DNA barcoding of some forest tree species of Western Ghats

A THESIS SUBMITTED TO

SAVITRIBAI PHULE PUNE UNIVERSITY

FOR AWARD OF DEGREE OF

DOCTOR OF PHILOSOPHY (Ph.D.)

IN THE FACULTY OF

BIOTECHNOLOGY

SUBMITTED BY

RASIKA M. BHAGWAT

ADVISOR-

DR. VIDYA S. GUPTA

PLANT MOLECULAR BIOLOGY GROUP,

DIVISION OF BIOCHEMICAL SCIENCES,

CSIR-NATIONAL CHEMICAL LABORATORY,

PUNE- 411008, INDIA

DECEMBER, 2015

सीएसआयआर–राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग, पुणे – 411 008. भारत
CSIR-NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411008. India

CERTIFICATE

This is to certify that the work included in the thesis, **"DNA barcoding of some forest tree species of Western Ghats"** submitted by Ms. Rasika M. Bhagwat, for the degree of **Doctor of Philosophy**, was carried out under my supervision in the Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune-411008, India. The information used from other resources has been duly acknowledged in the thesis.

Dr. Vidya S. Gupta

(Research Advisor)

December 2015

Pune

Communications Channels
NCL Level DID : 2590
NCL Board No. : +91-20-25902000
Four PRI Lines : +91-20-25902000
FAX
Director's Office : +91-20-25902601
COA's Office : +91-20-25902660
SPO's Office : +91 20 25902664
WEBSITE
www.ncl-india.org

Declaration

I hereby declare that the thesis "**DNA barcoding of some forest tree species of Western Ghats**" submitted for the degree of Doctor of Philosophy to Savitribai Phule Pune University has not been submitted by me for degree/ diploma to any other University.

R Bhagwat

Rasika Bhagwat

4th December 2015

Pune

Dedicated to My Parents

&

Family

TABLE OF CONTENTS

Acknowledgement

People henceforth will claim this research to be my work. I take this opportunity to refute that claim as a work of this magnitude which spans over 5 years cannot be the achievement of a single person. I was supported by a team of highly talented and passionate individuals who cared about Biotechnology and, for me as well.

I am deeply indebted to my advisor Dr. Vidya S. Gupta, for her guidance and encouragement during my research work. She has not only guided me but has also helped in my conceptual understanding. Her enthusiasm, ideas, patience, and caring nature have always inspired me. Her long discussions on research topic provoked me to think in different directions. Those lengthy discussions and brainstorming sessions, with her, on our results shaped up my work. A special note is required here. Many a times I got so engrossed in the work that I would lose context of the bigger picture and wonder if my direction was correct, whether my goals were worthy enough? Dr. Gupta would handle these situations adroitly and help me re-focus. I wholeheartedly thank her for everything.

My sincere thanks to Dr. Meena Lagu for planning our collection tours. Her punctilious way in this work and other lab related issues has always helped me in conducting research.

I express my appreciation to Dr. Narendra Kadoo for his help at various stages of research work. His contribution warrants even more appreciation because, even while he was miles away in Australia, he had a complete grasp of my work. His suggestions on our telephonic conversations and on email were succinct, vivid and lucid. His suggestions in the sequencing experiments have always worked for me. His critical views helped me to perform better. I had the privilege of learning everything from wet lab to computers from him.

I thank Dr. B.G. Kulkarni (Botanical Survey of India, Western Circle, Pune) for the help in collection tours. I would like to thank Dr. M. Balasundaran (Kerala Forest Research Institute, Trichur, Peechi, Kerala) for providing few Dalbergia species and their accessions.

I am grateful to Director, and Head of the Biochemical Sciences Department, CSIR-NCL, Pune for providing laboratory and other facilities required for research work. I also acknowledge the financial support from the Department of Biotechnology, New Delhi and the Council for Scientific and Industrial research, New Delhi for research and fellowship.

I am thankful to Dr. Mohini Sainani and Dr. Ashok Giri for extending their help unconditionally at all times.

My first research project was with Dr. Sofia Banu and in hindsight I'm really glad to have had it that way. Right from tissue collection to writing manuscript she taught me a lot. Her organization skills, punctuality and dedication to research has always inspired me. I would like to thank Dr. Sofia Banu and Dr. Ram Kulkarni for making our field trips successful.

Another name that I would like to mention is that of Dr. Bhushan Dholakia for his help in sample collection and experimentation. I thank him for his views, ideas and constant support. Our long debates on research topics have always provided food for thought.

Company of Dr. Gayatri Gurjar and Dr. Ashwini Rajwade will as always be cherished. Both of them have helped me in learning many laboratory techniques and advised me whenever needed. Ashwini and Gayatri have accompanied me throughout, even to my breaking point and have ensured that I return safely. I am thankful for their friendship and mentoring without which this uphill task would not have been possible to complete.

I thank Dr. Sagar Pandit, Dr. Varsha Pardeshi, Dr. Ramya Prashant, Richa, Aditi, Dr. Hemangi Chidley and Ashish. I learnt many things directly or indirectly from them and their company was always enjoyable.

I am thankful to Dr. Sachin Punekar (Biosphere, Pune, India), Dr. Aparna Watve (Tata Institute of Social Sciences), Dr. P. Tetali (Temple Rose Construction, Private Ltd, Pune), Amol K and Amol J for their help rendered during sample collection. I am also Thankful to Dr. Ankur Patwardhan (Head, Department of Biodiversity), Dr. Siddharthan (Centre for Ecological Sciences, Indian Institute of Sciences, Bangalore), Dr. Neelesh Dahanukar (Indian Institute of Science Education and Research, Pune) and Dr. Shobha Rao (Research & Training Society for Initiatives in Nutrition and Development, Pune) for the help in data analysis; and Dr. Anargha Wakhare (Department of Geography, Nowrosjee Wadia College, Pune) for her help in preparing the map.

I thank Charu, Sheetal and Indira for their help in all administration work. I am also thankful to Jagtap. Sincere thanks to all my labmates, Ajit, Elan, Neeta, Radhika, Manje, Ajay, Medha, Neha K, Neha M, Yojana, Priyanka, Vishal, Yashwant Pranjali, Amey, Sonal, Sheon, , Radhika, Vaiju, Manasi M, Prabhash, Arun, Atul, Vitthal, Sandip, Priya, Sheetal, Yashashree for creating a healthy environment in the lab.

I thank my teachers Dr. Pendse and Prof. Mrs. Chirputkar (S. P. College, Pune) as they inspired me to carry out research right from college level.

This work required me to travel a lot and meet a lot of good folks. Unfortunately, being the mortal that I'm, I would have missed out a lot of their names. However, I am thankful to students and professors whom I met at their institutes, locals who helped me with collections and accommodation and others who in some or the other way have touched my heart and this research as well.

Last but not the least I would like to thank my parents and my extended family members. My graduation will be their dream fructifying. My Mom, Dad, Kaka and Aatya have been a constant source of love, encouragement and support for me. Coming from a family that places high value on education, it was natural that I would pursue higher academic accolades. However, that 'family pressure' was not felt at all. My parents led by example and allowed me the freedom to choose and pursue goals of my choice. They instilled higher standards and ethical values in me.

They challenged my ideas, encouraged me to take the leap of faith and also lent a shoulder to rest on. Thank you Aai-Baba-Kaka-Aatya! My brothers, Suresh dada Yogesh dada and Suhas dada and my sisters-in-law Anjali, Varsha and Lalita always helped me wherever and whenever they were needed.

My in-laws, Aai and Kaka as we fondly call them, have been extremely supportive in this endeavour. My work required me juggle between daily chores and research activity. I also had to travel between Mumbai and Pune a lot. My in-laws supported me during the past one year which gave me a lot of space to push ahead with this work.

I would like to give a special thanks to my husband Dr. Vikram Shete who stood like wall by me from day one. There were times when I wanted to drop the ball. There were times when I would wonder about the worth of this work. He would jump in with those small pep talks and put me back on track. More importantly he believed in me and my work! Thank you Vikram.

Finally I would like to thank the divine powers, science or spiritual, for who I am!

Rasika

List of abbreviations

AD- *Anno Domini*
AFLP- Amplified Fragment Length Polymorphism
AMOVA- Analysis of Molecular Variance
AMP-PCR- Anchored Microsatellite Primed PCR
APFORGEN- Asia Pacific Forest Genetic Resource Program
APG- Angiosperm Phylogeny Group
AP-PCR- Arbitrarily Primed PCR
ASA- Allele-specific amplification
ASSR-Anchored Simple Sequence Repeat
BC- Before Christ
BLAST- Basic Local Alignment Search Tool
bp- base pair
BSI- Botanical Survey of India
CAPS- Cleaved Amplified Polymorphic Sequence
cDNA- Complementary Deoxy ribonucleic acid
CH- chlorotype
CITES- Convention on International Trade in Endangered Species of Wild Fauna and Flora
cp- chloroplast
CTAB- Cetyl-trimethyl-ammonium bromide
°C- Degree Celsius
DAF- DNA Amplification Fingerprinting
DALP-Direct Amplification of Length Polymorphism
DAMD-Directly Amplified Minisatellite DNA
DAMD-PCR- Direct Amplification of Microsatellite DNA by PCR
DFLP- DNA Fragment Length Polymorphism
DNA- Deoxy ribonucleic acid
dNTPs- Deoxy ribonucleotide triphosphate
dRAMP- Digested Randomly amplified microsatellite polymorphism
EDTA- Ethylene diamine tetra acetic acid
EG- Eastern Ghats
EH- Eastern Himalayas
hr- hour
HSTE- High salt tris EDTA
IFLP- Intron Fragment Length Polymorphism
IM-PCR- Inter-microsatellite PCR
In-del- insertion/deletion
IRAP- Inter-Retrotransposon Amplified Polymorphism
ISA- Inter-SSR amplification
ISSR- Inter Simple Sequence Repeats

IUCN- International Union for Conservation of Nature
KFRI- Kerala Forest Research Institute
km^2- square kilometer
Km-kilometer
m- meter
M –molar
MAAP- Multiple arbitrary amplicin profiling
min- minute
mL- milliliter
mM- millimolar
mya- million years ago
MoEF- Ministry of Environment
MP-PCR-Microsatellite-primed PCR
MSN- Minimum Spanning Network
mt- mitochondrial
MT- mitotype
Mt. Everest- Mount Everest
NEI- North East India
ng- nanogram
nm- nanometer
NRB- number of rare bands
NSB- number of shared bands
NSiB- number of similar bands
OLA- Oligonucleotide ligation assay
PCA- Principal Coordinate Analysis
PCR- Polymerase Chain Reaction
PCR-RFLP- Polymerase Chain Reaction-Restriction Fragment Length Polymerase
PEC- primer enzyme combinations
PI- probability of identical match
PIC- polymorphism information content
PPB- percent polymorphic bands
RAHM- Randomly Amplified Hybridizing Microsatellites
RAMP- Randomly Amplified Microsatellite Polymorphism
RAMS-Randomly Amplified Microsatellites
RAPD- Randomly Amplified Polymorphic DNA
RBIP- Retrotransposon-based insertion polymorphism
RE- Restriction enzyme
REMAP- Retrotransposon Microsatellite Amplified Polymorphism
RFLP- Restriction Fragment Length Polymorphism
rpm- revolutions per minute
RT- room temperature

SAMPL- Selective Amplification of Microsatellite Polymorphic Loci
SCAR- Sequence Characterized Amplified Region
sec- second
SI- South India
SPAR- Simple Primer Amplification Reaction
SPI- ISSR primer index
SRAP- Sequence Related Amplified Polymorphism
S-SAP- Sequence Specific Amplification Polymorphism
SSCP- Single Strand Confirmation Polymorphism
SSLP- Simple Sequence Length Polymorphism
SSRs- Simple Sequence Repeats
STAR- Sequence tagged amplified region
STMS- Sequence tagged microsatellites
STR- Short tandem repeat
STS- Sequence-tagged-site
TAE- Tris-acetate-EDTA
TBE- Tris-borate-EDTA
TDEF- Tropical dry evergreen forests
TE- Tris-EDTA
TRAP-Target Region Amplification Polymorphism
UBC- University of British Columbia
UPGMA- Unweighted pair group method with arithmetic mean
VNTR-Variable Number of Tandem Repeat
WG- Western Ghats
µL- microliter
µM- micromolar

List of tables

Literature review

Section 1

DNA barcoding of *Dalbergia* species collected from Western Ghats, India

Section 2

Genetic diversity studies in *Symplocos* species

List of figures

Section 1

DNA barcoding of *Dalbergia* species collected from Western Ghats, India

Section 2

Genetic diversity studies in *Symplocos* species

Section 2A

Genetic diversity studies in *S. laurina* using ISSR markers

Thesis abstract

The present work is divided into two objectives, to develop DNA barcode for differentiation of *Dalbergia* species, and to study the genetic diversity of *Symplocos* spp. Initially, the objective of the thesis was to develop DNA barcodes for important forest tree species in Western Ghats (WG). To fulfil this, we selected two genera namely, *Dalbergia* (valued for its timber) and *Symplocos* (medicinally important). In case of *Dalbergia*, multiple accessions of ten different species from various geographical locations were collected. However, in case of *Symplocos*, only two species *viz*. *S. laurina* and *S. racemosa* could be collected from different geographical locations, in spite of several collection attempts. Hence, the *Dalbergia* samples were further subjected to DNA barcoding studies; while *Symplocos* was subjected to genetic diversity studies, due to less sample numbers. The thesis is divided into two sections:

Section 1: DNA barcoding studies of *Dalbergia* and section 2: Genetic diversity studies in *Symplocos* species

Section 1: DNA barcoding of *Dalbergia* species collected from Western Ghats, India

In this study, DNA barcodes were identified for discrimination of *Dalbergia* species. The study included 7-26 accessions of ten *Dalbergia* species each, collected from different geographic locations in WG region of India. To identify a potential barcode, total 37 primer pairs were initially screened from several nuclear and plastid genes. Based on various parameters, four loci (*rbcL, matK, trnH-psbA* and nrITS) were selected and these and their combinations were further evaluated with five different analyses such as inter and intra specific distances, Neighbor Joining (NJ) trees, Best Match (BM) and Best Close Match, (BCM) character based approach and Wilcoxon signed rank (WSR) test. These results were ranked based on their performance and the best barcoding loci that could discriminate the various *Dalbergia* species were selected. The overall performance of these loci and their combinations after ranking all the results identified *matK*, followed by *matK+rbcL* as the most suitable barcoding loci to discriminate the *Dalbergia* species.

