiaceae	*rbcL, matK, trnH-psbA, rpoB*
79	*Tulipa edulis*	Liliaceae	*matK*
80	*Uncaria*	Rubiaceae	*ITS2*
81	*Uyghur*	Apiaceae	*ITS2*
82	*Vitex spp.*	Lamiaceae	*matK*

CHAPTER 3. MATERIALS AND METHODS

Key methods adopted during this study have been presented in the following flow chart. In the subsequent paragraphs each materials and methods are discussed in detail.

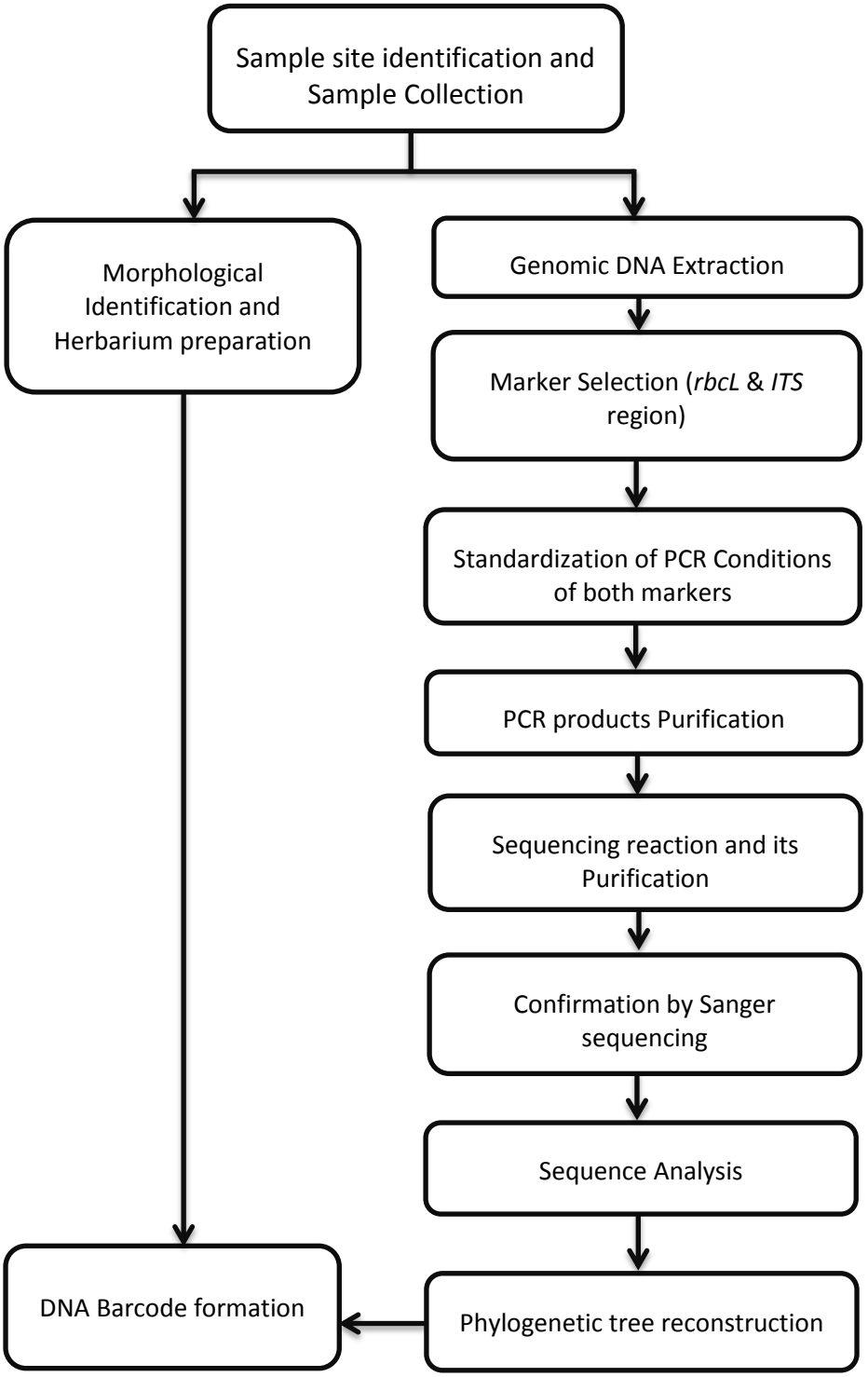


Figure 3.1 flow chart of the study.

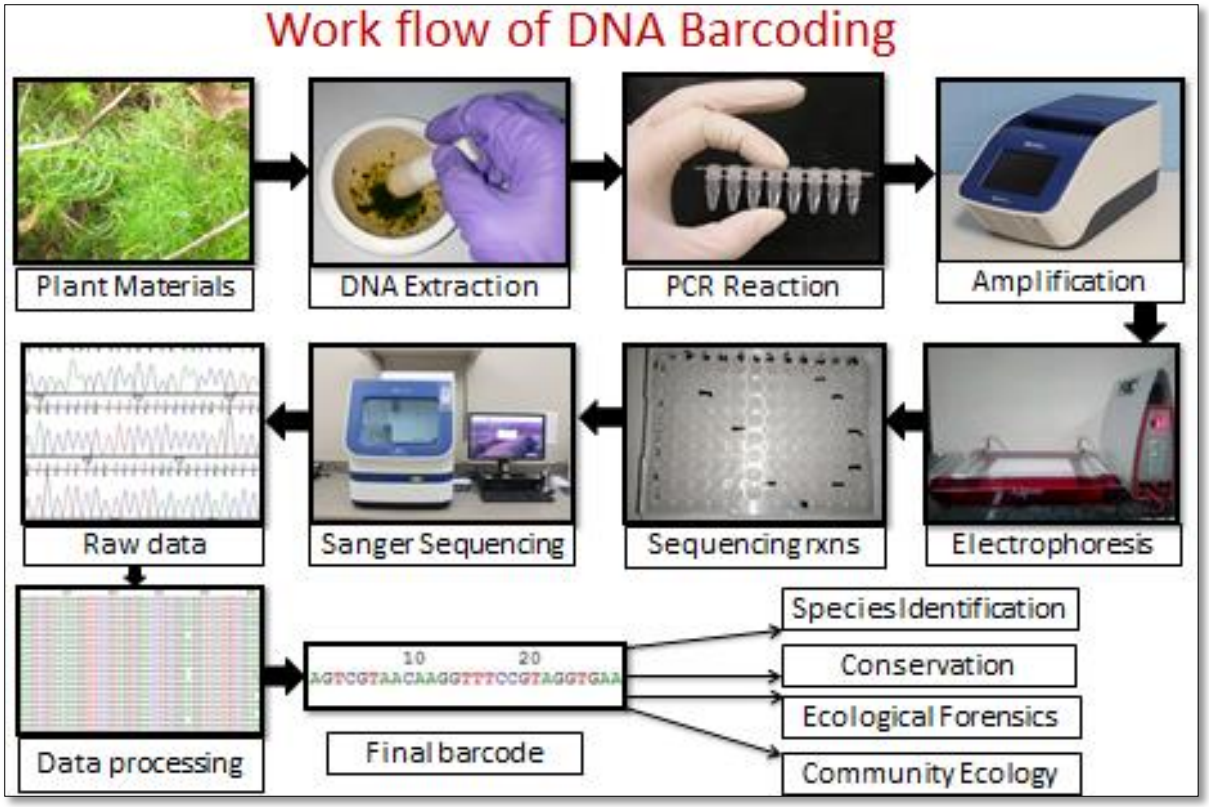


Figure 3.2 flow chart of work plan (graphical)

3.1 Chemical, Reagents and Kits

Chemicals

All the molecular biology (MB) or analytical grade (AR) chemicals were used in the study. The major chemicals along with their suppliers/manufacturers are listed in the Table 3.1

Table 3.1 List of chemicals used in this study and its manufactures/Suppliers.

Chemicals	Manufactures/ Suppliers
Tris [2-amino-2-(hydroxymethyl) propane-1,3-diol]	Sigma-Aldrich, USA
CTAB [Hexadecyltrimethyl ammonium bromide]	Sigma-Aldrich, USA
EDTA (Ethylenediaminetetracetic acid disodium salts dehydrate)	Sigma-Aldrich, USA
sodium chloride	Fisher Scientific, USA
PVP (Polyvinylpyroolidone)	Himedia, India
Isopropanol	Merck, Germany
Absolute Ethanol	Changshu Hongsheng fine chemical, China
Hydrochloric acid	Fisher Scientific, USA
Silica gel	Fisher Scientific, USA
Bovine serum albumin	Promega, USA
2x premix	Promega, USA
Gene ruler 100bp plus	Thermo Scientific, USA
Loading dye	Thermo Scientific, USA
Ethidium bromide	Promega, USA
Taq. DNA polymerase	Thermo Scientific, USA
Chloroform	Fisher scientific, USA
Isoamylalcohol	Merk, Germany
Agarose	Bioneer, South Korea

Reagents

The following stock solutions were prepared for the isolation and purification of DNA (modified CTAB method, Doyle and Doyle, 1987), and for the amplification of targeted regions, polymerase chain reaction (PCR; Mullis et al, 1986) was adopted.

a) 1 M Tris-HCl Buffer (pH 8.0)
b) 0.5 M Na_2EDTA (pH 8.0)
c) 5 M NaCl
d) 10% CTAB
e) 50X T.A.E. Buffer (Tris-HCl, Acetic acid, Na_2EDTA, pH 8.0)
f) 70% Ethanol
g) Chloroform : Isoamyl alcohol (24:1)
h) 5 mg/ml Ethidium bromide

All the reagents were prepared in double distilled deionized nuclease free Milli-Q water (MQ). The stock solutions, glassware and plasticware were sterilized by autoclaving for 15 lb pressure at 121°C for 20 min. Details of reagents preparation protocol was given in the Appendix 4.

Kits

The following kits were used for isolation of DNA, extraction of amplicons from the gel, purification of PCR products and finally sequencing of the amplicons.

1. DNeasy® plant Mini Kit (Qiagen, Germany).
2. PCR products clean up kit (Exosap, Affymetrix, USA).
3. Bigdye™ Terminator V3.1 cycle sequencing kit (Applied Biosystems, USA).
4. BigDye® Xterminator™ Purification Kit (Applied Biosystems, USA)

3.2 Bioinformatics tools

The software used for analyzing the chromatograms of amplicons generated after sequencing by ABI 3500XL DNA analyzer (Applied Biosystems, USA) were as follows:

1. Sequencer v. 4.1.4 (Gene Codes Corporation, USA).
2. BioEdit version 7.0.5.3 (Hall 1999).
3. MEGA version 7.0.14 (Kumar et al., 2007).
4. PAUP Version 4.0b10 (Swofford, 2003).
5. TaxonDNA version 1.8 (Meier et al., 2006)

3.3 Study Area

Nepal is a small landlocked country, roughly rectangular in shape, with a land area of 147,181 km^2. Ecologically, Nepal is divided into three main regions; the Hills, the Mountains and the Terai region (Shrestha & Dhillion, 2003). The study was carried out at Mardi-Himal region (83^0 50' E to 83^0 56' E and 28^0 19' N to 28^0 N) located in Kaski district, a mid to high hill of the Gandaki province (province no. 4) of Nepal (Figure 3.3). Altitude of the research area ranges from 900-5587 m above sea level. The study area is a part of Annapurna Conservation Area (ACA) and is located towards the North West of Pokhara, the Headquarter of State no.4.

In Mardi-Himal region, forest is the dominant land cover (58%) followed by cultivated land (22%) and the remaining 16% covers by others as reported by Awasthi et al. (2002). Major forests types in this area are Mixed Hardwood forest (MHF), Oak forest (OF) and High Mountain Mixed Forest (HMMF) as classified by Stainton (1972). Schima-Castanopsis are dominant trees in mixed MHF but *Alnus nepalensis* is major species in all community plantation sites. Species of *Quercus* dominates in OF, whereas Rhododendron is major species in HMMF. *Daphniphyllum* is found frequently between 1500-2500 m (Pokharel, 2013).

Figure 3.3 Map of Mardi-Himal; underlined part showing study area and black bold arrow indicate entry of the sampling site.

3.4 Site Selection

Mardi-Himal is a single Mountain range that ascends from low altitude subtropical forest to high altitude alpine vegetation in single vertical transects. This area is recently open for tourist thus has standard logistic and other facilities and is easily accessible from

Pokhara city. Gurung is a dominant ethnic tribe however other communities are also found in good numbers. Agriculture is predominant occupation of local people followed by animal husbandry, tourism and other forest based activities. Local people use plant for daily needs and medicinal purposes.

3.5 Plant Selection

Plants were selected based on the prioritized list of Nepal Government for protection, farming and research and development. Emphases were also given to include International Union for Conservation of Nature (IUCN) red listed species, Convention on International Trade in Endangered Species of Wild Flora and Fauna (CITES) appendices enlisted species and plant used by local people for their daily uses including traditional medication commonly practiced by the local communities.

3.6 Plant Collection and Identification

Plant collection permit was obtained from National Trust for Nature Conservation (NTNC) and Annapurna Conservation Area Project (ACAP). Field survey was carried out by NAST team in two times beginning on 2072-07-17 and 2073-6-11 for about 20 days in each trip. During field trip voucher herbarium specimens along with fresh leaves preserved in silica gel for DNA study were prepared. Information such as altitude, longitude and latitude readings of the plant collection sites were taken by using Geographical positioning system (GPS). Medicinal plants in the area were collected with the help of local traditional healers. Most of plant samples were collected from above 2000 m. Fresh leaves sample of economically important and endangered plants were preserved in Whatman lens cleaning paper (GE Healthcare, UK) pouches with Silica gel beads in air tight plastic container. The collected plants were properly dried for further identification. All these samples were brought to Molecular Biotechnology Unit Laboratory of NAST for further analysis. The Herbarium specimens were prepared for the collected economically important and endangered plants. All the collected herbarium vouchers were identified to species level through tallying with herbariums deposited at National Herbarium and Plant Laboratories (KATH), Godawari, and or consulting with professional plant taxonomist. Species identification and plant nomenclature was verified through consulting available standard taxonomic literature (Flora of Nepal Volume 3, Flora of China and Flora of Bhutan) and websites (http://www.efloras.org, http://www.theplantlist.org). All the specimens collected during the field trip were provided an accession code according to the criteria of Molecular Biotechnology Laboratory at Nepal academy of Science and Technology (NAST). Herbarium specimens of all the studied species are mounted in standard Herbarium sheets and stored properly at Plant curation and management unit at NAST (Appendix 6, Appendix 7 and Appendix 8). Details of samples collected from Mardi-Himal region is given in the Appendix 1.

3.7 Isolation of Total Genomic DNA

Two standard methods were adopted for the extraction of DNA from Silica gel preserved plant samples. Details of each method are given below.

Method1: Modified CTAB Method

Plants DNA was extracted using widely used modified CTAB protocol (Doyle, 1991). The protocol is followed as below.

Protocol:

The CTAB buffer (2% CTAB [Sigma-Aldrich, USA]). 1.4 M NaCl, 0.02 mM EDTA, 100 mM Tris-HCl [pH 8.0], 1% PVP) was preheated in 50ml falcon tube at 60-65^0 C in a water bath. About 50-100 mg silica gel preserved plant leaves were taken in lysing matrix A tube and homogenized in TissueLyser (FastPrep-24™ 5G, MP Biomedicals, USA) at 6.5 m/sec for 50-55 sec or alternatively, About 50-100 mg silica gel preserved plant leaves were grounded in chilled sterile mortar and pestle with liquid Nitrogen. 1 ml CTAB buffer was added and mixed and transferred in 2 ml Eppendorf (Ep) tube and then incubated at 60-65^{0}C in a water bath for 30-60 minutes with shaking in every 10 minutes. After incubation, the tubes were allowed to cool at room temperature and added ~800µl Chloroform: Isoamylalcohol (24:1) and inverted the tubes slowly for 4-5 minutes (mixed properly). All the Eppendorf tubes were centrifuged at 9000 rpm for 8 minutes at room temperature (RT). Carefully, Upper aqueous layer was transfer to new Eppendorf Tubes and re-extracted from above aqueous layer, by adding 1 ml Chloroform: Isoamylalcohol (24:1) followed by gentle mixing and then centrifuged at 9000 rmp for 8 minutes. Upper aqueous layer was transferred to the new Eppendorf Tubes. To the new solution 2/3 volume of chilled isopropanol was added and mixed or inverted and then incubated at -20^0 C for 1 hour. The Ep. tubes were centrifuged at 10,000 rpm for 8 minutes at RT. The supernatant was discarded and washed the pellet twice with 200-300 µl of 70% ethanol and centrifuged at 10,000 rpm for 1-2 minutes. Again, the pellet was washed with 100% ethanol and centrifuged at 10,000 rpm for 1-2 minutes. The ethanol was discarded completely, and dried the pellet in paper towel or tissue paper. 50-100 µl of T.E. was added and re-suspended the pellet and then stored at -20^{0}C for further down streaming process.

Method2: Genomic DNA extraction using Qiagen kit

Samples (limited samples no. 10-15) (≤100 mg wet weight or ≤20 mg lyophilized tissue) were disrupted using the TissueLyser (FastPrep-24 ™ 5G, MP Biomedicals, USA) at 6m/sec for 40 sec or a sterile mortar and pestle using liquid nitrogen. 400 µl Buffer AP1 and 4 µl RNase A were added and mixed by vortex and then incubated for 10-30 min at 65^0 C in water bath. All the Ep. Tubes were inverted 2-3 times during incubation. 130 µl buffer P3 was added, mixed and then incubated for 5 minutes on ice. The lysate was centrifuged for 5 minutes at 20,000x g (14,000 rpm). The lysate was pipetted into a QIAshredder spin column placed in a 2ml collection tube. The tubes were centrifuged at 20,000xg (14,000 rpm) for 2 minutes. The flow-through was transferred into a new tube without disturbing the pellet if present, 1.5 volumes of buffer AW1 was added and mixed by pipetting. 650µl of the mixture was transferred into a DNeasy Mini spin Column placed in a 2ml collection tube. The tubes were centrifuged at ≥6,000x g (≥8,000 rpm) for 1 minute. The flow-through was discarded and repeated this step with the remaining sample. The spin column was placed into a new 2ml collection tube and then added 500 µl buffer AW2 and centrifuged at ≥6,000xg (≥8,000 rpm) for 1minute. The flow-through was discarded. Another 500 µl buffer AW2 was added and centrifuged for 2 minutes at 20,000xg (14,000 rpm) (Note**:** The spin column was removed carefully from the collection tube, so that the column does not came into contact with the flow-through). The spin column was transferred to a new 1.5 or 2 ml microcentrifuge tube. 50-100 ul Buffer AE was added for elution and incubated for 5 min at RT (15-25^0 C) and then centrifuged at ≥6,000xg (≥8,000 rpm) for 1 minute. The step 11 was repeated. DNA sample was stored in -20^{0}C for further purposes.

3.8 Amplification of Selected Barcode loci

Table 3.2 Components of 2X Master Mix obtained from Promega

Components	Concentration
Taq DNA polymerase	50 U/ml
dNTPs (dATP, dGTP, dCTP, dTTP)	400 µm (each)
$MgCl_2$	3 mM
Reaction Buffer	pH 8.5

PCR experiments were carried out in thermal cycler (Applied Biosystems, USA). For optimization of the PCR conditions, PCR reactions were carried out using different concentration of PCR ingredients. Optimization was performed using the master mix (2X Premix Promega, USA) as constant parameter and variable parameters included $MgCl_2$ concentration, primer concentration, Taq DNA polymerase concentration and addition of Bovine Serum Albumin (BSA). (BSA) was added to all reactions, it acts as stabilizer for enzymes, to decrease problems with secondary structure, and improve annealing temperature (Palumbi 1996).

Both *rbcL* and *ITS* regions were amplified by using 2X PCR Master Mix (Promega, US). The components of master mix (Promega, USA) were as listed in Table 3.2.

Standard PCR composition and conditions has been followed for both *rbcL* and *ITS* region as given in Table 3.3 and Table 3.4. However, during the optimization process, concentration of PCR components was varied depending upon the plant material being used.

Table 3.3 Standard PCR components composition of both *rbcL* and *ITS* markers

Components	Volume 20 µl/ Reaction	Final Concentration
2X PCR mix (Promega)	10 µl	1X
25mm $Mgcl_2$	0.8 µl	0.5 mM
10 µm Primer-F	0.8 µl	400 pmol
10 µm Primer-R	0.8 µl	400 pmol
1 mg/ml BSA (Promega)	2 µl	0.1 mg/ml
DNA Template	2 µl	
5U/ µl Taq DNA polymerase (Promega)	0.1 µl	0.5 U/ 20 µl
Nuclease Free water (Mili Q)	3.5 µl	

Note: Some reaction that failed to amplify from above protocol was treated with 2% DMSO.

Amplification for *rbcL* and *ITS* region

PCR amplification was done using chloroplast and nuclear gene specific primers for rbcL and *ITS* regions respectively for all samples. The primers used in amplification of both *rbcL* and *ITS* regions are listed in the Table 3.4. PCR conditions for the amplification of both *rbcL* and *ITS* regions are given in the Table 3.4.

3.9 PCR product Cleanup

PCR Product Cleanup by using ExoSap-IT™ Kit (Product no. 75001/75002; Applied Biosystems, USA). The following procedure was followed for the purification of the PCR products of both barcode markers *rbcL* and *ITS*.

Protocol:

ExoSap-IT™ Express reagent was removed from -20^{0}C freezer and kept on ice throughout the procedure. 2µl of ExoSap-IT™ Express reagent was added in 5µl of a post PCR reaction product and mixed thoroughly by gentle vortexing and quick spin to bring the contents to the bottom of the tube. The reaction was incubated in thermal cycler (Applied Biosystems, USA) with a heated lid following program Table 3.5. Now PCR product is ready for further processing. The treated PCR product was used instantly for the sequencing reactions or stored in -20^{0}C.

Table 3.4 Details of primers and their amplification condition used in this study.

Region	Primer	Sequence (5'-3')	Thermocycling conditions	Reference
rbcL	rbcLa-F	5'-ATGTCACCACAGACAGAGACTAAAGC-3'	94^{0}C, 2 min; (35 cycles: 94^{0}C 40 sec; 52^{0}C 45 sec; 72^{0}C 1 min; 72^{0}C 7 min) 4^{0}C ∞	(Levin, 2003; Kress & Erickson, 2007)
	rbcLa-R	5'-GTAAAATCAAGTCCACCRCG-3'		
ITS	*ITS*-p5	5'-CCTTATCAYTTAGAGGAAGGAG-3'	94^{0}C, 4 min; (35 cycles: 94^{0}C 40 sec; 55^{0}C 40 sec; 72^{0}C 1 min 10 sec; 72^{0}C 7 min) 4^{0}C ∞	(Cheng et al., 2015)
	ITS-u4	5'-RGTTTCTTTTCCTCCGCTTA-3'		

Table 3.5 PCR program of PCR product cleanup of the both markers.

Step	Temperature	Time
Degradation	37^{0}C	4 min
Inactivation	80^{0}C	1 min
Hold	4^{0}C	∞

3.10 Gel Electrophoresis

Gel electrophoresis was done to confirm whether DNA extraction and PCR optimization was successful. The Extracted DNA was analyzed on a 0.8% agarose gel electrophoresis in 1X TAE Buffer. Amplified PCR product of both markers *rbcL* and *ITS* were visualized in Agarose gel electrophoresis. Total volume of 6 µl was loaded in each well containing 5 µl DNA and 1 µl gel loading dye (Thermo Scientific, USA). The PCR clean up product was also loaded in Agarose gel (1% in 1X TAE) in AHoefer (Holliston, MA, USA) Gel Tank. For staining and visualisation purpose, 0.5µg/ml Ethidium bromide (promega) was added to the molten gel. In case of PCR clean up product 1µl was loaded into wells with 1µl of 6X loading dye. A constant current of 70V was applied for 60-90 minutes. Then, the gels were visualized using Syngene (Synegen Bio-imaging, UK) gel documentation system.

3.11 Sequencing

3.11.1 Cycle Sequencing

The sequencing reactions were performed by using high quality single band of PCR purified product after gel visualization. The sequencing reactions were prepared by following BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). The contents of the BigDye™ Terminator v3.1 Cycle Sequencing Kit and primer was thawed completely and kept on ice. The tubes were vortexed 2 or 3 seconds, then short spin in microcentrifuge to collect contents at the bottom of the tubes. The sequencing reaction mixture were prepared separately for both forward and reverse primers of both markers *rbcL* and *ITS* based on Table 3.6. The plate was sealed with MicroAmp™ Clear adhesive film (Applied Biosystems, USA). The plate was then vortexed for 2 or 3 Seconds, short spin in a swinging bucket centrifuge (ST 8R, Thermo Scientific, USA) to collect the content to the bottom of the wells (5 to 10 seconds) at 1000 x g. The plate was loaded in Thermal cycler (Veriti 96 well thermal cycler, Applied Biosystems, USA) and run with sequencing program is given in the Table 3.7.

Table 3.6 Standard sequencing reactions mixture of the both markers.

Components	**Quantity/per reaction**	**Final Concentration**
Ready Reaction mix (2.5X)	1.0 µl	0.25X
Dilution Buffer (5X)	1.5 µl	0.75X
Template (200 ng/µl)	1.0 µl	20 ng/ µl
Primer 10 µm	1.5 µl	1.5 µm
MQ water	5 µl	
Total	10 µl	

Table 3.7 PCR program for Sequencing.

Parameter	**Stage/ Step**				
	25/30 Cycles				
	Incubation	**Denature**	**Anneal**	**Extension**	**Hold**
Temperature	96^0C	96^0C	50^0C	60^0C	4^0C
Time (mm:ss)	01:00	00:10	00:05	04:00	Hold until ready to purify

3.11.2 Purification of Sequencing Reactions

Purification of sequencing reactions of the both markers (*rbcL* and *ITS*) were done using BigDye Xterminator™ kit (Applied Biosystems, USA).

Table 3.8 Mix of SAM/BigDye Xterminator™ beads

Component	**Volume per 10µl reaction**
SAM solution	45 µl
BigDye Xterminator™ bead solution	10 µl
Total	55 µl

Sequencing reactions contains salts, unincorporated dye terminators, and dNTPs. These interfere with the base calling process in the DNA analyzer. These sequencing reactions

were purified by using BigDye Xterminator™ Kit (Applied Biosystems, USA). After performing cycle sequencing reactions, the plate was short spin in centrifuge. The bottle of BigDye Xterminator™ beads were vortexed for 20 or 30 seconds before mixing with SAM solution. SAM/BigDye Xterminator™ beads were prepared for working solution as followed in the Table 3.8. The MicroAmp™ Clear Adhesive Film was removed from the sequencing plate. 55 µl mix of SAM/BigDye Xterminator™ beads working solution was transferred to each well of sequencing plate. The plate was sealed with MicroAmp™ Clear Adhesive Film and vortexed (Mixmate Eppendorf, USA) at 1800 rpm for 30 minute. The plate was then centrifuged at 1,000 x g for 2 minute. The MicroAmp™ Clear Adhesive Film was removed. The plate was sealed with MicroAmp™ 96-Well Full Plate Cover (Applied Biosystems, USA). The plate was loaded plate on the deck of sequencer, locked and linked to the 3500XL Genetic Analyzer (Applied Biosystems, USA) for capillary electrophoresis.

3.12 Data Analysis

3.12.1 Sequence Editing and Alignment

The chromatograms obtained for each region were base called using PHRED quality score (Ewing & Green, 1998b). To estimate the quality of generated sequence traces, the original forward and reverse raw sequences were assembled and edited in Sequencer v. 4.1.4 (GeneCodes Corporation, USA). Sequences were assembled based on the parameters minimum match percentage 70 and minimum overlap 20. Each Contig were viewed and manually edited (removal of gaps and dealing with ambiguous nucleotides). The aligned sequences were also edited by comparing with the reference sequence (www.ncbi.nlm.nih.gov/blast) by closely inspecting the peaks of chromatograms of forward and reverse sequence. The assembled consensus contigs were exported in text format and imported in Bioedit v.7. All candidate barcode sequences were aligned by ClusterW, (multiple sequence alignment tools) in Bioedit using default parameters. The both primer end was delineated from the alignment matrix. Primer excluded barcode sequences were exported for further analysis.

3.12.2 Extraction of *ITS2*

ITS2 sequences were annotated according to Keller et al., (2009). The each fasta file of *ITS* sequences of sample was imported in annotation tool (http://*ITS2*.bioapps.biozentrum.uni-wuerzburg.de) and choosing a model Hidden Markov Modeling (HMM), with parameters E<0.01<-- and minimum size of *ITS2* <150 bp <-- and then run query against Eukaryotes (571/416) (Ankenbrand et al., 2015).

3.12.3 Similarity search using Simple BLAST and Optimized BLAST

Similarity search was performed by following the protocol provided by Ghorbani et al., (2017). Two approaches were used for DNA barcode identification. The two methods are simple method taking the top hit and a modified optimized method putting extra weight on the identity score from query-reference comparison. For both methods sequences were sequentially queried using megablast (Altschul et al., 1990) in the online platform of NCBI database. For the simple BLAST method, all the top hits within 10 points deviation down of the max score were considered: if the max score (-10 points) included only a single species then a species level identification was assigned; if the max score (-10 points) included multiple species in the same genus then a genus level identification

was assigned; and if the max score (-10 points) included multiple species in different genera in the same family then a family level identification was assigned. However, the length of the query coverage and the identity between the query and the reference sequence influences the max score in BLAST, and hits with high identity but low query coverage revealed the chance of high identity for the species identification. For the modified optimized method a similarity score was calculated for up to 100 BLAST hits if the query cover was 70% or higher: max score*(identity/query cover). Subsequently all hits were ordered by this score, and the deviation for each similarity score value from the highest similarity score was calculated Identifications were assigned based on a combination of the identity score (High identity: i ≥ 95%; Medium identity: 90% ≤ i < 95%; Low identity: i < 90%) and the number of species within 1% deviation of the calculated similarity score. High identity and one species within 1% deviation was assigned species-level confidence; high identity and more than one species was assigned genus-level confidence; medium identity and one or more species within the same genus was assigned genus-level confidence; medium identity and species from more than one genus was assigned family-level confidence; and low identity was assigned family-level confidence.

3.12.4 Sequence variation analysis

The aligned ".fas" files for each locus (*rbcL* and *ITS*; without the reference from NCBI) were individually processed in MEGA v. 7.0 and parameter like number of constant sites (C), variable sites (V), parsimony informative sites (Pi) and singleton sites (S) were estimated.

3.12.5 Evaluation of genetic divergence by TaxonDNA

The barcode markers viz *rbcL*, *ITS2* and *ITS* were evaluated by TaxonDNA v.1.8 using aligned delineated consensus sequences generated from Sequencer v.4.1.4 for respective locus along with other sequences of the sampled species retrieved from NCBI. Analysis was done separately with core data (experimental) and with NCBI retrieved sequences. The pairwise distances present in the dataset were also calculated using Pairwise Summary explorer. 'Kimura 2-parameter corrected sequences' was used for calculating pairwise distance. The percentage distribution values of intraspecific and interspecific distances were used in Microsoft spreadsheet to plot the distribution graph.

Overall the genetic divergences of sampled species were evaluated by TaxonDNA v.1.8 (Species Identifier and SequenceMatrix respectively) using both *rbcL* and *ITS* markers at species and family level. To estimate intraspecies divergence of the species, species having more than two collections were evaluated based on K2P distance matrix. In family based analysis, only those species are evaluated which are listed as protected plants of Nepal, CITES listed and plant species prioritized for research and development.

For barcode gap analysis, overlap of 5% largest intraspecific and 5% smallest interspecific divergences are considered as cut off (Meier et al., 2008) value with the evaluation of the barcode gap for each marker and various combinations. The study was evaluated in 'Pairwise summary' of TaxonDNA v.1.8.

3.13 Phylogenetic Inference

Phylogenetic tree was constructed for each region (*rbcL*, *ITS* and *ITS2*) and with various combination of multilocus barcode using the respective aligned sequences. Phylogeny tree was reconstructed by Neighbor-joining (NJ) in MEGA v.7.0.14. NJ tree was constructed using K2P distance as genetic measure and setting negative branch length to zero with uniform distribution rates applied. Typically 1000 replicates of bootstrap were used to estimate tree reliability. Except *ITS* and *ITS2*, p-distance model is used as genetic measure while reconstructing NJ phylogeny. Node support was estimated based on the following scale: BS 50–74% (weak bootstrap support) and 75-100% for strong support (Hillis and Bull, 1993; Murphy et al. 2001; Daru et al., 2013). The efficiency of species resolution was considered successful only when all the single clades have at least ≥50% bootstrap value. In addition, the sampled species was considered successful when discriminate individuals of species formed a monophyletic group in the phylogenetic tree (Hollingsworth et al., 2009). The ratio of successfully identified species to all sampled species was calculated as the proportion of species that were discriminated.

3.14 Species Resolution

The efficiency of *rbcL* and *ITS* markers in species identification was carried out using three analytical methods. The first method was two sequence similarity-based methods estimated using online BLAST (Altschul et al., 1990) and Optimized BLAST (Ghorbani et al., 2017) features available in https://www.ncbi.nlm.nih.gov/blast/. The second method was to estimate K2P distances of each locus by utilizing software TaxonDNA v.1.8 (Species Identifier and SequenceMatrix) and the third method is tree method implemented in MEGA v.7.0.

Assigning plant samples to appropriate species

An identification of each voucher specimens collected from Mardi Himal area has followed two different but complementary methods as discussed above. The first morphological method consisted of tallying the collected herbarium voucher with standard herbarium specimens deposited at National Herbarium and Plant Laboratory (KATH), Godawari, Lalitpur and with the help of standard flora related literature. Where as the second molecular method was based on BLAST and optimised BLAST of the sequences generated through following series of steps like DNA extraction, PCR using an appropriate primer pairs and finally sequencing the successful PCR amplicons. In the present study molecular methods adopted to identify medicinal plant species of Mardi Himal are presented and discussed in detail.

CHAPTER 4. RESULTS

4.1 Sampling

All the collected specimens were pressed and prepared as herbarium and are deposited in molecular Biotechnology unit, NAST. Among identified voucher specimens, plants tissue samples from 89 species belonging to 63 genera were collected from Mardi-Himal region. Collected samples were identified morphologically up to species level. In studied data set, Angiosperms represented in higher number 97% (87) followed by Pteridophytes 2% (2) and Gymnosperms 1% (1) as shown in Figure 4.1. Among the 89 samples, 36% (32), herbs, 20% (18) shrubs, 35% (31) trees, 6% (5) climber, and 3% (3) ferns were subjected for molecular identification as shown in Figure 4.2. Studies samples belong to 41 different families. Among them, Ericaceae was the highest 17% (7) in number, followed by Asteraceae 12% (5), Polygonaceae 12% (5), Ranunculaceae 12% (5), and so on Figure 4.3.

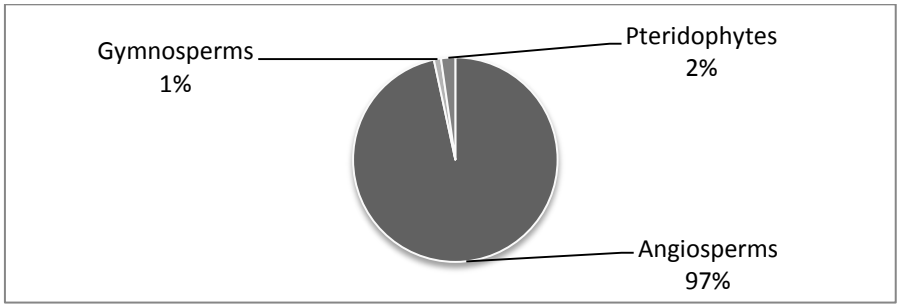


Figure 4.1 Graphical representation of studied plant groups.

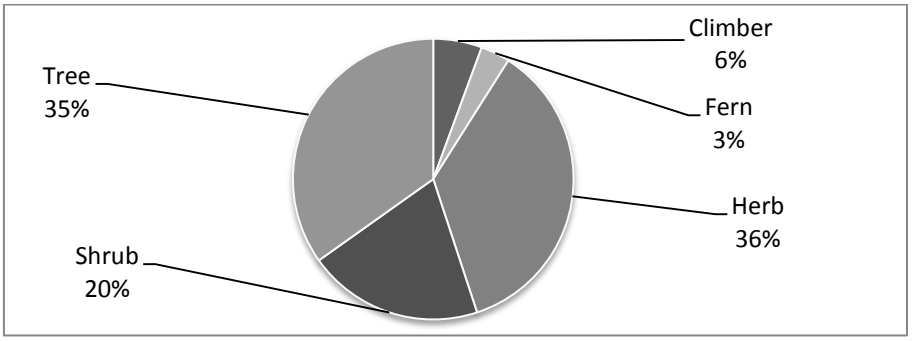


Figure 4.2 Graphical representations of plant life forms for studied species.

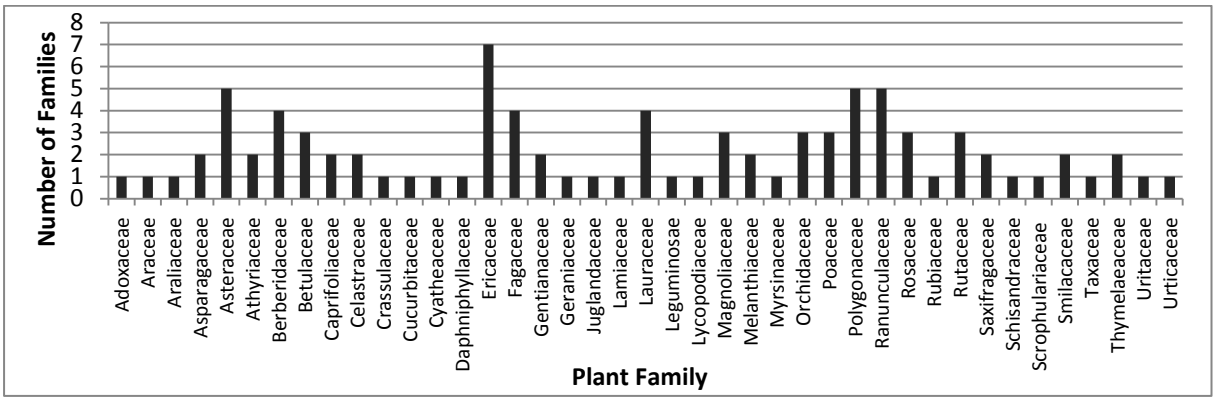


Figure 4.3 Graphical representations showing plant families included in this study.

4.2 DNA Extraction

The genomic DNA was successfully extracted from silica gel preserved leaves from a total of 89 samples. The quantities and quality of DNA visualised on gel electrophoresis is shown in Figure 4.4.

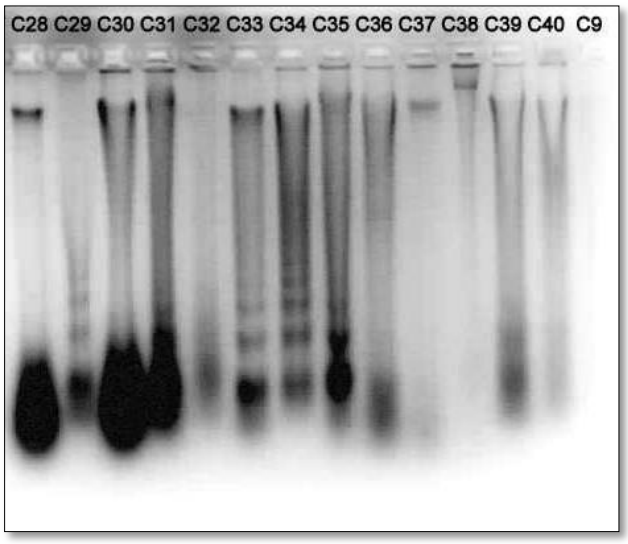


Figure 4.4 Representative diagram of agarose gel electrophoresis (0.8%) of genomic DNA extracted from plants samples (leaves).

4.3 Amplification of Barcode markers

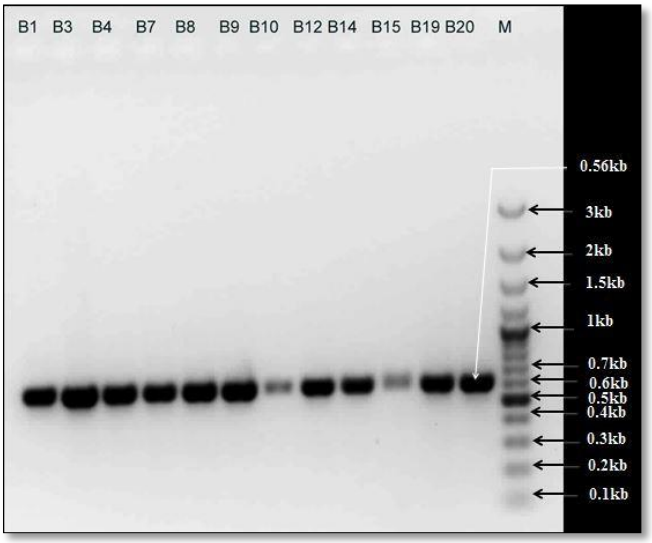
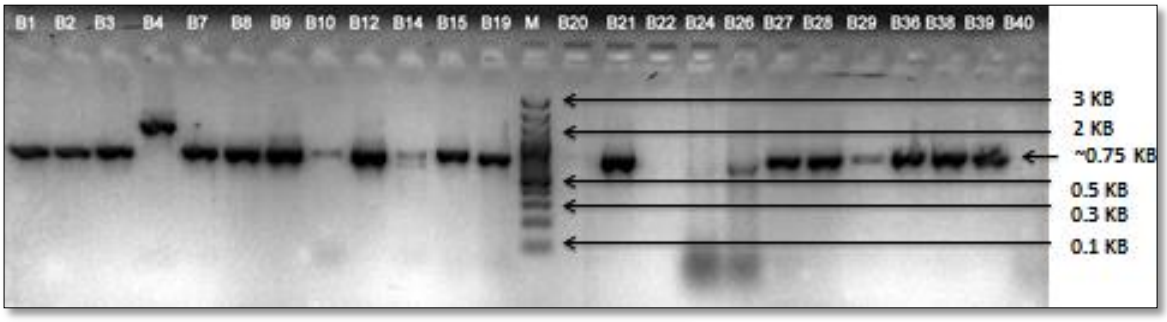


Figure 4.5 Representative diagram of agarose gel electrophoresis (1%) of PCR product of *rbcL* region, M= marker 100bp plus (Promega, USA).

PCR efficiency of the barcode markers *rbcL* and *ITS* were found to be 85.39% (76) and 83.15% (74) respectively. *rbcL* and *ITS* PCR product size were about 500-600 bp and 700-800 bp respectively except sample B4 as shown in Figure 4.5 and Figure 4.6. PCR product of sample B4 was about 1.2 kb in size. PCR product size of each barcode marker *rbcL* and *ITS* were estimated by standard marker 100 bp plus (Promega, USA).



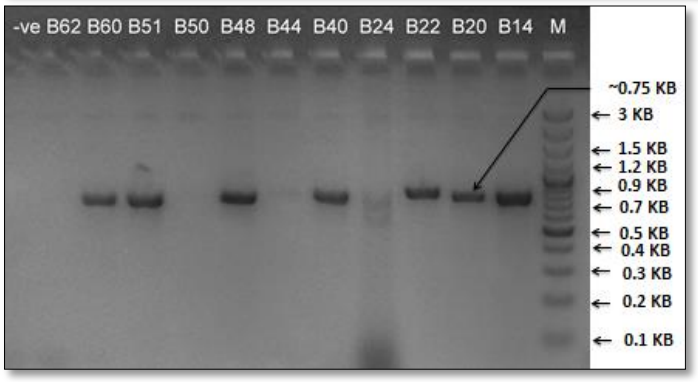


Figure 4.6 Representative diagram of agarose gel electrophoresis (1%) of PCR product of *ITS* region, M= marker 100bp plus (Promega, USA)

4.4 PCR Clean up

Successfully amplified PCR products of the both markers (*rbcL* and *ITS*) were purified by Exosap method (Product no. 75001/75002). After clean up, high quality single band was obtained. In case of double band or no single band visualized in the gel, PCR was repeated again for the subsequent samples. Visualization of purified PCR product on agarose gel electrophoresis is shown in Figure 4.7 and Figure 4.8.

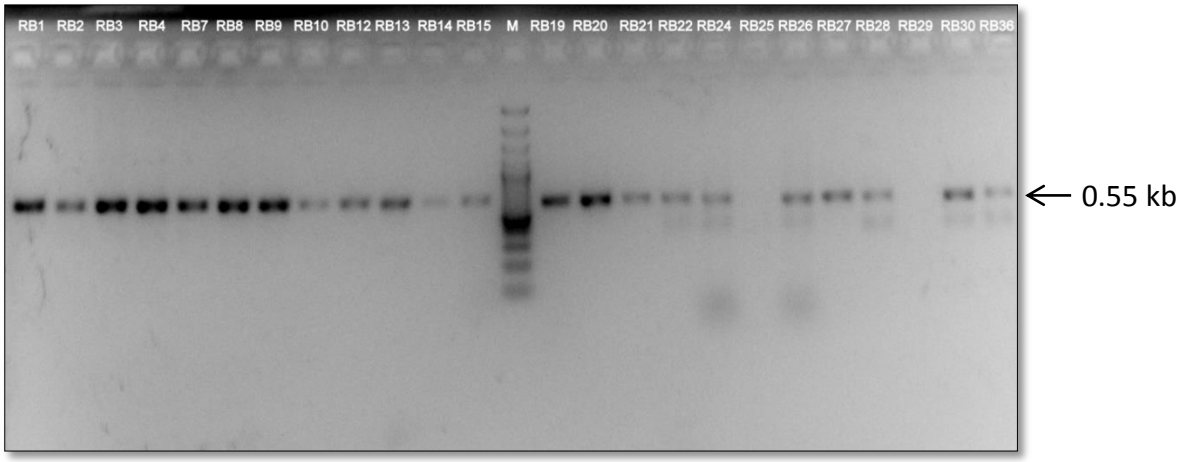


Figure 4.7 Diagrammatic representation of Agarose gel electrophoresis (1%) of purified PCR products of *rbcL* markers, M = marker 100bp plus (Promega, USA).

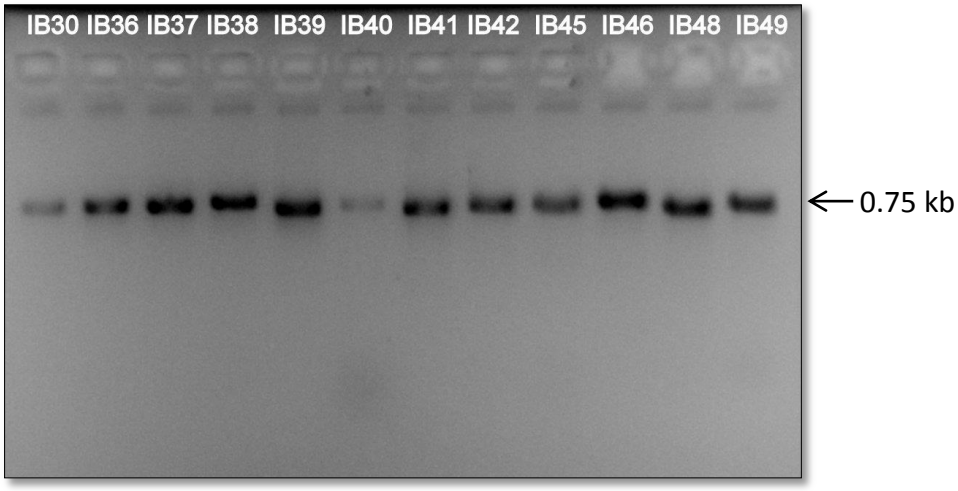


Figure 4.8 Diagrammatic representation of agarose gel electrophoresis (1%) purified PCR products of *ITS* markers**.**

4.5 Sequencing efficiency

Out of 89 samples, the efficiency of sequencing for the markers *rbcL* and *ITS* were 65 (73%) and 58 (65%) respectively. In case of *ITS*, two samples were found contaminated due to the handling errors. These contaminated samples and 4 additional samples with bad sequences quality were excluded from the analysis. A total of 72 (80%) of the samples have sequencing success for both markers (sample having both or single sequences of the marker). The efficiency of sequencing for both markers were 57% (51 samples have both sequences) and rest of 18.31% (13) were with only the *rbcL* sequences and 9.86% (7) were with only the *ITS* sequences as shown in Figure 4.9.

The examination of sequences quality and coverage indicated that *rbcL* and *ITS* routinely generated high quality bidirectional sequences. High quality sequences were found to be 84.61% (55) of *rbcL* and 85.93% (55) of *ITS*. Both *rbcL* and *ITS* Sequences of some high value plant species were given in the Appendix 5.

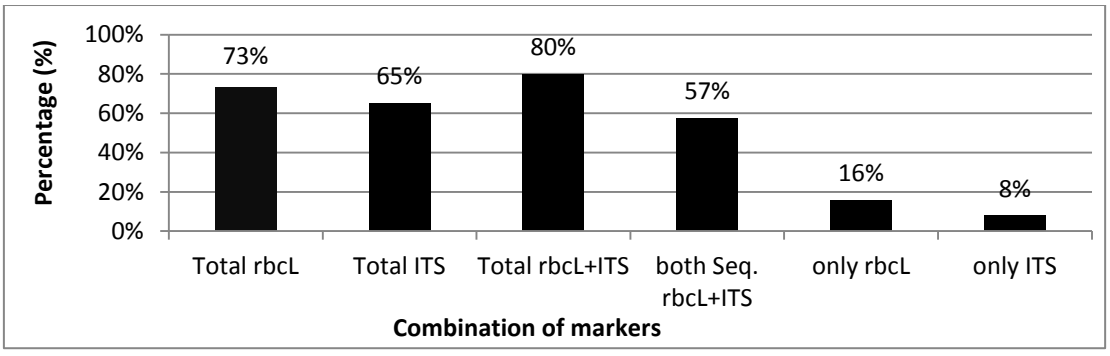


Figure 4.9 Efficiency of sequencing success with or without combination of barcode markers**.**

4.6 Nucleotide Composition

Based on the sequence data, the final lengths of *rbcL* for studied samples were 495-536 bp with delineated primer ends. Nucleotide composition of *rbcL* region of the samples is given in the Table 4.1. Similarly, the length of *ITS* was found to be 688-867 bp. Nucleotide composition of *ITS* region of the samples is given in the Table 4.2. *ITS2* region

extracted from the *ITS* sequences from high quality sequences of all the studied samples ranges from 258-303 bp. Nucleotide composition of each *ITS2* region of the samples is given in the Table 4.3.

Table 4.1 Nucleotide composition of each representative samples of barcode marker *rbcL* with total length of sequences**.**

Species name (lab code)	T(U)	C	A	G	Length (bp)
Aconitum spicatum (RB12)	29.5	19.6	29.3	21.6	495.0
Alnus nepalensis (RB53)	29.3	21.1	26.9	22.6	535.0
Anaphalis contorta (RB8)	30.1	19.6	27.3	23.0	535.0
Arisaema speciosum (RC33)	29.3	20.9	27.7	22.1	535.0
Aruncus dioicus (RC28)	29.9	20.2	26.2	23.7	535.0
Asparagus filicinus (RB38)	29.6	21.1	27.5	21.7	530.0
Berberis angulosa (RC7)	29.3	20.9	27.3	22.4	535.0
Berberis erythroclada (RB2)	29.3	20.9	27.3	22.4	535.0
Bergenia purpurascens (RC8)	29.3	21.1	27.1	22.4	535.0
Betula utilis (RB10)	29.5	21.1	27.5	21.9	535.0
Betula utilis (RC25)	29.5	21.1	27.5	21.9	535.0
Bistorta macrophylla (RC20)	28.6	21.5	27.1	22.8	535.0
Cassiope fastigiata (RC3)	29.2	20.6	27.5	22.8	535.0
Cotoneaster integrifolius (RC21)	30.3	20.2	26.7	22.8	535.0
Dactylorhiza hatagirea (RB7)	28.8	21.3	28.0	21.9	535.0
Dactylorhiza hatagirea (RB61)	28.8	21.3	28.0	21.9	535.0
Dactylorhiza hatagirea (RC11)	28.8	21.3	28.0	21.9	535.0
Danthonia cumminsii (RC13)	29.9	19.8	27.7	22.6	535.0
Daphne bholua (RC31)	28.4	21.5	27.1	23.0	535.0
Daphne papyracea (RB21)	28.4	21.5	27.1	23.0	535.0
Daphniphyllum himalayense (RB46)	28.6	21.9	27.1	22.4	535.0
Delphinium brunonianum (RB14)	29.0	20.5	28.3	22.2	527.0
Delphinium himalayai (RB15)	28.6	20.6	28.6	22.2	535.0
Delphinium himalayai (RC5)	28.6	20.6	28.6	22.2	535.0
Euonymus tingens (RB5)	29.5	21.3	26.2	23.0	535.0
Indigofera atropurpurea (RC36)	30.5	20.2	27.1	22.2	535.0
Leucosceptrum canum (RB42)	28.5	21.2	28.2	22.1	529.0
Lindera neesiana (RB54)	28.8	21.7	26.7	22.8	535.0
Lithocarpus grandifolius (RB48)	28.6	22.4	26.9	22.1	535.0
Lycopodium japonicum (RB50)	32.1	18.9	28.4	20.6	535.0
Machilus odoratissima (RB62)	28.8	21.7	26.5	23.0	535.0
Magnolia campbellii (RB20)	28.3	21.7	27.7	22.3	530.0
Mahonia nepalensis (RB51)	29.3	20.5	27.4	22.8	536.0
Michelia doltsopa (RB29)	28.2	21.7	27.1	23.0	535.0
Michelia kisopa (RB45)	28.2	21.5	27.4	22.9	536.0
Myrsine semiserrata (RB37)	30.0	20.1	26.9	22.9	536.0
Nardostachys grandiflora (RC2)	28.4	21.7	27.3	22.6	535.0
Neopicrorhiza scrophulariiflora (RC41)	28.2	21.1	27.7	23.0	535.0
Panax pseudoginseng (RC9)	29.3	20.9	27.1	22.6	535.0
Paris polyphylla (RB3)	28.4	20.7	28.0	22.8	535.0
Paris polyphylla (RC1)	28.4	20.7	28.0	22.8	535.0
Parnassia nubicola (RC19)	29.9	21.1	26.7	22.2	535.0
Persea duthiei (RB44)	28.8	21.9	26.5	22.8	535.0
Polygonum molle (RC23)	29.3	21.1	27.1	22.4	535.0
Polygonum molle (RB27)	29.5	20.9	27.3	22.2	535.0
Quercus lamellosa (RB28)	29.0	22.1	26.7	22.2	535.0
Quercus semecarpifolia (RB39)	29.2	21.9	26.7	22.2	535.0
Rheum australe (RC10)	28.6	22.1	27.1	22.2	535.0
Rhododendron anthopogon (RC4)	28.8	20.7	27.7	22.8	535.0
Rhododendron campanulatum (RB11)	28.8	20.4	27.9	23.0	535.0
Rubia manjith (RB41)	28.6	20.7	27.9	22.8	535.0
Schisandra grandiflora (RC32)	29.2	21.1	26.4	23.4	535.0
Senecio graciliflorus (RC22)	30.5	19.8	27.3	22.4	535.0
Smilax ferox (RC39)	28.8	20.4	28.4	22.4	535.0
Sorbus arachnoidea (RC24)	30.3	20.4	26.5	22.8	535.0
Sweretia chirayita (RB1)	28.4	22.2	27.1	22.2	535.0
Swertia teres (RC6)	28.6	22.1	27.3	22.1	535.0
Taraxacum officinale (RB9)	30.8	19.1	26.7	23.4	535.0
Taxus wallichiana (RB4)	27.9	22.8	28.8	20.5	531.0

Thamnocalamus nepalensis (RB60)	29.2	20.9	26.5	23.4	535.0
Trichosanthes wallichiana (RB40)	29.9	19.8	27.9	22.4	535.0
Yushania maling (RB19)	28.9	21.0	26.8	23.3	533.0
Zanthoxylum acanthopodium (RB36)	28.0	21.9	27.1	23.0	535.0
Zanthoxylum acanthopodium (RC38)	28.0	21.9	27.1	23.0	535.0
Zanthoxylum armatum (RB49)	28.0	21.9	26.9	23.2	535.0
Avg.	29.1	21.0	27.4	22.5	534.0

Table 4.2 Nucleotide composition of each representative samples of barcode marker *ITS* with total length of sequences.

Species name (lab code)	T(U)	C	A	G	Ambiguous bases	Length (bp)
Aconitum spicatum (IB12)	17.9	29.5	23.3	29.4		722.0
Aconitum spicatum (IC30)	17.9	29.5	23.3	29.4		722.0
Polygonum molle (IB27)	18.7	30.5	21.1	29.7	1.0	684.0
Polygonum molle (IC23)	17.5	32.4	20.7	29.4		680.0
Alnus nepalensis (IB53)	16.6	32.4	21.4	29.6		697.0
Anaphalis contorta (IB8)	27.1	22.9	25.0	24.9	1.0	720.0
Arisaema speciosum (IC33)	20.1	28.4	21.4	30.1		763.0
Asparagus filicinus (IB38)	18.0	29.4	20.7	31.9		749.0
Berberis angulosa (IC7)	24.9	24.5	24.4	26.2	2.0	700.0
Berberis asiatica (IB56)	24.9	24.7	24.0	26.4		700.0
Berberis erythroclada (IB2)	24.7	24.9	24.3	26.1		700.0
Bergenia ciliata (IC8)	26.2	22.8	24.8	26.2	1.0	754.0
Betula utilis (IB10)	19.4	29.7	21.5	29.4		697.0
Betula utilis (IC25)	19.4	29.7	21.5	29.4		697.0
Bistorta macrophylla (IC20)	17.4	31.7	21.3	29.6		685.0
Cassiope fastigiata (IC3)	22.7	28.1	23.0	26.2	3.0	729.0
Cotoneaster integrifolius (IC21)	17.8	32.5	18.7	31.0	1.0	687.0
Dactylorhiza hatagirea (IB61)	24.8	22.0	24.3	28.9		733.0
Dactylorhiza hatagirea (IB7)	24.8	22.0	24.3	28.9		733.0
Dactylorhiza hatagirea (IC11)	24.8	22.0	24.3	28.9		733.0
Danthonia cumminsii (IC13)	18.1	29.8	20.7	31.4	12.0	693
Daphne bholua (IC31)	27.4	20.7	25.5	26.3		685.0
Daphne papyracea (IB21)	27.4	20.7	25.5	26.3		685.0
Delphinium brunonianum (IB14)	13.0	33.7	20.5	32.7	3.0	709.0
Delphinium himalayai (IB15)	17.9	30.0	23.2	29.0	2.0	723.0
Delphinium himalayai (IC5)	17.9	29.9	23.1	29.1	1.0	723.0
Engelhardia spicata (IB55)	19.9	31.1	20.6	28.5		734.0
Euonymus tingens (IB5)	17.6	29.3	23.0	30.1		704.0
Leucosceptrum canum (IB42)	15.8	33.3	19.7	31.2		702.0
Lindera neesiana (IB54)	14.1	33.6	19.2	33.1	1.0	715.0
Lyonia ovalifolia (IC26)	22.3	28.6	21.5	27.6		744.0
Magnolia campbellii (IB20)	19.0	30.1	22.0	28.8	1.0	695.0
Mahonia nepalensis (IB51)	25.7	23.9	24.3	26.1		700.0
Myrsine semiserrata (IB37)	21.7	28.0	23.0	27.3		718.0
Nardostachys grandiflora (IC2)	12.6	35.7	20.4	31.3	1.0	707.0
Panax pseudoginseng (IC9)	18.7	30.5	22.4	28.5		702.0
Paris polyphylla (IB3)	23.2	22.6	23.7	30.5		727.0
Paris polyphylla (IC1)	23.2	22.6	23.7	30.5		727.0
Quercus semecarpifolia (IB39)	17.7	32.6	19.3	30.4	1.0	685.0
Rheum australe (IC10)	15.3	35.8	18.0	31.0		668.0
Rhododendron anthopogon (IC4)	24.9	26.3	22.3	26.5	1.0	740.0
Rhododendron arboreum (IB22)	24.7	26.4	22.5	26.4	1.0	735.0
Rhododendron campanulatum (IB11)	24.3	26.6	22.6	26.5		736.0
Schisandra grandiflora (IC32)	25.1	26.7	20.0	28.2		765.0
Senecio graciliflorus (IC22)	27.2	23.2	23.7	25.9		734.0
Sorbus arachnoidea (IC24)	18.0	31.7	19.5	30.8		682.0
Swertia chirayita (IB1)	17.3	28.9	23.9	29.9		712.0
Swertia teres (IC6)	18.4	27.9	23.9	29.7		716.0
Taraxacum officinale (IB9)	24.8	25.2	23.5	26.5		733.0
Thamnocalamus nepalensis (IB60)	14.2	34.4	19.6	31.7		688.0

Trichosanthes wallichiana (IB40)	19.0	33.9	19.4	27.7	1.0	706.0
Urtica dioica (IB25)	24.2	25.1	23.3	27.4	3.0	681.0
Viburnum mullaha (IB30)	18.3	31.9	19.6	30.1		698.0
Yushania maling (IB19)	14.8	34.5	19.3	31.4		690.0
Zanthoxylum acanthopodium (IB36)	18.5	32.0	20.5	29.0		713.0
Zanthoxylum acanthopodium (IC38)	18.5	32.0	20.5	29.0		713.0
Zanthoxylum armatum (IB49)	17.3	33.3	19.4	29.9		715.0
Taxus wallichiana (IB4)	17.1	31.0	22.8	29.1		867.0
Avg.	20.4	28.6	22.1	28.9		714.1

Table 4.3 Nucleotide composition of each representative samples of barcode marker *ITS2* with total length of sequences.

Species name (lab code)	T(U)	C	A	G	Ambiguous bases	Length (bp)
Aconitum spicatum (IB12)	17.8	34.1	17.4	30.7		270.0
Aconitum spicatum (IC30)	17.8	34.1	17.4	30.7		270.0
Alnus nepalensis (IB53)	16.8	35.1	16.5	31.5		279.0
Anaphalis contorta (IB8)	28.2	26.3	18.8	26.7		266.0
Arisaema speciosum (IC33)	14.9	31.8	19.0	34.3		289.0
Asparagus filicinus (IB38)	13.5	34.0	17.5	35.0		297.0
Berberis angulosa (IC7)	24.5	28.2	19.4	27.8		273.0
Berberis asiatica (IB56)	24.2	28.6	18.7	28.6		273.0
Berberis erythroclada (IB2)	23.8	28.9	19.0	28.2		273.0
Bergenia ciliata (IC8)	26.0	24.2	23.9	26.0		289.0
Betula utilis (IB10)	19.5	33.2	16.2	31.0		277.0
Betula utilis (IC25)	19.5	33.2	16.2	31.0		277.0
Bistorta macrophylla (IC20)	16.1	36.0	14.7	33.2		286.0
Cassiope fastigiata (IC3)	19.7	31.9	20.4	28.0	1.0	280.0
Cotoneaster integrifolius (IC21)	15.9	38.0	12.2	33.9		271.0
Dactylorhiza hatagirea (IB61)	24.6	21.8	22.2	31.4		293.0
Dactylorhiza hatagirea (IB7)	24.6	21.8	22.2	31.4		293.0
Dactylorhiza hatagirea (IC11)	24.6	21.8	22.2	31.4		293.0
Danthonia cumminsii (IC13)	16.7	31.4	15.9	36.0	2.0	266.0
Daphne bholua (IC31)	29.4	20.2	22.4	27.9		272.0
Daphne papyracea (IB21)	29.4	20.2	22.4	27.9		272.0
Delphinium brunonianum (IB14)	11.2	39.0	15.6	34.2	1.0	270.0
Delphinium himalayai (IB15)	17.5	35.1	17.2	30.2	2.0	270.0
Delphinium himalayai (IC5)	17.5	34.9	17.1	30.5	1.0	270.0
Engelhardia spicata (IB55)	19.0	35.2	15.4	30.4		273.0
Euonymus tingens (IB5)	14.8	35.0	17.9	32.3		263.0
Leucosceptrum canum (IB42)	13.9	38.7	14.3	33.1		266.0
Lindera neesiana (IB54)	10.5	37.3	15.7	36.6	1.0	288.0
Lyonia ovalifolia (IC26)	21.0	33.0	16.3	29.7		276.0
Magnolia campbellii (IB20)	16.9	33.9	18.6	30.6	1.0	302.0
Mahonia nepalensis (IB51)	25.3	27.5	19.0	28.2		273.0
Myrsine semiserrata (IB37)	21.3	31.5	17.2	30.0		267.0
Nardostachys jatamansi (IC2)	11.0	41.3	12.8	34.9	1.0	282.0
Panax pseudoginseng (IC9)	18.2	35.0	16.8	30.0		280.0
Paris polyphylla (IB3)	21.3	26.2	18.1	34.4		282.0
Paris polyphylla (IC1)	21.3	26.2	18.1	34.4		282.0
Polygonum molle (IB27)	17.4	35.2	15.7	31.7		293.0
Polygonum molle (IC23)	16.2	36.8	14.8	32.3		291.0
Quercus semecarpifolia (IB39)	17.2	38.3	12.3	32.2	1.0	262.0
Rheum australe (IC10)	11.2	40.3	13.2	35.3		258.0
Rhododendron anthopogon (IC4)	22.3	31.6	18.8	27.3		282.0
Rhododendron arboreum (IB22)	23.3	30.5	18.3	28.0		279.0
Rhododendron campanulatum (IB11)	22.9	30.8	18.3	28.0		279.0
Schisandra grandiflora (IC32)	24.6	30.6	12.8	32.0		281.0
Senecio graciliflorus (IC22)	27.4	24.8	19.3	28.5		274.0
Sorbus arachnoidea (IC24)	16.9	37.1	12.7	33.3		267.0
Swertia chirayita (IB1)	17.9	32.6	17.6	31.9		279.0
Swertia teres (*IC6*)	19.1	31.1	18.0	31.8		283.0
Taraxacum officinale (IB9)	25.4	26.9	19.0	28.7		279.0
Taxus wallichiana (IB4)	22.6	27.2	15.4	34.8		279.0
Thamnocalamus nepalensis IB60)	10.7	37.8	15.2	36.3		270.0
Trichosanthes wallichiana (IB40)	16.8	38.6	15.2	29.4		303.0
Urtica dioica (IB25)	23.2	27.3	18.7	30.8		289.0

Species name (lab code)	T(U)	C	A	G	Ambiguous bases	Length (bp)
Viburnum mullaha (IB30)	17.2	35.1	14.7	33.0		279.0
Yushania maling (IB19)	11.3	38.7	14.6	35.4		274.0
Zanthoxylum acanthopodium (IB36)	17.4	35.9	15.9	30.8		276.0
Zanthoxylum acanthopodium (IC38)	17.4	35.9	15.9	30.8		276.0
Zanthoxylum armatum (IB49)	14.8	39.0	14.8	31.4		277.0
Avg.	19.3	32.1	17.2	31.3		278.0

4.7 Sequence Analysis on MEGA

Sequences analysis of two barcode regions were performed in MEGA. *ITS2* barcode marker was with high G+C ratio content ratio. *rbcL* showed relatively highly conserved region (60%-70%) and with lowest number of variable sites (223/537). *ITS* barcode marker showed highly variable sites and with high discriminatory power (80-90%) as shown in Table 4.4.

Table 4.4 The characteristics of the three barcode marker loci.

DNA markers	Total aligned length	Sequences length (bp)	G+C ratio %	Conserved sites (C)	Variable sites (V)	Parsimony information sites (Pi)	Singleton sites (S)
ITS	833	668-765	57.5	189	630	568	57
ITS2	343	258-303	63.5	37	297	267	23
rbcL	537	495-536	43.5	314	223	180	42

4.8 Species Resolution

Species resolution of the barcoding markers was evaluated through three different methods. Results obtained from different methods are discussed below.

4.8.1 Similarity searches by Simple BLAST and Optimized BLAST

In BLAST method, the identification success of species depends on the marker and availability of reference sequences in the GenBank. For some species, reference sequences in GenBank were available for only one of the two markers. The BLAST sequence matching included 65 *rbcL* and 58 *ITS* query sequences. The simple and optimized BLAST results based on sequence matching as well as the putative species identification for each of the 72 samples are given in Table 4.5. The simple *rbcL* BLAST searches results identified 15.38% (10 samples) to species level, 84.61% (55 samples) to genus level and 100% (65 samples) to family level. Similarly the simple *ITS* BLAST searches results identified 27.59% (16 samples) to species level, 86.20% (50 samples) to genus level and 100% (58 Samples) to family level. The optimized BLAST searches provided comparatively better results for both markers. The optimized *rbcL* BLAST searches results identified 29.23% (19 samples) to species level, 80.00% (52 samples) to genus level and 100% (65 samples) to family level. The optimized *ITS* BLAST searches results identified 65.52% (38 samples) to species level, 94.82% (55 samples) to genus level and 100% (58 Samples) to family level. Interestingly species discrimination was found relatively higher in *ITS2* marker than in *ITS*. Similar result was obtained through optimized BLAST method too.

The simple *ITS2* BLAST searches results identified 39.66% (23 samples) to species level, 89.66% (52 samples) to genus level and 100% (58 samples) to family level. The optimized

ITS2 BLAST searches results identified 72.41% (42 samples) to species level and 100% (58 samples) to genus level and family level.

Species discrimination using combined markers slightly improved the efficacies of the markers. Combined data from three markers (*rbcL*, *ITS* and *ITS2*) using simple BLAST sequence matching method identified 31.94% (23 samples) to species level, 88.89% (64 samples) to genus level and 100% (72 samples) to family level. Combined data from three markers using optimized BLAST sequence matching method identified 62.50% (45 samples) to species level, 87.50% (63 samples) to genus level and 100% (72 samples) to family level. Similar finding were also obtained by combination of *rbcL+ITS* and *rbcL+ITS2* as given in the Table 4.5.

However, with the correlation of morphological data with BLAST similarity data, the final accurate identification of species based on simple BLAST (*rbcL+ITS+ITS2*) is 20.83% (15) to species level, 84.72% (61) to genus level and 100% (72) to family level. Similarly, the final correct identification of species based on modified optimized BLAST (*rbcL+ITS+ITS2*) are 36.11% (26) to species level, 83.33% (60) to genus level, and 100% (72) to family level. The final list of correctly identified species (correlation of morphological identification) based on both normal and optimized BLAST is given in Table 4.6.