Further, these two loci along with nrITS were validated using available sequences from NCBI database. Those having minimum two sequences of different accessions of each species were selected for analysis. The DNA sequences of the three regions were downloaded from the NCBI database. The same parameters as mentioned above except WSR were used for further analysis. The ranking also revealed highest scores for *matK*

followed by *matK+rbcL*. Moreover, the unauthenticated samples were also used and analyzed based on NJ trees. The results showed that, in all the three loci, one unauthenticated sample clustered with *D. melanoxylon* while five with *D. horrida*. One accession of *D. sissoides* collected and authenticated from local person clustered with one of the variants of *D. latifolia* in all the three loci. Similarly, another accession of *D. paniculata* collected and authenticated from Mysore region clustered with *D. lanceolaria* samples. The phylogenetic analysis was carried out by combining all the above four primers with sequences used in the development of barcodes and downloaded from NCBI database. The same species were found to be clustered together irrespective of their geographical locations. The study revealed mixed pattern of distribution irrespective of the Old World and the New World origin, indicating monophyletic nature of the genus *Dalbergia*.

Section 2: Genetic diversity studies in *Symplocos* spp.

This section is subdivided into two sections. Section 2A deals with the genetic diversity studies in *S. laurina* using ISSR markers while 2B deals with effect of anthropogenic activities on patterns of genetic diversity in *S. racemosa*

Section 2A: Genetic diversity studies in *S. laurina* using ISSR markers

In this study, we evaluated 208 accessions from 13 populations of *S. laurina* belonging to three ecoregions in India namely, Western Ghats (WG), Eastern Ghats (EG) and Northeast India (NEI), using inter simple sequence repeat (ISSR) primers. We screened 100 primers, of which eight were selected based on their clear and reproducible amplification and polymorphism patterns. These eight ISSR primers produced on an average 1,014 bands from 208 accessions. The ISSR primer, UBC880 produced the maximum number of rare bands and hence can be considered as the most suitable primer to identify a particular accession. Similarly, UBC835 with high SPI value (5.40), was highly efficient in detecting polymorphism in *S. laurina* populations. Being an indicator of altitude, *S. laurina* was analysed to evaluate the effect of altitudinal variation on genetic diversity. Pearson's correlation coefficient revealed significant negative correlation (r =-0.59) and indicated better adaptability to higher altitudes. The overall gene flow was slightly low (0.9890) compared to threshold value, however, within region it was more than one migrant per generation indicating good differentiation within populations and within regions. The unweighted pair group method with arithmetic mean (UPGMA) dendrogram revealed clear separation of the NEI populations from the WG and EG

populations. However, the Munnar population from WG, clustered with EG instead of WG. The Principal Coordinate Analysis (PCA) also showed similar pattern of clustering, explaining the presence of Palghat gap (a 30 Km discontinuity in WG), separating Munnar from other WG populations. Further, based on PCA analysis, individuals of Kolli hills population from EG were distributed close to the axis separating WG and EG clusters, indicating that the Kolli hills population could be a link between WG and EG. The total genetic diversity (H_T) and the average heterozygosity (H_S) values were 0.3407 and 0.2263, respectively; while the coefficient of genetic differentiation (G_{ST}) was observed to be 0.3358 and the genetic diversity within-populations was found to be 0.2263 which was similar with the long-lived, perennial and outcrossing species of late successional stages. The within-population variance was higher (68%) than among-population variance (17%) indicating high fitness and adaptability of the species to changing environmental conditions.

Section 2B: Effect of anthropogenic activities on patterns of genetic diversity in *S. racemosa*

The effect of anthropogenic activities on genetic diversity of *S. racemosa* collected from seven populations of protected, semi-protected and disturbed regions of WG was studied using PCR-RFLP approach in intergenic sequences of chloroplast (cp) and mitochondrial (mt) DNA. The study identified nine chlorotypes and eight mitotypes. The Agumbe population represented the highest number of haplotypes (six chlorotypes and five mitotypes) while the Tillari population represented the lowest number of haplotypes (one in each case). The diversity parameters (H_T – total diversity, H_S – average intrapopulation diversity, G_{ST} – the level of population subdivision of diversity using unordered alleles) and the frequency of haplotypes were calculated. High total diversity (cp=0.786 and mt=0.778) was revealed for both the genomes. The UPGMA dendrograms indicated that the genetic affinities among the sites were irrespective of the geographical distances. The genetic variation was found to be partitioned among the sites. This could be because these sites might have become susceptible to stochastic effects of genetic drift leading to reduction in heterozygosity and loss of alleles. The study identified *S. racemosa* populations with low diversity as well as with unique haplotypes that could be used for formulating conservation strategies for the species. Interestingly, among all the populations, the populations from protected areas exhibited high diversity while the disturbed areas showed less diversity. Since the Agumbe population depicted the highest

number of haplotypes for both cp and mt genomes it would be a good source of different haplotypes to be introduced in the other areas.

Overall, the present study established DNA barcodes in *Dalbergia* genus; and analysed the genetic diversity as well as the effect of anthropogenic activities in case of *Symplocos* species.

Literature Review

DNA BARCODES OF CLOSELY RELATED (BUT MORPHOLOGICALLY AND ECOLOGICALLY DISTINCT) SPECIES OF SKIPPER BUTTERFLIES (HESPERIIDAE) CAN DIFFER BY ONLY ONE TO THREE NUCLEOTIDES

tion of genetic variability in *Symplocos laurina* Wall. wo biodiversity hotspots of India

ding Methods for Land Plants

kas, Maria L. Kuzmina, Steven G. Newmaster, Hollingsworth

of Anthropogenic Activities on Patterns of ic Diversity in *Symplocos racemosa* Roxb. from rn Ghats, India

1. Taxonomy: The Most Primitive Discipline of Science

The history of taxonomy dates back to the days of Aristotle, a Greek philosopher (384–322 BC). However, its use started much before that. Classifying edible, medicinal and poisonous plants and communicating them within tribe was in practice before the classification process came on paper. The ancient taxonomy is always linked with the Greeks and Romans; however, one of the earliest pharmacopoeias was written by Shen Nung (Chinese emperor) around 3000 BC. Another reference from the eastern world was found at around 1500 BC in Egypt; where, medicinal plants were found to be illustrated on wall paintings. Recently, the interesting history of taxonomy with important milestones has been described by Rouhan and Gaudeul (2014) and Manktelow (2010) and references therein.

Development of the taxonomy field can be broadly divided into pre-Linnean era, Linnean era and post-Linnean era as described below:

1.1 Pre-Linnean era of taxonomy:

The pre-Linnean era is described as a time before modern taxonomy. This era illustrates the work of some early taxonomists whose contribution is recognized even today, after so many centuries. Aristotle was the first to classify living organisms and divide them into vertebrates (animals with blood) and invertebrates (animals without blood). As per Aristotle, the vertebrates were further divided into live and egg bearing, while the invertebrates were recognized today as insects, crustacean and testacea (molluscs). Similarly, Theophrastus (370–285 BC), a student of Aristotle and Platon, classified plants based on the growth forms into categories such as trees, shrubs and herbs and illustrated nearly 500 plants in *De Historia Plantarum*. Many genera (*Narcissus*, *Crocus* and *Cornus*) described by him are still in use. Dioscorides (40-90 AD), a Greek physician travelled a lot in Greek and Roman regions and acquired the knowledge of medicinal plants. On the basis of medicinal properties of these plants, he described around 600 species in *De Materia Medica*. The Roman army man, Pliny the Elder (23–79 AD), described many plants in *Naturalis Historia* and gave them Latin names; some of the genera described by him like *Populus alba*, *Populus nigra* are still in use.

1.1.1 Role of navigations in plant taxonomy

The major navigations or expeditions took place approximately between 1450 and 1550 AD. These navigators (Columbus, Vasco de Gama, Magellan, etc.) or the naturalists, who were part of these expeditions, brought the plants from various countries to Europe, which

were illustrated by various herbalists. Further, the preservation technique invented by Luca Ghini (1490–1556) revolutionised the study of taxonomy by allowing the drying, pressing and preservation of the plants thus assisting in their analysis. This was the first instance for creation of herbarium (permanent collection of dried plants from various regions), a hallmark achievement for plant taxonomy and its development.

Andrea Cesalpino (1519–1603) from Italy is considered as the first taxonomist who contributed to the concept of natural classification. He classified plants on the basis of growth habits along with fruit and seed form and described around 1500 species in *De Plantis* (1583). The two large plant families Asteraceae and Brassicaceae, discussed by him, are very well acknowledged even today. Two Swiss Bauhin brothers (1541-1631; 1560-1624), illustrated their work on 6,000 species in *Pinax Theatri Botanici* (1623). They included synonyms for species and also established genera and species as major taxonomic levels. The British naturalist John Ray (1627–1705) suggested grouping of the plants based on multiple characters mostly related to leaves, flowers and fruits. He published *Methodus Plantarum Nova* in 1682 and described 18,000 plant species. He differentiated between flowering and non-flowering plants and within flowering plants further distinguished the plants with single cotyledon and two cotyledons and coined the terms monocotyledons and dicotyledons, respectively. He established species as the ultimate unit of taxonomy and introduced first text-based dichotomous keys to classify plants. On the other hand, French botanist Joseph Pitton de Tournefort (1656–1708) explained classification of plants based on only few characters such as corolla of flower. This resulted into relatively easy classification of plants based on symmetry and number of corolla. On the basis of this he identified 698 genera and published *Institutiones Rei Herbariae* (1700).

1.2 Linnean era

The Swedish botanist Carl Linnaeus (1707–1778) brought in a new era of taxonomy. Till the middle of the 18th century, the plants were identified by Latin names with polynomials (i.e. containing descriptor name and generic name), which was a complicated system. Linnaeus developed a binomial system of nomenclature for both, plants and animals consisting of genus name followed by species name. This system of classification is still in use even after two centuries. He published his work in *Systema Naturae* (1735) and *Species Plantarum* in 1753. The *Species Plantarum* is considered as a starting point for setting rules in plant taxonomy. Linnaeus proposed his own system of classification, where he classified the plants on the basis of number and arrangement of stamens and

pistils and called it as “sexual system” which was contrary to that proposed by Tournefort (1656–1708).

1.3 Post-Linnean era

The Linnaean system of classification was accepted by all except the French scientists. Georges-Luise Leclerc de Buffon (1707-1788) criticized Linnaeus work and dealt with species development and acquired inherited characters in species. On the other hand, Michel Ananson (1727-1806) appraised Linnaeus and suggested modification with consideration of more characters instead of giving weightage to some specific characters. Antoine Laurent de Jussieu (1748–1836) divided the plants into acotyledons, monocotyledons and dicotyledons, established the concept of family in the hierarchy and ranked it between genus and class. Bernard and A.L. de Jussieu specified that plants showing character constancy should be grouped together. However, all the characters could not be used to classify the same level and hence, characters showing higher variability were given less weightage than the conserved ones. As a result of this, floral characters were considered less suitable at higher levels as compared to seed and embryo characters.

Simultaneous to Linnean and post-Linnean era of taxonomy, Jean-Baptiste de Lamarck (1744–1829) introduced the theory of evolution during the eighteenth and nineteenth century. This was the first theory of evolution; however, with several misleading assumptions. Later, Charles Darwin (1809–1882) published the theory of evolution in his work entitled “*On the origin of species*” in year 1859. In this theory, he put forth the idea that every life form on the Earth has descended from a single common ancestor. This theory was explained by the “Tree of Life” where he explained that single-celled organism after many generations developed into new and more complex living forms. This unicellular organism was represented as the root of the tree while the new forms that developed from it were the branches. According to him, the branches were the organisms which formed due to change in existing organisms. This mechanism was described as ‘Natural Selection’. His other observation was variability in traits of the individuals of the same group. These variations were proposed to be responsible for the survival of the organism.

2. Modern Taxonomy

Taxonomy based on the development of cladistics theory and phylogeny reconstruction changed the style of taxonomy. Till then, all the traits were morphological and floral and

environment dependent. The cladistics era was established by German biologist Willi Hennig (1913-1976), which was based on synapomorphies i.e. the characters that are only inherited from the last common ancestor and not the primitive characters. According to him, the overall similarity does not necessarily reflect the evolutionary relationship. He published a book "*Grundzüge einer Theorie der Phylogenetischen Systematik*" in 1950 and its English translation entitled "*Phylogenetic Systematics*" in 1966 (Rouhan and Gaudeul 2014).

Later, Robert Sokal and Peter Sneath developed the concept of numerical taxonomy to reconstruct phylogenies (Sokal and Sneath 1963). It involved large number of characters and those which vary among individuals were considered more informative. Moreover, all the characters were given an equal weightage. It involved binary or multistate coding which resulted into character by taxa data matrix. Thus, on the basis of character analysis, similarity between the individuals was established to classify the organisms. However, this method was not based on any evolutionary theory (Rouhan and Gaudeul 2014).

3. Molecular Era

The discovery of DNA by James Watson, Francis Crick and Maurice Wilkins followed by the Sanger sequencing method (Sanger et al. 1977) and Polymerase Chain Reaction (PCR) invention by Karry Mullis (Mullis and Faloona 1987) radically changed the field of biology. Further, restriction enzymes and PCR facilitated work at DNA level in various organisms. The use of sequence data along with cladistics method resulted in formation of the 'Angiosperm Phylogeny Group' (APG), which is a group of systematics botanists working for development of robust classification system of angiosperms on the basis of molecular phylogeny. The APG first published in 1998 (APG I) followed by two revisions; one in 2003 (APG II) and another in 2009 (APG III) (Gupta 2011). APG III included 56 orders and 450 families (Cole 2015).

The use of molecular data in the form of molecular markers/nucleotide sequence of a DNA fragment has become a landmark of taxonomy. Molecular marker is a heritable DNA sequence with definite localization on a chromosome (Pourmohammad 2013). DNA based markers are phenotypically neutral, environmentally stable and all organisms of a species have similar genomic DNA. These markers are found to be more versatile in fields of taxonomy, physiology, genetic engineering and embryology (Joshi et al. 1999). An ideal molecular marker must be frequently occurring in the genome, highly polymorphic in nature to display sufficient genetic diversity and expected to show

codominant inheritance for determination of homozygous and heterozygous states of diploid organisms. Along with this, they are anticipated to have easy access and fast assay, high reproducibility and easy data exchange between laboratories. However, it is difficult to have all the properties in a single marker hence they are chosen based on their suitability of applications. These markers are further classified as PCR based and non-PCR based as described below.