Table 4.5 Putative species level Identification rate for Simple BLAST and modified Optimized BLAST of the three markers.

Sequencing success (out of 72)		**Simple BLAST**		
Marker	Numbers	Species level	Genus level	Family level
rbcL	65	20.00% (13)	84.61% (55)	100% (65)
ITS	58	27.59% (16)	86.20% (50)	100% (58)
ITS2	58	39.66% (23)	89.66% (52)	100% (58)
rbcL+ITS	72	26.39% (19)	86.11% (62)	100% (72)
rbcL+ITS2	72	31.94% (23)	88.89% (64)	100% (72)
rbcL+ITS+ITS2	72	31.94% (23)	88.89% (64)	100% (72)
Sequencing success (Out of 72)		**Optimized BLAST**		
Markers	Number	Species level	Genus level	Family level
rbcL	65	29.23% (19)	80.00% (52)	100% (65)
ITS	58	65.52% (38)	94.82% (55)	100% (58)
ITS2	58	72.41% (42)	100% (58)	100% (58)
rbcL+ITS	72	56.94% (41)	88.89% (64)	100% (72)
rbcL+ITS2	72	59.72% (43)	91.67% (66)	100% (72)
rbcL+ITS+ITS2	72	62.50% (45)	87.50% (63)	100% (72)

Note: Sample C9 (Psuedo-ginseng) and C26 (Lyonia) are identified through non-modified optimized method (max score*(query cover/Identity).

Table 4.6 Correctly identified species by similarity searches (simple BLAST + Optimized BLAST) based on correlation with morphology.

Collection code	Lab Code	Name in Field	Scientific name based on morphology	Family	Normal BLAST	Optimized BLAST	Identification/DNA barcoding
CNRCP 01	B1	Chiraiyito	*Swertia chirayita*	Gentianaceae	*Swertia chirayita*	*Swertia chirayita*	*Swertia chirayita*
CNRCP 68	B56	Chutro	*Berberis asiatica*	Berberidaceae	*Berberis asiatica*	*Berberis asiatica*	*Berberis asiatica*

CNRCP 04	B4	Silinge	*Taxus wallichiana*	Taxaceae	*Taxus wallichiana*	*Taxus wallichiana*	*Taxus wallichiana*
CNRCP 015	B8	Buki phool	*Anaphalis contorta*	Asteraceae	*Anaphalis contorta*	*Anaphalis contorta*	*Anaphalis contorta*
CNRCP 016	B9	Taraxacu m	*Taraxacum officinale*	Asteraceae		*Taraxacum officinale*	*Taraxacum officinale*
CNRCP 017	B10	Bhoj patra	*Betula utilis*	Betulaceae		*Betula utilis*	*Betula utilis*
CN42	C25	bhoj patra	*Betula utilis*	Betulaceae		*Betula utilis*	*Betula utilis*
CNRCP28	B20	Champ	*Magnolia campbellii*	Magnoliaceae	*Magnolia campbellii*	*Magnolia campbellii*	*Magnolia campbellii*
CNRCP35	B27	Thotne	*Polygonum molle*	Polygonaceae	*Polygonum molle*	*Polygonum molle*	*Polygonum molle*
CNRCP49	B37	Kali kath	*Myrsine semiserrata*	Myrsinaceae		*Myrsine semiserrata*	*Myrsine semiserrata*
CNRCP 51	B39	Kharsu	*Quercus semecarpifolia*	Fagaceae	*Quercus semecarpifolia*	*Quercus semecarpifolia*	*Quercus semecarpifolia*
CNRCP 54	B42	Dhurseli	*Leucosceptrum canum*	Lamiaceae	*Leucosceptrum canum*	*Leucosceptrum canum*	*Leucosceptrum canum*
CNRCP 61	B49	Ankhe Timur	*Zanthoxylum armatum*	Rutaceae	*Zanthoxylum armatum*	*Zanthoxylum armatum*	*Zanthoxylum armatum*
CNRCP 63	B51	Bhutro	*Mahonia nepalensis*	Berberidaceae		*Mahonia nepalensis*	*Mahonia nepalensis*
CNRCP 65	B53	Utis	*Alnus nepalensis*	Betulaceae		*Alnus nepalensis*	*Alnus nepalensis*
CNRCP48	B36	Boke timur	*Zanthoxylum acanthopodium*	Rutaceae		*Zanthoxylum acanthopodium*	*Zanthoxylum acanthopodium*
CN56	C38	Boke Timur	*Zanthoxylum acanthopodium*	Rutaceae		*Zanthoxylum acanthopodium*	*Zanthoxylum acanthopodium*
CNRCP18	B11	Gurash	*Rhododendron campanulatum*	Ericaceae		*Rhododendron sp.*	*Rhododendron sp.*
CNRCP42	B30	Malo	*Viburnum mullaha*	Sambucaceae		*Viburnum mullaha*	*Viburnum mullaha*
CNRCP 67	B55	Mauwa	*Engelhardia spicata*	Juglandaceae	*Engelhardia spicata*	*Engelhardia spicata*	*Engelhardia spicata*
CN2	C2	Jatamansi	*Nardostachys grandiflora/jataman si*	Valerianaceae	*Nardostachys grandiflora/jataman si*	*Nardostachys grandiflora/jataman si*	*Nardostachys grandiflora/jataman si*
CN5	C3	Cassiope-Ericaceae	*Cassiope fastigiata*	Ericaceae		*Cassiope fastigiata*	*Cassiope fastigiata*
CN14	C8	Bergenia	*Bergenia purpurascens*	Saxifragaceae	*Bergenia purpurascens*	*Bergenia purpurascens*	*Bergenia purpurascens*
CN16	C9	Pseudo-ginseng	*Panax pseudo-ginseng*	Araliaceae		*Panax pseudoginseng*	*Panax pseudoginseng*
CN23	C13	Salim grass	*Danthonia cumminsii*	Poaceae	*Danthonia cumminsii*	*Danthonia cumminsii*	*Danthonia cumminsii*
CN43	C26	Lynoia	*Lyonia ovalifolia*	Ericaceae	*Lyonia ovalifolia*	*Lyonia ovalifolia*	*Lyonia ovalifolia*
CN45	C28	Aruncus	*Aruncus dioicus*	Rosaceae	*Aruncus dioicus*	*Aruncus dioicus*	*Aruncus dioicus*

4.8.2 Genetic divergence

4.8.2.1 Divergence of Core data set

Genetic divergences among different samples for each locus and various combinations of loci (with core data set) were evaluated by TaxonDNA v1.8 using K2P pairwise distance model. We found that genetic divergence of *ITS* locus has relatively higher than *rbcL* locus. Overall interspecific divergence (K2P) within different taxonomic categories of tested samples of the markers *rbcL* and *ITS* is shown in Figure 4.11. In addition, Interspecific and Intraspecific of each locus and various combinations were evaluated for the samples that have more than one individual. The genetic divergences of samples having single individuals were evaluated by comparing with reference sequences retrieved from NCBI (Appendix 3). In the present analysis, total 13 species that have more than one individual thus their intra- and interspecific divergence were estimated.

In the same way mean intra- and interspecific genetic distances for all the tested loci was also estimated based on pairwise distribution.

Based on genetic divergence, *ITS2* exhibited highest mean intra- (0.0891) and interspecific (0.0970) genetic distances followed by *ITS* marker. *rbcL* exhibited lowest mean intra and interspecific divergences. In this study, no single or multiple loci revealed clear barcode gaps; all the markers overlapped between the intra- and interspecific distances. Analysis revealed that interspecific distance of single or multiple loci was significantly higher than intraspecific distance of single or multiple loci (Table 4.7), that indicates barcode gap exists. Based on the frequency distribution of pairwise distances, all the loci exhibited clear barcoding gaps between range of intra- and interspecific distances (Figure 4.10). The numbers of species with a barcoding gap have shown substantial sequence variation (Mishra et al., 2017) that indicates effective discrimination of the studied species.

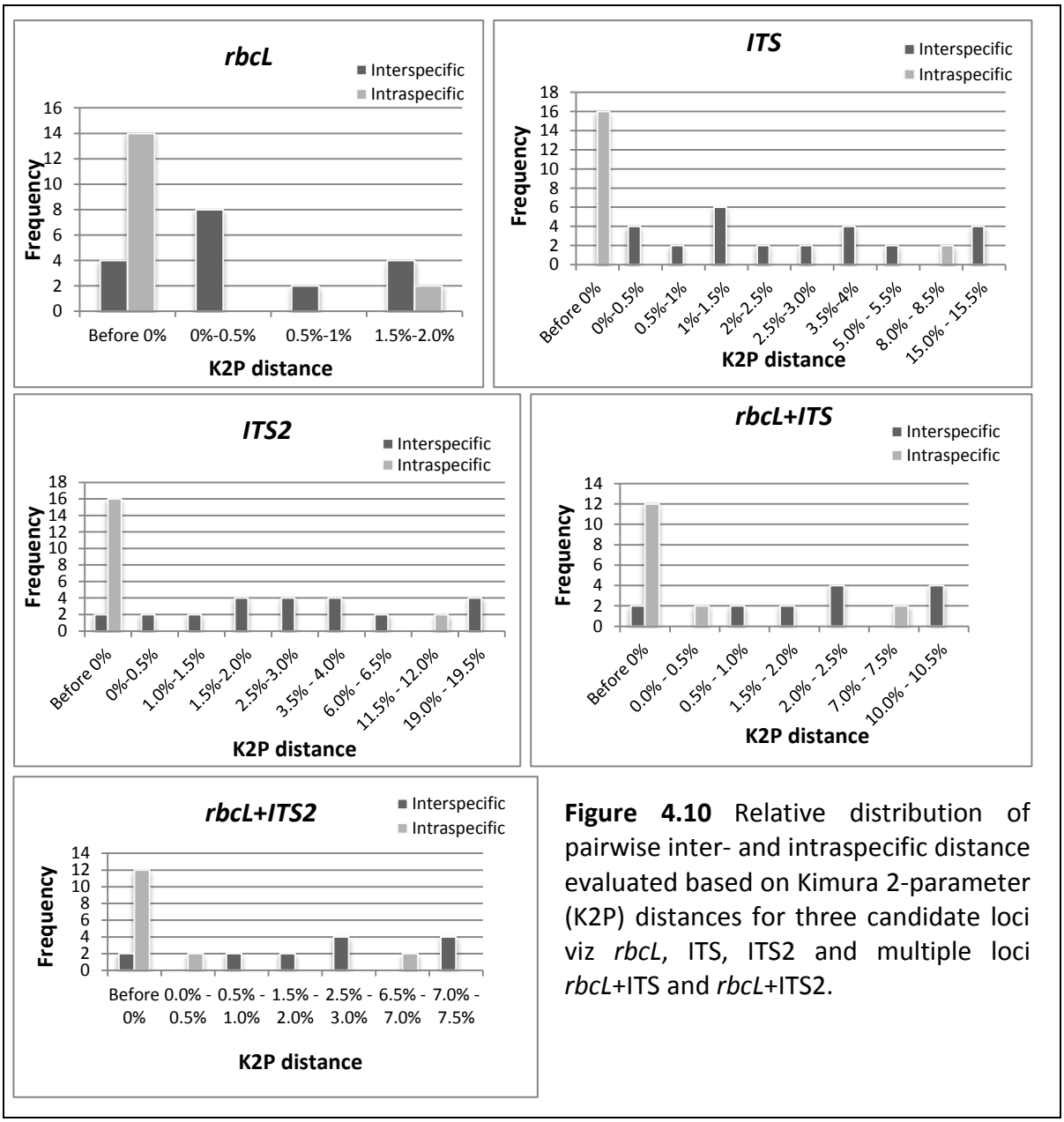


Figure 4.10 Relative distribution of pairwise inter- and intraspecific distance evaluated based on Kimura 2-parameter (K2P) distances for three candidate loci viz *rbcL*, ITS, ITS2 and multiple loci *rbcL*+ITS and *rbcL*+ITS2.

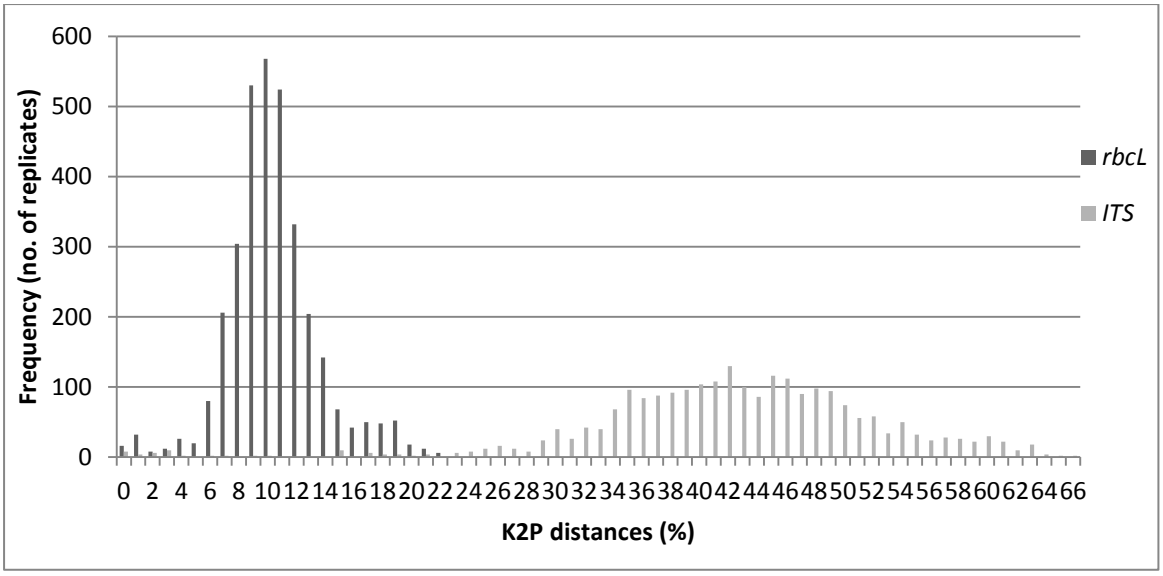


Figure 4.11 Relative distribution of inter-specific (K2P) divergence within different taxonomic categories for the 72 individuals of 63 species analysed using *rbcL* and *ITS* markers.

Genetic distance of species having more than one individual

Among tested loci, *ITS2* showed high interspecific divergence viz. *Delphinium brunonianum* (B14) and *Delphinium himalayai* (B15 &C5) that ranges from 19.0% to 19.5%. Similarly, the highest intraspecific variation was found within *ITS2* locus among tested sample (B27 & C23) of *Polygonum molle* that range from 11.5% to 12.0% as shown in Table 4.10. *rbcL* showed least inter- and intraspecific divergence in all the tested samples (Table 4.7 and Table 4.8). *ITS* and *ITS2* marker of some high value plant species viz. *Zanthoxylum acanthopodium* (B36 & C38) and *Zanthoxylum armatum* (B49); *Swertia chirayita* (B1) and *Swertia teres* (C6) have showed to be relatively higher interspecific divergence (3.0% to 5.0%) than other tested samples Similar finding found in *Berberis* spp (*B. angulosa*, *B. erythroclada* and *B. asiatica*) those exhibited 1.0% to 2.0% interspecific divergence among *ITS* and *ITS2* loci. Species of of *Rhododendron arboreum*, *R. campanulatum* and *R. anthopogon* have also shown to have interspecific divergence 0.5% to 1.0% for *rbcL* and *ITS* but of higher interspecific divergence 2.5% to 3.0% for *ITS2* markers as given in Table 4.8; Table 4.9 and Table 4.10. *ITS2* performed very well in the species discriminations for above mentioned species. Some high value plant species viz. *Paris polyphylla*, *Dactylorhiza hatagirea*, *Aconitum spicatum* exhibited no intraspecific variation for both the tested regions.

Table 4.7 Summary of genetic distances generated using K2P model for the candidate barcode and *ITS* various combinations.

Bacode locus/loci	Intraspecific distance			Interspecific distance			Overlap
	Minimum	Maximum	Mean	Minimum	Maximum	Mean	
rbcL	0.0000	0.0170	0.0085	0.0000	0.0173	0.0086	1.70%
ITS	0.0000	0.1050	0.0525	0.0000	0.1709	0.0854	10.36%
ITS2	0.0000	0.1782	0.0891	0.0000	0.1941	0.0970	17.82%
rbcL+ITS	0.0000	0.0747	0.03735	0.0000	0.1018	0.0509	7.47%
rbcL+ITS2	0.0000	0.0698	0.0349	0.0000	0.0719	0.0359	8.80%

Table 4.8 Interspecific and Intraspecific divergence of between species of rbcL marker of core data set.

Interspecific divergence between species of *rbcL* marker		
Replicates_1 (lab code)	**Replicates_2 (lab code)**	**Variation**
Berberis angulosa (RC7)	*Berberis erthoyclada* (RB2)	0.00%
Daphne bholua (RC31)	*Daphne papyracea* (RB21)	0.00%
Michelia doltsopa (RB29)	*Michelia kisopa* (RB45)	0.19%
Quercus lamellosa (RB28)	*Quercus semecarpifolia* RB39)	0.19%
Zanthoxylum acanthopodium (RB36)	*Zanthoxylum armatum* (RB49)	0.19%
Rhododendron anthopogon (RC4)	*Rhododendron campanulatum* (RB11)	0.93%
Delphinium brunonianum (RB14)	*Delphinium himalayai* (RB15)	1.73%
Delphinium himalayai (RB15)	*Delphinium brunonianum* (RB14)	1.73%
Intraspecific divergence between species of *rbcL* marker		
Betula utilis (RB10)	*Betula utilis* (RC25)	0.00%
Dactylorhiza hatagirea (RB61)	*Dactylorhiza hatagirea* (RB7)	0.00%
Dactylorhiza hatagirea (RB61)	*Dactylorhiza hatagirea* (RC11)	0.00%
Dactylorhiza hatagirea (RB7)	*Dactylorhiza hatagirea* (RC11)	0.00%
Delphinium himalayai (RB15)	*Delphinium himalayai* (RC5)	0.00%
Paris polyphylla (RB3)	*Paris polyphylla* (RC1)	0.00%
Zanthoxylum acanthopodium (RB36)	*Zanthoxylum acanthopodium* (RC38)	0.00%
Polygonum molle (RB27)	*Polygonum molle* (RC23)	1.68%

Table 4.9 Interspecific and Intraspecific divergence between species of *ITS* marker with core data set.

Interspecific divergence between species of *ITS* marker		
Replicates_1 (lab code)	**Replicates_2 (lab code)**	**Variation**
Daphne bholua IC31)	*Daphne papyracea* (IB21)	0.15%
Rhododendron campanulatum IB11)	*Rhododendron arboreum* (IB22)	0.68%
Berberis angulosa IC7)	*Berberis erythroclada* (IB2)	1.16%
Berberis angulosa IC7)	*Berberis asiatica* (IB56)	1.30%
Berberis asiatica IB56)	*Berberis erythroclada* (IB2)	1.30%
Rhododendron anthopogon IC4)	*Rhododendron campanulatum* (IB11)	2.36%
Rhododendron anthopogon IC4)	*Rhododendron arboreum* (IB22)	2.64%
Zanthoxylum acanthopodium IB36)	*Zanthoxylum armatum* (IB49)	3.62%
Zanthoxylum acanthopodium IC38)	*Zanthoxylum armatum* (IB49)	3.62%
Swertia chirayita IB1)	*Swertia teres* (IC6)	5.27%
Delphinium brunonianum IB14)	*Delphinium himalayai* (IC5)	15.09%
Intraspecific divergence between species of *ITS* marker		
Aconitum spicatum (IB12)	*Aconitum spicatum* (IC30)	0.00%
Betula utilis (IB10)	*Betula utilis* (IC25)	0.00%
Dactylorhiza hatagirea (IB61)	*Dactylorhiza hatagirea* (IB7)	0.00%
Dactylorhiza hatagirea (IB7)	*Dactylorhiza hatagirea* (IC11)	0.00%
Dactylorhiza hatagirea (IC11)	*Dactylorhiza hatagirea* (IB61)	0.00%
Delphinium himalayai (IB15)	*Delphinium himalayai* (IC5)	0.00%
Paris polyphylla (IB3)	*Paris polyphylla* (IC1)	0.00%
Zanthoxylum acanthopodium (IB36)	*Zanthoxylum acanthopodium* (IC38)	0.00%
Polygonum molle (IB27)	*Polygonum molle* (IC23)	8.28%

Table 4.10 Interspecific and Intraspecific divergence between species of *ITS2* marker with core data set.

Interspecific divergence between species of *ITS2* marker		
Replicates_1 (lab code)	**Replicates_2 (lab code)**	**Variation**
Daphne bholua (IC31)	*Daphne papyracea* (IB21)	0.00%
Rhododendron arboreum (IB22)	*Rhododendron campanulatum* (IB11)	0.36%
Berberis asiatica (IB56)	*Berberis erythroclada* (IB2)	1.49%
Berberis angulosa (IC7)	*Berberis asiatica* (IB56)	1.87%
Berberis angulosa (IC7)	*Berberis erythroclada* (IB2)	1.87%
Rhododendron anthopogon (IC4)	*Rhododendron arboreum* (IB22)	2.93%
Rhododendron anthopogon (IC4)	*Rhododendron campanulatum* (IB11)	2.93%
Zanthoxylum acanthopodium (IB36)	*Zanthoxylum armatum* (IB49)	3.75%
Zanthoxylum armatum (IB49)	*Zanthoxylum acanthopodium* (IB36)	3.75%
Swertia chirayita (IB1)	*Swertia teres* (IC6)	6.46%
Delphinium himalayai (IC5)	*Delphinium brunonianum* (IB14)	19.34%
Delphinium himalayai (IB15)	*Delphinium brunonianum* (IB14)	19.42%
Intraspecific divergence between species of *ITS2* marker		
Aconitum spicatum (IB12)	*Aconitum spicatum* (IC30)	0.00%
Betula utilis (IB10)	*Betula utilis* (IC25)	0.00%
Dactylorhiza hatagirea (IB61)	*Dactylorhiza hatagirea* (IB7)	0.00%
Dactylorhiza hatagirea (IB61)	*Dactylorhiza hatagirea* (IC11)	0.00%
Dactylorhiza hatagirea (IB7)	*Dactylorhiza hatagirea* (IC11)	0.00%
Delphinium himalayai (IB15)	*Delphinium himalayai* (IC5)	0.00%
Paris polyphylla (IB3)	*Paris polyphylla* (IC1)	0.00%
Zanthoxylum acanthopodium (IB36)	*Zanthoxylum acanthopodium* (IC38)	0.00%
Polygonum molle (IB27)	*Polygonum molle* (IC23)	11.58%

4.8.2.2 Analysis of core data & References data

Species that has single individual sequence were evaluated through comparing with the sequence of same species retrieved from GenBank. A total of 14 of *rbcL*, 17 of *ITS* and 19 of *ITS2* reference sequences of the corresponding species were retrieved from NCBI database. When retrieving sequences high priority was given to Nepalese species and the second preference was given to the species of China and India only. Based on genetic divergence, *ITS2* and *ITS* sequences of *Polygonum molle* (C23) revealed highest intraspecific variation (11.58% and 10.54%) with the retrieved reference sequences (EF653687.1) (Table 4.12 and Table 4.13). *Schisandra grandiflora* and *ITS* reference sequence (Accession: KP689672.1) of *ITS2* marker showed intraspecific sequence divergence of 10.66%. Lowest genetic divergence was found among core data set and those with retrieved reference sequences (Table 4.11).

Similarly *ITS* sequences comparison done between studies species viz *Rheum australe, Lyonia ovalifolia, Swertia chirayita, Bistorta macrophylla, Taraxacum* officinale with respective references sequences from GenBank exhibited 1.5% to 8% intraspecific variation. Other evaluated species are *Taxus wallichiana* that has <1% intraspecific variation (Table 4.12). Moreover, comparison of *ITS2* sequences of *Polygonum molle, Swertia chirayita, Engelhardia spicata, Bistorta macrophylla and Taraxacum officinale* have shown 1.5% to 5% intraspecific variation, while other species including high value species like *Taxus wallichiana*, *Rheum australe* and *Zanthoxylum armatum* exhibited <1% intraspecific variation (Table 4.13).

Table 4.11 Intraspecific divergence between core data set (14 sequences) of *rbcL* marker and and their subsequent reference sequences retrieved from GenBank.

Intraspecific Divergence				
Core data_*rbcL* (lab code)	**Reference sequences**	**Location**	**Accession**	**Variation**
Alnus nepalensis (RB53)	*Alnus nepalensis*	China	KF418930.1	0.00%
Cassiope fastigiata (RC3)	*Cassiope fastigiata*	China	JF941134.1	0.00%
Lycopodium japonicum (RB50)	*Lycopodium japonicum*	China	MF786611.1	0.00%
Myrsine semiserrata (RB37)	*Myrsine semiserrata*	China	MG950622.1	0.00%
Polygonum molle (RB27)	*Polygonum molle*	China	JF943512.1	0.00%
Rhododendron anthopogon (RC4)	*Rhododendron anthopogon*	China	KM606531.1	0.00%
Schisandra grandiflora (RC32)	*Schisandra grandiflora*	China	KP689892.1	0.00%
Zanthoxylum armatum (RB49)	*Zanthoxylum armatum*	India	KJ667668.1	0.00%
Bergenia purpurascens (RC8)	*Bergenia purpurascens*	India	KY986464.1	0.12%
Bistorta macrophylla (RC20)	*Bistorta macrophylla*	China	JF943511.1	0.19%
Aruncus dioicus (RC28)	*Aruncus dioicus*	China	KF154894.1	0.78%
Taxus wallichiana (RB4)	*Taxus wallichiana*	China	HM591034.1	0.79%
Leucosceptrum canum (RB42)	*Leucosceptrum canum*	China	KR608485.1	0.99%
Polygonum molle (RC23)	*Polygonum molle*	China	JF943512.1	1.77%

Table 4.12 Intraspecific divergences among core data set (18 sequences) of *ITS* marker and their subsequent reference sequences retrieved from GenBank**.**

Intraspecific Divergence				
Core data_*ITS* (lab code)	**Reference sequences**	**Location**	**Accession**	**Variation**
Rhododendron campanulatum (IB11)	*Rhododendron campanulatum*	China	KM605963.1	0.00%
Taxus wallichiana (IB4)	*Taxus wallichiana*	China	JX680623.1	0.00%
Schisandra grandiflora (IC32)	*Schisandra grandiflora*	China	JX680623.1	0.15%
Anaphalis contorta (IB8)	*Anaphalis contorta*	China	JQ895423.1	0.16%
Leucosceptrum canum (IB42)	*Leucosceptrum canum*	China	KR608738.1	0.17%
Myrsine semiserrata (IB37)	*Myrsine semiserrata*	China	MG877850.1	0.33%
Bergenia purpurascens (IC8)	*Bergenia purpurascens*	China	EU239674.1	0.46%
Cassiope fastigiata (IC3)	*Cassiope fastigiata*	China	AF393438.1	0.48%
Zanthoxylum armatum (IB49)	*Zanthoxylum armatum*	China	DQ016546.1	0.49%
Berberis angulosa (IC7)	*Berberis angulosa*	India	HM347898.1	0.68%
Polygonum molle (IB27)	*Polygonum molle*	China	EF653687.1	0.87%
Rheum australe (IC10)	*Rheum australe*	China	KF258683.1	1.63%
Lyonia ovalifolia (IC26)	*Lyonia ovalifolia*	China	KP092599.1	1.76%
Swertia chirayita (IB1)	*Swertia chirayita*	Nepal	JX569817.1	2.00%
Bistorta macrophylla (IC20)	*Bistorta macrophylla*	China	JN235093.1	2.11%
Magnolia campbellii (IB20)	*Magnolia campbellii*	China	KU853479.1	5.95%
Taraxacum officinale (IB9)	*Taraxacum officinale*	China	MG519289.1	7.92%
Polygonum molle (IC23)	*Polygonum molle*	China	EF653687.1	10.54%

Table 4.13 Intraspecific divergence between core data set (20 sequences) of *ITS2* marker and their subsequent reference sequences retrieved from GenBank.

Intraspecific Divergence				
Core data_*ITS2* (lab code)	**Reference sequences**	**Location**	**Accession**	**Variation**
Anaphalis contorta (IB8)	*Anaphalis contorta*	China	JQ895425.1	0.00%
Leucosceptrum canum (IB42)	*Leucosceptrum canum*	China	KR608738.1	0.00%
Myrsine semiserrata (IB37)	*Myrsine semiserrata*	China	MG877850.1	0.00%
Quercus semecarpifolia (IB39)	*Quercus semecarpifolia*	China	KY624385.1	0.00%
Rhododendron anthopogon (IC4)	*Rhododendron anthopogon*	China	KM605954.1	0.00%

Intraspecific Divergence				
Core data_*ITS2* (lab code)	**Reference sequences**	**Location**	**Accession**	**Variation**
Rhododendron campanulatum (IB11)	*Rhododendron campanulatum*	China	KM605963.1	0.00%
Taxus wallichiana (IB4)	*Taxus wallichiana*	China	KX981186.1	0.00%
Alnus nepalensis (IB53)	*Alnus nepalensis*	China	FJ825418.1	0.43%
Cassiope fastigiata (IC3)	*Cassiope fastigiata*	China	AF393438.1	0.43%
Rhododendron arboreum (IB22)	*Rhododendron arboreum*	China	KM605819.1	0.44%
Rheum australe (IC10)	*Rheum australe*	China	KF258683.1	0.86%
Lyonia ovalifolia (IC26)	*Lyonia ovalifolia*	China	MF785536.1	0.87%
Zanthoxylum armatum (IB49)	*Zanthoxylum armatum*	India	KM887380.1	0.89%
Polygonum molle (IB27)	*Polygonum molle*	China	EF653687.1	1.29%
Swertia chirayita (IB1)	*Swertia chirayita*	Nepal	JX569818.1	1.34%
Engelhardia spicata (IB55)	*Engelhardia spicata*	China	KR532080.1	2.73%
Bistorta macrophylla (IC20)	*Bistorta macrophylla*	China	JN235093.1	3.58%
Taraxacum officinale (IB9)	*Taraxacum officinale*	China	MG519290.1	4.50%
Schisandra grandiflora (IC32)	*Schisandra grandiflora*	China	KP689672.1	10.66%
Polygonum molle (IC23)	*Polygonum molle*	China	EF653687.1	11.58%

4.8.2.3 Divergence among families

In this analysis, species enlisted by Government of Nepal for protection and conservation, CITES listed species and threatened species have been evaluated to understand genetic divergence among species to their family level. Genetic divergence between families of different species for *ITS* marker revealed higher differences than for *rbcL* marker. The highest genetic variation among *ITS* and *rbcL* marker have exhibited 63.29% and 18.78% among families of Asparagaceae with Orchidaceae and Polygonaceae with Taxaceae respectively. The lowest genetic variation among *ITS* and *rbcL* marker have exhibited among families of Caprifoliaceae with Rutaceae (29.01%) and Asparagaceae with Orchidaceae (3.88%). In *rbcL* sequence matrix, genetic divergence of Taxaceae exhibited highest (14.5% - 19%) divergence with rest of the species (highlighted in Table 4.14), other families have relatively lower genetic distance. In *ITS* genetic divergence analysis, genetic distance of Orchidaceae and rest of other sampled families except Polygonaceae has exhibited >55%; Asparagaceae with Berberidaceae and Saxifragaceae also exhibited >55% genetic distance (highlighted in Table 4.15). Other families replicate have relatively lower than 55% genetic distance.

Table 4.14 Genetic divergence among families (*rbcL* data set) including only species those considered as protected plants of Nepal, CITES and research and developments.

Replicates family_1	**Replicates family_2**	**Variation**
Asparagaceae	Orchidaceae	3.88%
Asparagaceae	Melanthiaceae	6.10%
Asparagaceae	Saxifragaceae	7.16%
Asparagaceae	Berberidaceae	8.98%
Asparagaceae	Gentianaceae	9.22%
Asparagaceae	Caprifoliaceae	9.53%
Asparagaceae	Rutaceae	9.87%
Asparagaceae	Scrophulariaceae	9.89%
Asparagaceae	Rubiaceae	10.13%
Asparagaceae	Ranunculaceae	10.31%
Asparagaceae	Polygonaceae	10.70%

Replicates family_1	Replicates family_2	Variation
Asparagaceae	Taxaceae	14.24%
Berberidaceae	Ranunculaceae	6.61%
Berberidaceae	Saxifragaceae	7.29%
Berberidaceae	Rutaceae	8.40%
Berberidaceae	Orchidaceae	8.82%
Berberidaceae	Caprifoliaceae	9.26%
Berberidaceae	Gentianaceae	10.19%
Berberidaceae	Scrophulariaceae	10.24%
Berberidaceae	Melanthiaceae	10.54%
Berberidaceae	Polygonaceae	11.04%
Berberidaceae	Rubiaceae	12.11%
Berberidaceae	Taxaceae	17.86%
Caprifoliaceae	Saxifragaceae	7.11%
Caprifoliaceae	Scrophulariaceae	8.11%
Caprifoliaceae	Gentianaceae	8.56%
Caprifoliaceae	Polygonaceae	9.08%
Caprifoliaceae	Rutaceae	9.13%
Caprifoliaceae	Orchidaceae	9.66%
Caprifoliaceae	Rubiaceae	10.26%
Caprifoliaceae	Ranunculaceae	10.69%
Caprifoliaceae	Melanthiaceae	12.27%
Caprifoliaceae	Taxaceae	16.52%
Gentianaceae	Scrophulariaceae	7.61%
Gentianaceae	Rubiaceae	7.70%
Gentianaceae	Saxifragaceae	7.73%
Gentianaceae	Rutaceae	9.54%
Gentianaceae	Polygonaceae	10.34%
Gentianaceae	Orchidaceae	10.41%
Gentianaceae	Ranunculaceae	11.53%
Gentianaceae	Melanthiaceae	12.84%
Gentianaceae	Taxaceae	16.80%
Melanthiaceae	Orchidaceae	6.26%
Melanthiaceae	Saxifragaceae	9.39%
Melanthiaceae	Rutaceae	11.73%
Melanthiaceae	Ranunculaceae	11.98%
Melanthiaceae	Polygonaceae	12.58%
Melanthiaceae	Scrophulariaceae	12.87%
Melanthiaceae	Rubiaceae	14.49%
Melanthiaceae	Taxaceae	16.93%
Orchidaceae	Saxifragaceae	8.35%
Orchidaceae	Scrophulariaceae	9.80%
Orchidaceae	Rutaceae	10.44%
Orchidaceae	Ranunculaceae	10.97%
Orchidaceae	Polygonaceae	11.76%
Orchidaceae	Rubiaceae	12.23%
Orchidaceae	Taxaceae	16.06%
Polygonaceae	Saxifragaceae	8.45%
Polygonaceae	Rutaceae	10.45%
Polygonaceae	Scrophulariaceae	11.36%
Polygonaceae	Rubiaceae	11.87%

Replicates family_1	Replicates family_2	Variation
Polygonaceae	Ranunculaceae	12.38%
Polygonaceae	Taxaceae	18.78%
Ranunculaceae	Saxifragaceae	8.69%
Ranunculaceae	Rutaceae	9.93%
Ranunculaceae	Scrophulariaceae	9.98%
Ranunculaceae	Rubiaceae	12.34%
Ranunculaceae	Taxaceae	18.19%
Rubiaceae	Scrophulariaceae	9.37%
Rubiaceae	Saxifragaceae	10.08%
Rubiaceae	Rutaceae	12.58%
Rubiaceae	Taxaceae	17.77%
Rutaceae	Saxifragaceae	5.97%
Rutaceae	Scrophulariaceae	9.07%
Rutaceae	Taxaceae	17.35%
Saxifragaceae	Scrophulariaceae	8.11%
Saxifragaceae	Taxaceae	15.09%
Scrophulariaceae	Taxaceae	16.68%

Table 4.15 Genetic divergence among families (*ITS* data set) including only species those considered as protected plants of Nepal, CITES and research and developments.

Replicates family_1	Replicates family_2	Variation
Asparagaceae	Ranunculaceae	47.50%
Asparagaceae	Rutaceae	47.82%
Asparagaceae	Gentianaceae	52.71%
Asparagaceae	Melanthiaceae	54.27%
Asparagaceae	Berberidaceae	55.22%
Asparagaceae	Saxifragaceae	55.77%
Asparagaceae	Orchidaceae	63.29%
Berberidaceae	Gentianaceae	40.27%
Berberidaceae	Caprifoliaceae	41.28%
Berberidaceae	Rutaceae	42.75%
Berberidaceae	Melanthiaceae	47.71%
Berberidaceae	Orchidaceae	55.15%
Caprifoliaceae	Rutaceae	29.01%
Caprifoliaceae	Polygonaceae	32.01%
Caprifoliaceae	Gentianaceae	32.50%
Caprifoliaceae	Ranunculaceae	34.23%
Caprifoliaceae	Saxifragaceae	43.45%
Caprifoliaceae	Asparagaceae	45.18%
Caprifoliaceae	Melanthiaceae	50.68%
Caprifoliaceae	Orchidaceae	61.62%
Gentianaceae	Rutaceae	35.66%
Gentianaceae	Melanthiaceae	49.49%
Gentianaceae	Orchidaceae	58.84%
Melanthiaceae	Polygonaceae	48.47%
Melanthiaceae	Rutaceae	49.55%
Melanthiaceae	Ranunculaceae	54.49%
Melanthiaceae	Saxifragaceae	54.77%
Melanthiaceae	Orchidaceae	59.19%
Orchidaceae	Polygonaceae	53.13%

Replicates family_1	Replicates family_2	Variation
Rutaceae	Orchidaceae	60.14%
Orchidaceae	Ranunculaceae	60.38%
Orchidaceae	Saxifragaceae	64.77%
Polygonaceae	Rutaceae	30.06%
Polygonaceae	Gentianaceae	36.06%
Polygonaceae	Ranunculaceae	39.33%
Polygonaceae	Berberidaceae	39.57%
Polygonaceae	Saxifragaceae	40.64%
Polygonaceae	Asparagaceae	46.40%
Ranunculaceae	Gentianaceae	38.69%
Ranunculaceae	Berberidaceae	38.88%
Ranunculaceae	Saxifragaceae	48.16%
Rutaceae	Ranunculaceae	38.34%
Saxifragaceae	Rutaceae	39.89%
Saxifragaceae	Berberidaceae	42.52%
Saxifragaceae	Gentianaceae	44.40%

4.8.3 Phylogenetic Inference

The phylogenetic tree was reconstructed using three loci viz. *rbcL*, *ITS*, and *ITS2* with their combinations viz. *rbcL+ITS* and *rbcL+ITS2* loci. Sixtyfive species having *rbcL* sequences were used for tree reconstruction, followed by 58 species for *ITS* and *ITS2* and 51 species of both *rbcL+ITS* and *rbcL+ITS2*. Generally, species generated separate clusters (monophyletic clades) in the tree with bootstrap value >50% were considered to be discriminated successfully.

Phylogenetic reconstruction of *rbcL* sequences placed different species of the same family in single clade of the tree. Despite of low bootstrap supports for few clades (>50%) tree topology was almost similar to the global tree of APG IV of *rbcL* locus. Average bootstrap value (when considered >50%) of *rbcL* based phylogeny was found to be 84.61%. In *ITS*, the tree was also supported with the global tree of APG IV of *ITS* locus. However in both phylogenetic trees, families were almost in similar positions. Average bootstrap value (only considered >50%) of *ITS* and *ITS2* based phylogeny was found to be 74.13% and 79.31% respectively. In combined locus analysis, average bootstrap value (only considered >50%) of *rbcL+ITS* and *rbcL+ITS2* were found to be 90.19% and 88.23% respectively. In combined locus analysis of *rbcL* with *ITS* and *ITS2*, despite of low bootstrap supports for few clades (>50%) tree topology was almost similar to the global tree of APG IV (Figure 4.14 and Figure 4.15).

Based on the NJ tree, multi-loci combinations viz. *rbcL+ITS* and *rbcL+ ITS2* tree topology were supported with the global tree of APG IV *rbcL*/*ITS*. *rbcL+ITS2* combinations have shown highest species discrimination representing 71.91% of monophyletic clades among tested loci, followed by 68.53% in *rbcL+ITS* combination. Among single locus, *rbcL* marker has discriminated 62.92% of samples at species level, followed by 51.68% by *ITS2* and 48.31% by *ITS* (Table 4.16). Each barcode marker successfully separated up to genus level by forming monophyletic clades (those are successfully discriminated), in which each clade appeared distinctly distant from other clades (Figure 4.12; Figure 4.13; Figure 4.14 and Figure 4.15).

Table 4.16 Summary of species discrimination by tree method.

Markers (No. of Sequences)	Bootstrap support (≥50%)	Species discrimination (out of 89)
	NJ method	NJ method
rbcL (65)	84.61% (55)	61.79% (55)
ITS (58)	74.13% (43)	48.31% (43)
ITS2 (58)	79.31% (46)	51.68% (46)
rbcL+ITS (51)	90.19% (46)	68.53% (61)
rbcL+ITS2 (51)	88.23% (45)	71.91% (64)

4.9 Species discrimination summary

Among total of 89 studied species of plants belonging to 63 genera and 41 families, 72 species could be successfully discriminated via one or more analytical methods by single locus or multi-locus combinations (Table 4.17). However, 17 species were failed to be identified at species level and or any combinations between studied loci. In this study, multi-locus combination *rbcL+ITS2* showed highlest discrimination efficiencies among all three analytical methods.

Table 4.17 Summary of species resolution by three analytical methods.

Markers	Species discrimination		
	Similarity method (%)	Distance method (%)	Tree method (%)
rbcL	30	88.23±5	61.79±2
ITS	65.52	52.38±5	51.68±2
ITS2	72.41	85.71±5	51.68±2
rbcL+ITS	55.56	86.66±5	68.53±2
rbcL+ITS2	61.11	95.00±5	71.91±2

Note: markers (no. of sequences) viz. *rbcL* (65), *ITS* and *ITS2* (58), *rbcL+ITS* and *rbcL+ITS2* (72) discriminated by similarity method, while species having more than one individuals were discriminated by distance method.

Combining all three methods (BLAST, Distance and Tree method), a total of 24 of high value (only included protected plants of Nepal, CITES listed and prioritized species listed for research and development) species such as *Aconitum spicatum, Berberis angulosa, B. asiatica, B. erthoyclada, Bergenia purpurascens, Dactylorhiza hatagirea, Delphinium brunonianum, D. himalayai, Engelhardia spicata, Mahonia nepalensis, Michelia doltsopa, M. kisopa, Magnolia campbellii, Nardostachys grandiflora, Neopicrorhiza scrophulariiflora, Paris polyphylla, Rheum australe, Rubia manjith, Swertia chirayita, S. teres, Taxus wallichiana, Zanthoxylum armatum and Z. acanthopodium* were correctly identified at species level.

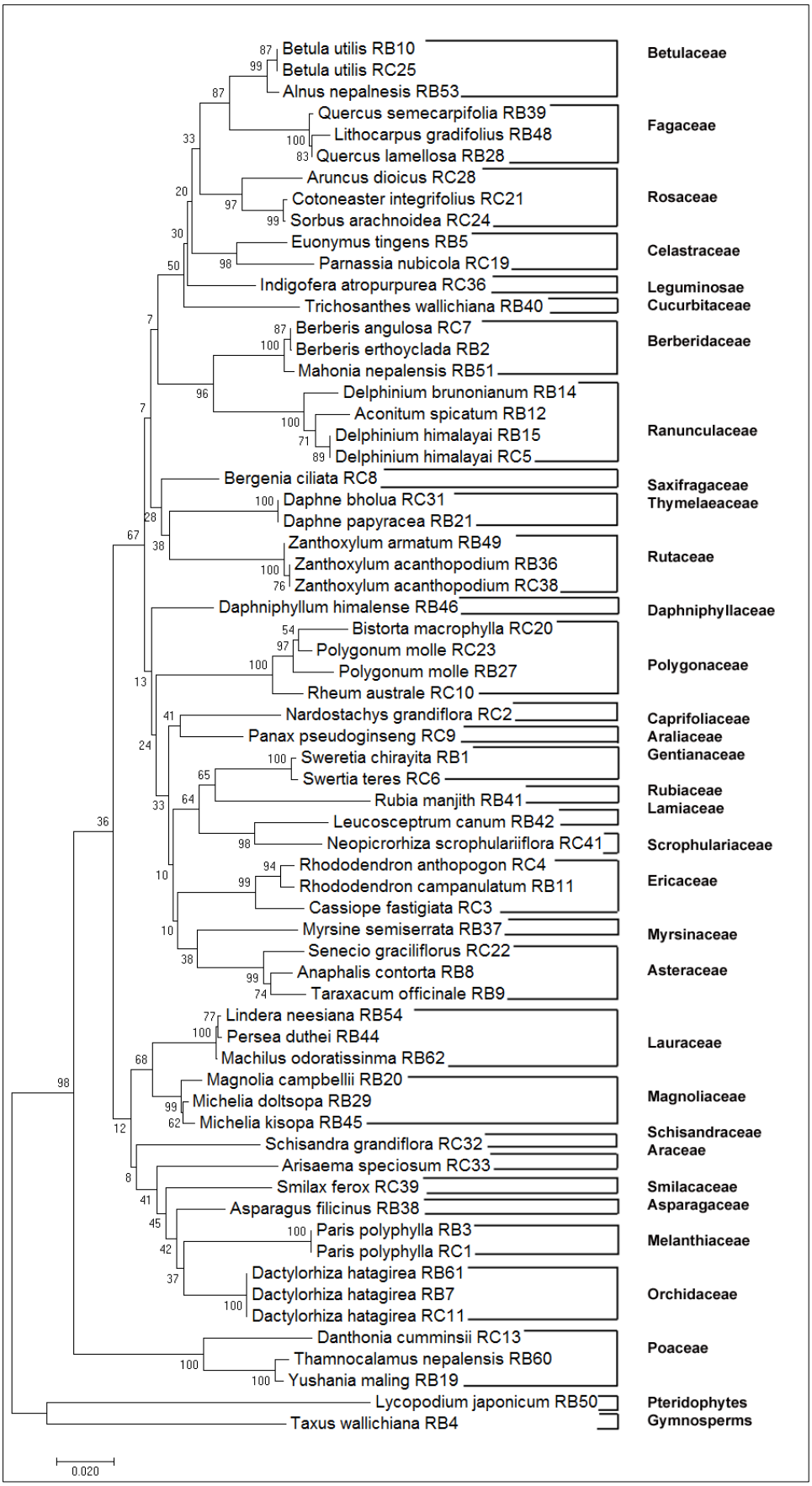


Figure 4.12 Phylogenetic tree reconstructed for the 65 sampled species of *rbcL* sequences using neighbour-joining method (K2P distance model, original tree).

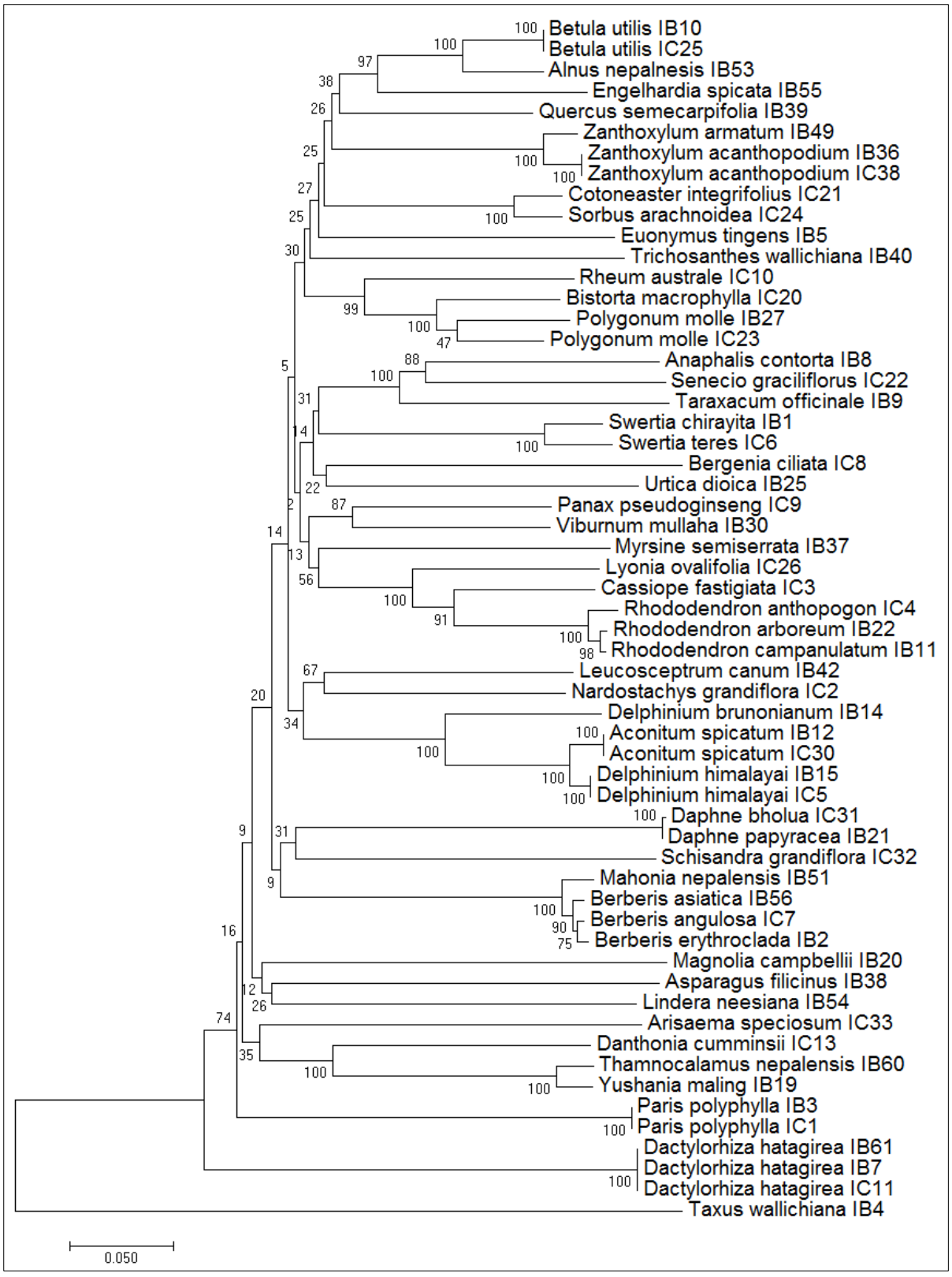


Figure 4.13 Phylogenetic tree reconstructed for 58 sampled species of *ITS* sequences using neighbour-joining method (p-distance model, original tree).

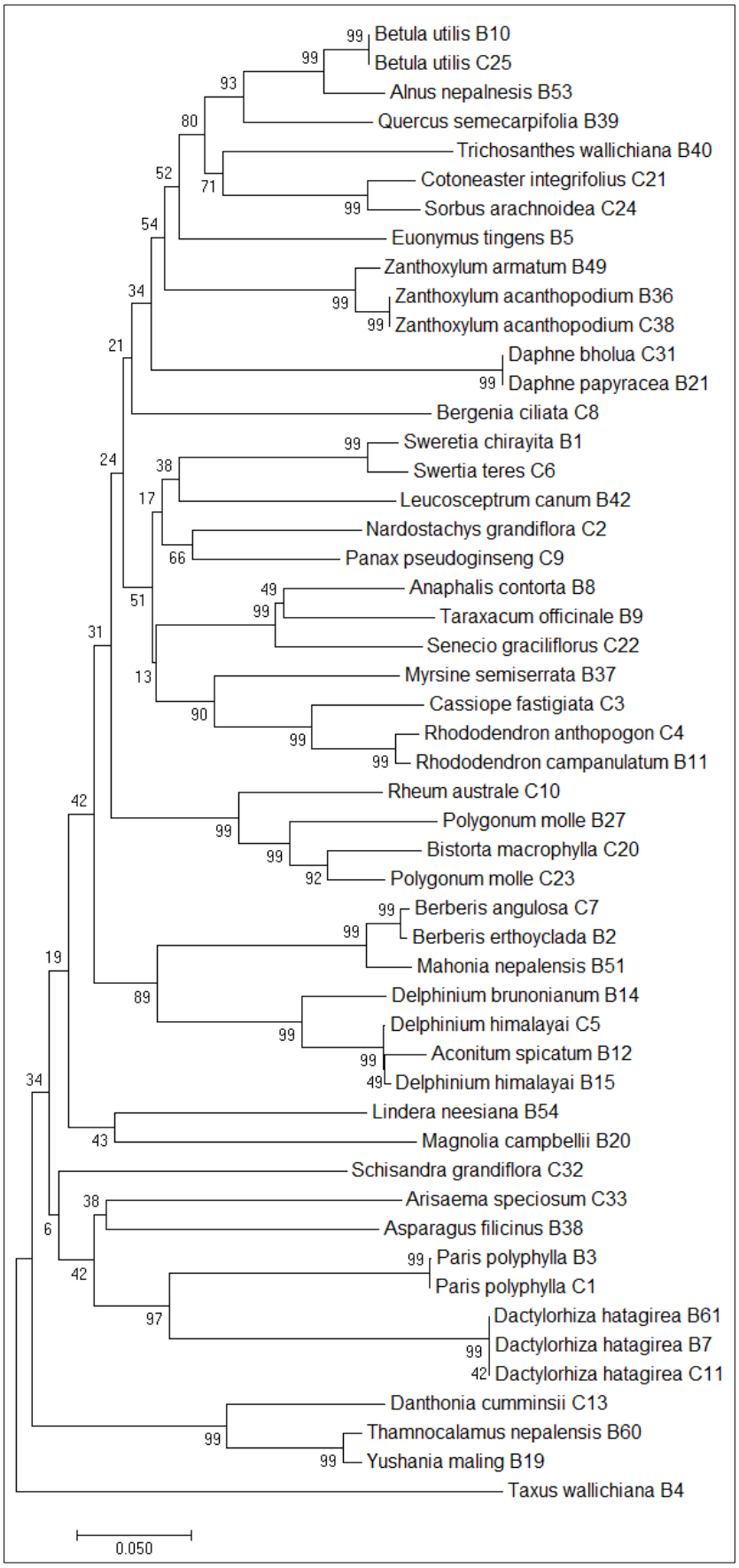


Figure 4.14 Phylogenetic tree reconstructed for 51 sampled species having both sequences of *rbcL+ITS2* markers using neighbour-joining method (K2P distance model, original tree).

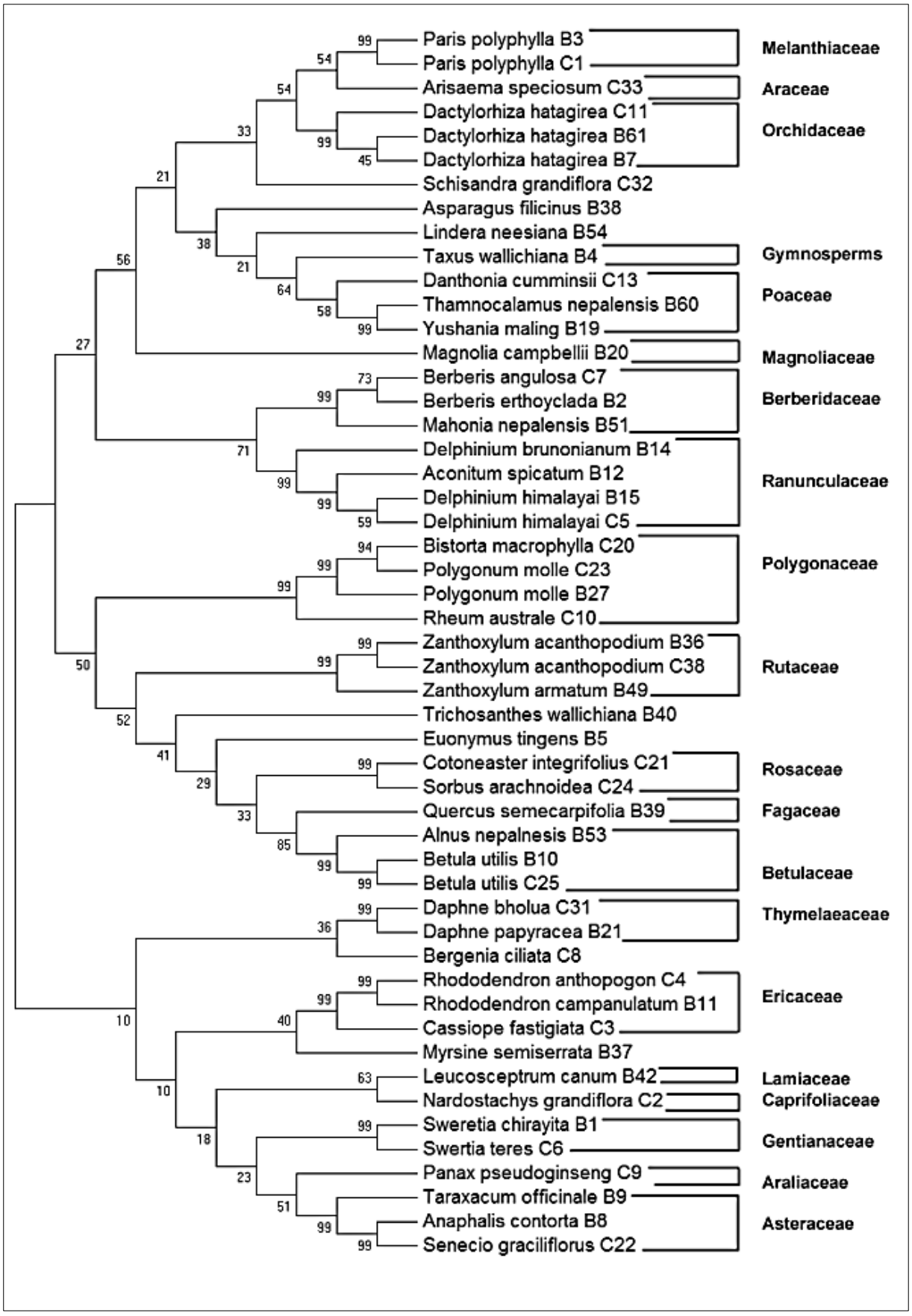


Figure 4.15 Phylogenetic tree reconstructed for the combination of 51 sampled species having both sequences of *rbcL*+*ITS* markers (only monophyletic clusters are labeled to their respective families; bootstrap tree).

CHAPTER 5. DISCUSSION

Since long rural people of Nepal are relying upon plants and its products for livelihoods. Almost 90% of Nepalese rural communities depend on agriculture, with plants providing arable crops, fodder, fruits, vegetables, fuel, building materials and medicines. Nepal has high diversity of plants with comparison to its land area. However, rapid increase of urbanization, adoption of modern agricultural practices, loss of indigenous and endemic plants and related indigenous knowledge on sustainable use and overexploitation of highly demanded species (www.floraofnepal.org) etc possess major challenge in biodiversity conservation. In addition, plant biodiversity are threatened by several other factors which are contributing directly to the loss of biodiversity viz habitat loss, fragmentation, habitat degradation, overexploitation of forest, invasive species, air pollution, nitrogen deposition and climate change (Corlett, 2016). Demands of Non timber forest products including medicinal and aromatic plants are increasing both in the national and international markets. Due to increasing need of MAPs, they are over-harvested, collected at premature and illegal traded thus causing extinction of such species from several of their natural habitats (Chaudhary et al., 2016). Therefore, based on scientific and reliable data effective planning for sustainable conservation of our invaluable plant genetic resources is urgently needed.

The term 'DNA barcoding' was coined by Herbert et al., (2003). Herbert initially successfully adopted this modern molecular tool of species identification on lepidopteran insects, fishes and birds. Now, this molecular tool is widely used for species level identification of all the eukaryotes. Various DNA barcode programmes are regularly generating DNA sequences and enriching DNA barcode library of important flora and fauna across the globe. DNA barcoding is not only applicable to identify species and reveal cryptic species but also applied in the prevention of potential adulteration and illegal trades. In animal, mitochondrial gene encoding *cytochrome c oxidase subunit 1* (*CO1*) region has been established as practical and standardized portion of DNA as barcode marker for species level identification. In plants, the events of hybridization, introgression and allopolyploidy are more pronounced than in animals and the species identification based on one locus was considered to be insufficient (Chase et al. 2005). CBOL plant group (2011) has recommended two plastid based barcode viz. *rbcL* and *matK* for species level identification. Now, most of the research group recommends combination of two coding plastid region viz. *matK, rbcL,* and non-coding plastid region *trnH-psbA* with nuclear internal transcribed spacer (*ITS*) for accurate and reliable barcoding of plants.

5.1 Sample Collection, Preservation, Identification and Documentation.

Present investigation has explored the barcode efficacies in high value plant of Mardi-Himal area. Some of the most important aspects of DNA Barcoding of plant is sample collection, its correct identification and documentation of relevant information associated with the plant species under study. To enhance the utility value of the Barcode database, species included should preferentially of wild origin and be in use for local communities or relevant to their livelihood. Many high value medicinal and aromatic plants are in high risk of unsustainable collection and overexploitation. Collection of plants, those are even for scientific studies should follow the rules and

regulations specified by the authorized government institutions. Getting permits for scientific collection from National parks and community forest/national forest is another most important obligation that everyone should follow strictly during the research period. Adoption of proper collection techniques and after collection procedures should follow standard specimen handling and long term preservation protocols to facilitate easy retrieval of the data in the future. Therefore, preservation of collected samples in silica gel, drying and pressing herbarium vouchers and maintaining good herbarium also determines the quality of DNA barcode database (Smith & Chinnappa, 2015). To ensure long term preservation of plant samples collected for genetic study, they are dessicated in "Silica Gel". Silica gel is an inexpensive, easy to handle, and relatively easy for preservation that could be used for further molecular studies in future. Silica gel dessicate (remove water materials) the plant materials at room temperature. In many cases dried plant materials are stable for 1 months (Chase & Hills, 1991) to 6 months (Adams et al., 1992) in room temperature for DNA isolation. In present study plant samples are preserved in silica gel which can be used upto 100 years. However, silica gel must be changed regularly (within 3-6 months). Lahaye et al., (2008) reported that DNA can be extraction from herbarium materials but, the quantities and quality of genomic DNA will vary. There are number of factors that will affect the quanity and quality of genomic DNA viz. unsatisfactory preservation of plant materials, DNA isolation length time, stressed plant materials (accumulation of phenolic compounds) etc may limit the utility and shelf-life of materials (Till et al., 2015).

Moreover, initial identification of plant species based on morphology is a crucial step and many cases could be very challenging too. For morphological identification, highly experienced plant taxonomic experts are needed. Since correct initial identification ensure the reliable and accurate labeling of DNA barcode of particular species, this step should be performed sincerely and with caution. In many instances DNA sequences of several species are missing thus, references sequences in NCBI and BOLD databases could not assign proper species and gives high probability of identification to the sister species if present in the database (Ghorbani et al., 2017). In such cases, inputs from morphology based identification plays important role in the species identity determination decision process. Therefore both the morphological and molecular studies are complements of each other. Taking consideration of this fact this study has prepared leaf samples preserved in Silica gel and Herbarium specimen of each samples. Both Silica preserved samples and Herbarium specimens are deposited in the 'Plant Curation Facility' of Molecular Biotechnology Laboratory, NAST. These preserved plant materials are regularly curated by the laboratory.

5.2 DNA Extraction

Plants DNA was extracted using widely used modified CTAB protocol. This method has been used with success on wide variety of plant groups (Doyle, 1991) and even in animals (Thomas et al., 1997). This method is used to isolate total genomic DNA (nuclear, chloroplast and mitochondrial). It is relatively simple, quick, inexpensive method that is suitable for use in combination with other protocols. The modified CTAB method of DNA isolation uses high salt concentrations and polyvinyl pyrrolidone (PVP) to remove polysaccharides and phenols respectively (Porebski et al., 1997). Some investigated samples viz *Nardostachys grandiflora* (B13), *Urtica dioica* (B25), *Trichosanthes wallichiana* (B40), *Rubia manjith* (B41), *Leucosceptrum canum* (B42), *Persea duthiei* (B44), *Zanthoxylum armatum* (B49), *Cotoneaster integrifolius* (C21),

Smilax aspera (C37) and *Zanthoxylum acanthopodium* (C38) were difficult to isolate good quantities and quality of genomic DNA from their respective samples. These medicinal plants contain high amount of secondary metabolites such as polysaccharides and polyphenols (Moyo et al., 2008). From above mentioned species after the isolation of genomic DNA, T.E buffer used to dissolve their pellet was observed viscous, jelly like and some were reddish or pinkish in color (B41) and dark brown in color (B13 & B25). The viscous nature and various color present in the samples is due to presence of secondary metabolites such as polyphenols, tannins and polysaccharides. Polysaccharides present in the samples are visually apparent in DNA extracted by their viscous nature, glue-like texture those cause difficulty in pipetting and inhibit polymerase activities during PCR (Fang et al., 1992). To remove the viscous nature of extracted DNA, samples were processed in Qiagen kit and in CTAB protocol precipitation step chloroform: Isoamylalcohol was performed twice to three times.

In the present study, genomic DNA was isolated from both one and and two year old leaves samples. The quality of DNA from one year old leaves was relatively good than two years old samples. The quantities and quality of genomic DNA of the extracted sample was evaluated on agarose gel electrophoresis. DNA isolation from the samples (viz B31, B35, B43, B47, B52 and B57) that have two years old leaves was found very difficult to isolate by both the methods. This could be due to the degradation of DNA in the old samples (Doyle & Dickson, 1987). The quality of DNA is also dependent on the time period of preservation during storage and temperature at which it is stored (Doyle & Dickson, 1987). Majority of the plant samples used in this study were collected not more than three years before thus obtaining low quality DNA was not the serious issue for the present study.

5.3 Selection of Barcode Loci

Barcode loci selected for plants could be one or few standard loci from nuclear regions or plastid regions. However, a suitable barcode should possesses some characteristics such as conserved flanked regions to enable routine amplification among highly divergent taxa using universal primers, sufficient internal variability to discriminate between species, short length easy to amplify (500-600 bp), single copy number that allows direct PCR and sequencing without cloning, easy to align (sequence quality) and allows the use of character-based data analysis, lack of several microsatellites that reduces sequence quality (Hollingsworth et al., 2009; 2011) and easy recovering DNA sequences from herbarium samples, parts/fragments and market samples (Särkinen et al., 2012). Kress et al., (2005) suggested the use of *ITS* from nuclear regions and *trnH-psbA* spacer from non-coding plastid regions for discriminating flowering plants. Chase et al., (2005) performed *in silico* approaches to test the potentiality of *rbcL* and *ITS* sequences available in GenBank for species identification and suggested combination of multi-locus barcode markers from plastid regions and one from nuclear regions as universal markers to sufficiently discriminate species for land plants. Newmaster et al., (2006) also suggested rbcL gene as a core DNA barcode for identifying plants as well as discriminating species of land plants, due to the presence of coding regions in all land plants except few parasitic plants. Chase et al., (2007) again proposed multilocus barcode and recommended using either two or three-locus combinations, *matK*+*rpoB*+*rpoC1* or *matK*+*rpoC1*+*trnH-psbA*, used as universal barcodes for plants. On the other hand, Kress and Erickson (2007) recommended the use of coding *rbcL* in combination with a non-coding intergenic spacer, *trnH-psbA* as a two-locus global DNA

barcode for land plants. CBOL plant working group, (2009) recommended the two-locus combination of *rbcL* + *matK* as the core plant barcode for land plants due to sufficient recovery rate of *rbcL* and high discriminatory power of *matK*. Now, most researcher groups use one or few plastid regions viz. protein coding *rbcL* and *matK* and the non-coding *trnH-psbA* with combination of the internal transcribed spacers (*ITS*) regions of ribosomal nuclear DNA (*ITS*-either full length or *ITS2* region) for specimen based plant barcoding studies (Kress et al., 2009, Hollingsworth et al., 2009; 2011; Chen et al., 2010). Chen et al., (2010) suggested *ITS2* of nuclear ribosomal regions as standard barcode for prevention of international trade or illegal trade and safe use of medicinal plants. *ITS2* barcode have highly variable internal region and easily to differentiated of closely related species. Like several barcoding studies the present research to test the efficiencies of chloroplast and nuclear barcode found two locus *rbcL*+*ITS* or *rbcL*+*ITS2* sufficient for identifying many high value of plants of central Nepal.

5.4 Amplification of Barcode Markers

Amplification of the barcode loci (*rbcL* & *ITS*) were successfully achieved from 89 species of high value plants representing 63 genera under 41 families. The data was analysed by comparing their amplification efficiency and sequencing success rates among the tested species. In this study, out of 89 samples, PCR efficiency of barcode marker *rbcL* and *ITS* were 85.39% (76) and 83.15% (74) respectively. The remaining samples have failed to amplify and some had very poor quality of amplicons. The efficiency of PCR depends upon a number of factors such as quality of genomic DNA, PCR reactions and PCR program. In this study, Bovine Serum Albumin (BSA), enzyme (DNA Taq. polymerase) and $MgCl_2$ were added in to the PCR mix (Promega, USA) for optimization (Figure 4.5 and Figure 4.6). BSA is used as enhancer for amplification. In some PCR reactions, DMSO is also added with conjugation of BSA, this conjugation is used to amplify the high GC content samples (Farell & Alexandre, 2012). $MgCl_2$ act as cofactor, stabilizes the primer and template annealing temperature, increase fidelity, specificity and high yields (Lorenz, 2012). In this study, 0.5mm $MgCl_2$ is added in to the PCR mix, that enhances and optimized for the amplification of the both barcode markers.

An efficiency of PCR amplification of flowering plant using *rbcL* and *ITS* markers reported 95% and ≤ 88% success (although high-quality sequence data were not obtained from all *ITS* amplifications) by Kress et al., (2005). Kress & Erickson, (2007) reported 92.7% (43 of 48 genera) efficiency for the amplification of *rbcL* marker of the land pants. CBOL Plant working group (2009) compiled some reports and reported 90-98% efficiency of PCR amplification of *rbcL* marker for DNA barcoding of angiosperms group or land plants. Another study, de Vere et al., (2012) achieved 97.7% efficiency of PCR amplification for *rbcL* marker on flowering plants and conifer from Wales. Cheng et al., (2016), showed 82.2% PCR efficiency using newly designed plant specific primer pairs of full *ITS* in 313 plant samples representing 270 families (angiosperms: 222 samples from 219 families: gymnosperms: 24 samples from 11 families; ferns and lycophytes; 44 samples from 24 families respectively; bryophytes: 23 samples from 16 families). In the present study, amplification rate of *rbcL* marker is some what lower than previous studies and amplification rate of *ITS* marker is relatively higher than (Cheng et al., 2016) but lower than (Kress et al., 2005).