3.1 Non-PCR based molecular markers

RFLP (Restriction Fragment Length Polymorphism)

RFLP is one of the first molecular markers developed in early 1980. David Botstein and others claimed the presence of large amount of DNA variation in human populations, part of which was detected in the form of variable length DNA fragments produced by restriction enzymes (Botstein et al. 1980). RFLPs follow Mendelian inheritance and originate at the time when DNA rearrangements occur due to evolutionary processes, insertions or deletions within fragments, unequal crossing over and point mutations at the recognition site of restriction enzyme. The RFLP protocol involves digestion of genomic DNA by restriction enzyme (RE) followed by gel electrophoresis to resolve DNA and then blotting on to a nitrocellulose membrane. The banding patterns are then visualised by southern hybridization with labelled probe. The probes are obtained from cDNA or genomic DNA library (Joshi et al. 1999). The RFLPs are highly reproducible. They can detect coupling phase of DNA molecules due to their co-dominant nature and are very reliable markers in linkage analysis and breeding and can easily determine a state in individual (homozygous or heterozygous) where a linked trait is present. However, on the negative side, they require large quantity (1–10 μg) of purified, high molecular weight DNA and the assay is time and labour intensive. Further, RFLPs are unable to detect single base changes and this restricts their applicability (Kumar et al. 2009). The RFLPs were first used in mapping of genes responsible for Huntington disease (Botstein et al. 1980), but later were widely used for many applications in plants. They have also been majorly used in preparation of genetic maps in crops and conifers (Neale et al. 1992). They were also found to be useful in genetic diversity studies of Pea (Lu et al. 1990), *Acacia mangium* (Butcher et al. 1998), wild and cultivated rice (Sun et al. 2001) etc.

3.2 PCR based markers

RAPD (Randomly Amplified Polymorphic DNA)

These are the dominant markers, developed by Welsh and McClelland (1990). RAPDs detect polymorphism using a single primer of arbitrary nucleotide sequence which anneals to complementary strands of genomic DNA at two different sites. PCR based amplification produces discrete DNA bands that are originated from various loci in the genome. The standard RAPD primers are short synthetic oligonucleotides approximately 10 bases long. The advantages of this method are that RAPD requires only small amount (~20 ng) of DNA, is fast and efficient, and species specific probe libraries and hybridization are not required (Joshi et al. 1999; Chawla 2009). However, being dominant markers, they can cause the loss of information compared to co-dominant markers. In addition, due to the short primer length, mismatch of even a single nucleotide in the primer binding site prevents the primer from annealing the target DNA and leads to loss of amplification. Further, RAPD also faces a problem of poor reproducibility (Joshi et al. 1999). In previous studies, West African and Asian genotypes of okra (*Abelmoschus caillei* and *A. esculentus*) were assessed successfully for genetic distinctiveness and relationships using RAPD (Aladele et al. 2008). RAPDs also effectively identified the *Glycyrrhiza glabra* L. from its adulterant *Abrus precatorius* L. in the study of authentication of medicinal plants thus, serving RAPD as a complementary tool for quality control (Khan et al. 2009). RAPD markers were also employed to evaluate genetic diversity of mandarin accessions (*Citrus* spp.) (Coletta Filho et al. 1998), Barberton daisy (*Gerbera jamesonii*) along with morpho-agronomic characters (da-Mata et al. 2009), *Plantago* spp. (Singh et al. 2009) and endangered plant species such as *Anisodus tanguticus* (Zheng et al. 2008), *Neolitsea sericea* (Wang et al. 2005) and *Heptacodium miconioides* (Lu et al. 2006). Ecological characteristics, spatial distribution and genetic features of the *Heptacodium miconioides* Rehd were also studied using RAPD markers explaining the effects of habitat fragmentation (Liu et al. 2007). However, due to its less reproducibility and availability of more precise molecular technologies, the use of RAPD markers finds fewer applications in recent years. Conversion of specific RAPD markers into Sequence Characterized Amplified Region (SCAR) markers is considered more precise (Theerakulpisut et al. 2008; Yang et al. 2013).

AFLP (Amplified Fragment Length Polymorphism)

This tool was developed by Zabeau and Vos (1993). It is a combination of RFLP and PCR and is a powerful DNA fingerprinting technology. They are abundantly found in the

genome; however, AFLP requires purified and high molecular weight DNA. The protocol involves restriction of genomic DNA with a mix of two enzymes followed by ligation of adaptors and selective PCR amplification with limited sets of AFLP primers without any prior knowledge of DNA sequence. These fingerprints are visualised after gel electrophoresis through autoradiography or fluorescence methodology (Vos et al. 1995; Jones et al. 1997). AFLP is a robust technique as primer annealing conditions are stringent. The AFLP fragments are the unique positions on the genome and hence can be used as landmark for genetic and physical mapping (Joshi et al. 1999). They can also be converted into SCAR markers for more specificity. The technology is majorly used for evaluating the degree of variability among closely related genotypes. AFLP when used for *Solanum* taxonomy, was found to be an encouraging tool for evolutionary studies (Kardolus et al. 1998). They have also been employed in phylogenetic analysis of *Croton alabamensis* (Van et al. 2006) and conservation genetics studies of *Pulsatilla vernalis* (Ronikier 2001) and *Leucopogon obtectus* (Zawko et al. 2001). Recently, plants such as *Jatropha curcas* (Tatikonda et al. 2009), *Rhodiola rosea* (Elameen et al. 2008), wild populations of *Agave angustifolia* and endangered herb such as *Dendrobium officinale* (Li et al. 2008) have been characterized by AFLP based molecular characterization and genetic diversity studies.

SSRs (Simple Sequence Repeats)/ Short tandem repeats/ micro satellites

SSRs are PCR-based, co-dominant sequence tagged microsatellites (STMS) markers (Joshi et al. 1999). They are short mono, di-, tri-, tetra- and penta- nucleotide repeats (2-7 bp), widely distributed in the plant genome, which can be used in detecting variation between and within species. These repeat regions at the respective loci between the genotypes may vary in length, but their flanking region is sufficiently conserved. Hence, the SSR primers are synthesized from these conserved regions which can work in several genotypes of a species. Because of the high variability, SSRs are used to distinguish closely related genotypes and favoured in population studies (Smith and Devey 1994). The main advantage of SSR is the requirement of low quantity (10–100 ng per reaction) and moderate quality of template DNA. SSR markers are highly reproducible and easy to assay. However, initial development costs are required and there is a possibility of detection of non-homologous similar sized fragments during the assay. SSR markers have been successfully used in assessment of genetic diversity in crops such as rice (Ghneim et al. 2008) and barley (Wang et al. 2010a) and development and characterization of endemic plant *Psudolarix amabillis* (Geng et al. 2015).

Other microsatellite markers are Randomly Amplified Microsatellite Polymorphism (RAMP), Simple Primer Amplification Reaction (SPAR), Sequence Related Amplified Polymorphism (SRAP), and Target Region Amplification Polymorphism (TRAP) that are variations of SSR system (Kharkwal and Jain 2004).

ISSR (Inter Simple Sequence Repeats)

ISSRs were first reported by Zietkiewicz et al. (1994). These are approximately 100-3000 bp fragments located adjacent to oppositely oriented microsatellite regions. They are dominant markers and highly sensitive, reproducible and cost effective compared to other PCR-based markers (Reddy et al. 2002). The microsatellite repeats are used as primers in case of ISSRs. These are either anchored at 3' or 5' end or unanchored. ISSR based assay is a single primer PCR reaction amplifying a region between two microsatellites. ISSRs do not require prior DNA sequence information and can work with small quantity (5–50 ng per reaction) template DNA detecting very low level of genetic variation effectively. However, amplification of the non-homologous similar sized fragments can take place and may also have reproducibility problems. In spite of these problems, ISSRs have been successfully employed in genetic diversity studies in many forest plants such as *Primula obconica* (Nan et al. 2003), *Eurya nitida* (Bahulikar et al. 2004), and *Gmelina arborea* (Naik et al. 2009); medicinal plants such as *Glycyrrhiza uralensis* (Yao et al. 2008), flax (Rajwade et al. 2010), *Salvia miltiorrhiza* (Song et al. 2010), *Cunila* species (Agostini et al. 2008) and *Pinus nigra* (Rubio-Moraga et al. 2012) and economically important plants such as mango (Pandit et al. 2007) and Brazilian cultivated *Jatropha curcas* L. accessions (Grativol et al. 2011).

Minisatellites

The term minisatellites was introduced by Sir Allec Jeffreys (Jeffreys et al. 1985) . These loci contain tandem repeats that vary in the number of repeat units between genotypes (e.g. Variable Number of Tandem Repeats, Directly Amplified Minisatellite DNA etc.). Many more marker systems have been reported and listed in Table 1. Such recent tools are proved to be more useful when morphological characters are insufficient (Bradford and Barnes 2001; Widmer et al. 2000). As mentioned above, molecular markers such as RAPD, RFLP, ISSRs are used to study diversity and are also useful in addressing ecological issues and taxonomical problems (Arif et al. 2009). Moreover, DNA sequencing and fingerprinting data using plastid and nuclear microsatellite markers along with AFLP are also used to study population dynamics (Fay and Krauss 2003), species delimitation (Richardson et al. 2003) and hybridization (Clarkson et al. 2004).

Table 1: List of molecular markers

AFLP	Amplified Fragment Length Polymorphism
AMP-PCR	Anchored Microsatellite Primed PCR
AP-PCR	Arbitrarily Primed PCR
ASA	Allele-Specific Amplification
ASSR	Anchored Simple Sequence Repeat
CAPS	Cleaved Amplified Polymorphic Sequence
DAF	DNA Amplification Fingerprint
DALP	Direct Amplification of Length
DAMD-PCR	Direct Amplification of Microsatellite DNA by PCR
DFLP	DNA Fragment Length Polymorphism
dRAMP	Digested RAMP
IFLP	Intron Fragment Length Polymorphism
IM-PCR	Inter-Microsatellite PCR
IRAP	Inter-Retrotransposon Amplified Polymorphism
ISA	Inter-SSR Amplification
ISSR	Inter Simple Sequence Repeats
MAAP	Multiple Arbitrary Amplicon Profiling
MP-PCR	Microsatellite-Primed PCR
OLA	Oligonucleotide Ligation Assay
RAHM	Randomly Amplified Hybridizing Microsatellites
RAMPO	Randomly Amplified Microsatellite Polymorphisms
RAMP	Randomly Amplified Microsatellite Polymorphism
RAMS	Randomly Amplified Microsatellites
RAPD	Random Amplified Polymorphic DNA
RBIP	Retrotransposon-based Insertion Polymorphism
REMAP	Retrotransposon Microsatellite Amplified Polymorphism
RFLP	Restriction Fragment Length Polymorphism
SAMPL	Selective Amplification of Microsatellite Polymorphic loci
SCAR	Sequence Characterized Amplified Regions
SNP	Single Nucleotide Polymorphism
SPAR	Single Primer Amplification Reactions
S-SAP	Sequence Specific Amplification Polymorphism
SSCP	Single Strand Confirmation Polymorphism
SSLP	Simple Sequence Length Polymorphism
SSR	Simple Sequence Repeat
STAR	Sequence Tagged Amplified Region
STMS	Sequence Tagged Microsatellite Site
STR	Short Tandem Repeat
STS	Sequence-Tagged-Site
VNTR	Variable Number of Tandem Repeats

4. DNA Barcoding- Reinventing the Traditional Taxonomy

The plants are of immense importance, they provide food (30,000 species of plant are edible and 80% of plant based calorie intake comes from 12 domesticated plant species including 8 cereals and 4 tubers), medicine (7,000 plant species are used for medicinal

purpose in India and 10,000 species in China; 75% of the world's population relies on traditional medicines), firewood (http://www.fao.org/forestry/27939-4743b4f1653a2e73766199b4c184b5c5.pdf; http://www4.total.fr/pdf/AMO/Kenya/ecochallenge/the%20expert%20corner/T-Pearce-Kew-Kews-Millennium-Seed-Bank-Partnership.pdf) etc. Besides fulfilling human needs, plants also play fundamental role in sustaining ecosystem by balancing climatic conditions. However, around 60,000 to 100,000 plant species are currently threatened with extinction (http://www.unesco.org/mab/doc/iyb/scConf/PaulSmith.pdf). One fifth of the world's tropical forests were destroyed between 1960 and 1990, resulting in the loss of diversity. Current plant extinction rate has increased to 100-1000 times (http://www.fao.org/forestry/27939-4743b4f1653a2e73766199b4c184b5c5.pdf).

Correct identification of plants is a prerequisite in various disciplines to report findings or to access available information. The traditional methods of identification involve morphological, anatomical, physiological and embryological characters. However, there are a few major limitations of this strategy, such as phenotypic plasticity where different individuals of the same species may exhibit variation due to natural conditions or local adaptations. Alternatively, two different species from the same genus may exhibit the same morphological characters (Duminil and Di Michele 2009). Similarly the problem of cryptic taxa and long maturity periods are the prevalent problems in plant identification system. Also, the lack of taxonomic keys for immature specimens and high level of specific expertise restrict the identification process with classical taxonomical approach. In fact, it has been mentioned that the science of taxonomy is facing the major crisis mainly because the lack of expertise (Pires and Marinoni 2010). According to previous studies, no doubt, molecular markers have been used in taxonomic studies; however, it is difficult to use them at species level identification. To revive this, complementary approaches such as DNA barcoding supported with information based approaches and cybertools (Wheeler et al. 2004; Pyle et al. 2008; La Salle et al. 2009) have been suggested to be implemented for precise taxonomical analysis.

5. What is DNA Barcoding?

The DNA barcode is a short stretch of DNA from standardized regions of the genome to identify species similar to supermarket barcodes which identifies different products. The idea behind it is the short sequence stretch from a reference gene being unique to one specific species might distinguished it from the other species (Hebert and Gregory 2005). A small region of DNA sequence contains more than enough information to resolve 10

or even 100 million species. It is illustrated best by an example of 600 bp protein-coding region. The region contains 200 nucleotides at third position, where substitutions are neutral and mutations accumulate through random drift. Even if we consider only A or T or G or C at the third position, there will be 4^{200} possible sequences based on only third position. This theory has been successfully applied for species determination in the animal kingdom based on the analysis of the mitochondrial cytochrome c oxidase subunit I (*cox I*) sequence and found to be highly effective in closely related species as well as diverse phyla such as birds, fish, flies, butterflies and other animal groups (Hebert et al. 2003a). It has also shown the accordance between morphology based taxonomy and DNA barcode analysis. Attempts have been made to apply the same for plant systems also.

5.1 Desired characteristics of a DNA barcode

An ideal DNA barcode is required to be short enough (approx 700bp) so that it can be processed quickly for thousands of specimens and analysed by computer programs and easily recoverable with a single primer pair. The primer should serve conserved flanking sites for development of universal DNA barcode and the region must be amenable to bidirectional sequencing with little requirement for manual editing of the sequence traces. Substantial divergence at species level is expected from the barcodes so that they can identify variation at lower taxonomic units. The sequences with less in-dels are preferred with respect to their easy alignments. The barcoding locus must be readily recoverable from the degraded samples also. This may help in case of forensic material and also in case of herbaria or preserved samples (Hollingsworth et al. 2009; Kress et al. 2005; Chase et al. 2007).

5.2 Challenges in plant DNA barcoding

There is a lot of debate regarding the distinctness of plant species. From many years it is of concern, that reticulate evolution facilitated by hybridization and genome duplication and apomixis rule out the possibility of single species concept (Stebbins Jr 1950; Levin and Funderburg 1979). The plastid genome is inherited maternally in plants once the hybridization has occurred. Backcrossing of the hybrid with the other parental species may lead to introgression of plastid genome from one species to the other. After backcrossing for several generations, it would result into almost pure nuclear genome of one species, but with the plastid genome of the other species. In such cases, the use of plastid loci would give erroneous results. This has been observed in many plant species like *Orchis* (Fay et al. 2007), *Heuchera* (Soltis and Kuzoff 1995), *Juglans* spp. (Hoban et al. 2009) and *Brassica* spp. (Hansen et al. 2003; Haider et al. 2009). In case of polyploids,

due to reproductive isolation, some variations may accumulate in plastid genome which might lead to failure in species discrimination using plastidial DNA barcoding. Hence in allopolyploids, combination of morphology and DNA barcoding might give meaningful results. However, autopolyploids and parental diploids are always indistinguishable on the basis of gross morphology and considered as cytotypes of the same species (Soltis et al. 2007). In some groups, apomictic individuals and polyploids are very common (e.g., in *Rubus*, *Sorbus*, *Taraxacum* and *Limonium*). In such cases, the plastid genome in the apomicts is likely to be unchanged, therefore old polyploids might not be distinguishable from one of their progenitors (Chester et al. 2007). In such cases, the use of plastid DNA barcodes may fail to distinguish between the maternal diploids and their polyploid derivatives.

Rieseberg et al. (2006) studied taxonomically recognized plant species and concluded that <60% species were phenotypically distinct and 70% correlated to reproductively isolated groups. Also, the reports by Fazekas et al. (2009) confirmed that plant species are not well separated genetically and observed that species monophyly is less common in plants. Thus, the study concluded that plants exhibit reproductive discontinuity and hence, discriminating plant species using only a single type of genome (plastid) is more challenging. Various regions in the organellar as well as plastid genomes in the plant systems are, therefore attempted to select an appropriate barcoding region at family or genus or species level which are discussed below.

5.3 Selection of loci for plant DNA barcoding

DNA barcoding has been implemented successfully for species determination of various phyla in animal kingdom using *cox I* sequence. However, the *cox I* region is not effective in plants as the rate of nucleotide substitution in plant mitochondrial (mt) genome is very slow perhaps because of hybridization and introgression, frequent genome rearrangements and transfer of genes between genomes and across species resulting in very little variation in mt genome (Palmer 1985; Mower et al. 2004; Stoeckle 2003). Therefore, in plants chloroplast and nuclear genomes are commonly used. Initial DNA barcoding studies in plants proposed few plastid coding as well as non-coding regions as promising candidates, which included *rbcL* and *trnH-psbA* (Kress et al. 2005), *matK*, *rpoC1*, *rpoB* and *trnH-psbA* (Chase et al. 2007), *matK*, *atpF/H*, *trnH-psbA* and *psbK/I* (Hollingsworth et al. 2011) and *trnL* intron and its shorter P6 loop (10-143 bp) (Taberlet et al. 2007) as suitable plant barcodes. However, the slow evolving plastidial coding regions might not possess enough variation to discriminate closely related plant species

and this could lower their potential as effective barcodes (Pettengill and Neel 2010). Hence, along with the plastid coding and non-coding regions, the recently evolving nuclear region, i.e. nuclear transcribed spacer from ribosomal gene (nrITS), has also been proposed as a potential barcode (Chen et al. 2010). Nuclear DNA detects hybridization and provides more information on species identification. However, there are difficulties in PCR amplification of low copy genes. Also, obtaining a good quality sequence from degraded samples is difficult. Till now only ITS region from ribosomal DNA is explored for DNA barcoding (Vijayan and Tsou 2010).

5.3.1 Single locus DNA barcodes

Many plastid loci such as *rbcL*, *matK* and *trnH-psbA* and nuclear locus (nrITS) have been successfully used on many taxa at species level although the DNA barcoding communities mostly rely on plastid loci as compared to nuclear ones (Dong et al. 2012). These loci are discussed below.

matK

matK codes for Maturase K protein in chloroplast. The gene is highly conserved and involved in Group II intron splicing. It has high evolutionary rate and sufficient interspecific divergence; hence, higher discrimination (close to *cox I*) and low transition/transversion rate. In DNA barcoding, universal amplification is difficult using current set of primers. Various case studies revealed difference in the rate of success. The CBOL Plant Working Group (Hollingsworth et al. 2009) showed 90% amplification success in angiosperms using a single primer pair while 83% with gymnosperms and 10% with cryptogams. Lahaye et al. (2008) achieved 100% success rate when 1667 samples of angiosperms were studied. In addition they also analysed more than 1,000 mesoamerican orchids with *matK* and revealed cryptic species and proved useful for identification of species as listed in CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora).

rbcL

It codes for the large subunit of the enzyme ribulose-1, 5-biphosphate carboxylase/oxygenase (rubisco) (Gielly and Taberlet 1994) and is widely used in phylogenetics. This gene is easy to amplify and sequence in most land plants and it is free from length mutations. It offers better resolution at the family and genus level; however, shows modest discriminatory power at the species level (Doebley et al. 1990; Fazekas et al. 2008; Lahaye et al. 2008; Hollingsworth et al. 2009; Chen et al. 2010). Also, the length

of the gene can be an issue as the sequencing of both the strands of the gene may require more than two primers. Regardless of these limitations, based on the recovery of the sequences and moderate discriminatory power, *rbcL* has been suggested as the best potential candidate for plant barcodes (Hollingsworth et al. 2009; Hollingsworth et al. 2011). It is further suggested that *rbcL* can be used in combination with various loci to improve the discrimination of species (Kress and Erickson 2007; Hollingsworth et al. 2009; Chase et al. 2007).

trnH-psbA

TrnH-psbA is one of the most widely used plastid intergenic non-coding spacer for barcoding. The region is characterized by good universality, higher discrimination, highly conserved coding sequences (which makes the design of universal primers feasible) (Shaw et al. 2005), good priming sites (Kress et al. 2005) but length variability and frequent termination of sequencing reads by SSRs (Kress and Erickson 2007). The alignment of *trnH-psbA* spacer can be problematic due to length variations (because of high rates of in-dels in majority of angiosperms) (Chase et al. 2007) and its complicated molecular evolution (Chang et al. 2006). Furthermore, the presence of homopolymers and inversions and insertion of *rps19* gene (Starr et al. 2009; Whitlock et al. 2010; Pang et al. 2012) also creates problems. It is observed that the length of *trnH-psbA* is >1000 bp in some conifers and monocots (Hollingsworth et al. 2009; Chase et al. 2007) while it is very short (less than 300 bp) in other groups (Kress et al. 2005) and shorter than 100 bp in bryophytes (Stech and Quandt 2014) which may not possess enough sequence variation to distinguish between the species. However, this intergenic spacer has been effectively used in barcoding of *Hydrocotyle* (Van De Wiel et al. 2009), *Dendrobium* (Yao et al. 2009), Pteridophytes (Ma et al. 2010), Amazonian trees (Gonzalez et al. 2009) and tropical forest plants in Panama (Kress et al. 2009).

Additionally, there are many plastid loci such as *rpoB, rpoC1, accD, ycf5, ndhJ, atpF–atpH* intergenic sequence, *psbK–psbI* intergenic sequences and *trnL (UAA)–trnF* (*GAA)* (including genic, intron and intergenic sequences) attempted for various plants at family, genus or species level (Vijayan and Tsou 2010).

nrITS

The ITS (18S-5.8S-26S) is the most commonly used spacer in plant molecular systematics at species level. Synonymous substitution rates are far greater in ITS than plastid genes which are three times greater than plant mitochondrial genes. Due to its broad utility and high interspecific divergence at lower taxonomic levels, it has been

proposed as a possible candidate for plant DNA barcoding (Alvarez and Wendel 2003; Stoeckle 2003). The region is more useful in parasitic plants where plastid regions offer less resolution. However, there are some advantages and limitations associated with this non-coding spacer. nrITS can be amplified in two smaller fragments (ITS1 and ITS2) enabling the recovery of entire ITS from degraded and poor quality samples. ITS2 is considered as more conserved in nature (Kress et al. 2005; Gao et al. 2010b; Chen et al. 2010). However, fungal contamination, paralogous gene copies and concerted evolution are the major problems which prevent it from being a core barcode (Hollingsworth et al. 2011). The presence of multiple copies in the tree genomes creating high intra-specific variation (Yamaguchi et al. 2006) might lead to inaccurate or misleading results (Alvarez and Wendel 2003). This variation can be explained because of recent hybridization events, lineage sorting, recombination events, high mutation rates and pseudogene formation (Vijayan and Tsou 2010). Secondary structure problems (Kress et al. 2005) associated with ITS also lead to poor quality sequence generation. Taking these problems into account, it has been suggested to find nuclear genes specific to order or family. This might help in identification of taxonomically complex groups that may lack sufficient variation in plastid DNA. In spite of the above mentioned drawbacks, CBOL has considered ITS as a supplementary locus (Hollingsworth et al. 2009). Furthermore, Chase et al. (2005) and Cowan et al. (2006) have suggested the use of a large number of short regions with pyrosequencing platform (Pacey-Miller and Henry 2003), which would allow targeting of many genes in a single reaction (Margulies et al. 2005). Thus, these techniques may facilitate the use of nuclear regions for future DNA barcoding efforts in plants (Fazekas et al. 2009).

5.3.2 Multi locus barcode

To overcome the problems caused by single loci, use of a two-locus barcode [based on combination of *trnH-psbA* and *rbcL* (Kress and Erickson 2007) and *matK*+*rbcL* (Hollingsworth et al. 2009)] was proposed (Chase et al. 2005). Kress and Erickson (2007) studied 48 genera using nine putative barcodes including coding and non-coding regions singly and in pairs. The results revealed that a single locus was able to discriminate the species not more than 79% of genera. However, when *trnH-psbA* was coupled with coding regions, the success rate increased to 88%. *In silico* analyses also supported the combination of *trnH-psbA* and *rbcL* and led to correct identification. de Groot et al. (2011) suggested *rbcL* and *trnL-F* as two-locus barcode for fern identification. The study by Pang et al. (2012) based on 586 species belonging to 71 genera and 47 families also

concluded that *trnH-psbA* and its combination with *ITS2* performed better as a plant DNA barcode. Similarly, Chase et al. (2007) and Pennisi (2007) proposed three loci combination involving *matK+rpoB+rpoC1* and *matK+atpF-atpH+psbK-psbI* or *matK+atpFatpH+ trnH-psbA*, respectively. The Southern African Combrataceae study by Gere et al. (2013) revealed that combination of *rbcL+matK+trnH-psbA* performed better. CBOL Plant Working Group (Hollingsworth et al. 2009) proposed *matK+rbcL* as the core barcode for plants which was also supported by Li et al. (2011).

5.4 Applications of DNA barcoding

In spite of many challenges in the path of DNA barcoding as discussed earlier, there are many applications which are useful in biodiversity conservation, forensic science and drug authentication methods etc. which have been discussed below.

1. In case of plants where morphology based identification (usually the reproductive organs) is limited, DNA barcoding is helpful. Thus, at juvenile stages, barcode based identification can be carried out effectively (Hollingsworth et al. 2011). This approach was successfully tested in identification of Amazonian trees where 72% juveniles were identified with the help of morphology. However, morphology and molecular approach in combination increased the identification rate to 96% (Gonzalez et al. 2009).
2. Chip based identification using microarrays can be established for species recognition. It represents the next step of automation. This method has been successfully used in identification of fishes from European seas (Kochzius et al. 2010) and microalgae from southern coastal region in Korea (Lee et al. 2012).
3. Cryptic species identification is possible using DNA barcoding tool especially in case of bryophytes where they lack the problematic processes such as hybridization, polyploidy, seed dispersal etc. (Shaw 2001; Vanderpoorten and Goffinet 2006; Hollingsworth et al. 2011). The studies by Miwa (Miwa et al. 2009) successfully differentiated six reproductively isolated cryptic species of *Conocephalum* by *rbcL* sequences. Similarly, three closely related species of the liverwort genus *Anastrophyllum* were discriminated by four plastid loci (*psbA–trnH*, *trnH–trnK*, *trnL* intron and *psbC– trnS*) (Long et al. 2006).
4. Ecological forensics: It is a type of DNA metabarcoding in which trophic interactions among species, within communities are resolved by genotyping complex mixtures of individuals. Such studies were successfully performed by Jones et al. (2011), in which 33 species of roots from 12 soil cores were identified.

Such studies are helpful in areas where identification of plant roots, seedlings or cryptic life stages is important but the areas are inaccessible (Schneider and Schuettpelz 2006; Kesanakurti et al. 2011). The DNA sequences can be utilized to find out plant components from the diet of animals by extracting DNA of their gut content, fecal samples (Deagle et al. 2009) or honey (Valentini et al. 2010), thus enabling the construction of complete food web.