Cheng et al., (2016) suggested *ITS* marker of ribosomal DNA which, is one of the most commonly used DNA markers in plant phylogenetic and DNA barcoding analysis and

recommended as a core plant DNA barcode but the universality and specificity of old PCR primers (White et al., 1990) for the *ITS* region were not satisfactory, resulting in amplification and sequencing difficulties and fungal contamination. The new universal *ITS* primers including *ITS1* and *ITS2* which are also used in this study are plant specific primer pairs. This new primer pair eliminate the chance of PCR amplification of nonplant templates thus, significantly improve the quality of *ITS* sequence information in plant molecular systematics and DNA barcoding (Cheng et al., 2016).

In addition, the amplification success rate are dependable upon quality of DNA and age of preserved samples (Lahaye et al., 2008). The poor quantities and quality of DNA significantly affects the success rate of amplification. As discussed in the previous section the quality of DNA was relatively poor for the samples those have high quantities of secondary metabolites. Although this study has used one to three years old samples preserved in Silica gel, minor problems associated with regular curation of the preserved samples might not be ignored. Some of the samples preserved in Silica gel comprises old leaves, outer bark of stem, parts of young shoots etc. These materials were collected as fresh and healthy leaves were not available during the collection period. Therefore, low quantity and quality of DNA recovery and amplification failures to such samples could be attributable to a insufficient post harvest condition of preservation and, state of Silica preserved materials used in the study.

5.5 Purification of PCR Product

Following amplification, the amplicons are generally sequenced directly. Properly optimized PCR amplicons normally have no or very low impurities. Generally PCR products contain excess amount of template DNA, primers, and dNTPS and other impurities viz. enzymes and buffers. These impurities hinder the sequencing process and reduce the quality of the sequence. Therefore, these surplus ingredients should be removed (http://www.biology.ualberta.ca). There are several methods to clean up of PCR products such as Column purification, Ethanol precipitation, Gel purification, PEG precipitation and enzymatic purification (ExoSAP method). Column, Ethanol and Enzymatic purification are applied only if the PCR products of markers have single bands. Gel and PEG purification methods are applied for more than single bands of PCR products of particular markers. In the present study, ExoSAP method was applied, Exo-sap method is referred to as gold standard method for enzymatic purification. This method is relatively simple, quick and easy to perform. This method is eliminating all gel or column purifications, sedimentations, filtrations and, beads or magnetic separations (Dugan et al., 2002).

5.6 Sequencing and Quality check

The modern automated method ABI 3500XL genetic analyzer (Sanger et al., 1977) is based on Sanger's di-deoxy chain termination method which, can be used to sequence up to 1000 bp long DNA molecule in a single reactions (Chan, 2005). In the present investigation, genetic analyzer based on Sanger's di-deoxy chain termination method is used to sequences the amplicons of two barcode markers (*rbcL* and *ITS*).

In this study wide range of families were evaluated for their barcoding success using two candidate makers (Appendix 2). Sequencing success rate of *rbcL* was highest 73% (65) in compared to *ITS* 65% (58). The efficiency of *rbcL* sequencing rate is consistent with Kang et al., (2017) (75.26% ± 3.65%) and indicates a promising universality for the *rbcL*

barcode for Nepalese medicinal plants as well. China Plant BOL group, (2011) has shown about 71.7% sequencing success for *ITS* region. The finding in an evaluation of *ITS* in the present study is similar to Mega analysis of *ITS* sequences performed by Chinese scientists. Although, *ITS* sequences occasionally encounter fungal contamination, fortunately this study did not have such incidences because new set of plant specific *ITS* primers designed by Cheng et al., (2016) was used. The efficiency of sequencing depends upon the recovery rate of PCR amplicon and post purification of PCR products. In the present study, only high quality purified PCR products were subjected for sequencing. Some samples viz *rbcL* (B13, B22, B30, B50, and B59) and *ITS* (B13, B24, B26, B40, and C39) were not recovered and other impurities could not completely removed that, slightly reduces the rate of sequencing efficiency in the present dataset.

High quality sequences were obtained from 84.61% (55) of *rbcL* and 85.93% (55) of *ITS*. The problems were faced during assembly of the birectional sequences with few ambiguous base in *ITS* as well as low quality of some sequences of both markers. The high quality sequences were edited only in end to end portion.

The divergence among species based on DNA barcoding depends upon variation in the nucleotide sequences of the tested loci. Therefore, DNA sequences used for this purpose are required to be of high fidelity with unambiguous identification of each base (Kress and Erickson 2007). To ensure correct identification of each base in the trace sequence (Ewing et al. 1998), electropherograms were subjected to base-calling using software, PHRED (Ewing & Green, 1998b). PHRED assigns a quality value (QV) or a score to each nucleotide after analyzing the corresponding peaks (Ewing et al., 1998). This score (Q) commonly called to as Phred score is logarithmically related to the error probabilities in basecalling by the software (Ewing & Green, 1998b).

The quality score of 20 assigned to a base means that the chance of this base called to be incorrect is only 1%. As it is a logarithmic scale, a quality score of 30 implies that the possibility of wrong base calling is 0.1% (Ewing and Green, 1998). A quality score of 20 and above is considered acceptable for DNA barcoding (CBOL Plant Working Group, 2009). Thus, in the present study, sequences of quality value below PHRED quality score of 20 was not considered for further analysis. To ensure retrieval of high number of sequences suitable for further analysis each sequence were trimmed from both the ends to remove the bases below 20 (CBOL Plant Working Group, 2009, Chen et al., 2010) that generally corresponds to primer sequences.

5.7 Species discrimination

An ideal barcode should be short, easily recoverable, can discriminate closely related species and provide maximum information (Kress et al., 2005; CBOL Plant Group, 2009). There are different analytical methods used to access high species discrimination and generate reliable reference DNA barcode library of these evaluated species.

5.7.1 Simple BLAST and Optimized BLAST

The most common and widely used method for species identification is Basic Local Alignment Search Tool (BLAST) followed by distance matrix computations. The possibility that the phylogenetic tree-based method (like Neighbor-joining (NJ) or Maximum Likelihood) gives the lowest accuracy of identification may be due to unavailability of homologs sequences in NCBI database (Fan et al., 2014). Methods viz. Sequence similarity, tree-based criteria, or character are utilizing reference database to identify

unknown query sequences for DNA barcode identification (Ghorbani et al., 2017). However, the efficiency of any method to assign sequences to a certain taxon is dependent on the taxonomic coverage of the reference database (de Boer et al., 2014). BLAST and Optimized BLAST similarity based identification methods gives more significance to the top hits, identity value of the query-reference comparison and the deviation of other from top hit (Ghorbani et al., 2017). In the present investigations, out of 72, 31.94% (23) species were identified at species level using simple BLAST method with combinations of *rbcL*+*ITS*+*ITS2* markers. While, out of 72, 62.50% (45) species were identified at species level using optimized BLAST method with combinations of all three markers. In case of single locus, *ITS2* has the highest identification success at species level in both methods. In simple BLAST, out of 58; 39.66% (23) species were identified at species level. In optimized BLAST, out of 58; 72.41% (42) species were identified at species level. Species level identification success in this study is significantly higher than Ghorbani et al., (2017), where they were able to identify 32% (22) species by simple BLAST and, 38% (26) species at species level by optimized BLAST using two marker combinations (*ITS*+*trn*L-F). In addition, with correlation of morphology, only 20.83% (15) species were identified to species level. Ghorbani et al., (2017) suggested species identification by similarity method is limited because no such complete database exists and that reduces the discrimination efficiencies when compared to morphology. (Taylor & Harris, 2012) argued that no single method could be applied universally for species identification in plants. In plants, species identification becomes more reliable when molecular information is correlated with morphology, geography, ecology, behaviour and other sets of available information (Hickerson et al., 2006). Furthermore, Smith et al., (2006) and Creer et al., (2010) also suggested DNA barcoding requires integration of ecological, morphological and other relevant data to identify species correctly.

Based on combined results of BLAST and morphological information some high value plants (only included protected plants of Nepal, CITES listed plants and species prioritized for research and development, farming and conservation) can be successfully identified at species level. These species are *Taxus wallichiana, Nardostachys grandiflora, Zanthoxylum armatum, Z. acanthopodium, Bergenia purpurascens, Berberis asiatica, Swertia chirayita, Magnolia campbellii* and *Mahonia nepalensis.*

In the present investigations, *ITS2* marker is superior among tested markers for the identification of species as, suggested by previous studies (Liu et al., 2014; Yu et al., 2017). The present study is also consistent with Chen et al., (2010) and China Plant BOL Group, (2011). Some complex species such as *Aconitum* spp. (Ranunculaceae), *Dactylorhiza hatagirea* (Orchidaceae) *Rhododendron* spp. (Ericaceae), *Yushania maling* (Poaceae) are relatively difficult to identify at species level using BLAST methods. The efficiency of identifying these species by BLAST method depends on availability of their reference sequence in GenBank. Most importantly another convincing reason for their difficulties to get hundred percent species identification through universal markers is attributed to their evolutionary history, where speciation is driven by hybridization, slow mutation rate relative to speciation rate and recent and rapid divergences of species in the group (Hollingsworth et al., 2016).

5.7.2 Genetic divergence

The evaluation of intra and inter-specific divergence is one of the major characteristics for the correct identification of species and evaluation of DNA barcodes. The efficiency

of good barcode has ability to differentiate the closely related species. The species is identified correctly when enough genetic differentiation is present within species i.e. when interspecific distance is significantly higher than intraspecific distance (Hebert et al. 2003; Lahaye et al., 2008). In this study also the single locus and multi-locus combinations were exhibited both intra- and interspecific divergence. Besides some overlap, single and multi-locus combinations have shown clear barcode gap, suggesting higher possibilities of species discrimination among high value plants. Barcode gap analysis provides the distribution of genetic distances within species and distance to the nearest neighbor of each species (Mishra et al., 2017). In this study, both single and multi-locus combinations discriminated species 80%-90% except *ITS* markers (52.38%±5). Multi-locus combinations *rbcL*+*ITS2* discriminated species upto 100%. Similar result was achieved by CBOL Plant Working Group, (2009); Chen et al., (2009) and Mankga et al., (2013).

Based on K2P distance analysis of core data set, some high value plants (only included protected plants of Nepal, CITES listed plants and species prioritized for research and development, farming and conservation) are successfully identified at species level such as *Aconitum spicatum, Berberis angulosa, B. erthoyclada, B. asiatica, Delphinium brunonianum, D. himalayai, Dactylorhiza hatagirea, Paris polyphylla, Swertia chirayita, S. teres, Zanthoxylum acanthopodium* and *Z. armatum.* In addition, the high value plant species (only included protected plants of Nepal, CITES listed plants and species prioritized plant for research and development) are identified at species level by K2P genetic distance analysis of core data (single individuals) with retrieved data from GenBank as well. The identified species are *Berberis angulosa*, *Bergenia purpurascens, Engelhardia spicata, Magnolia campbellii Rheum australe, Swertia chirayita, Taxus wallichiana* and *Zanthoxylum armatum.*

In distance analysis, *ITS* and *ITS2* markers have exhibited high K2P distances among core data set of the present study and with subsequent sequences of the GenBank. This is obvious in *Polygonum* species and also in *Delphinium* species (discussed in result section). Kress et al., (2005) and Kress and Erickson, (2007) have also reported *ITS* sequences with high interspecific divergences. The probable reason might be that *ITS* sequences are difficult to align among species with diverse taxonomic groups. Due to difficulties in alignment, a number of insertions and deletions (gap) are frequent or introduced that leads to variations in K2P distances and makes false-positive interpretations. Paralogous sequences can occur at many levels: within an individual, among individuals within a species, and among species. Single nucleotide polymorphisms (SNPs) generated in different positions within the sequence of *ITS* insertions/deletions and, concerted evolution will give rise to different *ITS* types. In the hybrid rRNA single stranded chains is formed as a secondary structure which contain stemmed regions and different loops correlating with base pairing which will be inferring phylogenies. During distance and phylogenetic analysis, it is difficult to determine whether a pseudogene or a paralogous sequence has occurred (Poczai & Hyvönen, 2010). In addition, *ITS* repeats are integrated within a single genome via hybridization (including allopolyploidy) or introgression. Such phenomena are very common in plants. Recent estimates suggest that 70% of all angiosperms have experienced one or more episodes of polyploidization. Stomatal size in fossil plants indicate the evidence for polyploidy in majority of angiosperms (Koch et al., 2003).

In other hand, *rbcL* marker was highly conserved among both present data set and retrieved data sets. This study is also consistent with several studies such as CBOL plant working group, (2009); China plant working group, (2011) and Mishra et al., (2017).

Genetic divergence among families

In this study, genetic divergence among family was strong in *ITS* than *rbcL* data sets. However, in both the datasets of *ITS* and *rbcL,* plants under different families have shown strong monophyly. Family level divergence between Taxaceae and others in *rbcL* data set was highest (18.78%). Similar difference was found with *ITS* data set too. Taxaceae belongs to gymnosperms thus genetic divergence of two clades viz. gymnosperms and angiosperms exhibited significant divergence. Torre et al., (2017) reported high rates of sequence divergence between these two groups. He also suggested molecular evolution of gymnosperms is slower than angiosperms, which indicates high rates of sequences divergence.

Genetic divergence among families for *ITS* marker was found high among Taxaceae, Orchidaceae, and Poaceae etc. Orchidaceae and Poaceae belong to monocot which is distinctly a monophyletic clade and differs significantly with eudicots (Soltis & Soltis, 2004; Soltis et al., 2005).

5.7.3 Phylogenetic Inference

Phylogenetic tree is reconstructed using sequences of the barcode marker. Species resolution/monophyly of species was estimated by cluster analysis (Lahaye et al., 2008, China Plant BOL Group, 2011). The species for which all the individuals clustered together in a single clade are considered as identified species/monophyletic and those which clustered (<50% bootstrap support) with the individuals of the other species were treated as unidentified. In this study, single locus could effectively discriminate or identify 50%-60% of species. Furthermore, multi-locus combination increases the discrimination up to 70%. In the present study, broader ranges of taxa representing angiosperm and gymnosperms have been analysed which resulted with low species discrimination. In addition the low level of discrimination observed in this study is attributed to incompletely sampled groups included in the analysis. Some internal groups are missing in phylogenetic tree of *ITS* and *ITS2* (Figure 4.13). Meyer & Paulay, (2005) reported low efficiency of identification in incompletely sampled groups. The Angiosperm Phylogeny Group, (2016) has also suggested such scenario in nuclear data which are poorly represented in a broader phylogenetic studies of angiosperms (Chase et al., 2016). The tree based approaches of species discrimination also provides a convenient method to know about the closely related group of the species under examination. *Dactylorhiza hatagirea* could not be identified at species level by BLAST method. However, tree based method grouped these taxa with other monocots taxa.

Despite of low bootstrap value in some clades, all the families are almost similar in position as in Global tree of APG IV. Moreover, NJ tree method also discriminated several taxa successfully. Many of them are high value plants (only included protected plants of Nepal, CITES listed plants and species prioritized for research and development, farming and conservation) viz. *Aconitum spicatum, Berberis angulosa B. asiatica, B. erythroclada Delphinium brunonianum D. himalayai, Dactylorhiza hatagirea*, *Engelhardia spicata, Mahonia nepalensis, Michelia doltsopa, M. kisopa, Magnolia*

campbellii, Nardostachys grandiflora, *Neopicrorhiza scrophulariiflora*, *Paris polyphylla*, *Rubia manjith, Swertia chirayita, S. teres*, *Zanthoxylum acanthopodium* and *Z. armatum.*

5.8 Species Identification and Revealing Cryptic Speciation

A total of 89 high value taxa 80% (72) were identified successfully using three different analytical methods comprising single or multi-locus combinations (Table 4.17). However, 20% species could not be identified using barcode markers which is mainly due to problems associated with PCR and sequencing. This result is consistent with the finding of CBOL Plant Working Group, (2009). All the samples used in this study were identified by taxonomic experts and also tallied with standard reference herbarium specimens deposited at National Herbarium and Plant Laboratory (KATH) Godavari. Although there are many benefits of morphology based identifications, *ITS* downfalls are difficulties that need to face while dealing with morphological identical and cryptic species. Specimen of collection code CN45 was morphologically identified as *Astilbe rivularis* of family Saxifragaceae. However with DNA barcoding, it was correctly identified as *Aruncus dioicus* (Walter) Fernald, which is in Rosaceae family. Both of these species are morphologically identical thus needs highly experienced and knowledgeable professional taxonomist to identify morphologically. Chen et al., (2010) highlighted that medicinal plants covers wide range of taxa, which probably morphologically very similar but genetically different or phylogenetically less related. Such similarity between species may lead to misidentification and inappropriate uses. Such misidentification is corrected by using DNA barcoding techniques. This is a good example in our case to demonstrate the usefulness of DNA barcoding and its crucial role to support traditional plant identification, cryptic species identification, proper documentation of our biodiversity and more importantly standardize our traditional knowledge of medicinal plants (Yessoufou, 2005) along with securing human health (Chen et al., 2010)

Rajbhandari et al., (2017), merged *Mahonia* species into *Berberis* as *Berberis nepalensis*. However, genetic divergence between *Berberis* and *Mahonia* was found to be 0.64% by *rbcL*, 1.5-2.5% by *ITS* and 2%-4.5% by *ITS2*. Previously Ahrendt, (1961) reported simple-leaved *Berberis* are derived from the compound-leaved *Mahonia* and *Mahonia* categorized as a distinct genus. However, Kim & Jansen, (1994) reported that *Berberis* and *Mahonia* share 11.5 kb expansion of the inverted repeat (IR) region of the chloroplast genome, which suggests a close phylogenetic relationship between two genus. Furthermore, our study reveals that *Mahonia* and *Berberis* are phylogenetically monophyletic as suggested by Schilling & Watson, (2014) and Rajbhandari et al., (2017).

Dactylorhiza hatagirea of Orchidaceae family was very difficult to identify at species level using BLAST method presumably due to lack of *ITS* reference data set in gene bank. But *Dactylorhiza hatagirea* is identified correctly by tree method and distance method with the assistance of morphological data. Another interesting example is morphologically identical specimens CNRCP35 and CN38 which are identified as *Polygonum molle.* However, intraspecific divergence between these species were 1.68%, 8.28%, and 11.58% for *rbcL*, *ITS* and *ITS2* respectively. *Polygonum molle* is correctly identified by tree method and with assistance of morphological data. Thus, morphological identification complemented with molecular identification using DNA barcodes provides accurate and reliable species information about the the species and other important facts.

5.9 Utility of DNA Barcoding in Conservation and Trade

DNA barcoding is an increasingly popular and emerging tool that has generated optimism in an assessment and documentation of biodiversity for conservation (Krishnamurthy & Francis, 2012). DNA barcodes are now being routinely used for the reliable identification and detection of wild collected plants, market adulterants in certified natural and commercial products, genetically modified crops etc. (Kress et al., 2014). DNA barcodes are utilized in the prevention of illegal trade of often protected and endangered species (Lahaye et al., 2008). DNA barcoding is seen as the highly sophisticated tool that can speed up the assessment of global biodiversity and disclose relationships of species including threatened and endangered populations worthy of conservation attention.

In the Nepalese context there is an urgent need of DNA reference library of high value medicinal and aromatic plants as well as other important species to effectively resolve the issues of illegal trade, adulteration and several other existing challenges in the sustainable utilization and conservation of our precious natural resources. DNA barcode can contribute to conservation policy in two important ways: by speeding up local biodiversity assessments to prioritise conservation areas or evaluate the success of conservation actions, and by providing information about evolutionary histories and phylogenetic diversity (Krishnamurthy & Francis, 2012).

Here, we have evaluated two barcodes and generated 72 DNA barcode of them which are mostly of wide range of high value plant including protected plant of Nepal, CITES appendices and prioritized plants for research and development (above species discrimination section). In this regard the amount of genetic data and associated information of MAPs generated for the first time in Nepal from Mardi Himal area has widened the possibility of establishment of DNA reference library. As practiced by several developed nations effective use of such database to overcome existing challenges (illegal trade, adulteration, and etc) in the conservation and sustainable utilization of high value plants of Nepal should be our sole objective.

5.10 Future Challenges of DNA Barcoding

Globally, since the time of the DNA barcode introduced into the botanical community, it has been applied to a variety of investigations in both basic and applied research in plants. However, Plant DNA barcodes have not yet universally accepted DNA barcoding as a core tool for identifying species, i.e. no single marker is able to completely discriminate among species in most taxonomic group. Reseachere's uses combination of DNA barcode loci to generate global plant DNA barcode library with 70-80% discrimination for universal uses (CBOL Plant Working Group, 2009; Kress, 2017). The discrimination success of Plant DNA barcodes is lower than many animal groups viz fishes, birds and butterflies (Hollingsworth et al., 2011). The assumption and prediction of future direction of plant DNA barcoding began with the initiation of studies applying DNA based markers to access questions regarding in taxonomy, evolution, and ecology, including the relationship between locus-based DNA barcodes and genomic approaches to species identification (Kress and Erickson, 2008). The need for both advanced sequencing technologies (HTS and NGS) as well as efficient database design for compiling DNA barcode and *ITS* analysis and search strategies for species identification was allowed.

DNA barcoding has been one of the routine activities in several countries of the world including our neighbouring country China. Multiple barcodes of thousands of plants from biodiversity hotspots of China are available in GenBank. Very few species from Nepal have been deposited in GenBank by the Nepalese students studying in Nepal and abroad. Systematic research on DNA barcoding of flora and fauna has not been done by the government and non government research institutions of Nepal. Therefore one of the major challenges to have a complete functioning of DNA barcoding in Nepal is the serious lack of the DNA sequences of Nepalese flora and fauna. If we will not start generating DNA sequences of our species we will be further pushed behind and their will more uncertainties in an efficient and effective implementation of this technique in our country. Well-planned and systematic studies need to be initiated as soon as possible. Two major activities can be initiated for this purpose. The first is to initiate collecting specimen of flora and fauna and preserve them in proper way so DNA extraction can be performed whenever necessary. Sampling must be both flexible and opportunistic and conducted across the entire growing season by expert taxonomists. The second parallel activity should be the PCR optimization and sequencing of those samples. Since we have well managed herbarium and updated publications on Nepalese flora. Preparing list of priority species and phase wise collection and generation of DNA sequences is not a big deal. Given the complex and diverse geography collecting specimens and barcoding them with our limited available finances could hinder our research activities. The success of the barcoding will depend upon the close collaboration among taxonomists, plant collectors, molecular biologists and with funding agencies.

CHAPTER 6. SUMMARY

A total of 89 plant species belonging to angiosperms 97% (87), Pteridophytes 2% (2) and Gymnosperms 1% (1) were chosen to be studied from Mardi Himal area. Out of 89 samples, the PCR efficiency of the barcode markers *rbcL* and *ITS* were found to be 85.39% (76) and 83.15% (74) respectively. *rbcL* and ITS PCR product size were about 500-600 bp and 700-800 bp (based on estimation of standard marker) respectively. Sequencing efficiency for *rbcL* was 73% (65) and *ITS* was 65% (58 samples). Total 80% (72) of the samples have sequencing success for both the markers (sample having both or single sequences of the markers). The efficiency of sequencing of both markers are 57% (51 samples) for having both loci, 18.31% (13) with only the *rbcL* sequences and 9.86% (7) with only the *ITS* sequences. High quality sequences were obtained 84.61% (55) and 85.93% (55) for *rbcL* and *ITS* respectively. Based on the sequence data, the final lengths of universal barcode markers *rbcL* and *ITS* for the plants of Mardi Himal area were 495-536 bp and 688-867 bp in size respectively. *ITS2* was found to be 258-303 bp in size. *rbcL* barcode marker was found to be relatively highly conserved (60%-70%) and has lowest number of variable sites (223/537). ITS barcode marker locus showed high variability (80-90%). The simple BLAST searches results identified 15.38% (10 samples) to species level, 84.61% (55 samples) to genus level and 100% (65 samples) to family level for *rbcL*. However optimized BLAST searches for *rbcL* results identified 29.23% (19 samples) to species level, 80.00% (52 samples) to genus level and 100% (65 samples) to family level. Similarly the simple BLAST searches results identified 27.59% (16 samples) to species level, 86.20% (50 samples) to genus level and 100% (58 Samples) to family level for *ITS*. But, the optimized BLAST searches for *ITS* results identified 65.52% (38 samples) to species level, 94.82% (55 samples) to genus level and 100% (58 Samples) to family level. In addition to ITS, the simple BLAST searches results for *ITS2* identified 39.66% (23 samples) to species level, 89.66% (52 samples) to genus level and 100% (58 samples) to family level. The optimized *ITS2* BLAST searches results identified 72.41% (42 samples) to species level and 100% (58 samples) to genus level and family level. However, with the correlation of morphological data, species identification was improved and restricted precisely to species of our concern. The final correct identification of species in correlation with morphology based on modified optimized BLAST for all three loci (*rbcL*+*ITS*+*ITS2*) was 36.11% (26) to species level, 83.33% (60) to genus level, and 100% (72) to family level. Interspecies genetic divergence of *ITS* locus was found to be relatively higher than *rbcL* locus. Similarly, ITS2 exhibited highest mean intra- (0.0891) and inter specific (0.0970) genetic distances followed by *ITS* marker and *rbcL* exhibits lowest mean intra and interspecific divergences. Genetic divergence of high value plants those are enlisted in CITES appendices and Nepal Government conservation, farming and research and development priorities were evaluated. The phylogenetic tree was reconstructed for *rbcL*, *ITS* and *ITS2* and their various combinations. Generally, species generated separate clusters (monophyletic clades) in the tree and with bootstrap value >50% were considered to be discriminated successfully. The topology of phylogenetic tree reconstructed using *rbcL* and *ITS* were almost similar to the respective global tree of APG IV. Average bootstrap value (only considered >50%) of *rbcL*, *ITS* and *ITS2* based phylogeny was found to be 84.61, 74.13% and 79.31% respectively. Based on the NJ tree, multi-locus combinations viz. *rbcL*+*ITS2* combinations has highest species discrimination efficiency (71.91% monophyletic clades), followed by 68.53% of *rbcL*+*ITS*. However In single loci, *rbcL* marker has 62.92%

discrimination efficiency, followed by 51.68% of *ITS2* and 48.31% of *ITS*. Finally among 89 species of plant species belonging to 63 genera and 41 families, 72 species were successfully discriminated via one or more analytical methods using single locus or multi-locus combinations. However, 17 species were failed to be identified by all DNA regions used in this study (unsuccessful recovery of sequencing and PCR), multi-locus combination *rbcL+ITS2* showed highest discrimination of species by all three analytical methods. Based on DNA barcoding method, we have successfully identified some high value plants of Mardi-himal region including protected plants of Nepal including prioritized plant for research and development. These high value plant species are *Aconitum spicatum, Berberis angulosa, B. asiatica, B. erthoyclada, Bergenia purpurascens, Dactylorhiza hatagirea, Delphinium brunonianum, D. himalayai, Engelhardia spicata, Mahonia nepalensis, Michelia doltsopa, M. kisopa, Magnolia campbellii, Nardostachys grandiflora, Neopicrorhiza scrophulariiflora, Paris polyphylla, , Rheum australe, Rubia manjith, Swertia chirayita, S. teres, Taxus wallichiana, Zanthoxylum armatum* and *Z. acanthopodium.*

CHAPTER 7. CONCLUSIONS and RECOMMENDATIONS

CONCLUSIONS

The current study presented as thesis has evaluated the efficacies of barcode loci in wide range of high value plants including Medicinal and aromatic plants collected from 1200 m to 4500 m of Mardi-Himal area, central Nepal. Plant species selected for DNA barcoding are predominantly Nepal government protected, CITES listed and prioritized plant species for conservation, farming, research & development. Among 89 plant species initially chosen for barcoding, we generated high quality sequences of the two markers (*rbcL* and *ITS*). *ITS* included the portion of *ITS2* as well. Optimised BLAST method revealed that single locus *ITS2* marker as the best DNA barcode with high resolution of species discrimination. *ITS2* showed 70%-80% species discriminated successfully by combining of all three methods (BLAST, divergence and Tree method). In the multi-locus combination *rbcL+ITS2* revealed to be best DNA barcode with highest resolution of species discrimination. Multi-locus combination *rbcL+ITS2* increased species discrimination to 80%-90% by combination of all three methods. However, *rbcL* alone provided high discrimination between species by the tree method among tested single locus.

Combining all three methods (BLAST, Distance and Tree method) a total of 24 high value species (only included protected plants of Nepal, IUCN listed, CITES listed and prioritized listed for research and development) such as *Aconitum spicatum, Berberis angulosa, B. asiatica, B. erthoyclada, Bergenia purpurascens, Dactylorhiza hatagirea, Delphinium brunonianum, D. himalayai, Engelhardia spicata, Mahonia nepalensis, Michelia doltsopa, M. kisopa, Magnolia campbellii, Nardostachys grandiflora, Neopicrorhiza scrophulariiflora, Paris polyphylla, , Rheum australe, Rubia manjith, Swertia chirayita, S. teres, Taxus wallichiana, Zanthoxylum armatum* and *Z. acanthopodium* were correctly identified at species level (24 species). DNA barcoding was also effective to identify morphologically identical species like our *Astilbe rivularis* and *Aruncus dioicus*, Mahonia and *Berberis* and *Neopicrorhiza scrophulariiflora* and *Veronica* species.

Finally this study strongly demonstrates that DNA barcoding of Nepalese high value plants including medicinal and aromatic plants have tremendous advantages in species identification; identify adulteration, value addition, and authentication of herbal products, curbing illegal trade and in the sustainable utilization as well as conservation. Moreover DNA sequences generated by this study forms very important foundation for the creation of 'DNA Reference Library' of the MAPs of Nepal that have greater potentialities in the documentation and conservation of the biodiversity of Nepal.

RECOMMENDATIONS

Based on the finding of this study and existing scenario of DNA barcoding technique and its applications in Nepal few recommendations are given below.

- Uses of multiple markers in species discrimination are in practice for some species therefore more markers should be added for reliable and accurate species identification.
- In addition to the data set of Mardi Himal more (at least two) collections from eastern and western Nepal added to the data set can improve the species discrimination and that would also light important knowledge on the population

genetics, evolutionary history, speciation (intra- and interspecific) and other information of the species.

- Standardization of markers by evaluating the corresponding species from market or traded samples available in different forms (powder, raw materials, and dried plant parts) provides greater confidence on species identification.
- Conduct DNA barcoding as a routine program by authorized government institutions and research institutions.
- Enrichment of DNA Reference Library with additional information on morphology, ecology, plant distribution and other studies including ethnobotanical information.
- Like in other biodiversity rich countries, DNA barcoding based biodiversity assessment studies should be initiated in Nepal. Such study has greater potentialities to identify cryptic species.
- Promotion of DNA barcoding in plant forensics should be initiated on time in Nepal.

REFERENCES

Acharya K, Chaudhary R, Vetaas O (2009) Medicinal plants of Nepal: Distribution pattern along an elevational gradient and effectiveness of existing protected areas for their conservation. *Banko Janakari*. **19**(1): 16–22. https://doi.org/10.3126/banko.v19i1.2178

Adachi J & Hasegawa M (1996) *MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood*. Tokyo: Institute of Statistical Mathematics

Adamowicz SJ, Hollingsworth PM Ratnasingham S & van der Bank M (2017) International Barcode of Life: Focus on big biodiversity in South Africa. *Genome*. **60**(11): 875–879. https://doi.org/10.1139/gen-2017-0210

Adhikari B (2001) Socio-economic heterogeneity and income distribution: evidence from common property resource management. *Journal of Forestry and Livelihoods*, **1**: 22–24

AEC/FNCCI (2006*) Compilation and prioritization of ten important NTFPs of Nepal for commercial promotion through private sector investment. Kathmandu*: Agro Enterprise Centre and Federation of Nepalese Chambers of Commerce and Industry

Adams RP Do N & Ge-lin C (1992) Preservation of DNA in plant specimens from tropical species by desiccation. *In Conservation of Plant Genes* (pp. 135–152). Elsevier. https://doi.org/10.1016/B978-0-12-044140-2.50013-2

Ahrendt LW A (1961) *Berberis* and *Mahonia*. *Botanical Journal of the Linnean Society*, **57**(369): 1–410. https://doi.org/10.1111/j.1095-8339.1961.tb00889.x

Alam N, Shinwari ZK, Ilyas M, & Ullah Z (2011) Indigenous knowledge of medicinal plants of Chagharzai valley, district Buner, Pakistan. *Pak. J. Bot*. **43**(2): 773–780

Althoff DM, Gitzendanner MA, Segraves KA, & Anderson F (2007) The utility of amplified fragment length polymorphisms in phylogenetics: A comparison of homology within and between genomes. *Systematic Biology*. **56**(3): 477–484. https://doi.org/10.1080/10635150701427077

Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*. **215**(3): 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Altschul S, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. **25**(17): 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Alvarez I & Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*. **29**(3): 417–34

Ankenbrand MJ, Keller A, Wolf M, Schultz J, & Förster F (2015) ITS2 database V: twice as much: Table 1. *Molecular Biology and Evolution.* **32**(11): 3030–3032. https://doi.org/10.1093/molbev/msv174

Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Bahkali AH, Al Sadoon M, Shobrak M (2010) A brief review of molecular techniques to assess plant dversity. *International Journal of Molecular Sciences*. **11**(5): 2079–2096. https://doi.org/10.3390/ijms11052079

Avise J (2004) *Molecular markers, natural history and evolution*. Springer Science and Business Media.

Baral SR & Kurmi PP (2006) *A compendium of medicinal plants in Nepal*. Rachana Sharma.

Barcaccia G, Albertini E, Rosellini D, Tavoletti S, & Veronesi F (2000) Inheritance and mapping of 2n-egg production in diploid alfalfa. *Genome*. **43**(3): 528–537. https://doi.org/10.1139/g00-017

Bhattarai KR & Ghimire M (2006) Commercially important medicinal and aromatic plants of Nepal and their distribution pattern and conservation measure along the elevation gradient of the Himalayas. *Banko Janakari*. **16**(1): 3–13. https://doi.org/10.3126/banko.v16i1.357

Bhattarai KR, Måren IE, & Chaudhary RP (2011) Medicinal plant knowledge of the Panchase region in the Middle Hills of the Nepalese Himalayas. *Banko Janakari*. **21**(2): 31–39. https://doi.org/10.3126/banko.v21i2.9127

Bhuju UR, Shakya PR, Basnet TB, & Shrestha S (2007) *Nepal biodiversity resource book: protected areas, Ramsar sites, and World Heritage sites. Nepal biodiversity resource book: protected areas, Ramsar sites, and World Heritage sites.* Kathmandu, Nepal: International Centre for Integrated Mountain Development (ICIMOD)

Botstein D, White RL, Skolnick M, & Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*. **32**(3): 314–31

Bromham L (2008) *Reading the story in DNA: a beginner's guide to molecular evolution*. Oxford University Press on Demand

Brown TA (2002) *Molecular Phylogenetics*. Wiley-Liss.