5. Processed material: Barcoding analysis proved to be useful in processed material to find out the adulterants in herbal medicines (Song et al. 2009; Chen et al. 2010; Gao et al. 2010a; Gao et al. 2011), to study species diversity in gut contents of animals to identify complex food webs (Soininen et al. 2009), food products (Jaakola et al. 2010; Valentini et al. 2010) and in analysis of herbivore's diet components (Valentini et al. 2009; Staudacher et al. 2011). The polygonaceae samples and their adulterants were identified using *trnH-psbA* in Chinese pharmacopeia (Song et al. 2009). Similarly, *trnH-psbA* also identified six different species of *Phyllanthus* from the raw drug samples collected in South India (Srirama et al. 2010).
6. DNA barcoding plays an important role in identification of invasive species (Bleeker et al. 2008; Van De Wiel et al. 2009). In case of armyworm *Spodoptera* species in Florida, DNA barcodes were used to assess the threat by this unknown specimen for the region. The identification in turn helped to take preventive measures against the pest (Nagoshi et al. 2011).
7. This tool can be applied to the species which are in illegal trade. Ramin is trafficked for its timber, often used in children furniture and window blinds. Many of the Ramin species have become endangered due to over exploitation. These species are protected by CITES. Ogden et al. (2009) developed a SNP genotyping approach based on the *matK* DNA barcodes to distinguish the products of Ramin (*Gonystylus*).
8. Biodiversity:
    a. Meusnier et al. (2008) developed the mini-barcodes in Coleophora (Order Lepidoptera) and demonstrated its effectiveness in ancient samples. The mini-barcodes were successfully used on collection of Coleophora which were difficult to identify because of their small size and cryptic morphology. Thus, this significantly widened the application of DNA barcoding in biodiversity studies.

b. Soil contains a wide range of organisms such as archaea, bacteria, fungi, nematodes, insects and earthworms. However, the soil biodiversity has remained unexplored due to number of methodological and logistical issues. DNA metabarcoding gives an opportunity to study such uncharted diversity and gain importance in biodiversity studies (Orgiazzi et al. 2015).

c. Ant diversity: The diversity of ants in Madagascar was studied using DNA barcoding. The study demonstrated that the DNA barcoding addressed the failure of current inventory methods and provided an effective substitute for time intensive morphological analyses (Smith et al. 2005).

5.5 Reports on plant DNA barcodes

Till date many studies have been carried out on various plant systems using a large number of loci to find out a universal barcode. These studies have been conducted at family, genus or species level. CBOL has proposed *matK*+*rbcL* as a universal DNA barcode for plants; however, it has been shown that it does not work well in all the plant systems and hence many other loci are proposed for various plant systems as shown in Table 2.

Table 2: Data of barcoding efforts at family, genus or species level

Plant system (family/genus/species)	**Loci used**	**Proposed loci**	**Reference**
Asteraceae	ITS, *psbA-trnH*, *matK*, *rbcL*, ITS2	ITS2	Gao et al. (2010b)
Podostemaceae	*matK*, *rpoB*, *rpoC1*, *trnH-psbA*	*matK*	Kelly et al. (2010)
Euphorbiaceae	*rbcL*, *matK*, ITS, and ITS2	ITS/ITS2	Pang et al. (2010)
Lemnaceae	*rpoB*, *rpoC1*, *rbcL* *matK*, *atpF-atpH*, *psbK-psbI*, *trnH-psbA*	*atpF-atpH*	(Wang et al. 2010b)
Fabaceae	ITS2	ITS2	Gao et al. (2010a)
Fabaceae	*matK*	*matK*	Gao et al. (2011)
Palms (Arecaceae)	*matK*, *rbcL*, *psbA-trnH*, nrITS2	nrITS2+*rbcL*+ *matK*	Jeanson et al. (2011)
Meliaceae	*rpoC1*, *rpoB*, *accD*, *psbB*, *psbN*, *psbT*	nrITS	Muellner et al. (2011)

	exons, *trnS-trnG* spacer, nrITS		
Cactaceae	*matK*, nrITS	Additional regions need to be explored	Yesson et al. (2011)
Podocarpaceae	*matK*, *rbcL*, nrITS2	Additional regions need to be explored	Little et al. (2013b)
Araliaceae	*matK*, *rbcL*, ITS2, *psbA-trnH* and *ycf5*	ITS2	Liu et al. (2012)
Zingiberaceae	*matK*	*matK*	Selvaraj et al. (2008)
Asteraceae	*rbcL*, *matK*, ITS, ITS2, *psbA-trnH*	ITS2	Gao et al. (2010b)
Myristicaceae	UPA, *rpoB*, *rpoc1*, *accD*, *rbcL*, *matK*, *trnH-psbA*	*matK*+*trnH-psbA*	Newmaster et al. (2008)
Potamogetonaceae	*rbcL*, *matK*, ITS, *trnH–psbA*	ITS	Du et al. (2011)
Combrataceae	*rbcLa* + *mat*K, nrITS and *trnH-psbA*	*rbc*La + *mat*K+*trnH-psbA*	Gere et al. (2013)
Cycades	nrITS, *accD*, *ndhJ*, *matK*, *trnH-psbA*, *rpoB*, *rpoC1*, *ycf5*	nrITS	Sass et al. (2007)
Epimedium	*psbA-trnH*, ITS, *rbcL*, *matK*	*psbA-trnH*	Jiang et al. (2011)
Aspalathus	nrITS, *psbA-trnH*, *trnT-trnL*	*psbA-trnH*	Edwards et al. (2008)
Hedyotis	*matK*, *trnH–psbA*, *petD*, *rbcL*, ITS	ITS+*petD*	Guo et al. (2011)
Quercus	*matK*, *trnH-psbA*, ITS2	Further analysis required	Simeone et al. (2013)
Parnassia	*rbcL*, *matK*, *trnH-psbA* and ITS	ITS+*trnH-psbA*	Yang et al. (2012)
Lotus	nrITS	nrITS	Sandral et al. (2010)
Lotus	*matK*+*rbcL*, *trnH-psbA*, *rpoC1*, *rpoB*, nrITS	nrITS	Ojeda et al. (2014)
Non flowering seed plants	*psbA-trnH*, *matK*, *rbcL*, *rpoB*, *rpoC1*, ITS1 and ITS2	ITS2	Pang et al. (2012)

Viburnum	*rbcL* + *matK*, *rpl32-trnL*, *trnH-psbA*, *trnK*, nrITS	*trnH-psbA* and nrITS are effective however, additional regions need to be explored	Clement and Donoghue (2012)
Dendrobium	*trnH-psbA*	*trnH-psbA*	Yao et al. (2009)
Primula	*matK*, *rbcL*, *rpoB*, *rpoC1*, *trnH-psbA*, *psbK-psbI*, *atpF-atpH*, ITS, *rbcL+matK*	ITS, further analysis required	Yan et al. (2011)
Paphiopedilum	*rpoB*, *rpoC1*, *rbcL*, *matK*, nrITS	*matK*	Parveen et al. (2012)
Dioscorea	*rbcL*, *matK*, and *psbA-trnH*	*matK* is strong but not perfect	Sun et al. (2012)
Taxus	*rbcL*, *matK*, *trnH-psbA*, *trnL-F*, ITS	*trnL-F* and ITS, separately or combined	Liu et al. (2011)
Pterygiella	*rbcL*, *matK*, and ITS	ITS	Dong et al. (2011)
Gossypium	*matK*, *rbcL*, ITS2	*matK*+*rbcL*+ITS2; however no single combination of barcodes could differentiate all the *Gossypium* species, and tetraploid species	Ashfaq et al. (2013)
Carex	*matK*, *rbcL*, *rpoC1*, *rpoB*, *trnH-psbA*	*matK*	Starr et al. (2009)
Sisyrinchium	*matK*, *trnH-psbA*, ITS	ITS can be used at preliminary stages	Alves et al. (2014)
Alnus	ITS, *rbcL*, *matK*, *trnH-psbA*	ITS and *trnH-psbA*	Ren et al. (2010)
Ligustrum	*rbcL*, *matK*, and *trnH–psbA*, nrITS	nrITS and *trnH-psbA* as a candidate	Gu et al. (2011)
Acacia	*rbcL*, *trnH-psbA*, *matK*	*rbcL*, *trnH-psbA*, *matK*	Newmaster and Ragupathy (2009)

NW-European Ferns	*rbcL* + *trnL-F*	*rbcL* + *trnL-F*	de Groot et al. (2011)
Dalbergia	nrITS, *trnL* intron, *matK*-designed, *matK*-barcode, *psbA-trnH*	nrITS	Phong et al. (2014)

Genesis of thesis

Plants play an important role in our day to day life as they form the major portion of our food and other life supporting properties/ utilities including medicinal value. Plants also play a crucial role in ecosystem balance. Correct identification and taxonomy of plants is, therefore, very important. Till 2010, approximately 1.7 million plant and animal species have been classified, which is less than 25% of the total estimated number of the species. Hence, there is a need for rapid identification and classification of plants using amalgamation of classical approach and modern molecular methods. Our lab has been engaged in studying the biodiversity and phylogeography of various plant species with the ultimate objective of conservation of WG ecology. The present study includes the DNA barcoding studies which will complement the classical taxonomic studies for accurate identification and biodiversity analysis of relict shola species from various locations distributed in WG, India.

The main objective of the present study was to develop the DNA barcodes for important forest tree species in WG. For this study we selected *Dalbergia* (valued for its timber) and *Symplocos* (one of the medicinally valued plant genus) as they are economically important and are exploited immensely. In case of *Dalbergia* nearly 30-35 species are reported in India and 10-15 in WG. We could collect ten different species and their multiple accessions from various geographical locations in WG and hence, they were further subjected to DNA barcoding analysis. Likewise, in case of *Symplocos*, total 68 species have been reported from India. However, even after several collection efforts (more than 10 visits to various reported regions along with the team of Botanical Survey of India) we could collect only two species *viz*. *S. laurina* and *S. racemosa* from different geographical locations. Due to the unavailability of enough number of species, *Symplocos* was further subjected to genetic diversity studies using various populations collected from WG, Eastern Ghats (EG) and North East India (NEI). The studies were extended to understand the population structure of *S. laurina* and the effect of anthropogenic activities on *S. racemosa* in order to further plan for sustenance of these species in WG. Based on this, following objectives were set for the thesis.

Objectives

- To evaluate chloroplast and nuclear regions as potential DNA barcodes for *Dalbergia* species
- To identify a single locus/ combination of loci as the DNA barcodes
- To implement these barcodes for phylogenetic analysis

- To analyse the genetic diversity in 13 populations of *S. laurina* collected from three ecological regions in India using ISSR markers and comment upon the population structure
- To analyse the effects of anthropogenic activities on the genetic diversity and population structure of *S. racemosa* and generate preliminary data for conservation purposes

Organization of thesis

Considering above objectives, the thesis has been organized into two sections, the first section of DNA barcoding in *Dalbergia* and the second one is on genetic diversity analysis in *Symplocos* using various molecular markers. These sections are preceded by the combined review of literature followed by combined summary and future prospects. The details are as given below.

Thesis abstract

Review of literature (Current chapter)

This includes the history of taxonomy with major contribution of various botanists. The taxonomy in this section is divided into three era: 1. pre-Linnean era, 2. Linnean era and 3. post-Linnean era followed by modern taxonomy. Achievements at molecular level explain various marker systems and their applications in taxonomy. DNA barcoding is a recently developed approach to supplement the classical taxonomy and identification process. This has been explained in details with its application in various plant species.

Section 1: DNA barcoding of *Dalbergia* species collected from Western Ghats, India

In the present study, we analysed various plastid (*matK*, *rbcL*, *trnH-psbA*) and nuclear ribosomal (nrITS) loci to develop the DNA barcodes for identification of *Dalbergia* species. The study evaluated various distance based and character based parameters. Considering the overall performance of these loci and their ranking with various approaches, *matK* and *matK+rbcL* were suggested as the most suitable barcodes to unambiguously differentiate *Dalbergia* species.

Section 2: Genetic diversity studies in *Symplocos* species

This section is subdivided into two sections: 2A and 2B. Section 2A deals with the genetic diversity studies in *S. laurina* using ISSR markers while 2B deals with effect of anthropogenic activities on patterns of genetic diversity in *S. racemosa*

Section 2A: Genetic diversity studies in *S. laurina* using ISSR markers

This section describes genetic diversity studies in 13 populations of *S. laurina* collected from three ecoregions in India (WG, EG and NEI) using ISSR markers. These studies identified the role of Palghat gap along with other geographical factors in shaping the genetic diversity of *S. laurina* in South India.

Section 2B: Effect of anthropogenic activities on patterns of genetic diversity in S. racemosa

The present study analyses the effects of anthropogenic activities on the genetic diversity and population structure of *S. racemosa*. The variations in intergenic sequences of chloroplast and mitochondrial genomes from seven sites of *S. racemosa* sampled from protected, semi-protected, and disturbed areas of WG were analysed using PCR-RFLP approach. The protected sites were found to be highly diverse, while the disturbed areas possessed less genetic diversity.

Thesis summary and future prospects

Bibliography

Curriculum vitae

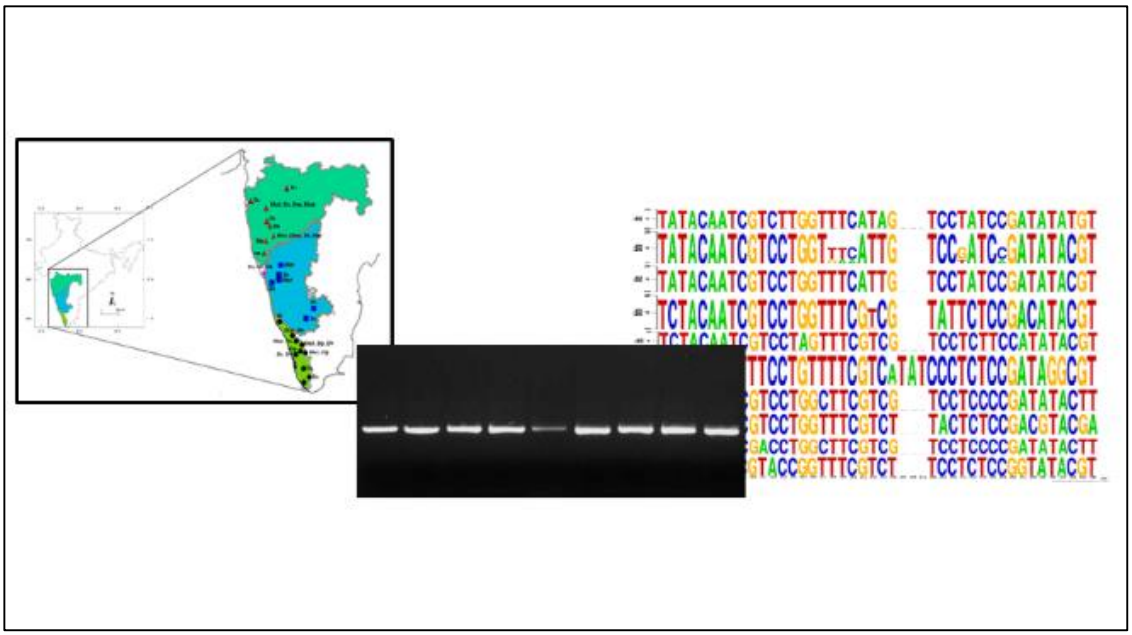
Appendix in the form of compact disc (CD): This includes 1. list of all samples with collection details and gene bank accession numbers (Table1.5), 2. Character based approach for validation (Table 1.17) and 3 sequences of ten *Dalbergia* species using *matK, rbcL, trnH-psbA* and nrITS used for analysis.

Section 1: DNA barcoding of *Dalbergia* species collected from Western Ghats, India

Publication

Two New Potential Barcodes to Discriminate *Dalbergia* Species (2015)

PLoS ONE (DOI: 10.1371/journal.pone.0142965)



S 1.1 Introduction

Dalbergia Linn. F. is a fabaceous genus (tribe Dalbergiae) of shrubs, lianas and trees. It is confined to the tropical regions of the world with Amazonia, Madagascar, Africa and Indonesia as the centers of diversity (Hiremath and Nagasampige 2004; Ribeiro et al. 2005; Vatanparast et al. 2013). The genus is recorded with more than 300 species in the world. Within the South-east Asia, India is the dwelling place for approximately 30 species representing more than 30% of the total South-east Asian *Dalbergia* species (Nair 1986).

Tribe Dalbergieae

The tribe comprises trees, shrubs or lianas. Generally, it is characterised by imparipinnate leaves and flowers having the two upper calyx lobes joined higher, sometimes spathaceous, 2-lipped or sub truncate, keel petals connate at apex, connate stamens 9-10, monoadelphous with uniform, dorsifixed anthers dehiscing by slits or pores or diadelphous (9+1 or 5+5) with versatile anthers, ovary sessile to long stipitate, 1- to few ovuled, glabrous style and samaroid and indehiscent fruits with globose to oblong-reniform seeds (Thothathri 1987; Dezaho et al. 2010). The tribe occupies the middle place in the phylogenetic relationship of family Fabaceae. The tribe has n=10 and may vary from 10-11 (Goldblatt 1981). In terms of phylogeny among paplionaceous species, n=10 is a comparatively high basic chromosome number and characteristic of major tropical groups indicating a separate origin from the polyploid tribe Sophorea of the Papilionaceae family.

The tribe Dalbergieae contains two genera:

1. *Dalbergia* with erect trees to climbing shrubs, small, white to reddish flowers and pods oblong to elliptic
2. *Pterocarpus* with trees, large yellow flowers, pods suborbicular to orbicular in shape with wings (Thothathri 1987).

Genus *Dalbergia*

The genus *Dalbergia* was created by younger Linnaeus in 1781 in memory of Nicholas Dalberg (Swedish botanist). Though the genus was created in 1781, the tribe Dalbergieae was founded in 1822 by Brongniart (Thothathri 1987). The genus was subdivided many times for taxonomic convenience and many people contributed for that, few of them are listed below; Roxburgh (1798), Willdenow (1800),Wight and Arnott (1834) Graham, Nimmo (1839), Dalzell (1850), George Bentham (1852 and 1860), Miquel (1855), Baker (1876-1878), Taubert (1890), Brandis (1907), Merril (1910), Gamble (1918), Haines

(1922), Ridley (1922), Parkinson (1923), Kanjilal et al. (1931), Backer et Bakhuzien (1963), Santapau (1967), Nasir and Ali (1977) and Hara et al. (1979). They contributed species from India, Pakistan, Nepal and other countries of South-East Asia. George Bentham (1852 and 1860) made an important contribution in the studies of genus *Dalbergia* (Bentham 1860). His preliminary work in 1852 recognised three major sections based on the stamens and fruit characters, namely, *Sissoa* (18 species), *Selenolobium* (2 species) and *Dalbergaria* (8 species). Further, he added three new species (*D. confertiflora*, *D. velutina*, both from Sylhet and *D. assamica* from Assam) and also validated the Wallichian names with detailed descriptions. Based on his previous work, he published a monograph on the tribe Dalbergieae in 1860 considering fruit characters. However, for further classification of genus he considered other characters also. He added one more section i.e. *Triptolemea* to his previously classified three sections and all the four sections further divided into series as follows: *Triptolemeae Americanae, T. Gerontogeae, Sissoae Americanae, S. Gerontogeae* and *Dalbergariae Gerontogeae*. Thus, the American species were separated from Asian and African ones. According to him, this classification was convenient for identification rather than phylogenetic considerations.

After Bentham, David Prain (1904) published taxonomic revision of the genus *Dalbergia* and wrote the monograph entitled '*The species of Dalbergia in South-East Asia*'. He made a remarkable contribution to the Asian species of *Dalbergia* including all available taxonomic information and proposing revised classification by evaluating the theories of Bentham and others. He subdivided the genus considering South-East Asian species and broadly classified it into two subgenera namely Sissoa and Amerimnon. Sissoa was further subdivided into subsections *Triptolemea*, *Podiopetallum* and *Endespermum*, while *Amerimnon* into *Miscolobium* and *Dalbergaria*. Thothathri (1983) proposed the revised classification and reduced *Ecastaphyllum* P. Br. (an American genus) with unifoliate leaves and orbicular pods to *Dalbergia* and rest was in line with Bentham (1860). *Ecastaphyllum* was allotted as a sectional status, under *Dalbergia*, along with sections *Sissoa*, *Dalbergia* and *Selenolobia*. In 1986, Thothathri critically reviewed the tribe Dalbergieae and the Asiatic genera for their taxonomic status and systematic position based on androecium and fruit types and illustrated in '*Hortus Indicus Malabaricus*' described by Rheede. The Brazilian species of *Dalbergia* were studied intensively by Carvalho (1989, 1997). In general, *Dalbergia* species are morphologically variable and they possess a wide range of habitat preferences, which previously made it difficult to classify the New and Old World species into natural groups (Bentham 1860;

Carvalho 1989; Prain 1904). The genus previously was identified by six different names. They were *Ecastaphyllum* P. Br. (1756), *Amerimnon* P. Br. (1756), *Salken* Adans (1763), *Solori* Adans (1763), *Pterocarpus* Berg. (1769) and *Acouroa* Aubl.

Thus, as mentioned above, the present classification system divides the genus into four sections based on stamens, leaves and pods characters:

1. *Sissoa* Benth. Emend. Thoth.
2. *Dalbergia*
3. *Selenolobia* Benth
4. *Ecastaphylla* (P. Br.) Thoth. Stat. nov.

Key

1a. Stamens monoadelphous; vexillum erect; pod thin to coriaceous — Sect. *Sissoa* 1

1b. Stamens mono to diadelphous; vexillum erect to reflexed; pod coriaceous

2a. Stamens only diadelphous; pods oblong to orbicular

3a. Leaf multifoliolate; vexillum reflexed; pod oblong — Sect. *Dalbergia* 2

3b. Leaf mostly unifoliolate; vexillum erect; pods round to orbicular — Sect. *Ecastaphylla* 3

2b. Stamens mono to diadelphous; pods falcately oblong and at times lunar — Sect. *Seleolobia* 4

Cytology

2n=20 chromosomes (Atchison 1951; Darlington and Wylie 1956). The studies were carried out on species from Kerala, namely *D. lanceolaria, D. latifolia, D. melanoxylon, D. paniculata, D. sissoo, D. spinosa* and *D. volubilis*, both tree species and climbers and the chromosome number was found to be consistent.

Palynology:

The palynological studies are carried out on Indian legumes including *D. volubilis, D. sissoo, D. latifolia, D. horrida* and *D. lanceolaria* (Mittre and Sharma, 1962). The pollens are characterized by 3-zonocolporate grains. They are further divided into two types on the basis of exine pattern:

1. *Crotalaria albida- D. volublis* Roxb. *D. sympathetica* Nimmo (*D. horrida*)
2. *Lathyrus pratensis- D. latifolia* Roxb., *D. lanceolaria* Linn.f. and *D. sissoo* Roxb. Ex DC (Thothathri 1987).

Distribution

The genus has a pantropical distribution. The species included in the study have a wide range of distribution. Nearly 35 and 10-15 species are reported from India and WG, respectively (Hiremath and Nagasampige 2004; Vatanparast et al. 2013). Among these, seven species are endemic to WG with overall high species diversity (http://wgbis.ces.iisc.ernet.in/biodiversity/sahyadri_enews/newsletter/issue38/article/index.htm). *D. candenatensis* is distributed in Asia, Malaysia and China while *D. lanceolaria* and *D. volubilis* are well distributed in India, Sri Lanka, Burma and former in Indo-China and later in Bangladesh. *D. melanoxylon* is originally African species but well documented in Western India. *D. sissoo* is observed in cultivated form in Burma and Malaya but it is found naturally in India, Afghanistan and Baluchistan. *D. horrida*, *D. tamarindifolia*, *D. latifolia* and *D. rubiginosa* are native of India proper.

Economic importance

The genus is economically important for the quality timber. The wood of different *Dalbergia* species is used for specific purposes such as making furniture (*D. latifolia*, *D. sissoo*), boat building (*D. sissoo*), making musical instruments (*D. melanoxylon*), etc. (Hiremath and Nagasampige 2004). *Dalbergia* species (*D. horrida*, *D. latifolia*, *D. melanoxylon*, *D. paniculata* and *D. sissoo*) also play an important role in nitrogen fixation as they form root nodules during symbiosis with rhizobia and thus, are useful in improving soil fertility (Rasolomampianina et al. 2005; Allen and Allen 1981). *D. lanceolaria* is a medicinally important species and its timber is used to prepare agricultural implements (Thothathri 1987).

Morphology based classification of *Dalbergia*

To address the primary need of accurate species identification, we attempted augmentation of the molecular efforts in genus *Dalbergia* by developing potential barcodes. We have included ten *Dalbergia* species distributed in WG. The key characters of all the species are described below

Stamens 9 (rarely 10), monoadelphous; staminal tube slit along the upper side only.

Tall trees unarmed

Leaflets acuminate……..*D. sissoo*

Leaflets obtuse……*D. latifolia*

A small tree armed with spines……...*D. melanoxylon*

Scadent shrubs

Leaflets 5-7

Pod staright……..*D. rubiginosa*

Pod crescent……..*D. candenatensis*

Leaflets 11-15……*D. horrida*

Leaflets 25-41……*D. tamarindifolia*

Stamens 10; staminal-tube slit on the upper and lower sides, so that the stamens are in 2 bundles of 5 each

Trees

Standard 0.25 inches broad, with a callosity at the

base of the limb……..*D. lanceolaria*

Standard 0.125 inches broad, without a callosity at the

base of the limg……*D. paniculata*

A climbing shrub; leaflets 11-15……..*D. volubilis*

***D. candenatensis* Dennst.**

It is a climbing shrub with purplish-black bark. Leaves are approximately 4-7.5 cm long, while leaflets 1.5-2.4 X 1-1.5 cm. Leaflets are elliptic-oblong, apex obtuse or emarginate, glabrous above and puberulous below. Flowers are in axillary panicles, white coloured and approximately 0.6-0.8 cm long. Calyx is glabrous. Pods are 2.4 cm long, brown, polished and with short stalks. The flowering and fruiting season is between June to August (Kulkarni 1988).

***D. horrida* Dennst.**

D. horrida is a large climbing shrub with strong, blunt and curved spines on trunk. Branches are twisted. Leaves are 8-10 cm long. Leaflets are 1.2-2.8 X 0.6-1.5 cm, oblong with obtuse or emarginated apex. Leaflets are pubescent on both the surfaces. Flowers are in axillary, cymose-panicles white and 0.6-0.8 cm long. Calyx is pubescent. Pods are

1-2 seeded and 4.5-6 cm long. Season of flowering and fruiting starts in November and extends up to May (Kulkarni 1988).

***D. latifolia* Roxb.**

These are the 8-10 m tall trees with grey coloured bark. Leaves are 8-16 cm long. Leaflets are 5-7 in number rarely 9, 3-5.5 X 2.5-4.5 cm and glabrous on both sides. Flowers are in lax, axillary panicles, creamy white in colour and 0.6-0.8 cm long. Calyx is glabrous. Pods are strap shaped, 1-3 seeded with narrowed ends and 3-5.5 cm long. The flowering and fruiting season is from February to November (Kulkarni 1988).

***D. paniculata* Roxb.**

It is a medium sized approximately 10-15 m tall tree with grey coloured bark. Leaves are around 10-15 cm long with 9-13, 1.5-2.4 X 0.8-1.5 cm long leaflets. Leaflets are glabrous, subcoriaceous. Flowers are in terminal subcoymbose panicles, bluish-white in colour, subsessile and 0.8-1 cm long. Calyx is densely silky. Pods are glabrous, oblong-lanceolate and 3.5-7.5 cm long. The start season of flowering is February, followed by fruiting and ends in April (Kulkarni 1988).

***D. sissoo* Roxb.**

D. sissoo is a moderate sized, approximately 8-10 m tall tree with spreading branches. Leaves are approximately 6-10 cm long with zigzag rachis. Leaflets are 2.5-8.7 X 1.5-3 cm in dimension with terminal one largest in size. Flowers are 0.6-0.8 cm long, in short panicles; sessile or subsessile and yellowish-white in colour. Calyx is hairy with ciliated teeth. Pods are thin, strap shaped and 2.5-6.0 cm long. The flowering and fruiting season is from February to October. Majority of *D. sissoo* trees are used for plantations and rarely found in their natural habitats (Kulkarni 1988).

***D. volubilis* Roxb.**

These are the large, shrubby climbers with thick glabrescent branches. Leaves are 8-10cm long with oblong or obovate-oblong shaped leaflets, Obtuse or truncate apex and 7-13, 2-3.5 X 1.5-2.4 in size. The leaflets are glabrous and pale green beneath. Flowers are in axillary or terminal panicles; lilac or pale bluish in colour and 0.6-0.8 cm long. Calyx is puberulous. Pods are linear-oblonged, glabrous and 3-5.5 cm long. Flowering season starts in February, followed by fruiting and ends in May (Kulkarni 1988).

***D. melanoxylon* Guill and Perr. Fl. Seng. Tent.**

These are the large shrubs. Branches are armed with stout spines. Bark is white in color. Leaflets per leaf are 9-15 in number and obovate. The inflorescence is axillary panicle with large number of flowers. Corolla yellow coloured. Stamens are 10 in bundles of 5, 4, and 1, respectively. Pods are 1-2 seeded and narrowed into a stalk. The flowering and fruiting season is from January to May (Kothari and Moorthy 1993; Yadav and Sardesai 2002).

***D. lanceolaria* L.f.**

D. lanceolaria are deciduous trees attaining height of 5-10 m with greyish-white coloured bark. Leaves are 9-17 foliolate, leaflets oblong-obovate or elliptic 2.2-7.7 X 2-4.2 cm, glabrous above and puberulous below. Flowers are subsessile and in copious terminal subcymose panicles. Corolla is bluish white in colour, standard with a claw which is as long as calyx. Stamens are 10, diadelphous with 5+5 arrangement. Pods are strap shaped (narrow at base and apex), glabrous with long stalk and 1-3 seeded. Flowering season starts in April, followed by fruiting and ends in June (Kothari and Moorthy 1993; Yadav and Sardesai 2002).