Bruno WJ, Socci ND, & Halpern AL (2000) Weighted Neighbor Joining: A likelihood-based approach to distance-based phylogeny reconstruction. *Molecular Biology and Evolution*. **17**(1): 189–197. https://doi.org/10.1093/oxfordjournals.molbev.a026231

Cavalli-Sforza LL & Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*. **21**(3): 550–570. https://doi.org/10.1111/j.1558-5646.1967.tb03411.x

Cavender-Bares J, Ackerly DD, Baum DA, & Bazzaz FA (2004) Phylogenetic overdispersion in floridian oak communities. *The American Naturalist*. **163**(163): 823–843.

CBOL Plant Working Group CPW, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, … Little DP (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(31), 12794–7. https://doi.org/10.1073/pnas.0905845106

Chan EY (2005) Advances in sequencing technology. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. **573**(1–2): 13–40. https://doi.org/10.1016/j.mrfmmm.2005.01.004

Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, … Chaw SM (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, Soltis D, Olmstead R, … DMA (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Annals of the Missouri Botanical Garden*. 528–580.

Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, … Stevens PF (2016) An update of the angiosperm phylogeny group classification for the orders

and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*. **181**(1): 1–20. https://doi.org/10.1111/boj.12385

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, … Wilkinson M (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*. **56**(2): 295–299.

Chase MW & Fay MF (2009) Barcoding of plants and fungi. *Science.* **325**(5941): 682–683. https://doi.org/10.1126/science.1176906

Chase MW & Hills HH (1991) Silica Gel: An ideal material for field preservation of leaf samples for DNA studies. *Taxon.* **40**(2): 215. https://doi.org/10.2307/1222975

Chaudhary RP, Uprety Y & Joshi SP (2016*)* Plant biodiversity in Nepal: Conservation and legal status. *Frontiers of Botany*, Tribhuvan University.

Chaudhary RP (1998) *Biodiversity in Nepal: status and conservation.* Know Nepal series. **17**.

Chauhan RS (2010) Socioeconomic Improvement through Medicinal and Aromatic Plants (MAPs) Cultivation in Uttarakhand, India. *Journal of Sustainable Agriculture*. **34**(6): 647–658. https://doi.org/10.1080/10440046.2010.493390.

Chhetri DB (1999) *Diversity of Medicinal and Aromatic Plants in Manang, Central Nepal with Emphasis on Ecology and Essential Oil Variation of Jatamansi (Natdostachys grandiflora DC.)*. M.Sc. Thesis. Central Department of Botany, Tribhuvan University, Kathmandu, Nepal.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, … Leon C (2010) Validation of the ITS2region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE.* **5**(1): e8613. https://doi.org/10.1371/journal.pone.0008613.

Chen Y, Ye W, Zhang Y, & Xu Y (2015) High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Research*. **43**(16): 7762–8. https://doi.org/10.1093/nar/gkv784

Cheng T, Xu C, Lei L, Li C, Zhang Y, & Zhou S (2016) 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

China Plant BOL Group CPB, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, … Duan GW (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America*. **108**(49): 19641–6. https://doi.org/10.1073/pnas.1104551108

Cloutier S & Landry BS (1994) Molecular markers applied to plant tissue culture. *In Vitro Cellular & Developmental Biology Plant*. **30**(1): 32–39. https://doi.org/10.1007/BF02632117

Cope JS, Corney D, Clark JY, Remagnino P, & Wilkin P (2012) Plant species identification using digital morphometrics: A review. *Expert Systems with Applications*. **39**(8): 7562–7573. https://doi.org/10.1016/J.ESWA.2012.01.073

Corlett RT (2016) Plant diversity in a changing world: Status, trends, and conservation needs. *Plant Diversity*. **38**(1): 10–16. https://doi.org/10.1016/J.PLD.2016.01.001

Costion C, Ford A, Cross H, Crayn D, Harrington M, & Lowe A (2011) Plant DNA barcodes can accurately estimate species richness in poorly known floras. *PLoS ONE*. **6**(11): e26841. https://doi.org/10.1371/journal.pone.0026841

Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, … Thomas WK

(2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*. **19**: 4–20. https://doi.org/10.1111/j.1365-294X.2009.04473.x

de Boer HJ, Ouarghidi A, Martin G, Abbad A, & Kool A (2014) DNA Barcoding reveals limited accuracy of identifications based on folk taxonomy. *PLoS ONE*. **9**(1): e84291. https://doi.org/10.1371/journal.pone.0084291

de Vere N, Rich TCG, Ford CR, Trinder SA, Long C, Moore CW, … Wilkinson MJ (2012) DNA barcoding the native flowering plants and conifers of Wales. *PloS One*. **7**(6): e37945. https://doi.org/10.1371/journal.pone.0037945

de Vere N, Rich TCG, Trinder SA, & Long C (2015) DNA barcoding for plants. In *Methods in molecular biology (Clifton, N.J.).* **1245**: 101–118. https://doi.org/10.1007/978-1-4939-1966-6_8

Desper R & Gascuel O (2002) Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *Journal of Computational Biology*. **9**(5): 687–705. https://doi.org/10.1089/106652702761034136

Després L, Gielly L, Redoutet B, & Taberlet P (2003) Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution*. **27**(2): 185–196. https://doi.org/10.1016/S1055-7903(02)00445-1

Devey DS, Chase MW, & Clarkson JJ (2009) A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. *Taxon*. **58**(1): 7–15.

DoF/MoFSC (2008) *Hamro Ban.* Department of Forests. Ministry of Forests and Soil Conservation.

Dong W, Liu J, Yu J, Wang L, & Zhou S (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE*. **7**(4): e35071. https://doi.org/10.1371/journal.pone.0035071

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, … Zhou S (2015) *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports*. **5**: 8348. https://doi.org/10.1038/srep08348

Dove CJ, Rotzel NC, Heacker M, & Weigt LA (2008) Using DNA barcodes to identify bird species involved in birdstrikes. *Journal of Wildlife Management*. **72**(5): 1231–1236. https://doi.org/10.2193/2007-272

Dowling TE, Moritz C, & Palmer JD (1990) *Nucleic acids II: restriction site analysis*. Molecular Systematics. Sinauer Associates Ltd., Sunderland. 250–317 pp

Doyle J (1991) DNA protocols for plants. In *Molecular Techniques in Taxonomy*. Berlin, Heidelberg: Springer Berlin Heidelber. 283–293. https://doi.org/10.1007/978-3-642-83962-7_18

Doyle JJ & Dickson EE (1987) Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon*. **36**(4): 715. https://doi.org/10.2307/1221122

DPR (2012) *Plant of Nepal: Fact sheet*. Kathmandu, Nepal: Department of Plant Resources.

Drummond AJ, Ho SYW, Phillips MJ, & Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*. **4**(5): e88. https://doi.org/10.1371/journal.pbio.0040088

Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology.* **7**(1): 214. https://doi.org/10.1186/1471-2148-7-

214

Dubey NK, Kumar R, & Tripathi P (2004) Global promotion of herbal medicine: India's opportunity. *Current Science*. Current Science Association. https://doi.org/10.2307/24109515

Dugan KA, Lawrence HS, Hares DR, Fisher CL, & Budowle B (2002) An improved method for post-PCR purification for mtDNA sequence analysis. *Journal of Forensic Sciences*. **47**(4): 811–8. Pubmed/12136989

Dunning LT & Savolainen V (2010) Broad-scale amplification of matK for DNA barcoding plants, a technical noteb oj_1071 1..9. *Botanical Journal of the Linnean Society*. **164**, 1–9

Ebihara A, Nitta JH, & Ito M (2010) Molecular Species Identification with Rich Floristic Sampling: DNA Barcoding the Pteridophyte Flora of Japan. *PLoS ONE*. **5**(12): e15136. https://doi.org/10.1371/journal.pone.0015136

Edwards DM (1996) The Trade in Non-Timber Forest Products from Nepal. *Mountain Research and Development*. **16**(4): 383. https://doi.org/10.2307/3673988

Edwards SV (2009) Is a new and general theory of molecular systematic emerging? *Evolution*. **63**(1): 1–19. https://doi.org/10.1111/j.1558-5646.2008.00549.x

Erickson DL, Jones FA, Swenson NG, Pei N, Bourg NA, Chen W, … Kress WJ (2014). Comparative evolutionary diversity and phylogenetic structure across multiple forest dynamics plots: a mega-phylogeny approach. *Frontiers in Genetics*. **5**, 358. https://doi.org/10.3389/fgene.2014.00358

Ewing B & Green P (1998b) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*. **8**(3): 186–94. pubmed/9521922

Ewing B, Hillier L, Wendl MC, & Green P (1998a) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research.* **8**(3): 175–85. https://doi.org/10.1101/GR.8.3.175

Fan L, Hui JHL, Yu ZG, & Chu KH (2014) VIP Barcoding: composition vector-based software for rapid species identification based on DNA barcoding. *Molecular Ecology Resources*. **14**(4): 871–881. https://doi.org/10.1111/1755-0998.12235

Fang G, Hammar S, & Grumet R (1992) A quick and inexpensive method for removing polysaccharides from plant genomic DNA. *BioTechniques.* **13**(1): 52–4, 56. pubmed/1503775

Farell EM & Alexandre G (2012) Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates. *BMC Research Notes*. **5**: 257. https://doi.org/10.1186/1756-0500-5-257

Farnsworth N (1988) *Screening plants for new medicines*. *Biodiversity*. Washington, USA: National Academy Press.

Farnsworth N & Soejarto D (1991) *Global importance of medicinal plants*. *conservation of medicinal plants*. Cambridge, UK: Cambridge University Press.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, … Barrett SCH (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE.* **3**(7): e2802. https://doi.org/10.1371/journal.pone.0002802

Felsenstein J (2005) *PHYLIP: Phylogenetic inference program*, version 3.6. University of Washington, Seattle.

Felsenstein J (1978) Cases in which Parsimony or compatibility methods will be positively misleading. *Systematic Biology*. **27**(4): 401–410.

https://doi.org/10.1093/sysbio/27.4.401

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*. **17**(6): 368–376. https://doi.org/10.1007/BF01734359

Fitch WM (1971) Toward Defining the Course of Evolution: Minimum change for a specific tree topology. *Systematic Biology.* **20**(4): 406–416. https://doi.org/10.1093/sysbio/20.4.406

Gaire BP & Subedi L (2011) Medicinal plant diversity and their pharmacological aspects of Nepal Himalayas. *Pharmacognosy Journal.* **3**(25): 6–17. https://doi.org/10.5530/PJ.2011.25.2

Gascuel O & Steel M (2006) Neighbor-Joining revealed. *Molecular Biology and Evolution*. **23**(11): 1997–2000. https://doi.org/10.1093/molbev/msl072

Ghimire SK (2008) Sustainable harvesting and management of medicinal plants in the Nepal Himalaya: current issues, knowledge gaps and research priorities, medicinal plants in Nepal. *An Anthology of Contemporary Research*. **25**(44).

Ghimire SK, Sapkota IB, Oli BR, & Parajuli RR (2008) *NTFP of Nepal Himalaya database of some important species found in the mountain protected areas and surrounding regions WWF Nepal* (1st edition). Kathmandu, Nepal: WWF Nepal.

Ghimire SK (2006) The endemic flora in Dolpo, North-West Nepal: distribution patterns, life forms, habitat specificity and conservation status. *Botanica Orientalis*, **5**: 30-39.

Ghorbani A, Saeedi Y, & de Boer HJ (2017) Unidentifiable by morphology: DNA barcoding of plant material in local markets in Iran. *PLOS ONE*. **12**(4): e0175722. https://doi.org/10.1371/journal.pone.0175722

GIZ (2011*) Medicinal and aromatic plants: Poverty impact assessment of proposed trade support measures in Nepal's MAPs sector*. Kathmandu: Deutsche Gesellschaft für Internationale Zusammenarbeit.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*. **11**(5): 759–769. https://doi.org/10.1111/j.1755-0998.2011.03024.x

Goldman N (1993) Statistical tests of models of DNA substitution. *Journal of Molecular Evolution.* **36**(2): 182–198. https://doi.org/10.1007/BF00166252

Goloboff PA, Farris JS, & Nixon KC (2008) TNT, a free program for phylogenetic analysis. *Cladistics*. **24**(5): 774–786. https://doi.org/10.1111/j.1096-0031.2008.00217.x

GoN/MoFSC (2014) *Nepal national biodiversity strategy and action plan: 2014-2020*. Ministry of Forests and Soil Conservation, Government of Nepal, Kathmandu.

Gregory TR (2005) DNA barcoding does not compete with taxonomy. *Nature*. **434**(7037): 1067–1067. https://doi.org/10.1038/4341067b

Gu W, Song J, Cao Y, Sun Q, Yao H, Wu Q, … Duan J (2013) Application of the ITS2 region for barcoding medicinal plants of selaginellaceae in pteridophyta. *PLoS ONE.* **8**(6): e67818. https://doi.org/10.1371/journal.pone.0067818

Guindon S, Gascuel O, & Rannala B (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. **52**(5): 696–704. https://doi.org/10.1080/10635150390235520

Hadi SII A, Santana H, Brunale PPM, Gomes TG, Oliveira MD, Matthiensen A, … Brasil BSAF (2016) DNA barcoding green microalgae isolated from neotropical inland waters. *PloS One*. **11**(2): e0149284. https://doi.org/10.1371/journal.pone.0149284

Hall BG (2008) *Phylogenetic trees made easy : a how-to manual*. Sinauer Associates.

Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. *Molecular*

Biology and Evolution. **30**(5): 1229–1235. https://doi.org/10.1093/molbev/mst012

Hall BG, Pikis A, & Thompson J (2009) Evolution and biochemistry of family 4 glycosidases: implications for assigning enzyme function in sequence annotations. *Molecular Biology and Evolution.* **26**(11): 2487–2497. https://doi.org/10.1093/molbev/msp162

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser*. **41**:95-98.

Hamilton A (2013) *Plant conservation: an ecosystem approach*. Routledge.

Hamilton AC (2004) Medicinal plants, conservation and livelihoods. *Biodiversity and Conservation*. **13**: 1477–1517.

Harry DE (1986) Inheritance and linkage of isozyme variants in incense-cedar. *Journal of Heredity*. **77**(4): 261–266. https://doi.org/10.1093/oxfordjournals.jhered.a110233

Hartigan JA (1973) Minimum mutation fits to a given tree. *Biometrics*. **29**(1): 53. https://doi.org/10.2307/2529676

Hebert PD & Barrett RD (2005) Reply to the comment by L. Prendini on "Identifying spiders through DNA barcodes" *Canadian Journal of Zoology*. **83**(3): 505–506. https://doi.org/10.1139/z05-026

Hebert PDN, Cywinska A, Ball SL, & deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences.* **270**(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert PDN, Penton EH, Burns JM, Janzen DH, & Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America*. **101**(41): 14812–7. https://doi.org/10.1073/pnas.0406166101

Heckenhauer J, Barfuss MHJ, & Samuel R (2016) Universal multiplexable *matK* primers for DNA barcoding of angiosperms. *Applications in Plant Sciences*. **4**(6): 1500137. https://doi.org/10.3732/apps.1500137

Hickerson MJ, Meyer CP, Moritz C, & Hedin M (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*. **55**(5): 729–739. https://doi.org/10.1080/10635150600969898

Hollingsworth Mi L, Andra Clark A, Forrest LL, Richardson J, Penninggton RT, Long DG., … Hollingsworth PM (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*. **9(**2): 439–457. https://doi.org/10.1111/j.1755-0998.2008.02439.x

Hollingsworth PM (2008) DNA barcoding plants in biodiversity hot spots: progress and outstanding questions. *Heredity*. **101**(1): 1–2. https://doi.org/10.1038/hdy.2008.16

Hollingsworth PM, Graham SW, Little DP, Beduschi G, & Vogler A (2011) Choosing and using a plant DNA barcode. *PLoS ONE.* **6**(5): e19254. https://doi.org/10.1371/journal.pone.0019254

Hollingsworth PM, Li DZ, van der Bank M, & Twyford AD (2016) Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. **371**(1702): https://doi.org/10.1098/rstb.2015.0338

Horiike T (2016) An introduction to molecular phylogenetic analysis. *Reviews in Agricultural Science*. **4**(0): 36–45. https://doi.org/10.7831/RAS.4.36-45

Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees.

Bioinfromatics Applications Note. **17**(8): 754–755. Retrieved from http://webpages.icav.up.pt/PTDC/BIA-BEC/104097/2008/23.pdf

Jarne P & Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*. **11**(10): 424–429. https://doi.org/10.1016/0169-5347(96)10049-5

Jones FA, Erickson DL, Bernal MA, Bermingham E, Kress WJ, Herre EA, … Turner BL (2011) The Roots of diversity: below ground species richness and rooting distributions in a tropical forest revealed by DNA barcodes and inverse modeling. *PLoS ONE*. **6**(9): e24506. https://doi.org/10.1371/journal.pone.0024506

Jukes TH & Cantor CR (1969) Evolution of protein molecules. *Mammalian protein metabolism*. **3**(21): 132.

Kang Y, Deng Z, Zang R, & Long W (2017) DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Scientific Reports*. **7**(1): 12564. https://doi.org/10.1038/s41598-017-13057-0

Karki M, Tiwari BK, Badony AK, & Bhattarai N (2003) Creating livelihoods enhancing medicinal and aromatic plants based biodiversity-rich production systems: preliminary lessons from South Asia; Paper presented: *3rd World Congress on Medicinal and Aromatic Plants for Human Welfare*. Chiang Mai: WOCMAP III.

Katoh K & Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics.* **9**(4): 286–298. https://doi.org/10.1093/bib/bbn013

Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, & Wolf M (2009) 5.8S-28S rRNA interaction and HMM-based *ITS2* annotation. *Gene*. **430**(1–2): 50–57. https://doi.org/10.1016/J.GENE.2008.10.012

Kim HK, Choi YH, & Verpoorte R (2011) NMR-based plant metabolomics: where do we stand, where do we go. *Trends in Biotechnology.* **29**(6): 267–275. https://doi.org/10.1016/j.tibtech.2011.02.001

Kim YD & Jansen RK (1994) Characterization and phylogenetic distribution of a chloroplast DNA rearrangement in theBerberidaceae. *Plant Systematics and Evolution.* **193**(1–4): 107–114. https://doi.org/10.1007/BF00983544

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution.* **16**(2): 111–120. https://doi.org/10.1007/BF01731581

Kingman JFC (1982) The coalescent. *Stochastic Processes and Their Applications*. **13**(3): 235–248. https://doi.org/10.1016/0304-4149(82)90011-4

Kishino H, Miyata T, & Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution.* **31**(2): 151–160. https://doi.org/10.1007/BF02109483

Koch MA, Dobeš C, & Mitchell-Olds T (2003) Multiple Hybrid Formation in Natural Populations: Concerted Evolution of the Internal Transcribed Spacer of Nuclear Ribosomal DNA (*ITS*) in North American Arabis divaricarpa (Brassicaceae). *Molecular Biology and Evolution*. **20**(3): 338–350. https://doi.org/10.1093/molbev/msg046

Kress WJ (2017) Plant DNA barcodes: Applications today and in the future. *Journal of Systematics and Evolution.* **55**(4): 291–307. https://doi.org/10.1111/jse.12254

Kress WJ, & Erickson DL (2007) A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLoS*

ONE. **2**(6): e508. https://doi.org/10.1371/journal.pone.0000508

Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, & Bermingham E (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences.* **106**(44): 18621–18626. https://doi.org/10.1073/pnas.0909820106

Kress WJ, Garcia-Robledo C, Uriarte M, & Erickson DL (2015) DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution.* **30**(1): 25–35. https://doi.org/10.1016/j.tree.2014.10.008

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, & Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*. **102**(23): 8369–74. https://doi.org/10.1073/pnas.0503123102

Krishnamurthy PK & Francis RA (2012) A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation*. **21**(8): 1901–1919. https://doi.org/10.1007/s10531-012-0306-2

Kumar S, Tamura K, & Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*. **5**(2): 150–163. https://doi.org/10.1093/bib/5.2.150

Kunwar RM, Mahat L, Acharya RP, & Bussmann RW (2013) Medicinal plants, traditional medicine, markets and management in far-west Nepal. *Journal of Ethnobiology and Ethnomedicine*. **9**: 24. https://doi.org/10.1186/1746-4269-9-24

Lahaye R, Bank M Van der … D. B.-… NA (2008) DNA barcoding the floras of biodiversity hotspots. *National Acad Sciences*, **105**(8): 2923–2928.

Lahaye R, Van der Bank M, Maurin O, Duthoit S, & Savolainen V (2008) A DNA barcode for the flora of the Kruger National Park (South Africa). *South African Journal of Botany*. **74**(2): 370–371. https://doi.org/10.1016/j.sajb.2008.01.073

Laikre L, Allendorf FW, Aroner LC, Baker CS, Gregovich DP, Hansen MM, … Waples R S (2010) Neglect of genetic diversity in implementation of the convention on biological diversity. *Conservation Biology.* **24**(1): 86–88. https://doi.org/10.1111/j.1523-1739.2009.01425.x

Larget B & Simon DL (1999) Markov Chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution.* **16**(6): 750–759

Li FW, Kuo LY, Rothfels CJ, Ebihara A, Chiou WL, Windham MD, & Pryer KM (2011) *rbcL* and *matK* earn two thumbs up as the core DNA barcode for ferns. *PLoS ONE.* **6**(10): e26597. https://doi.org/10.1371/journal.pone.0026597

Li H, & Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature.* **475**(7357): 493–496. https://doi.org/10.1038/nature10231

Li M, Cao H, But PPH, & Shaw PC (2011) Identification of herbal medicinal materials using DNA barcodes. *Journal of Systematics and Evolution.* **49**(3): 271–283. https://doi.org/10.1111/j.1759-6831.2011.00132.x

Li S, Pearl DK, & Doss H (2000) Phylogenetic tree construction using Markov Chain Monte Carlo. *Journal of the American Statistical Association.* **95**(450): 493–508. https://doi.org/10.1080/01621459.2000.10474227

Li X, Ding X, Chu B, Zhou Q, Ding G, & Gu S (2008) Genetic diversity analysis and conservation of the endangered Chinese endemic herb Dendrobium officinale Kimura et Migo (Orchidaceae) based on AFLP. *Genetica*. **133**(2): 159–166.

https://doi.org/10.1007/s10709-007-9196-8

Li X, Yang Y, Henry RJ, 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

Li Y, GAO LM, Poudel RC, Li DZ, & Forrest A (2011) High universality of matK primers for barcoding gymnosperms. *Journal of Systematics and Evolution*, *49*(3), 169–175. https://doi.org/10.1111/j.1759-6831.2011.00128.x

Liang H, & Hilu KW (1996) Application of the matK gene sequences to grass systematics. *Canadian Journal of Botany.* **74**(1): 125–134. https://doi.org/10.1139/b96-017

Liu J, Shi L, Han J, Li G, Lu H, Hou J, … Downie SR (2014) Identification of species in the angiosperm family Apiaceae using DNA barcodes. *Molecular Ecology Resources.* **14**(6): 1231–1238. https://doi.org/10.1111/1755-0998.12262

Liu Y, Yan HF, Cao T, & GE XJ (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

Lockhart P, Steel M, & Sullivan J (2005) A tale of two processes. *Systematic Biology*. **54**(6): 948–951. https://doi.org/10.1080/10635150500234682

Lolita GL, John DJM, Anthony RZ, & Fabian MD (2017) Chemical profiling and chemical standardization of Vitex negundo using 13C NMR. *Journal of Medicinal Plants Research*. **11**(1): 11–21. https://doi.org/10.5897/JMPR2015.6298

Lorenz TC (2012) Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *Journal of Visualized Experiments : JoVE*, **63**: e3998. https://doi.org/10.3791/3998

Luitel DR, Rokaya MB, Timsina B, & Münzbergová Z (2014) Medicinal plants used by the Tamang community in the Makawanpur district of central Nepal. *Journal of Ethnobiology and Ethnomedicine*. ***10***(5): https://doi.org/10.1186/1746-4269-10-5

Mahapatro G, Mishra D, Shaw K, Mishra S, & Jena T (2012) Phylogenetic tree construction for DNA sequences using clustering methods. *Procedia Engineering*. **38**, 1362–1366. https://doi.org/10.1016/j.proeng.2012.06.169

Manandhar NP (2002) *Plants and people of Nepal. Plants and people of Nepal.* Timber Press,USA.

Mankga LT, Yessoufou K, Moteetee AM, Daru BH, & van der Bank M (2013). Efficacy of the core DNA barcodes in identifying processed and poorly conserved plant materials commonly used in South African traditional medicine. *ZooKeys*. **365**: 215–33. https://doi.org/10.3897/zookeys.365.5730

Markert CL & Moller F (1959) Multiple forms of enzymes: tissue, ontogenetic and species specific patterns. *Proceedings of the National Academy of Sciences*. **45**(48): 215–199.

Maxam AM & Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences.* **74**(2)

Mba C & Tohme J (2005) Use of AFLP markers in surveys of plant diversity. *Methods in Enzymology.* **395**, 177–201. https://doi.org/10.1016/S0076-6879(05)95012-X

Meier R, Shiyang K, Vaidya G, & Ng PKL (2006) DNA barcoding and taxonomy in *Diptera*: A tale of high intraspecific variability and low identification success. *Systematic Biology*. **55**(5), 715–728. https://doi.org/10.1080/10635150600969864

Meier R, Zhang G, Ali F, & Zamudio K (2008) The use of mean instead of smallest interspecific distances eaggerates the size of the "Barcoding Gap" and leads to

misidentification. *Systematic Biology.* **57**(5): 809–813. https://doi.org/10.1080/10635150802406343

Meyer CP, & Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology.* **3**(12): e422. https://doi.org/10.1371/journal.pbio.0030422

Micales JA, Bonde MR, & Peterson GL (1992) Isoxyme analyssi in fungal taxonomy and molecular genetics. In *Handbook of applied myclology.* **4**: 57–79)

Milligan BG, LeebensMack J, & Strand AE (1994) Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology.* **3**(4): 423–435. https://doi.org/10.1111/j.1365-294X.1994.tb00082.x

Mishra P, Kumar A, Nagireddy A, Shukla AK, & Sundaresan V (2017) Evaluation of single and multilocus DNA barcodes towards species delineation in complex tree genus *Terminalia*. *PLOS ONE.* **12**(8): e0182836. https://doi.org/10.1371/journal.pone.0182836

Mizukami H, Ohbayashi K, Kitamura Y, & Ikenaga T (1993) Restriction fragment length polymorphisms(RFLPs) of medicinal plants and crude drugs. I. RFLP probes allow clear identification of *Duboisia* interspecific hybrid genotypes in both fresh and dried tissues. *Biological & Pharmaceutical Bulletin.* **16**(4): 388–390. https://doi.org/10.1248/bpb.16.388

MoAD (2017) *The state of Nepal's biodiversity for food and agriculture* (*BK Joshi, AK Acharya, D Gauchan, and P Chaudhary, eds*). Ministry of Agricultural Development, Kathmandu, Nepal.

MoFSC (2009) *Nepal fourth national report to the convention on biological diversity: Government of Nepal Ministry of Forests and Soil Conservation; Government of Nepal*. Kathmandu, Nepal

Mohammadi SA, & Prasanna BM (2003) Analysis of genetic diversity in crop plants salient statistical tools and considerations. *Crop Science.* **43**(4): 1235. https://doi.org/10.2135/cropsci2003.1235

Mondini L, Noorani A, & Pagnotta M (2009) Assessing plant genetic diversity by molecular tools. *Diversity.* **1**(1): 19–35. https://doi.org/10.3390/d1010019

Moyo M, Amoo SO, Bairu MW, Finnie JF, & Van Staden J (2008) Optimising DNA isolation for medicinal plants. *South African Journal of Botany.* **74**(4): 771–775. https://doi.org/10.1016/J.SAJB.2008.07.001

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, & Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology.* **51**: 263–73. https://doi.org/10.1101/SQB.1986.051.01.032

Nei M Kumar S (2000) *Molecular Evolution and Phylogenetics*. OXFORD University Press.

Neupane S, Jaishree S, Bhattarai T, & Shrestha S (2017) Genetic diversity in Nepalese population of *Swertia chirayita* (Roxb. Ex Fleming) H. Karst based on inter-simple sequence repeats (ISSR) markers. *African Journal of Biotechnology.* **16**(16): 895–907. https://doi.org/10.5897/AJB2016.15858

Newmaster SG, Fazekas AJ, & Ragupathy S (2006) DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Canadian Journal of Botany.* **84**(3): 335–341. https://doi.org/10.1139/b06-047

Newmaster SG, Fazekas AJ, Steeves RAD, & Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources.* **8**(3): 480–490. https://doi.org/10.1111/j.1471-8286.2007.02002.x

Palmer JD, Jansen RK, Michaels HJ, Chase MW, & Manhart JR (1988) Chloroplast DNA Variation and Plant Phylogeny. *Annals of the Missouri Botanical Garden.* **75**(4): 1180. https://doi.org/10.2307/2399279

Paterson AH, Tanksley SD, & Sorrells ME (1991) DNA Markers in Plant Improvement. *Advances in Agronomy*. **46**: 39–90. https://doi.org/10.1016/S0065-2113(08)60578-7

Paudel PK, Bhattarai BP, & Kindlmann P (2012) An overview of the biodiversity in Nepal. In *Himalayan biodiversity in the changing World,* Dordrecht: Springer Netherlands. 1–40. https://doi.org/10.1007/978-94-007-1802-9_1

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, … Fedak G (1993) Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods and Applications.* **2**(4): 341–5. https://doi.org/10.1101/GR.2.4.341

Pennisi E (2007) Taxonomy. Wanted: a barcode for plants. *Science (New York, N.Y.)*. **318**(5848): 190–1. https://doi.org/10.1126/science.318.5848.190

Pevsner J (2017) *Bioinformatics and functional genomics*. John Wiley & Sons.

Phondani PC, Maikhuri RK, & Saxena KG (2014) The efficacy of herbal system of medicine in the context of allopathic system in Indian Central Himalaya. *Journal of Herbal Medicine*. **4**(3): 147–158. https://doi.org/10.1016/j.hermed.2014.05.004

Poczai P & Hyvönen J (2010) Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Molecular Biology Reports.* **37**(4): 1897–1912. https://doi.org/10.1007/s11033-009-9630-3

Poczai P, & Varga I Bell N (2012) Genomics meets biodiversity: advances in molecular marker development and their applications in plant genetic diversity assessment. In *The molecular basis of plant genetic diversity*. 3–26

Polunin O, & Stainton A (1984) *Flowers of the Himalaya. Flowers of the Himalaya.* Oxford University Press.

Pond SL K & Muse SV (2005) HyPhy: Hypothesis testing using phylogenies. In *Statistical Methods in Molecular Evolution*. 125–181 https://doi.org/10.1007/0-387-27733-1_6

Porebski S, Bailey LG, & Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter.* **15**(1): 8–15. https://doi.org/10.1007/BF02772108

Poudel RC, Möller M, Gao LM, Ahrends A, Baral SR, Liu J, … Li DZ (2012) Using morphological, molecular and climatic data to delimitate Yews along the Hindu Kush-Himalaya and adjacent regions. *PLoS ONE*. **7**(10): e46873. https://doi.org/10.1371/journal.pone.0046873

Press JR, Shrestha KK, & Sutton DA (2000) *Annotated checklist of the flowering plants of Nepal. Annotated checklist of the flowering plants of Nepal.* Natural History Museum Publications.

Primack RB, Paudel PK, & Bhattarai BP (2013) *Conservation biology: A primer for Nepal.* Dreamland Publication, Kathmandu, Nepal.

Primmer CR (2009) The year in ecology and conservation biolgoy from conservation Genetics to Conservation Genomics Background: uses of genetic methods in conservation. *Annals of the New York Academy of Sciences.* **1162**: 357–368 https://doi.org/10.1111/j.1749-6632.2009.04444.x

Rajbhandari KR (2001) *Ethnobotany of Nepal. Ethnobotany of Nepal.* Kathmandu, Nepal: Ethnobotanical Society of Nepal.

Rajbhandari KR and Rai SK (2017) *A handbook of the flowering plants of Nepal*. Government of Nepal, Ministry of Forests and Soil Conservation, Department of Plant Resources, Kathmandu, Nepal. **1**

Rannala B & Yang Z (1996) Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *Journal of Molecular Evolution.* **43**(3): 304–311. https://doi.org/10.1007/BF02338839

Ratnasingham S & Hebert PDN (2007) Barcoding: bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*. **7**(3): 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rawal DS, Sijapati J, Rana N, Rana P, Giri A, & Shrestha S (2009) Some High Value Medicinal Plants of Khumbu Region Nepal. *Nepal Journal of Science and Technology*. **10**: 73–82.

Ren F, Tanaka H, & Gojobori T (1995) Construction of molecular evolutionary phylogenetic trees from DNA sequences based on minimum complexity principle. *Computer Methods and Programs in Biomedicine*. **46**(2): 121–130. https://doi.org/10.1016/0169-2607(94)01609-J

Rokaya MB, Münzbergová Z, Shrestha MR & Timsina B (2012) Distribution patterns of medicinal plants along an elevational gradient in central Himalaya, Nepal. *Journal of Mountain Science*. **9**(2): 201–213. https://doi.org/10.1007/s11629-012-2144-9

Rzhetsky A & Nei M (1992) A simple method for estimating and testing minimum-evolution trees. *Molecular Biology and Evolution*. **9**(5): 945-967

Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. **4**(4): 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Sanger F (1988) Sequences. *Annual Review of Biochemistry*. **57**(1): 1–29. https://doi.org/10.1146/annurev.bi.57.070188.000245

Sanger F, Nicklen S, & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences.* **74**(12)

Särkinen T, Staats M, Richardson JE, Cowan RS, & Bakker FT (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

Sarwat M, & Yamdagni MM (2016) DNA barcoding, microarrays and next generation sequencing: recent tools for genetic diversity estimation and authentication of medicinal plants. *Critical Reviews in Biotechnology*. **36**(2): 191–203. https://doi.org/10.3109/07388551.2014.947563

Sass C, Little DP, Stevenson DW, & Specht CD (2007) DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PLoS ONE*. **2**(11): e1154. https://doi.org/10.1371/journal.pone.0001154

Saunders GW, & McDevit DC (2012) Methods for DNA barcoding photosynthetic protists emphasizing the macroalgae and diatoms, *Methods Mol Biol.* **858**: 207-22. https://doi.org/10.1007/978-1-61779-591-6_10

Scandalios JG (1969) Genetic control of multiple molecular forms of enzymes in plants: A review. *Biochemical Genetics.* **3**(1): 37–79. https://doi.org/10.1007/BF00485973

Schilling T Watson M (2014) Clarifying the identities of two Nepalese Mahonia. *The Plantsman, The Royal Horticultural Society*. 94–100.

Semagn K, Bjørnstad Å, & Ndjiondjop M (2006) An overview of molecular marker methods for plants. *African Journal of Biotechnology*, **5**(25): 2540–2568.

Shaw J, Lickey EB, Schilling EE, & Small RL (2007) 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

Sheng-Ji P (2001) Ethnobotanical approaches of traditional medicine studies: Some experiences From Asia. *Pharmaceutical Biology.* **39**(1): 74–79. https://doi.org/10.1076/phbi.39.s1.74.0005

Shrestha PM & Dhillion SS (2003) Medicinal plant diversity and use in the highlands of Dolakha district, Nepal. *Journal of Ethnopharmacology*. **86**(1): 81–96. https://doi.org/10.1016/S0378-8741(03)00051-5

Shrestha TB & Joshi RM (1996*). Rare, Endemic and Endangered Plants of Nepal.* WWF Nepal Program, Kathmandu, Nepal.