***D. rubiginosa* Roxb.**

It is a scandent shrub with terete and glabrous branches. Leaves are imparipinnate, 3-4 inches in length, rachis appressedly pubescent. Leaflets are 5-7 in number, coriaceous and 1-2.5 X 0.5-1.25 inches. Leaflets are elliptical-oblong and obtuse in shape. They are emarginate, glabrous and shining above, pale in colour, minutely puberulous and inconspicuously reticulately veined beneath. Base of a leaflet is rounded or subacute and midrib is prominent on the underside. Petiolules are 0.1 inches long and pubescent. The inflorescence is axillary panicles and flowers are crowded, 1-2 inches long with pubescent branches. Pedicels are approximately 0.08 inches long; bracts are ovate-oblong, subacute, pubescent on both the sides and ciliolate. Bracteoles are suborbicular-oblong, concave, rounded at the apex, pubescent on both the sides and ciliolate. Calyx is approximately 0.125 inches long, brown-silky, calyx teeth are obtuse, obovate-oblong and 0.0625 inches broad. Stamens are monoadelphous (nine or ten). Ovary is stalked with 2-4 ovules. Pods are strongly veined with 1-2 seeds (Cooke 1901).

D. tamarindifolia* Roxb. *(D. malabarica)

It is a scandent shrub growing up to 15-40 feet with densely rufous-pubescent young branches. Leaves are 4-6 inches long with densely puberulous rachis. Stipules are

lanceolated, 0.2 inches long. Leaflets are 25-41 in number, moderately firm, caduceus, 0.37-0.75 X 0.18-0.37 inches, trapezoid-oblong in shape, truncate, rounded and sometimes emarginated at the apex, pubescent on both the surfaces, paler beneath with short petioles. Flowers are in congested, sessile corymbosely-branched axillary panicles. They are upto 2 inches long. The branches of the panicles are densely brown-pubescent. Pedicels and bracts are 0.06 inches long. Bracts are ovate, subacute, pubescent and persistent. Calyx is 0.16 inches long and glabrous. Calyx teeth are one third as long as tube, short, subequal and subobtuse. Corolla is approximately 0.37 inches long and 0.1 inch broad, white coloured and not thickened above the claw. Stamens are usually nine, however, rarely ten, monoadelphous. Ovary is glabrous with 2-3 ovules. Pods are 1.5-3 X 0.33-0.5 inches, strap shaped, thin, and greenish while become reddish-brown on drying. They are glabrous, shining and slightly reticulate (Cooke 1901).

Threats to *Dalbergia*:

The African blackwood (*D. melanoxylon*) is highly prized for the properties of its heartwood. Because of its wide applicability as a timber, it has become the major source of income. Therefore, the processed wood has become very costly and costs around US$13,000 /$m^3$. The studies on tropical dry evergreen forests (TDEF) of India have pointed out that the logging is one of the major threats to commercial tree species decreasing its biodiversity. It is noticeable for the species which are enlisted in Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) Appendix II (Nithaniyal et al. 2014). The Red list of IUCN (International Union for Conservation of Nature) has >30 *Dalbergia* species under endangered category (http://www.iucnredlist.org) including *D. cochinchinensis* and *D. latifolia* as vulnerable while APFORGEN (Asia Pacific Forest Genetic Resource Programme) has identified *D. latifolia* of prior concern from conservation point of view. According to these reports, *Dalbergia* species are traded illegally in some countries and it is difficult to prove their identity and take action against it (Nithaniyal et al. 2014). Moreover, *D. cochinchinensis* has been listed as vulnerable by IUCN. It is a major concern that uncontrolled exploitation of *Dalbergia* species will lead to their extinction.

The need of DNA barcoding in *Dalbergia*

Dalbergia species are morphologically variable and possess a wide range of habitat preferences, which previously made it difficult to classify the Old (Africa, Asia and Europe) and the New World (North and South America) species into natural groups (Bentham 1860; Carvalho 1989). Many revisions of including and deleting certain species

from the *Dalbergia* genus over the past several decades have made the taxonomic speciation of this genus also quite challenging. Very little information is available on the molecular taxonomy of the *Dalbergia* genus in the literature. This emphasizes the need for molecular characterization of the genus along with its morphological discriminating parameters. Till date, several provincial revisions based on morphological characters have been made in *Dalbergia* (Bentham 1860; Prain 1904; Thothathri 1987; Carvalho 1997; Sunarno and Ohashi 1997; Niyomdham 2002; Ribeiro et al. 2005). The genus was included in the study of evolutionary analysis of Leguminosae (Lavin et al. 2001). The relationship of *Machaerium* and *Aeschynomene* was studied using sequencing of *trnL* and nuclear ribosomal DNA (Ribeiro et al. 2007). So far, there is only one report on phylogeny of *Dalbergia* species (Vatanparast et al. 2013) suggesting it to be of monophyletic nature. Only a few reports on molecular analysis of Indian *Dalbergia* species are available (Mohana et al. 2001; Rout et al. 2003; Hiremath and Nagasampige 2004; Arif et al. 2009; Bakshi and Sharma 2011). It is, therefore, important to study the genus for various aspects such as phylogeny, diversity and end-use quality using molecular approaches such as DNA markers and sequence based polymorphism in suitable genomic regions.

S 1.2 Materials and Methods:

S 1.2.1 Development of DNA barcodes:

Sample collection

The study included 166 accessions from ten *Dalbergia* species representing three sections, *Sissoa* (*Dalbergia latifolia*, *D. melanoxylon*, *D. sissoo*, *D. rubiginosa*, *D. horrida* and *D. tamarindifolia*), section *Dalbergia* (*D. volubilis*, *D. paniculata* and *D. lanceolaria*) (Hiremath and Nagasampige 2004) and section *Selenolobia* (*D. candenatensis*) (Thothathri 1987) (Figure S1.1). Our study was focused on locations in WG, one of the biodiversity hotspots (Table S1.1 and Figure S1.2). The samples were authenticated by the Kerala Forest Research Institute (KFRI) and the Botanical Survey of India (BSI, Western Circle, Pune, India) and voucher specimens from each species were deposited in their respective herbaria. *Pterocarpus marsupium*, which is native to WG, was used as an out-group in the present study, as the genus *Pterocarpus* falls outside the *Dalbergia* clade (Vatanparast et al. 2013)


D. lanceolaria


D. melanoxylon


D. rubiginosa


D. candenatensis


D. latifolia


D. sissoo


D. paniculata


D. volubilis


D. horrida

D. tamarindifolia

Figure S1.1: Photos of 10 *Dalbergia* species used in present study

Table S1.1: List of all samples with collection details

Species with accession numbers	**Locality**
D. rubiginosa 1- *D.rubiginosa* 10	Methunganam, Kasargod, Kerala
D. rubiginosa 11- *D. rubiginosa* 20	Thekkal, Kasargod, Kerala
D. rubiginosa 21- *D. rubiginosa* 22	Anshighat, Karnataka-Goa border
D. candenatensis 1- *D. candenatensis* 5	Kollam, Kerala
D. candenatensis 6- *D. candenatensis* 7	Alleppy, Kerala
D. candenatensis 8- *D.* candenatensis 9	Cherai, Paravoor, Ernakulam, Kerala
D. candenatensis 11, *D. candenatensis* 13- *D. candenatensis* 15, *D. candenatensis* 17- *D. candenatensis* 22	Anachal, Paravoor, Ernakulam, Kerala
D. latifolia 2- *D. latifolia* 4	Near Malabar cement factory, Walayar, Palakkad, Kerala
D. latifolia 6	Mangalam Dam Palakkad, Kerala
D. latifolia 9	Mukkali, Attappady, Palakkad, Kerala
D. latifolia 11, *D. latifolia* 21, *D. latifolia* 31	Chinnar Wildlife Sanctuary, Kerala
D. latifolia 21A, *D. latifolia* 22, *D. latifolia* 23	Kurichiad, Wayanad, Kerala
D. latifolia B	Gadhinglaj, Dist. Kolhapur, Maharashtra
D. latifolia D, *D. latifolia* E	Radhanagari, Maharashtra
D. latifolia F	Kerle, Kolhapur, Maharashtra
D. latifolia G, *D. latifolia* H	Panhala, Kolhapur, Maharashtra
D. latifolia I, *D. latifolia* J	Jotiba, Kolhapur, Maharashtra
D. latifolia K	Kolhapur, Maharashtra
D. latifolia L, *D. latifolia* P	Karnataka
*D. latifolia*Q	Pirangut, Pune, Maharashtra
D. latifolia R	University of Agricultural Sciences, Dharwad, Karnataka
D. latifolia S	Forestry college, Sirsi, Karnataka

D. latifolia T	Ulvi, Anshi National Park
D. melanoxylon 1- *D. melanoxylon* 5	Vetaltekadi, Kanchangalli, Pune, Maharashtra
D. melanoxylon 6- *D. melanoxylon* 8	Empress garden, seasonal flower nursery, Pune, Maharashtra
D. melanoxylon 9- *D. melanoxylon* 22	Pune, Maharashtra
D. melanoxylon 23- *D. melanoxylon* 24	Panhala road, Kolhapur, Maharashtra
D. paniculata 9, *D. paniculata* 11- *D. paniculata* 13	Chinnar Wildlife Sanctuary, Kerala
D. paniculata 14- *D. paniculata* 18	Parambikulam Wildlife Sanctuary, Kerala
D. paniculata 19- *D. paniculata* 23	KFRI, Kerala
D. paniculata 602	Near Malabar cement factory, Walayar, Palakkad, Kerala
D. volubilis 1, *D. volubilis* 5, *D. volubilis* 6, *D. volubilis* 29	KFRI, Kerala
D. volubilis 2, *D. volubilis* 3	KFRI Nilambur sub centre, Kerala
D. volubilis 9	Mukkali, Attappady, Palakkad, Kerala
D. volubilis 20- *D. volubilis* 22	Quilon, Thenmala, Kerala
D. volubilis 23- *D. volubilis* 25	Kasaragod, Parappa, Kerala
D. volubilis 26- *D. volubilis* 28	Kurichiad, Wayanad, Kerala
D. volubilis 30	Kottappara, Ernakulam, Kerala
D. volubilis 31	Naduvathumuzhi, Konni, Kerala
D. volubilis 32, *D. volubilis* 33	Thannithode-Thekkuthode;Pathanamthitta, Kerala
D. lanceolaria D, *D. lanceolaria* E	Radhanagari, Maharashtra
D. lanceolaria F, *D. lanceolaria* G	Kerle, Kolhapur, Maharashtra
D. lanceolaria CRD	Canal road, Pune, Maharashtra
D. lanceolaria DIRC	DIRC, NCL campus, Pune, Maharashtra
D. lanceolaria KERALA	Kerala
D. sissoo 1, *D. sissoo* 2, *D. sissoo* 6	Pune, Maharashtra
D. sissoo 3	Empress garden, Pune, Maharashtra
D. sissoo 4	Bhandarkar road, Pune, Maharashtra
D. sissoo 5	Pravrangar, Maharashtra
D. sissoo 7	BARC colony, Mumbai, Maharashtra
D. sissoo 8, *D. sissoo* 10	Maharashtra Cultural Centre garden, Dharavi, Mumbai, Maharashtra
D. sissoo 9	Parks and Gardens, BARC, Mumbai, Maharashtra
D. sissoo 11, *D. sissoo* 12	Pasaranighat, Mahableshwar, Maharashtra
D. sissoo 13	Gadhinglaj, Dist. Kolhapur, Maharashtra
D. sissoo 14	Atith, NH4 highway, Maharashtra
D. sissoo 15	Radhanagari Road, Maharashtra
D. sissoo 16, *D. sissoo* 17	Jotiba road, Kolhapur, Maharashtra
D. sissoo 18	Lalbagh garden, Banglore, Karnataka
D. sissoo 19	Ranganthittu, Mysore, Karnataka
D. sissoo 20, *D. sissoo* 21	Mysore University Guest House, Karnataka
D. sissoo 22	Aurangabad, Maharashtra
D. tamarindifolia 1, *D. tamarindifolia* 2	Sahasralinga, Karnataka

D. tamarindifolia 3	Kumta, Karnataka
D. tamarindifolia 4- *D. tamarindifolia* 7	Nuggi nursery, Anshi National Park, Goa
D. horrida A	Gaganbavada, Maharashtra
D. horrida 1	Koyananagar, Maharashtra
D. horrida 3- *D. horrida* 5, *D. horrida* 6	Amboli, Maharashtra
D. horrida 7- *D. horrida* 9	Anshi National Park, Goa
Pterocarpus marsupium	NGCPR, Satara

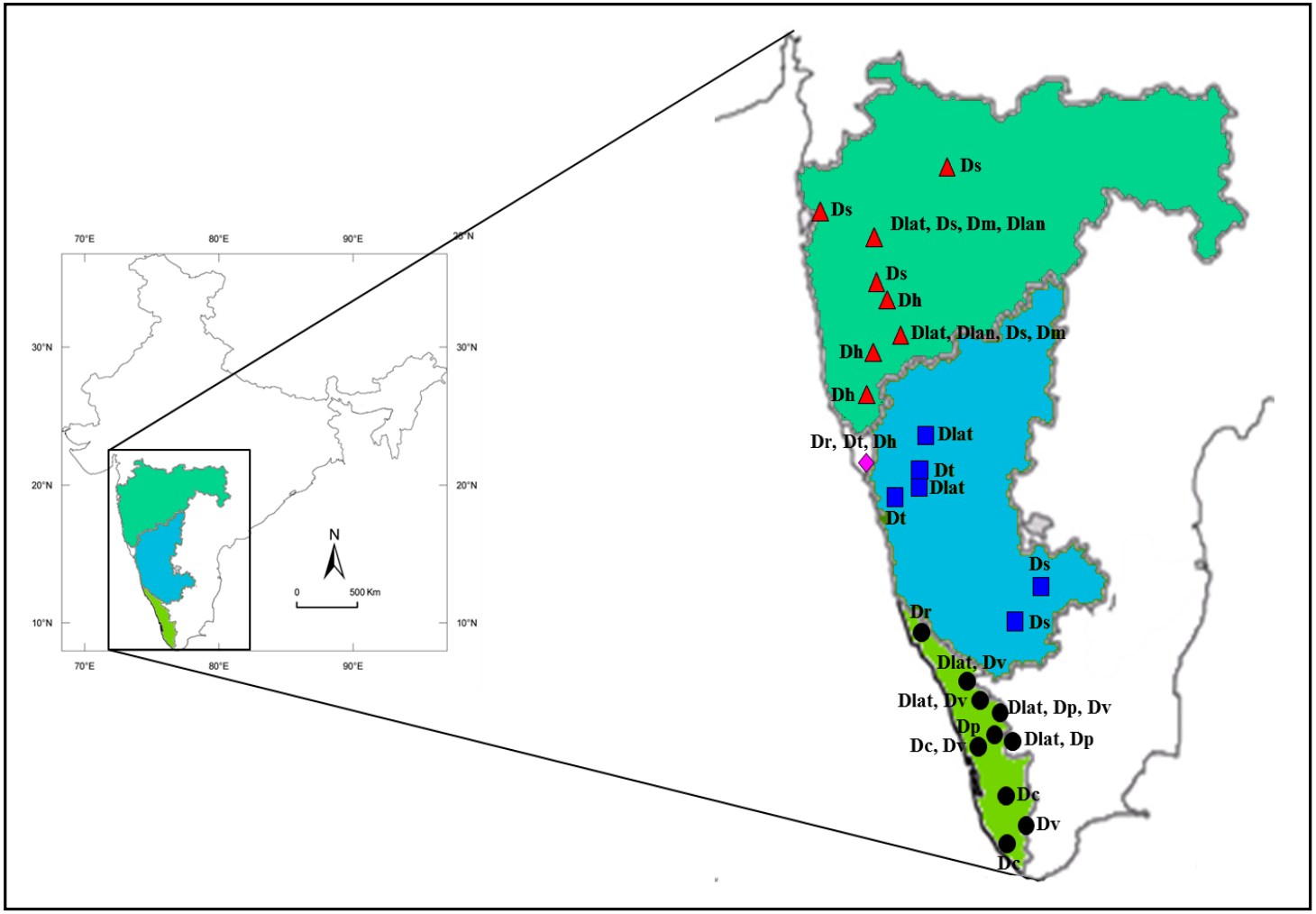


Figure S1.2: Map of India. Map showing the sites of collection from Western Ghats, India