Shrestha KK, Tiwari NN, & Ghimire SK (2000) *MAPDON-Medicinal and aromatic plant database of Nepal. In: The Himalayan Plants, Can They Save Us*. Society for the Conservation and Development of Himalayan Medicinal Resources (SCDHMR), Tokyo, Japan. 53-74 pp.

Smith B & Chinnappa CC (2015) *Plant collection, identification, and herbarium procedures.* In: Yeung E., Stasolla C., Sumner M., Huang B. (eds) Plant Microtechniques and Protocols. Springer, https://doi.org/10.1007/978-3-319-19944-3_30

Smith MA, Rodriguez JJ, Whitfield JB, Deans AR, Janzen DH, Hallwachs W, & Hebert PDN (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proceedings of the National Academy of Sciences of the United States of America*. **105**(34): 12359–64. https://doi.org/10.1073/pnas.0805319105

Smith MA, Woodley NE, Janzen DH, Hallwachs W, & Hebert PDN (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of the United States of America* **103**(10): 3657–62. https://doi.org/10.1073/pnas.0511318103

Soltis DE & Soltis PS (1998) *Choosing an approach and an appropriate gene for phylogenetic analysis*. Molecular Systematics of Plants II. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-5419-6_1

Soltis DE, Soltis PS, Endress PK, & Chase MW (2005). *Phylogeny and evolution of angiosperms.* Sinauer Associates Incorporated.

Soltis PS & Soltis DE (2004) The origin and diversification of angiosperms The origin and diversification of angiosperms. *American Journal of Botany.* **91:** https://doi.org/10.3732/ajb.91.10.1614

Song J, Shi L, Li D, Sun Y, Niu Y, Chen Z, … Chen S (2012) Extensive pyrosequencing reveals frequent intra-genomic variations of internal transcribed spacer regions of nuclear ribosomal DNA. *PLoS ONE.* **7**(8): e43971. https://doi.org/10.1371/journal.pone.0043971

Srivastava, S, Sanchita, Bhargava, M, & Sharma, A (2016) DNA barcoding of medicinal plants. In Tsay HS, Shyur LF, Agrawal DC, Wu YC, & Wang (Eds.) *Medicinal Plants. Recent Advances in Research and Development*. Springer Singapore. https://doi.org/10.1007/978-981-10-1085-9_4

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. **22**(21): 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Staub JE, Box J, Meglic V, Horejsi TF, & McCreight JD (1997) Comparison of isozyme and random amplified polymorphic DNA data for determining intraspecific variation in Cucumis. *Genetic Resources and Crop Evolution.* **44**(3): 257–269. https://doi.org/10.1023/A:1008639616331

Stech M & Quandt D (2010) 20,000 species and five key markers: the status of molecular bryophyte phylogenetics. *Phytotaxa.* **9**(1): 196. https://doi.org/10.11646/phytotaxa.9.1.11

Štorchová H & Olson MS (2007) The architecture of the chloroplast *psbA-trnH* non-coding region in angiosperms. *Plant Systematics and Evolution*. **268**(1–4): 235–256. https://doi.org/10.1007/s00606-007-0582-6

Taberlet P, Coissac E, Pompanion F, Brochmann C, & Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*. **21**(8): 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, … Willerslev E (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*. **35**(3): 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

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, & Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*. **28**(10): 2731–2739. https://doi.org/10.1093/molbev/msr121

Taylor HR & Harris WE (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, Crockett S, Khan I, & Scheffler B (2004) Authentication of medicinal plants using molecular biology techniques to compliment conventional methods. *Current Medicinal Chemistry.* **11**(11): 1391–1401. https://doi.org/10.2174/0929867043365206

Thomas JC, Khoury R, Neeley CK, Akroush AM, & Davies EC (1997) A fast CTAB method of human DNA isolation for polymerase chain reaction applications. *Biochemical Education*. **25**(4): 233–235. https://doi.org/10.1016/S0307-4412(97)00122-2

Thompson JD, Higgins DG, & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.* **22**(22): 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Till BJ, Jankowicz-Cieslak J, Huynh OA, Beshir MM, Laport RG, & Hofinger BJ (2015) *Sample collection and storage*. In l*ow-cost methods for molecular characterization of mutant plants.* Springer Cham. 9–11 pp. https://doi.org/10.1007/978-3-319-16259-1_3

Torre DL, AR, Li Z, Van de Peer Y, & Ingvarsson PK (2017) contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering pants. *Molecular Biology and Evolution*. **34**(6): 1363–1377.

https://doi.org/10.1093/molbev/msx069

Stoeckle M, Janzen D, Hallwachs W, Hanken J, & Baker J (2003) Taxonomy, DNA, and the barcode of life," *Draft conference report. In Rubinoff DNA Barcodes and Conservation Barcode Conference.*

Subedi BP (2006*) Linking plant-based enterprises and local communities to biodiversity conservation in Nepal Himalaya.* Akhil Book Distributors.

Swofford DL (2000) *PAUP*: Phylogenetic analysis by parsimony, version 4*. Sinauer, Massachusetts

Swofford DL (2003) PAUP*: phylogenetic analysis using parsimony, version 4.0 b10

UNEP (2012) *Green Economy Sectoral Study: BioTrade - Harnessing the potential for transitioning to a green economy – The Case of Medicinal and Aromatic Plants in Nepal*. http://re.indiaenvironmentportal.org.in/files/file/Bio Trade Medicinal and Aromatic Plants in Nepal.pdf

Uprety Y, Asselin H, Boon EK, Yadav S, & Shrestha KK (2010) Indigenous use and bio-efficacy of medicinal plants in the Rasuwa District, Central Nepal. *Journal of Ethnobiology and Ethnomedicine*. **6**(1): 3. https://doi.org/10.1186/1746-4269-6-3

Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F, … Taberlet P (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Resources*. **9**(1): 51–60. https://doi.org/10.1111/j.1755-0998.2008.02352.x

Vogler AP & Monaghan MT (2007) Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research*. **45**(1): 1–10. https://doi.org/10.1111/j.1439-0469.2006.00384.x

Welsh J & McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research.* **18**(24): 7213–7218. https://doi.org/10.1093/nar/18.24.7213

White T, Bruns T, Lee S, & Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications.* **18**(1): 315–322.

Whitlock BA, Hale AM, & Groff PA (2010) Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. *PLoS ONE.* **5**(7): e11533. https://doi.org/10.1371/journal.pone.0011533

WHO (2013) WHO traditional medicine strategy: 2014-2023 <http://www.who.int/medicines/publications/traditional/trm_strategy14_23/en/>

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, & Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*. **18**(22): 6531–6535. https://doi.org/10.1093/nar/18.22.6531

Wold S (1995) Chemometrics and intelligent laboratory systems Chemometrics; what do we mean with it, and what do we want from it? *Chemometrics and Intelligent Laboratory Systems.* **30**: 109–115.

Yang Z (1996) Phylogenetic analysis using parsimony and likelihood methods. *Journal of Molecular Evolution.* **42**(2): 294–307. https://doi.org/10.1007/BF02198856

Yang Z (2006) *Computational molecular evolution*. Oxford University Press.

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*. **24**(8): 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang Z & Rannala B (2012) Molecular phylogenetics: principles and practice. *Nature Reviews Genetics.* **13**(5): 303–314. https://doi.org/10.1038/nrg3186

Yessoufou K (2005) Ecological and ethnobotanical research on *Irvingia gabonensis* and *Blighia sapida* in Plateau Province, Eastern Benin. MSc thesis, University of Abomey-Calavi,Benin.

Yongyu Z, Shujun S, Jianye D, … W WQ C (2011) Quality control method for herbal medicine-chemical fingerprint analysis. *Intechopen.Com*. https://www.intechopen.com/download/pdf/23472

Yu N, Wei Y, Zhang X, Zhu N, Wang Y, Zhu Y, … Sun A (2017) Barcode ITS2: a useful tool for identifying Trachelospermum jasminoides and a good monitor for medicine market. *Scientific Reports.* **7**(1): 5037. https://doi.org/10.1038/s41598-017-04674-w

Yuan QJ, Zhang B, Jiang D, Zhang WJ, Lin TY, Wang NH, … Huang LQ (2015) Identification of species and materia medica within Angelica L. (Umbelliferae) based on phylogeny inferred from DNA barcodes. *Molecular Ecology Resources.* **15**(2): 358–371. https://doi.org/10.1111/1755-0998.12296

Zhang Z, Schwartz S, Wagner L, & Miller W (2000) A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology.* **7**(1–2): 203–214. https://doi.org/10.1089/10665270050081478

Zheng W, Wang L, Meng L, & Liu J (2007) Genetic variation in the endangered Anisodus tanguticus (Solanaceae), an alpine perennial endemic to the Qinghai-Tibetan Plateau. *Genetica.* **132**(2): 123–129. https://doi.org/10.1007/s10709-007-9154-5

Zhu JK, Neuteboom L, Pinas J, Chardonnes A, Schat H, Verkleji J, … Guerinot M Lou (2000) Genetic analysis of plant salt tolerance using Arabidopsis. *Plant Physiology*. **124**(3): 941–8. https://doi.org/10.1104/pp.124.3.941

Zuo Y, Chen Z, Kondo K, Funamoto T, Wen J, & Zhou S (2011) DNA barcoding of *Panax* species. *Planta Medica*. **77**(02): 182–187. https://doi.org/10.1055/s-0030-1250166

Zurawski G, Clegg M, Genetics AB (1984) The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics Soc America*. http://www.genetics.org/content/106/4/735.short

Zwickl D (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion (Doctoral dissertation)*. The University of Texas, Austin

http://www.floraofnepal.org/countryinformation

APPENDICES

Appendix 1 studied plant species, habit, altitude, latitude, longitude, family, local name, collection code, collectors, and lab code.

SN	Identification	Name in Field	Family	Collection_no	Lab code	Collection_date	Collectors	Habit	Altitude (m)	Latitude			Longitude		
1	***Viburnum mullaha*** Buch.-Ham. ex D. Don	Malo	Adoxaceae	CNRCP42	B30	2072-07-17	(RCP)	Tree	2092	28	23	10.4	83	52	7.6
2	***Arisaema speciosum*** (Wall.) Mart. ex Schott & Endl.	Dhakayo	Araceae	CN50	C33	2073-6-11		Herb	3014	28	24	20.9	83	51	30
3	***Panax pseudo-ginseng*** Wall.	Ginseng	Araliaceae	CN16	**C9**	2073-6-9	RCP, DK, MB	Herb	3261	28	24	51.3	83	51	19.9
4	***Asparagus filicinus*** Buch.-Ham. ex D. Don	Kurilo	Asparagaceae	CNRCP50	*B38*	2072-07-17	RCP	Shrub	2092	28	23	10.4	83	52	7.6
5	***Maianthemum oleraceum*** (Baker) LaFrankie	Badure	Asparagaceae	CN51	C34	2073-6-11	RCP, DK, MB	Herb	3014	28	24	20.9	83	51	30
6	***Anaphalis contorta*** (D. Don) Hook. f.	Buki Phool	Asteraceae	CNRCP15	B8	2072-07-14	RCP	Herb	3208	28	21	39.9	83	57	37.5
7	***Artemisia dubia*** Wall. ex Besser	Artemisia	Asteraceae	CN52	C35	2073-6-11	RCP, DK, MB	Shrub	2092	28	24	51.4	83	51	11.7
8	***Artemisia indica*** Willd.	Pati	Asteraceae	CNRCP77	*B59*	2072-07-19	RCP	Shrub	1238	28	24	51.4	83	51	11.7
9	***Senecio graciliflorus*** DC	Senecio	Asteraceae	CN37	C22	2073-6-10	RCP, DK, MB	Shrub	3537	28	26	1.7	83	52	5.5
10	***Taraxacum officinale*** Wigg.		Asteraceae	CNRCP16	B9	2072-07-14	RCP	Herb	3208	28	23	10.4	83	52	7.6
11	***Diplazium maximum*** (D. Don) C. Christensen	Neuro	Athyriaceae	CNRCP59	*B47*	2072-07-17	RCP	Fern	2092	28	23	10.4	83	52	7.6
12	***Diplazium spectabile*** (Wall. ex Mett.) Ching	Lakhuto	Athyriaceae	CNRCP55	*B43*	2072-07-17	RCP	Fern	2092	28	23	10.4	83	52	7.6
13	***Berberis angulosa*** Wall. ex Hook. f. & Thomson	Berberis	Berberidaceae	CN13	C7	2073-6-9	RCP, DK, MB	Shrub	3993	28	24	12	83	51	25
14	***Berberis asiatica*** Roxb. ex DC.	Chutro	Berberidaceae	CNRCP68	B56	2072-07-17	RCP	Shrub	1847	28	23	4.4	83	52	29.4
15	***Berberis erythroclada*** Ahrendt	Chutro	Berberidaceae	CNRCP02	B2	2072-07-13	RCP	Shrub	2998	28	23	4.4	83	52	29.4
16	***Mahonia napaulensis*** DC.	Bhutro	Berberidaceae	CNRCP63	*B51*	2072-07-17	RCP	Tree	1847	28	27	12	83	52	35.4
17	***Alnus nepalensis*** D. Don	Utis	Betulaceae	CNRCP65	*B53*	2072-07-17	RCP	Tree	1847	28	25	9.5	83	51	25.8
18	***Betula utilis*** D. Don	Bhoj Patra	Betulaceae	CNRCP17	B10	2072-07-14	RCP	Tree	3360	28	23	4.4	83	52	29.4
19	***Betula utilis*** D. Don	Betula	Betulaceae	CN42	C25	2073-6-11	RCP, DK, MB	Tree	3358	28	25	13.6	83	57	29.2
20	***Nardostachys grandiflora*** DC.	Jatamansi	Caprifoliaceae	CNRCP20	B13	2072-07-15	RCP	Herb	3732	28	26	35.3	83	52	32.3
21	***Nardostachys grandiflora*** DC.	Jatamansi	Caprifoliaceae	CN2	*C2*	2073-6-9	RCP, DK, MB	Herb	3732	28	26	35.3	83	52	32.3
22	***Euonymus tingens*** Wall.	Kesari	Celastraceae	CNRCP08	B5	2072-07-13	RCP	Tree	3014	28	24	20.9	83	51	30
23	***Parnassia nubicola*** Wall. ex Royle	Parnasia	Celastraceae	CN33	C19	2073-6-10	RCP, DK, MB	Herb	3537	28	26	1.7	83	52	5.5
24	***Rhodiola fastigiata*** (Hook. f. & Thomson) S. H. Fu	Crasssulaceae	Crassulaceae	CN20	C12	2073-6-10	RCP, DK, MB	Herb	3537	28	26	1.7	83	52	5.5
25	***Trichosanthes wallichiana*** (Ser.) Wight,	Indreni	Cucurbitaceae	CNRCP52	B40	2072-07-17	RCP	Climber	2092	28	23	10.4	83	52	7.6
26	***Cyathea spinulosa*** Wall. ex Hook.	Rukha Unyu	Cyatheaceae	CNRCP72	*B57*	2072-07-18	RCP	Shrub	1847	28	23	4.4	83	52	29.4
27	***Daphniphyllum himalense*** (Benth.) Mull. Arg.	Rachan	Daphniphyllaceae	CNRCP58	*B46*	2072-07-17	RCP	Tree	2092	28	23	10.4	83	52	7.6
28	***Cassiope fastigiata*** (Wall.) D. Don	Cassiope	Ericaceae	CN5	C3	2073-6-9	RCP, DK, MB	Herb	3993	28	25	9.5	83	51	25.8
29	***Lyonia ovalifolia*** (Wall.) Drude	Angeri	Ericaceae	CNRCP43	B31	2072-07-17	RCP	Tree	2092	28	23	56.5	83	51	21.6
30	***Lyonia ovalifolia*** (Wall.) Drude	Lyonia	Ericaceae	CN43	C26	2073-6-11	RCP, DK, MB	Tree	3358	28	23	10.4	83	52	7.6
31	***Rhododendron anthopogon*** D. Don	Sunpati	Ericaceae	CN6	C4	2073-6-9	RCP, DK, MB	Shrub	3993	28	25	13.6	83	57	29.2
32	***Rhododendron arboreum*** Sm.	Gurans	Ericaceae	CNRCP30	B22	2072-07-17	RCP	Tree	2827	28	25	45.8	83	51	53.5
33	***Rhododendron barbatum*** Wall. ex G. Don	Gurans	Ericaceae	CN44	C27	2073-6-11	RCP, DK, MB	Tree	3462	28	27	12	83	52	35.4
34	***Rhododendron campanulatum*** D. Don	Guras	Ericaceae	CNRCP18	B11	2072-07-14	RCP	Tree	3360	28	27	12	83	52	35.4
35	***Castanopsis tribuloides*** (Sm.) A. DC.	Musure katus	Fagaceae	CNRCP64	*B52*	2072-07-17	RCP	Tree	1847	28	23	38.9	83	51	38.7
36	***Lithocarpus elegans*** (Blume) Hatus. ex Soepadmo	Arkhalo	Fagaceae	CNRCP60	*B48*	2072-07-17	RCP	Tree	2092	28	23	10.4	83	52	7.6
37	***Quercus lamellosa*** Sm.	Phalat	Fagaceae	CNRCP38	B28	2072-07-17	RCP	Tree	2400	28	23	10.4	83	52	7.6
38	***Quercus semecarpifolia*** Sm.	Kharsu	Fagaceae	CNRCP51	B39	2072-07-17	RCP	Tree	2092	28	23	4.4	83	52	29.4
39	***Swertia chirayita*** (Roxb. ex Fleming) Karsten	Chirayito	Gentianaceae	CNRCP01	B1	2072-07-13	RCP	Herb	2998	28	24	12	83	51	25

SN	Identification	Name in Field	Family	Collection_no	Lab code	Collection_date	Collectors	Habit	Altitude (m)	Latitude			Longitude		
40	***Swertia teres*** (G. Don) J. Shah	Swertia	Gentianaceae	CN10	C6	2073-6-9	RCP, DK, MB	Herb	3993	28	27	12	83	52	35.4
41	***Geranium wallichianum*** D. Don ex Sweet	Geranium	Geraniaceae	CN46	C29	2073-6-11	RCP, DK, MB	Herb	3462	28	25	45.8	83	51	53.5
42	***Engelhardia spicata*** Lesch. ex Blume	Mauwa	Juglandaceae	CNRCP67	*B55*	2072-07-17	RCP	Tree	1847	28	23	4.4	83	52	29.4
43	***Leucosceptrum canum*** Sm.	Dhurseli	Lamiaceae	CNRCP54	*B42*	2072-07-17	RCP	Tree	2092	28	23	10.4	83	52	7.6
44	***Lindera neesiana*** (Wall. ex Nees) Kurz.	SilTimur	Lauraceae	CNRCP66	*B54*	2072-07-17	RCP	Tree	1847	28	23	10.4	83	52	7.6
45	***Lindera pulcherrima*** (Nees) Benth. ex. Hook.f.	Seto lokar	Lauraceae	CNRCP47	B35	2072-07-17	RCP	Tree	2092	28	23	10.4	83	52	7.6
46	***Machilus odoratissima*** (D Don) S.N. Biswa	Bhatekaulo	Lauraceae	CNRCP70	*B62*	2072-07-17	RCP	Tree	1847	28	23	4.4	83	52	29.4
47	***Persea duthiei*** (King) A.J.G. H. Kosterm	Kathe Kaulo	Lauraceae	CNRCP56	*B44*	2072-07-17	RCP	Tree	2092	28	23	4.4	83	52	29.4
48	***Indigofera atropurpurea*** Buch.-Ham. ex Hornem	Indigofera	Leguminosae	CN54	C36	2073-6-12	RCP, DK, MB	Shrub	2092	28	23	10.4	83	52	7.6
49	***Lycopodium japonicum*** Thunb	Nagbeli	Lycopodiaceae	CNRCP62	*B50*	2072-07-17	RCP	Fern	1847	28	23	4.4	83	52	29.4
50	***Magnolia campbellii*** Hook. f. & Thomson	Champ	Magnoliaceae	CNRCP28	B20	2072-07-17	RCP	Tree	2827	28	23	56.5	83	51	21.6
51	***Michelia doltsopa*** Buch.-Ham. ex DC.	Chapo	Magnoliaceae	CNRCP41	B29	2072-07-17	RCP	Tree	2300	28	23	38.9	83	51	38.7
52	***Michelia kisopa*** Buch.-Ham. ex DC.	Chanp	Magnoliaceae	CNRCP57	B45	2072-07-17	RCP	Tree	2092	28	23	10.4	83	52	7.6
53	***Paris polyphylla*** Sm.	Satuwa	Melanthiaceae	CNRCP03	B3	2072-07-13	RCP	Herb	2998	28	24	12	83	51	25
54	***Paris polyphylla*** Sm.	Satuwa	Melanthiaceae	CN1	*C1*	2073-6-6	RCP, DK, MB	Herb	2700						
55	***Myrsine semiserrata*** Wall.	Kalikath	Myrsinaceae	CNRCP49	B37	2072-07-17	RCP	Tree	2092	28	23	10.4	83	52	7.6
56	***Dactylorhiza hatagirea (D. Don) Soo***	Panchaunle	Orchidaceae	CNRCP21	*B61*	2072-07-15	RCP	Herb	3732	28	26	35.3	83	52	32.3
57	***Dactylorhiza hatagirea (D. Don) Soo***	Panchaule	Orchidaceae	CNRCP14	B7	2072-07-14	RCP	Herb	3208	28	24	51.4	83	51	11.7
58	***Dactylorhiza hatagirea*** (D. Don) Soo	Panchaunle	Orchidaceae	CN18	C11	2073-6-10	RCP, DK, MB	Herb	3787	28	26	35.9	83	52	32.4
59	***Danthonia cumminsii*** J. D. Hooker	Salim Grass 1	Poaceae	CN23	C13	2073-6-10	RCP, DK, MB	Herb	3537	28	23	56.5	83	51	21.6
60	***Thamnocalamus nepalensis*** (Stapleton) Stapleton	Jarbuto	Poaceae	CNRCP05	*B60*	2072-07-13	RCP	Shrub	2998	28	24	12	83	51	25
61	***Yushania maling*** (Gamble) R.B. Majumdar	Malingo	Poaceae	CNRCP27	B19	2072-07-17	RCP	Shrub	2827	28	26	1.7	83	52	5.5
62	***Polygonum molle*** D. Don	Thotne	Polygonaceae	CNRCP35	B27	2072-07-17	RCP	Shrub	2523	28	23	38.9	83	51	38.7
63	***Polygonum molle*** D. Don	Thotne	Polygonaceae	CN38	C23	2073-6-10	RCP, DK, MB	Herb	3537	28	23	38.9	83	51	38.7
64	***Bistorta macrophylla*** (D. Don) Sojak	Bistorta	Polygonaceae	CN34	C20	2073-6-10	RCP, DK, MB	Herb	3537	28	26	35.3	83	52	32.3
65	***Rheum australe*** D. Don	Padamchal	Polygonaceae	CN17	C10	2073-6-10	RCP, DK, MB	Herb	3732	28	26	1.7	83	52	5.5
66	***Rumex nepalensis*** Spreng.	Halhale	Polygonaceae	CNRCP34	B26	2072-07-17	RCP	Herb	2523	28	26	1.7	83	52	5.5
67	***Aconitum spicatum*** (Bruhl) Stapf	Bish	Ranunculaceae	CNRCP19	B12	2072-07-14	RCP	Herb	3211	29	24	40.1	83	51	23.4
68	***Aconitum spicatum*** (Bruhl) Stapf	Aconitum	Ranunculaceae	CN47	C30	2073-6-11	RCP, DK, MB	Herb	3462	28	26	35.3	83	52	32.3
69	***Delphinium brunonianum*** Royle	Bish	Ranunculaceae	CNRCP22	B14	2072-07-15	RCP	Herb	3732	28	26	35.9	83	52	32.4
70	***Delphinium himalayai*** Munz	Nirmasi	Ranunculaceae	CNRCP23	B15	2072-07-15	RCP	Herb	3787	28	25	45.8	83	51	53.5
71	***Delphinium himalayai*** Munz	Nirmasi	Ranunculaceae	CN7	*C5*	2073-6-9	RCP, DK, MB	Herb	3993	28	27	12	83	52	35.4
72	***Aruncus dioicus*** (Walter) Fernald	Aruncus	Rosaceae	CN45	C28	2073-6-11	RCP, DK, MB	Herb	3462	28	26	1.7	83	52	5.5
73	***Cotoneaster integrifolius*** (Roxb.) Klotz	Cotoneaster	Rosaceae	CN36	C21	2073-6-10	RCP, DK, MB	Shrub	3537	28	25	13.6	83	57	29.2
74	***Sorbus arachnoidea*** Koehne	Sorbus	Rosaceae	CN41	C24	2073-6-11	RCP, DK, MB	Tree	3358	28	25	45.8	83	51	53.5
75	***Rubia manjith*** Roxb. ex Fleming	Machheto	Rubiaceae	CNRCP53	*B41*	2072-07-17	RCP	Climber	2092	28	23	10.4	83	52	7.6
76	***Zanthoxylum acanthopodium*** DC.	Boke timur	Rutaceae	CNRCP48	B36	2072-07-17	RCP	Tree	2092	28	23	10.4	83	52	7.6
77	***Zanthoxylum acanthopodium*** DC	Boke Timur	Rutaceae	CN56	C38	2073-6-12	RCP, DK, MB	Tree	2092	28	23	4.4	83	52	29.4
78	***Zanthoxylum armatum*** DC.	Ankhe Timur	Rutaceae	CNRCP61	*B49*	2072-07-17	RCP	Tree	1847	28	23	10.4	83	52	7.6
79	***Bergenia ciliata*** (Haw.) Sternb	Bergenia	Saxifragaceae	CN59	C40	2073-6-12	RCP, DK, MB	Herb	1800	28	21	37.2	83	52	31.5
80	***Bergenia purpurascens*** (Hook. f. & Thomson) Engl	Bergenia	Saxifragaceae	CN14	C8	2073-6-9	RCP, DK, MB	Herb	3993	28	27	12	83	52	35.4
81	***Schisandra grandiflora*** (Wall.) Hook. f. & Thomson	Nalkim	Schisandraceae	CN49	C32	2073-6-11	RCP, DK, MB	Climber	3014	28	24	20.9	83	51	30
82	***Neopicrorhiza scrophulariiflora (Pennell)*** D.Y. Hong	Kutki	Scrophulariaceae	CN4	C41	2073-6-9	RCP, DK, MB	Herb	3993	28	27	12	83	52	35.4
83	***Smilax aspera*** L.	Smilax	Smilacaceae	CN55	C37	2073-6-12	RCP, DK, MB	Climber	2092	28	23	10.4	83	52	7.6
84	***Smilax ferox*** Wall. ex Kunth	Kukur daino	Smilacaceae	CN57	C39	2073-6-12	RCP, DK, MB	Climber	2092	28	23	10.4	83	52	7.6

SN	Identification	Name in Field	Family	Collection_no	Lab code	Collection_date	Collectors	Habit	Altitude (m)	Latitude			Longitude		
85	***Taxus wallichiana*** Zucc.	Silinge	Taxaceae	CNRCP04	B4	2072-07-13	RCP	Tree	2998	28	24	12	83	51	25
86	***Daphne bholua*** Buch.-Ham. ex D. Don	Lokatha	Thymelaeaceae	CN48	C31	2073-6-11	RCP, DK, MB	Shrub	3014	28	23	56.5	83	51	21.6
87	***Daphne papyracea*** Wall. ex Steud.	Lokta	Thymelaeaceae	CNRCP29	B21	2072-07-17	RCP	Shrub	2827	28	24	20.9	83	51	30
88	***Urtica dioica*** L.	Sisnu	Uritaceae	CNRCP33	B25	2072-07-17	RCP	Shrub	2523	28	23	38.9	83	51	38.7
89	***Girardinia diversifolia*** (Link) Friis	Chalne sisnu	Urticaceae	CNRCP32	B24	2072-07-17	RCP	Shrub	2523	28	23	38.9	83	51	38.7

Note: Poudel RC (RCP), Khadka, D (DK) and Bhattarai, M (MB)

Appendix 2 Status of PCR and Sequencing efficiency of each sampled species from Mardi-Himal, central Nepal.

SN	Identification	Name in Field	Family	Collection_no	Lab code	PCR		Sequencing	
						rbcL	*ITS*	*rbcL*	*ITS*
1	***Viburnum mullaha*** Buch.-Ham. ex D. Don	Malo	Adoxaceae	CNRCP42	B30	+ve	+ve	-ve	+ve
2	***Arisaema speciosum*** (Wall.) Mart. ex Schott & Endl.	Dhakayo	Araceae	CN50	C33	+ve	+ve	+ve	+ve
3	***Panax pseudo-ginseng*** Wall.	Ginseng	Araliaceae	CN16	C9	+ve	+ve	+ve	+ve
4	***Asparagus filicinus*** Buch.-Ham. ex D. Don	Kurilo	Asparagaceae	CNRCP50	*B38*	+ve	+ve	+ve	+ve
5	***Maianthemum oleraceum*** (Baker) LaFrankie	Badure	Asparagaceae	CN51	C34	-ve	-ve	-ve	-ve
6	***Anaphalis contorta*** (D. Don) Hook. f.	Buki Phool	Asteraceae	CNRCP15	B8	+ve	+ve	+ve	+ve
7	***Artemisia dubia*** Wall. ex Besser	Artemisia	Asteraceae	CN52	C35	-ve	-ve	-ve	-ve
8	***Artemisia indica*** Willd.	Pati	Asteraceae	CNRCP77	*B59*	+ve	-ve	-ve	-ve
9	***Senecio graciliflorus*** DC	Senecio	Asteraceae	CN37	C22	+ve	+ve	+ve	+ve
10	***Taraxacum officinale*** Wigg.		Asteraceae	CNRCP16	B9	+ve	+ve	+ve	+ve
11	***Diplazium maximum*** (D. Don) C. Christensen	Neuro	Athyriaceae	CNRCP59	*B47*	-ve	-ve	-ve	-ve
12	***Diplazium spectabile*** (Wall. ex Mett.) Ching	Lakhuto	Athyriaceae	CNRCP55	*B43*	-ve	-ve	-ve	-ve
13	***Berberis angulosa*** Wall. ex Hook. f. & Thomson	Berberis	Berberidaceae	CN13	C7	+ve	+ve	+ve	+ve
14	***Berberis asiatica*** Roxb. ex DC.	Chutro	Berberidaceae	CNRCP68	B56	+ve	+ve	-ve	+ve
15	***Berberis erythroclada*** Ahrendt	Chutro	Berberidaceae	CNRCP02	B2	+ve	+ve	+ve	+ve
16	***Mahonia napaulensis*** DC.	Bhutro	Berberidaceae	CNRCP63	*B51*	+ve	+ve	+ve	+ve
17	***Alnus nepalensis*** D. Don	Utis	Betulaceae	CNRCP65	*B53*	+ve	+ve	+ve	+ve
18	***Betula utilis*** D. Don	Bhoj Patra	Betulaceae	CNRCP17	B10	+ve	+ve	+ve	+ve
19	***Betula utilis*** D. Don	Betula	Betulaceae	CN42	C25	+ve	+ve	+ve	+ve
20	***Nardostachys grandiflora*** DC.	Jatamansi	Caprifoliaceae	CNRCP20	B13	+ve	+ve	-ve	-ve
21	***Nardostachys grandiflora*** DC.	Jatamansi	Caprifoliaceae	CN2	*C2*	+ve	+ve	+ve	+ve
22	***Euonymus tingens*** Wall.	Kesari	Celastraceae	CNRCP08	B5	+ve	+ve	+ve	+ve
23	***Parnassia nubicola*** Wall. ex Royle	Parnasia	Celastraceae	CN33	C19	+ve	+ve	+ve	-ve
24	***Rhodiola fastigiata*** (Hook. f. & Thomson) S. H. Fu	Crasssulaceae	Crassulaceae	CN20	C12	-ve	-ve	-ve	-ve
25	***Trichosanthes wallichiana*** (Ser.) Wight,	Indreni	Cucurbitaceae	CNRCP52	B40	+ve	+ve	+ve	+ve
26	***Cyathea spinulosa*** Wall. ex Hook.	Rukha Unyu	Cyatheaceae	CNRCP72	*B57*	+ve	+ve	-ve	-ve

SN	Identification	Name in Field	Family	Collection_no	Lab code	PCR		Sequencing	
						rbcL	*ITS*	*rbcL*	*ITS*
27	***Daphniphyllum himalense*** (Benth.) Mull. Arg.	Rachan	Daphniphyllaceae	CNRCP58	*B46*	+ve	+ve	+ve	-ve
28	***Cassiope fastigiata*** (Wall.) D. Don	Cassiope	Ericaceae	CN5	C3	+ve	+ve	+ve	+ve
29	***Lyonia ovalifolia*** (Wall.) Drude	Angeri	Ericaceae	CNRCP43	B31	-ve	-ve	-ve	-ve
30	***Lyonia ovalifolia*** (Wall.) Drude	Lyonia	Ericaceae	CN43	C26	+ve	+ve	-ve	+ve
31	***Rhododendron anthopogon*** D. Don	Sunpati	Ericaceae	CN6	C4	+ve	+ve	+ve	+ve
32	***Rhododendron arboreum*** Sm.	Gurans	Ericaceae	CNRCP30	B22	+ve	+ve	-ve	+ve
33	***Rhododendron barbatum*** Wall. ex G. Don	Gurans	Ericaceae	CN44	C27	-ve	-ve	-ve	-ve
34	***Rhododendron campanulatum*** D. Don	Guras	Ericaceae	CNRCP18	B11	+ve	+ve	+ve	+ve
35	***Castanopsis tribuloides*** (Sm.) A. DC.	Musure katus	Fagaceae	CNRCP64	*B52*	-ve	-ve	-ve	-ve
36	***Lithocarpus elegans*** (Blume) Hatus. ex Soepadmo	Arkhalo	Fagaceae	CNRCP60	*B48*	+ve	+ve	+ve	-ve
37	***Quercus lamellosa*** Sm.	Phalat	Fagaceae	CNRCP38	B28	+ve	+ve	+ve	-ve
38	***Quercus semecarpifolia*** Sm.	Kharsu	Fagaceae	CNRCP51	B39	+ve	+ve	+ve	+ve
39	***Swertia chirayita*** (Roxb. ex Fleming) Karsten	Chirayito	Gentianaceae	CNRCP01	B1	+ve	+ve	+ve	+ve
40	***Swertia teres*** (G. Don) J. Shah	Swertia	Gentianaceae	CN10	C6	+ve	+ve	+ve	+ve
41	***Geranium wallichianum*** D. Don ex Sweet	Geranium	Geraniaceae	CN46	C29	+ve	+ve	-ve	-ve
42	***Engelhardia spicata*** Lesch. ex Blume	Mauwa	Juglandaceae	CNRCP67	*B55*	+ve	+ve	-ve	+ve
43	***Leucosceptrum canum*** Sm.	Dhurseli	Lamiaceae	CNRCP54	*B42*	+ve	+ve	+ve	+ve
44	***Lindera neesiana*** (Wall. ex Nees) Kurz.	SilTimur	Lauraceae	CNRCP66	*B54*	+ve	+ve	+ve	+ve
45	***Lindera pulcherrima*** (Nees) Benth. ex. Hook.f.	Seto lokar	Lauraceae	CNRCP47	B35	-ve	-ve	-ve	-ve
46	***Machilus odoratissima*** (D Don) S.N. Biswa	Bhatekaulo	Lauraceae	CNRCP70	*B62*	+ve	+ve	+ve	-ve
47	***Persea duthiei*** (King) A.J.G. H. Kosterm	Kathe Kaulo	Lauraceae	CNRCP56	*B44*	+ve	+ve	+ve	-ve
48	***Indigofera atropurpurea*** Buch.-Ham. ex Hornem	Indigofera	Leguminosae	CN54	C36	+ve	-ve	+ve	-ve
49	***Lycopodium japonicum*** Thunb	Nagbeli	Lycopodiaceae	CNRCP62	*B50*	+ve	+ve	+ve	-ve
50	***Magnolia campbellii*** Hook. f. & Thomson	Champ	Magnoliaceae	CNRCP28	B20	+ve	+ve	+ve	+ve
51	***Michelia doltsopa*** Buch.-Ham. ex DC.	Chapo	Magnoliaceae	CNRCP41	B29	+ve	+ve	+ve	-ve
52	***Michelia kisopa*** Buch.-Ham. ex DC.	Chanp	Magnoliaceae	CNRCP57	B45	+ve	+ve	+ve	-ve
53	***Paris polyphylla*** Sm.	Satuwa	Melanthiaceae	CNRCP03	B3	+ve	+ve	+ve	+ve
54	***Paris polyphylla*** Sm.	Satuwa	Melanthiaceae	CN1	*C1*	+ve	+ve	+ve	+ve
55	***Myrsine semiserrata*** Wall.	Kalikath	Myrsinaceae	CNRCP49	B37	-ve	-ve	+ve	+ve
56	***Dactylorhiza hatagirea (D. Don) Soo***	Panchaunle	Orchidaceae	CNRCP21	*B61*	+ve	+ve	+ve	+ve
57	***Dactylorhiza hatagirea (D. Don) Soo***	Panchaule	Orchidaceae	CNRCP14	B7	+ve	+ve	+ve	+ve
58	***Dactylorhiza hatagirea*** (D. Don) Soo	Panchaunle	Orchidaceae	CN18	C11	+ve	+ve	+ve	+ve
59	***Danthonia cumminsii*** J. D. Hooker	Salim Grass 1	Poaceae	CN23	C13	+ve	+ve	+ve	+ve
60	***Thamnocalamus nepalensis*** (Stapleton) Stapleton	Jarbuto	Poaceae	CNRCP05	*B60*	+ve	+ve	+ve	+ve
61	***Yushania maling*** (Gamble) R.B. Majumdar	Malingo	Poaceae	CNRCP27	B19	+ve	+ve	+ve	+ve
62	***Polygonum molle*** D. Don	Thotne	Polygonaceae	CNRCP35	B27	+ve	+ve	+ve	+ve

SN	Identification	Name in Field	Family	Collection_no	Lab code	PCR		Sequencing	
						rbcL	*ITS*	*rbcL*	*ITS*
63	***Polygonum molle*** D. Don	Thotne	Polygonaceae	CN38	C23	+ve	+ve	+ve	+ve
64	***Bistorta macrophylla*** (D. Don) Sojak	Bistorta	Polygonaceae	CN34	C20	+ve	+ve	+ve	+ve
65	***Rheum australe*** D. Don	Padamchal	Polygonaceae	CN17	C10	+ve	+ve	+ve	+ve
66	***Rumex nepalensis*** Spreng.	Halhale	Polygonaceae	CNRCP34	B26	+ve	+ve	-ve	-ve
67	***Aconitum spicatum*** (Bruhl) Stapf	Bish	Ranunculaceae	CNRCP19	B12	+ve	+ve	+ve	+ve
68	***Aconitum spicatum*** (Bruhl) Stapf	Aconitum	Ranunculaceae	CN47	C30	+ve	+ve	-ve	+ve
69	***Delphinium brunonianum*** Royle	Bish	Ranunculaceae	CNRCP22	B14	+ve	+ve	+ve	+ve
70	***Delphinium himalayai*** Munz	Nirmasi	Ranunculaceae	CNRCP23	B15	+ve	+ve	+ve	+ve
71	***Delphinium himalayai*** Munz	Nirmasi	Ranunculaceae	CN7	*C5*	+ve	+ve	+ve	+ve
72	***Aruncus dioicus*** (Walter) Fernald	Aruncus	Rosaceae	CN45	C28	+ve	+ve	+ve	-ve
73	***Cotoneaster integrifolius*** (Roxb.) Klotz	Cotoneaster	Rosaceae	CN36	C21	+ve	+ve	+ve	+ve
74	***Sorbus arachnoidea*** Koehne	Sorbus	Rosaceae	CN41	C24	+ve	+ve	+ve	+ve
75	***Rubia manjith*** Roxb. ex Fleming	Machheto	Rubiaceae	CNRCP53	*B41*	+ve	+ve	+ve	-ve
76	***Zanthoxylum acanthopodium*** DC.	Boke timur	Rutaceae	CNRCP48	B36	+ve	+ve	+ve	+ve
77	***Zanthoxylum acanthopodium*** DC	Boke Timur	Rutaceae	CN56	C38	-ve	+ve	+ve	+ve
78	***Zanthoxylum armatum*** DC.	Ankhe Timur	Rutaceae	CNRCP61	*B49*	+ve	+ve	+ve	+ve
79	***Bergenia ciliata*** (Haw.) Sternb	Bergenia	Saxifragaceae	CN59	C40	+ve	+ve	-ve	-ve
80	***Bergenia purpurascens*** (Hook. f. & Thomson) Engl	Bergenia	Saxifragaceae	CN14	C8	+ve	+ve	+ve	+ve
81	***Schisandra grandiflora*** (Wall.) Hook. f. & Thomson	Nalkim	Schisandraceae	CN49	C32	+ve	+ve	+ve	+ve
82	***Neopicrorhiza scrophulariiflora (Pennell)*** D.Y. Hong	Kutki	Scrophulariaceae	CN4	C41	+ve	-ve	+ve	-ve
83	***Smilax aspera*** L.	Smilax	Smilacaceae	CN55	C37	-ve	-ve	-ve	-ve
84	***Smilax ferox*** Wall. ex Kunth	Kukur daino	Smilacaceae	CN57	C39	+ve	+ve	+ve	-ve
85	***Taxus wallichiana*** Zucc.	Silinge	Taxaceae	CNRCP04	B4	+ve	+ve	+ve	+ve
86	***Daphne bholua*** Buch.-Ham. ex D. Don	Lokatha	Thymelaeaceae	CN48	C31	+ve	+ve	+ve	+ve
87	***Daphne papyracea*** Wall. ex Steud.	Lokta	Thymelaeaceae	CNRCP29	B21	+ve	+ve	+ve	+ve
88	***Urtica dioica*** L.	Sisnu	Uritaceae	CNRCP33	B25	-ve	+ve	-ve	+ve
89	***Girardinia diversifolia*** (Link) Friis	Chalne sisnu	Urticaceae	CNRCP32	B24	+ve	+ve	-ve	-ve

Note: +ve= Positive and –ve= Negative

Appendix 3 Studied plant species and their corresponding accessions retrieved as reference from NCBI database for *rbcL*, *ITS* and *ITS2* makers.

***rbcL* data set (Lab code)**	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**	**Location**
Polygonum molle RB27	987	987	99%	0	100%	JF943512.1	China
Alnus nepalensis RB53	985	985	99%	0	100%	KF418930.1	China
Aruncus dioicus RC28	948	948	98%	0	99%	KF154894.1	China
Bergenia purpurascens RC8	983	983	100%	0	99%	KY986474.1	India
Cassiope fastigiata RC3	979	979	99%	0	100%	JF941134.1	China
Leucosceptrum canum RB42	928	928	99%	0	98%	KR608485.1	China
Lycopodium japonicum RB50	989	989	100%	0	100%	MF786611.1	China
Myrsine semiserrata RB37	944	944	96%	0	99%	MG950622.1	China
Paris polyphylla RB3	976	976	100%	0	99%	GU178930.1	China
Rhododendron anthopogon RC4	989	989	100%	0	100%	KM606531.1	China
Schisandra grandiflora RC32	970	970	98%	0	100%	KP689892.1	China
Swertia chirayita RB1	983	983	100%	0	99%	KU859995.1	Nepal
Taxus wallichiana RB4	957	957	98%	0	99%	HM591034.1	China
Zanthoxylum armatum RB49	989	989	100%	0	100%	KJ667668.1	India
Bistorta macrophylla RC20	981	981	99%	0	99%	JF943511.1	China
***ITS* data set (Lab code)**	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**	**Location**
Polygonum molle IB27	1066	1066	86%	0	99%	EF653687.1	China
Anaphalis contorta IB8	1146	1146	86%	0	99%	JQ895423.1	China
Berberis angulosa IC7	1098	1098	80%	0	99%	HM347898.1	India
Bergenia purpurascens IC8	1317	1317	96%	0	99%	EU239674.1	China
Cassiope fastigiata IC3	1164	1164	88%	0	99%	AF393438.1	China
Leucosceptrum canum IB42	1214	1214	94%	0	99%	KR608738.1	China
Lyonia ovalifolia IC26	1251	1251	98%	0	98%	KP092599.1	China
Magnolia campbellii IB20	798	798	79%	0	93%	KU853479.1	China
Myrsine semiserrata IB37	1133	1133	86%	0	99%	MG877850.1	China
Paris polyphylla IB3	1086	1086	90%	0	96%	KX146541.1	China
Rheum australe IC10	1005	1005	86%	0	98%	KF258683.1	China
Rhododendron campanulatum IB11	1310	1310	96%	0	99%	KM605963.1	China
Swertia chirayita IB1	1240	1240	100%	0	98%	JX569817.1	Nepal
Schisandra grandiflora Ic32	1232	1232	87%	0	99%	KP689672.1	China
Taraxacum officinale IB9	1182	1182	99%	0	96%	MG519289.1	China

***ITS* data set (Lab code)**	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**	**Location**
Zanthoxylum armatum IB49	1134	1134	87%	0	99%	DQ016546.1	China
Taxus wallichiana IB4	2135	2135	99%	0	99%	JX680623.1	China
Bistorta macrophylla Ic20	1182	1182	99%	0	98%	JN235093.1	China
***ITS2* data set (Lab code)**	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**	**Location**
Polygonum molle IB27	477	477	91%	0	99%	EF653687.1	China
Alnus nepalensis IB53	468	468	91%	0	99%	FJ825418.1	China
Anaphalis contorta IB8	438	438	89%	0	100%	JQ895425.1	China
Cassiope fastigiata IC3	475	475	93%	0	99%	AF393438.1	China
Engelhardia spicata IB55	460	460	100%	0	97%	KR532080.1	China
Leucosceptrum canum IB42	459	459	93%	0	99%	KR608738.1	China
Lyonia ovalifolia IC26	472	472	100%	0	97%	MF785536.1	China
Magnolia campbellii IB20	207	207	60%	0	88%	KU853479.1	China
Myrsine semiserrata IB37	451	451	91%	0	100%	MG877850.1	China
Quercus semecarpifolia IB39	481	481	100%	0	99%	KY624385.1	China
Rhododendron anthopogon IC4	521	521	100%	0	100%	KM605954.1	China
Rheum australe IC10	424	424	91%	0	99%	KF258683.1	China
Rhododendron arboreum IB22	512	512	100%	0	99%	KM605819.1	China
Rhododendron campanulatum IB11	516	516	100%	0	100%	KM605963.1	China
Schisandra grandiflora IC32	468	468	90%	0	100%	KP689672.1	China
Swertia chirayita IB1	484	484	100%	0	98%	JX569818.1	Nepal
Taraxacum officinale IB9	429	429	97%	0	95%	MG519290.1	China
Taxus wallichiana IB4	499	499	100%	0	99%	KX981186.1	China
Zanthoxylum armatum Ib49	473	473	95%	0	99%	KM887380.1	India
Bistorta macrophylla IC20	484	484	100%	0	97%	JN235093.1	China

Appendix 4 Reagents and Extraction Buffers Preparation Protocol

TAE (50X) Stock Buffer preparation (Tris-Acetate-EDTA)

Tris base (242 g Sigma-Aldrich, USA) was dissolved in approximately 750 ml double distilled water (D/W). To this solution, glacial acetic acid (57.1 ml) was added and followed by 0.5 M EDTA (pH 8.0, 100 ml, Sigma-Aldrich, USA). The final volume made up to 1L. The final stock solution was autoclaved and stored in room temperature. The working solution of 1X TAE buffer was prepared by diluting the stock solution (50X) in double distilled water.

1M Tris Buffer (pH 8.0)

Tris base 12.11 g (Sigma-Aldrich, USA) was dissolved in 75 ml D/W. The pH was adjusted to 8.0 by using conc. HCl (approx. 4-5 ml added). The final volume was then made up to 100 ml, autoclaved and stored at room temperature.

0.5 M EDTA (pH 8.0)

Disodium Ethylene Diamine Tetra Acetate.$2H_2O$ (18.6 g, Sigma-Aldrich, USA) was added to 75 ml d/w. NaOH pallets were added to adjust the pH and dissolved the solution by magnetic stirrer. Th final volume was made upto 100 ml by adding D/W, autoclaved and stored at room temperature.

TE Buffer (pH 7.5)

EDTA (200 µl of 0.5M stock, pH 8.0) was added to falcon tube containing Tris (1 ml of 1M solution) and the final volume made up to 100 ml, autoclaved and stored at -20^{o} C.

NaCl (5M)

Sodium chloride (29.22 g; Fisher Scientific, USA) was added to reagent bottle (sterile) containing 75 ml D/W, mixed on a magnetic stirrer. The final volume was adjusted up to 100 ml with D/W, autoclaved and stored in room temperature.

CTAB Buffer (pH 5.0)

Tris buffer (10 ml of 1M solution, pH 8.0) was added in reagent bottle (sterile) and EDTA (4 ml of 0.5M solution, pH 8.0), NaCl (28 ml of 5M solution) and 1 g PVP40 were mixed and final volume was made up to 100 ml of D/W.

Agarose gel (1%)

Agarose (1 g; Bioneer, South Korea) was dissolved in TAE buffer (100 ml, 1X) in the microwave. It was then cooled to approx. 55^{o} C and poured on to the gel casting tray with an appropriate comb fixed in the place for well formation.

Appendix 5 Barcode Sequences of some high value plants species.

rbcL

>*Aconitum spicatum*_RB12

TTATGTCACCCCCATAGAGACTCAAGC
TGGTGTTAAAGATTACAAATTGAATTATTATACTCCGGAATATGCACCCAAAGATACTGATACCTTGGCGGCATTCCGAGTAACTCC
TCAACCTGGAGTTCCACCCGAAGAAGCAGGGGCTGCTGTAGCTGCCGAATCTTCTACAGGTACATGGACAACTGTGTGGACCGAT
GGACTTACCAACCTTGATCGTTACAAAGGACGATGCTACCACATTGAGCCCGTTGCTGGAGAAGAAAATCAATATATTTGTTATGT
AGCATATCCTTTAGACCTTTTTGGAAAAGGTTCTGTTACTAACATGTTTACTTCCATTGTGGATGATGTATTTGGGTTCAAGTTGCTT

ATGTTAAAACTTTCCAAGGCCCACCTCACGGCATCCAAGTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCACTATTGGGATGT
ACTATTAAACCAAAATTGGGATTATCTGCTAAGAACTATGGCAGAGCGTTTTATGAATGTCTG
CGYGGTGYTKKTATWTTTTWWAMA
←

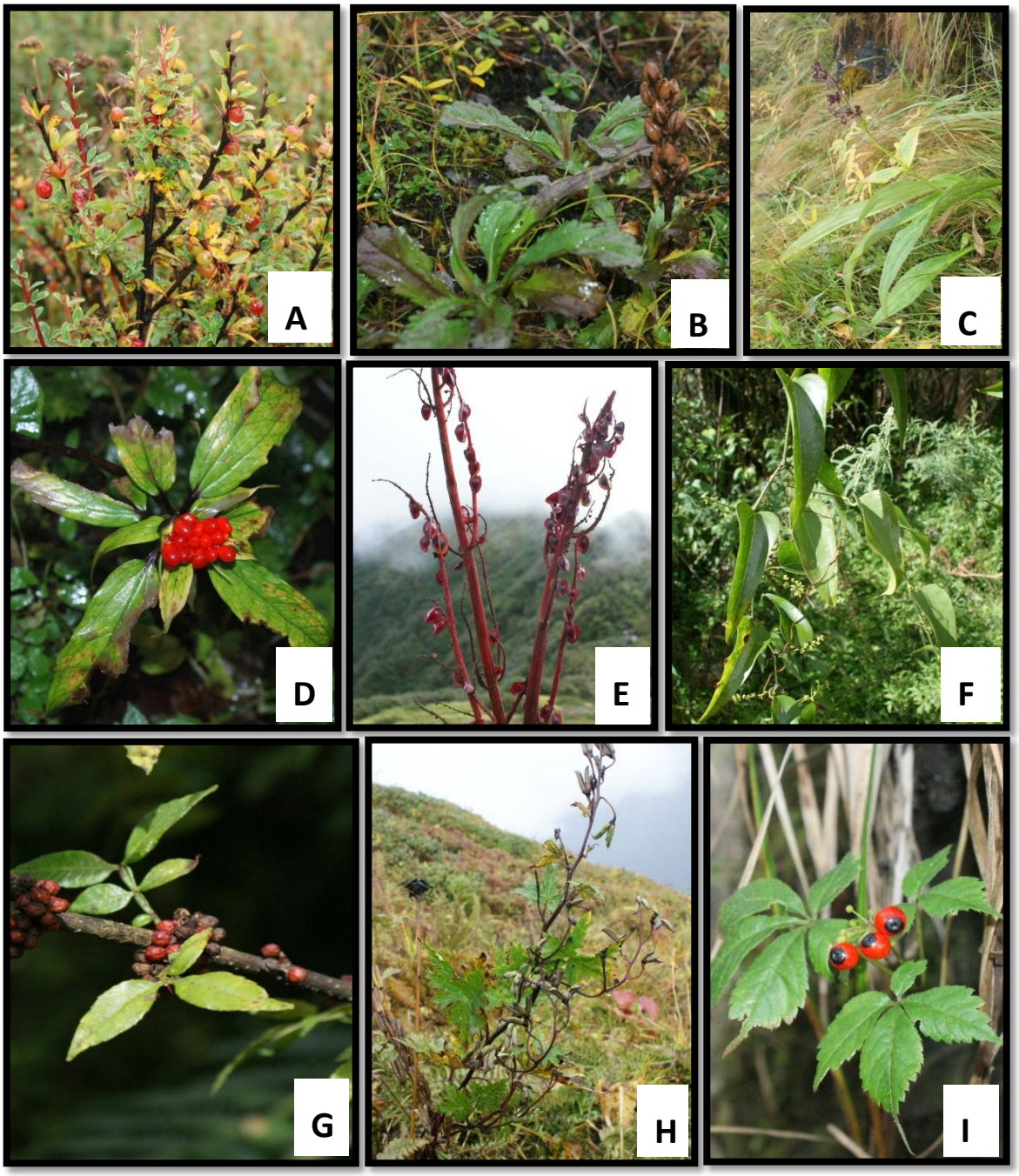
>*Dactylorhiza hatagirea*_RB61
TGGTGTTAAAGATTACAAATTGACTTATTATACTCCTGACTACGAAACCAAAGATACTGATATCTTGGCAGCATTCCGAGTAACTCC
TCAACCGGGAGTTCCGCCTGAAGAAGCAGGCGCTGCGGTAGCAGCCGAATCTTCTACTGGTACATGGACAACTGTGTGGACTGAT
GGACTTACCAGTCTCGATCGTTACAAAGGACGATGCTACCACATCGAGGCCGTTGTTGGGGAGGAAAATCAATATATTGCTTATGT
AGCTTATCCTTTAGACCTTTTTGAAGAAGGTTCTGTTACTAACATGTTTACTTCCATTGTGGGTAATGTTTTTGGTTTCAAAGCTCTG
CGAGCTCTACGTCTGGAAGATCTGCGAATTCCCCCTGCTTATTCCAAAACTTTCCAAGGTCCGCCTCATGGCATCCAAGTTGAAAGA
GATAAATTGAACAAGTACGGTCGTCCCCTATTGGGATGTACTATTAAACCAAAAT
>*Taxus wallichiana*_RB4
TGGTGTTAAAGATTACAGACTAACTTATTATACTCCACAATATCAGACCAAAGATACTGATATCTTGGCAGCATTCCGAGTCACTCC
TCAACCGGGAGTGCCCCCCGAGGAAGCGGGAGCAGCAGTAGCTGCCGAATCTTCCACTGGTACATGGACCACTGTTTGGACCGAT
GGACTTACCAGTCTTGATCGTTACAAGGGACGATGCTATGATATCGAACCCGTTCCTGGAGAGGAAAATCAATTTATTGCCTATGT
AGCTTACCCCTTAGATCTTTTCGAAGAAGGTTCTGTGACTAACCTGTTCACTTCCATTGTAGGTAATGTCTTTGGATTCAAAGCCCTA
CGAGCTCTACGTCTGGAAGATCTACGAATTCCTCCTGCTTATTCAAAAACTTTCCAAGGCCCACCACATGGTATCCAAGTGGAAAG
AGATAAATTAAACAAATATGGTCGTCCTTTGTTGGGATGTACTATAAAACCAAAATTGGGCCTATCTGCCAAAAATTATGGTAGAG
CCTTTAAAGTTTC TGGGATTATCCGCAAAAAACTACGGCAGAGCAGTTTATGAATGTCTA

ITS

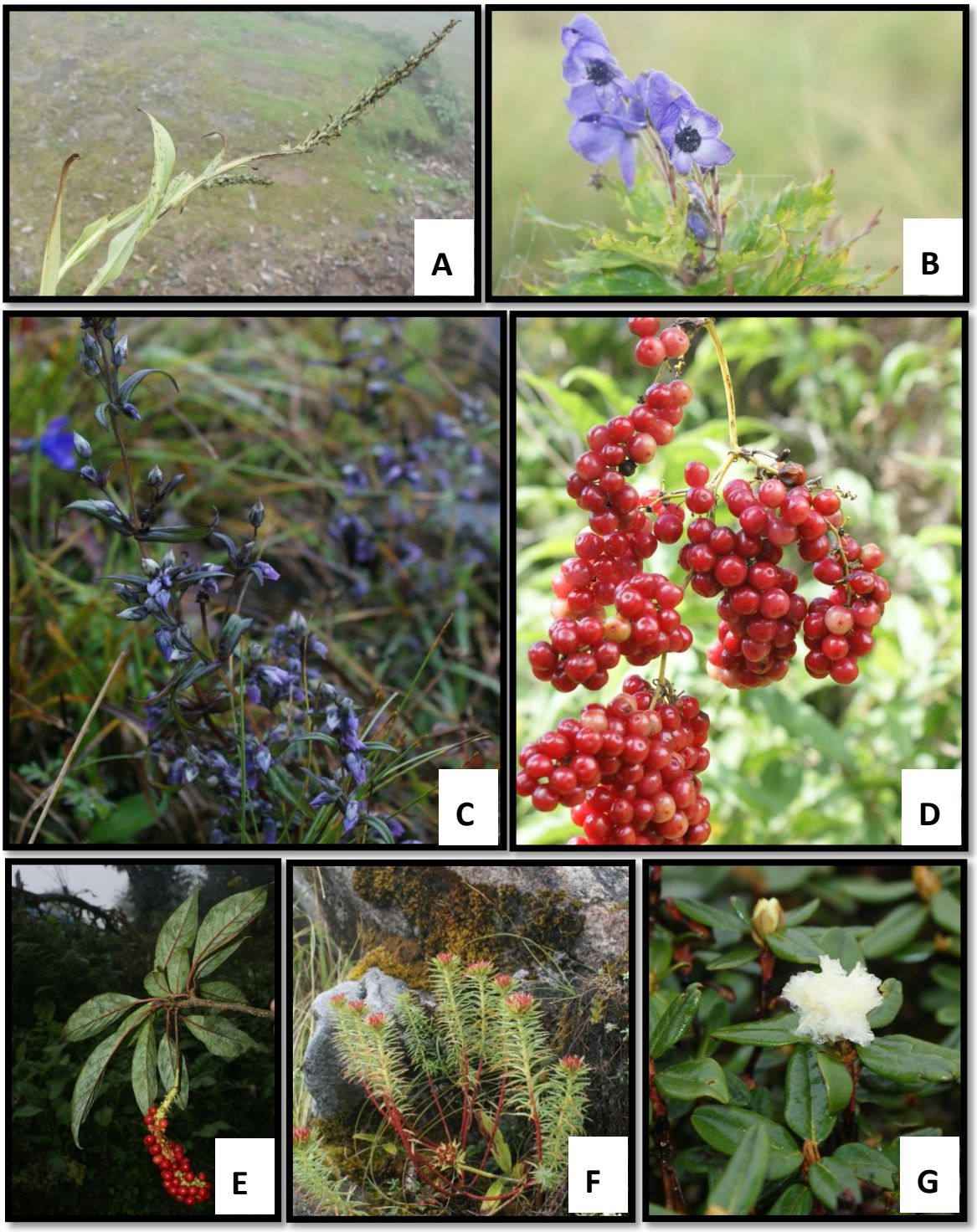
>*Aconitum spicatum*_IB12
→
TCCTTWYWYWCTTTTTGGAGGAAGGAG
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAGCGACCCGCGAACAAGTGA
AAACAAAGCTGGACGAACCGAAGAGGGGCGCATGCCCCCGATCGCTCGCCCGTCGGACCACGGCCACTTCTGCGACCGCACTGAT
CTGCGGGTGGAGGGGTGGTCGTTGTGTCCGCACAAAACCAAAAACCGGCGCGACAGGCGTCAAGGAAATCTTAGCGGAAAAAG
AGGGTTGCCCCGTCCGCGGAGGCAACCTTCAGAATCCGATACTCAAACGACTCTCGGCAACGGATATCTCGGCTCTTGCATCGATG
AAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAGGC
CATTAGGTCGAGGGCACGTCTGCCTGGGCGTCACACACAGCGTCGCAACCTGTCAACCACGTTGTCAGGGAGCGGAGATTGGCCC
CCCGGGCCCCTGCGGGCACGGTCGGCACAAATGTTTGTCCCCGGCGGCGAGCGTCGCGGTCAGTGGTGGTTGTATCTCTCATCCT
CCAAAGACATCAAGACGCGTCGTCCTCGTTGCATGTTGGGACACATCGACCCCAAGGAGCCGCTTCGCGCGGCATTCACCCTGCG
ACCCCAGGTCAGGCGGGATCACCCGCTGAGTTTAAGCATATCAA
TAAGCGGAAAAAGGAACTCTA
←
>*Dactylorhiza hatagirea*_IB61
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTGTCGAGACCCTGAAGAGATCGAGCGATTTGACAACTTGTG
AACTTCTTCAGCAGCGCGCATTAGCTGCTGTGCACCCGTCCATCTATTGCATGAACAACCCGATGGGAACATGTTATAGGCGGATG
GGAGATCAATTCGGCGCAGCTTTGTGCCAAGGCAAATATGCAGCATGAGCAGAGTTTCGACCACATATCCTCAAAGCGATTTGTTT
TGTGGAGTTATTTGCTCCTAAAGAGTTGTAGGGCTCTCGGCAATGGATATCTTGGCTCTCGCATCGATGAAGAGCGCAGCGAAAT
GCGATACGTGGTGCGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCTGAGGCCAGCTGGCCGAGGGCA
CGTCCGCCTGGGCGTCAAGCATTGAATCGCTCCATAAGACCTGTGGCGCAATGCAGTGGTCTTATCTAGGATGCGGAGAATGGCC
CGTCATGCGCTGATGTGTGGCAGGCTGAAGAGCGGGGATGATTTTCTCTTGGCAACAATCGATTAATGGGTGGGATGGAATCTCC
AGTTGATCGTCATCATCGTCAGGTTGCTTTGAGAATGCTGTGATATCCCGGGCTATCCCAACTCAATTGTTTGGAAGAAAATTGACA
TGCGACCCCAGGATGGGGCGGGGTGACCCGCTGAGTTTAAGCATATCAA
>*Taxus wallichiana*_IB4
GGGCCGCGAAGTTGGGACGCACCCTCCCGCGGAAGAGTCGCGACTAACAACGAGTAAAGTTCCGCACAAATCCGTGACCTCGTG
CCACAATCCAACGCCGACCAACATTGGCCACCGCTCATCGTGACATCGGCCACAAATCTCGTGCCCATTTAAGCCGACCGCGCCGC
TGGTGGGCACGGACAGCCATCTCCGATCCTCCCGCCGAGGAACGGGGGAAGTCGATTTTAGAGTGCGACGCCCAAGCAGACGTG
CCCTCGGCCGAGGCCTCGGGCGCAACTTGCGTTCAAAGACTCGATGATTCACGGGATTCTGCAATTCACACTAAGTATCGCATTTC
GCTACATTCTTCATCGTGGCGAGAGCCGAGATATCCGTTGCCGAGAGTCGTGTTTAATTGGCGCCGTGAGTCCGACGCGCGCATG
GCGCGCAGCCCGCACAGCTATTTATCAGATCCTTGGTGCAATCCGCACCGATGCGTTGAGTGTTATGACACGCACCTCCACTCGAC
GCCCCCCTGCAACCAGAGGCGATGCCCCGTGGGGGAAAATCGGAGAAAGGGTGTCCCCGCCCCGGAGTGTCGAAAGGCCGGAG
ACCGAACAGATGGGAACGGAACCAGTGGACGGTCGGCCCGCCCGGCGACGTACGCAGGTGCACCCGTCCAAAGGCACGAAATA
ATCTGTACGGCATTACCTGCCCGGCGACACTTCAGGGGGTTGCAGCTGCTTTGCAGACCGGCAAATGTTCAACGGTCTGCAAAGC
AGCCACGGGCGTTTGCAAGGAACCGCCTAGCAGGTGGAAATTTCTGGGACCCTTCGACCGCAAAGGAGCAGGACTGCCAGACGA
ACAGAGTCGCTGGTGCACCCA

Note: → = forward primer region and ← = Reverse primer region

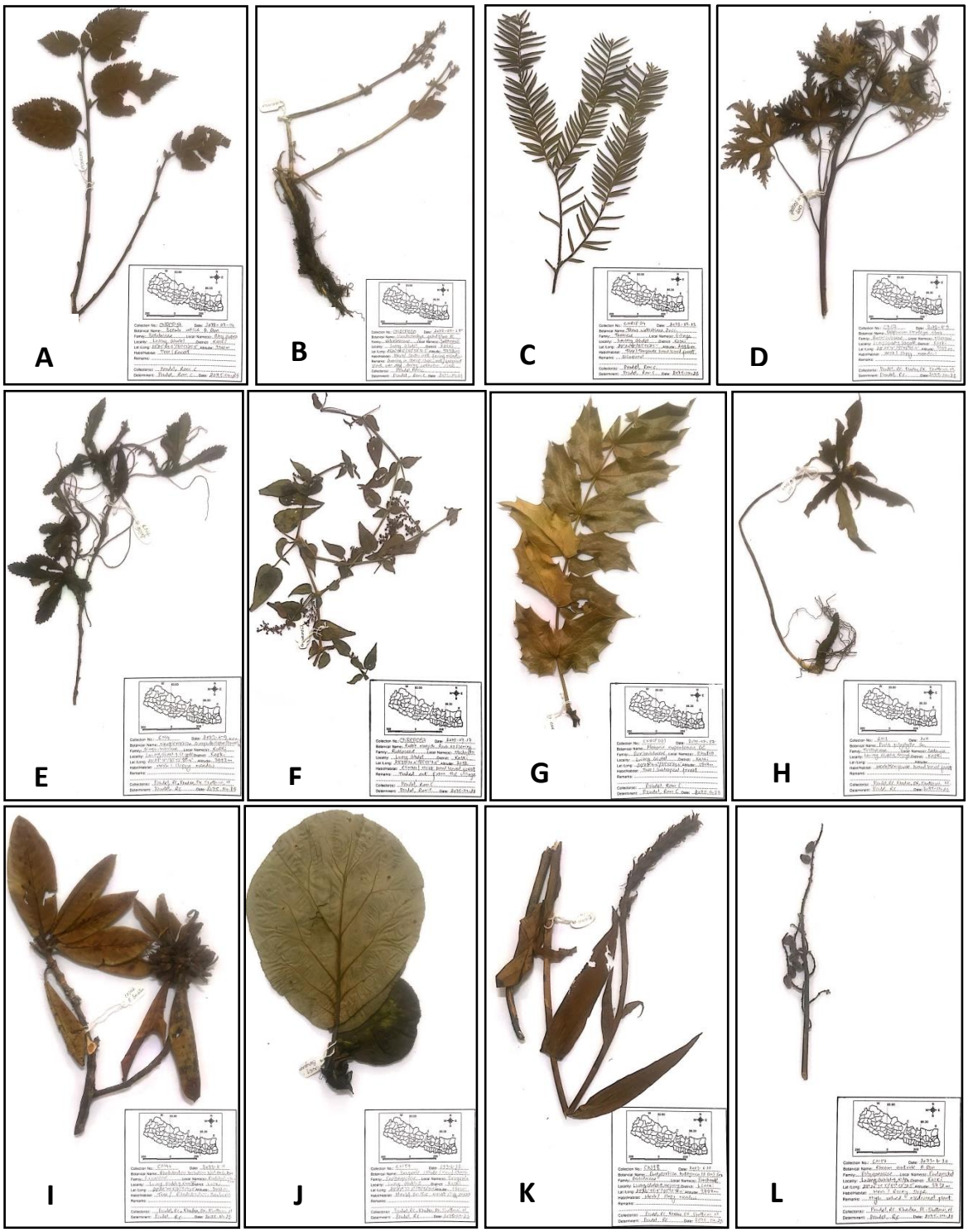
Photographs of selected high value plants



Appendix 6 photographs of studied some high value plants. A. *Berberis angulosa* B. *Neopicrorhiza scrophulariiflora*, C. *Nardostachys grandiflora*, D. *Paris polyphylla*, E. *Rheum australe*, F. *Smilax aspera*, G. *Zanthoxylum acanthopodium* and H. *Delphinium*



Appendix 7 photographs of studied some high value plants. A. *Dactylorhiza hatagirea*, B. *Aconitum spicatum*, C. *Swertia teres*, D. *Smilax ferox*, E. *Schisandra grandiflora*, F. *Rhodiola fastigiata*, and G. *Rhododendron anthopogon*.



Appendix 8 Photographs of herbariums prepared during present study. A. *Betula utilis,* B. *Nardostachys grandiflora,* C. *Taxus wallichiana*, D. *Delphinium himalayai*, E. *Neopicrorhiza scrophulariiflora*, F. *Rubia manjith*, G. *Mahonia nepalensis*, H. *Paris polyphylla*, I. *Rhododendron barbatum*, J. *Bergenia ciliata*, K. *Dactylorhiza hatagirea* and L. *Rheum australe*.