DNA extraction

Total genomic DNA was extracted from fresh or dried leaf samples using the modified CTAB method (Richards et al. 1994). After collection, leaves were cleaned and stored at -80°C. One gm of leaf tissue was crushed to fine powder using liquid nitrogen and transferred to polypropylene tube containing 15 mL pre-warmed extraction buffer (2% CTAB, 1.4 M NaCl, 100 mM tris with pH8.0, 20 mM EDTA, 100 mM sodium metabisulphite). The mixture was swirled gently and incubated at 65°C for 45 min-1 hr with intermittent shaking. After incubation tube was allowed to cool and then equal amount of chloroform: isoamyl alcohol (24:1) was added. The tube was shaken moderately, thoroughly mixed and centrifuged at 10000 rpm for 10 min at room temperature (RT). The aqueous layer was transferred to fresh polypropylene tube. CTAB ($1/10^{th}$ volume of 10%) was added to aqueous layer followed by equal volume chloroform: isoamyl alcohol. The mixture was swirled gently. The centrifugation was carried out at 10000 rpm for 10 min. The aqueous layer was removed carefully and transferred to new polypropylene tube followed by addition of $1/4^{th}$ volume of precipitation buffer (1% CTAB, 50 mM Tris, 10mM EDTA, 1% β-Mercaptoethanol, pH 8.0) and equal volume of sterile deionized water. After thorough mixing, the centrifugation was carried out at 10000 rpm at RT. The solution was decanted carefully and pellet was dissolved in High Salt Tris EDTA buffer (HSTE) (10mM Tris, 0.5mM EDTA, 1M NaCl, pH 8.0) as the plants contain high amount of polysaccharides. After dissolution the DNA was precipitated with double the volume of chilled ethanol. The tube was centrifuged at 4°C for 10000 rpm and 10 min. The pellet was then washed with 70% ethanol. The pellet was kept for air drying at RT. The dry pellet was then dissolved in Tris EDTA buffer (TE). RNase treatment was given to DNA and then it was quantified spectrophotometrically at 260/280 nm. The purity of the DNA was determined by calculating the ratio of absorbance at 260 and 280 nm and A260/A230 on the Nanodrop 1000 (Thermo Scientific, USA) as well as on 0.8% agarose gel and compared it with commercial samples of known DNA concentrations (Sambrook et al. 1989).

PCR amplification and sequencing

Since no specific region was suggested as universal plant barcode, we used various gene regions such as *matK* (7 primer pairs), *rpoC* (4 primer pairs), *rpoB* (5 primer pairs), *accD* (6 primer pairs), *ndhJ* (3 primer pairs), *ycf5* (4 primer pairs), *trnH-psbA* (5 primer pairs), nrITS (2 primer pairs) and *rbcL* (single primer pair) based on the available literature on plant barcodes (http://www.kew.org/barcoding/update.html; (Kress et al. 2005; Sass et al.

2007; Fazekas et al. 2008; Chen et al. 2010) (Table S1.2). These 37 primer pairs were tested to identify loci satisfying the criteria set for DNA barcoding. Four primer pairs (Table S1.3) corresponding to *matK*, *rbcL*, *trnH-psbA* and nrITS produced sharp bands with specific amplifications and produced good quality DNA sequences. Moreover, these were also recommended by the Consortium for the Barcode of Life (CBOL) as potential barcodes, hence these were used for further study. PCR amplifications were performed in final volume of 20 or 25µL (Table S1.4), and the amplicons were resolved on 1% agarose gels in 0.5X TAE buffer.

Sequencing was performed using Sanger chemistry in both the directions using MegaBACE DYEnamic ET dye terminator kit with MegaBACE 1000 DNA Analysis System (GE Healthcare, USA). Most of the samples produced sharp single bands, where in the PCR products were directly used as a template for sequencing reaction. Very few samples in addition generated nonspecific bands. In such cases the bands with expected size were eluted using PureLink® Quick Gel Extraction Kit (Invitrogen, USA) and used.

The sequencing reaction mixture contained 6 µL premix, 1 µL PCR product (75-100 ng), 1 µL primer (10 µM) and deionized water making the final volume of 10 µL. Forward and reverse reactions were carried out separately. The sequencing PCR was carried out in a PTC 200 thermal cycler (MJ Research, USA). The cycling conditions included initial denaturation of 30 sec. at 94ºC, followed by 34 cycles of 30 sec at 94ºC and combined primer annealing and extension step of 1.45 min at 60ºC. The cleaning up of sequencing products was further carried out. For 10 µL reaction, 1.0 µL of 7.5 M ammonium acetate was added followed by 1 µL glycogen. After each addition, brief spin was taken at 1000 rpm for 10 sec. To this, absolute ethanol (2.5 times of total volume) was added. The reaction was mixed well by thorough pipetting and incubated at -20°C for 1 hr. After incubation, centrifugation was carried out at 3700 rpm for 45 min to 1 hr and then the plate was decanted on a tissue paper and centrifuged followed by invert spin at 200 rpm for 10 sec. Ethanol wash (150 µL of 70%) was then given to the pellet. The pellets were air-dried. 10 µL of loading solution (formamide based) was added to each well and brief spin was taken at <2000 rpm for 10 sec. All these procedures were carried out in dark. The plate was stored at -20°C till sequencing was commenced.

Sequence analysis

The chromatograms obtained were base called using Phred software provided by the company (Amersham Biosciences, USA). For each sequence, the chromatograms were inspected and poor quality 5′ and 3′ DNA sequence ends were trimmed. Post trimming

lengths were maintained at least 60% of the original read length, subject to the minimum average quality score of Q20.

Table S1.2: Primers screened for *Dalbergia* species

Primers			
Sr No.		**Sr No.**	
1	*matK* x f + *matK* 5 r	20	*accD* LP1 f + *accD* LP4 r
2	*matk* 2.1 a f + *matK* 5 r	21	*accD*2 f+ *acc D*3 r
3	*matk* f + *mat k* r (Equi)	22	*ndhJ* 1 f + *ndhJ* 3 r
4	*matk* 2.1 a(f) + *matK* 3.2 (r)	23	*ndhJ* 1 f + *ndhJ* 4 r
5	*matK* X f + *matK* 3.2 r	24	*ndhJ* LP1 f + *ndhJ* LP4 r
6	*matK* f (ad) + *matK* r (eq)	25	*Ycf* 5 1 f + *Ycf* 5 3 r
7	*rpoC*1(1) f + *rpoC*1(3) r	26	*Ycf* 5 2 f + *Ycf* 5 4 r
8	*rpoC*1(1) f + *rpoC*1(4) r	27	*Ycf* 5 1 f + *Ycf*5 4 r
9	*rpoC*1(2) f + *rpoC*1(4) r	28	*Ycf*5 2 f + *Ycf* 5 3 r
10	*rpoC*1(LP1) f + *rpoC1*(LP5) r	29	*psbA* - *trnH* f + cycadales
11	*rpoB*1 f + *rpoB*3 r	30	*psbA* – *trnH* (protein coding region) + short frag r
12	*rpoB*2 f + *rpoB*3 r	31	*psbA* – *trnH* (protein coding region) + long frag r
13	*rpoB*2 f + *rpoB* (LP3) r	32	nr ITS 5a f + nr ITS 4 r
14	*rpoB* LP1.1 f + *rpoB* LP4.3 r	33	nr ITS 5a f + nr ITS 2c r
15	*rpoB* LP1.1 f + *rpoB* LP5.2r	34	*trnH*(GUG) f + *psbA* r
16	*accD*1 f+ *accD*3 r	35	*matK* F+ *matK* R (modified)
17	*accD*1 f+ *accD*4 r	36	*trnH-psbaA* F + *trnH-psbA* R (modified)
18	*accD*1 f+ *accD* LP3 r	37	*rbcL* a + *rbcL* ajf634
19	*accD*2 f + *accD*4 r		

Table S1.3: Primers used for DNA barcoding of *Dalbergia* species

Locus	Primer name	Primer sequence (5'→3')	Approximate size in genus *Dalbergia*	Reference
matK	matK2.1 a F	ATC CAT CTG GAA ATC TTA GTT C	900 bp	Royal Botanic Gardens, Kew
	matK3.2 R	CTT CCT CTG TAA AGA ATT C		
rbcL	rbcLa F	ATGTCACCACAAACA GAGACTAAA GC	650 bp	Fazekas et al. (2008)
	rbcLajf6 34 R	GAA ACG GTC TCT CCA ACG CAT		
trnH-psbA	trnH-psbA F	CGC GCA TGG TGG ATT CAC AAT CC	300 bp	Kress et al. (2005)
	trnH-psbA R	GTT ATG CAT GAA CGT AAT GCT C		
nrITS	nrITS 5aF	CCT TAT CAT TTA GAG GAA GGA G	700 bp	Chen et al. (2010)
	nrITS 4R	TCC TCC GCT TAT TGA TAT GC		

The sequences failing this criterion were rejected and re-sequenced. All the nucleotide variations were evaluated and confirmed by inspecting the chromatograms of forward and reverse strands. Forward and reverse sequences with 70% or more overlap were considered to create consensus sequence for each amplicon (Tripathi et al. 2013). Good quality sequences which satisfied the above mentioned criteria for all the individuals were assembled and aligned using CLUSTALW 1.83 (Thompson et al. 1994). Conserved, variable and parsimony informative sites were determined using MEGA 5.0 (Tamura et al. 2011). Distance matrices and Neighbor-Joining (NJ) trees were calculated and built in MEGA using the best fit nucleotide substitution model [chosen with Akaike information criterion (AICc)] (Akaike 1974). Nucleotide sequences of analysed loci from all the individuals were deposited in NCBI database (Accession numbers - *matK*: KM276475-KM276412; *rbcL*: KM100059-KM099987; *trnH-psbA*: KM276322-KM276250 and nrITS : KM276165-KM276104). All the sequences submitted to NCBI are listed in Table S1.5 (**Table is provided in a compact disc [CD] alongwith this dissertation**).

Table S1.4: PCR conditions of shortlisted loci

	Effective concentration of various components									**Cycling conditions**	
Locus	**DNA (ng)**	**dNTPs (mM)**	**Buffer**	**$MgCl_2$ (mM)**	**Forward primer (Pmoles)**	**Reverse primer (Pmoles)**	**DMSO (%)**	***Taq* DNA polymerase (units)**	**Total reaction volume (µL)**	**Details of steps**	**No. of cycles**
rbcL	50-60	0.2	10X	1	0.16	0.16	4	0.048	20	a- 95°C, 2m b- 95°C, 30s c- 64°C, 45s d- 72°C, 50s e- 72°C, 5m	35
trnH-psbA	60-70	0.2	10X	1	0.16	0.16	-	0.048	25	a- 94°C, 5m b- 94°C, 1m c- 65°C, 30s d- 72°C, 1m e- 72°C, 5m	
nrITS	60-70	0.2	10X	0.5	0.16	0.16	4	0.144	25	a- 94°C, 5m b- 94°C, 1m c- 56°C, 30s d- 72°C, 1m e- 72°C, 5m	
matK	50-60	0.2	10X	1	0.12	0.12	-	0.048 (AccuTaq LA DNA polymerase)	25	a- 94°C, 3m b- 94°C, 30s c- 48°C, 30s d- 68°C, 1m e- 68°C, 2m	

a- initial denaturation, b- cycle denaturation, c- annealing temperature, d- extension, e- final extension

Data analysis

Genetic distances were calculated using Kimura-2-Parameter (K2P) model (Kimura 1980). The interspecific divergence was studied among the species by three parameters *viz.* (i) average inter specific distance; (ii) average theta prime (θ'), where θ' is the mean pairwise distance within species, thus eliminating the biases associated with different numbers of individuals among species and (iii) minimum inter specific distance. Three additional parameters were studied for the intraspecific divergence: (i) average intraspecific divergence, (ii) theta (θ) and (iii) average coalescent depth (Chen et al. 2013).

Wilcoxon signed rank tests were performed to check existence of significant divergence between inter and intra specific variability among the pairs of barcoding loci (Kress and Erickson 2007). Consensus sequences were generated for all the ten *Dalbergia* species using TaxonDNA (Meier et al. 2006) with 1000 bootstraps. To analyze inter and intraspecific variation, sequence variants were generated with DnaSP 5.0 (Rozas et al. 2003) using consensus sequences and further NJ trees were constructed in MEGA 5.0 using 1000 bootstraps. Based on the distance method using K2P parameter and a minimum sequence overlap of 300 bp, accurate species identification was carried out by TaxonDNA or SpeciesIdentifier 1.7.7 software (Meier et al. 2006) using two approaches, (i) Best match (BM) and (ii) Best close match (BCM). In these approaches, each sequence from the dataset was used as a query against the remaining sequences from the same dataset. With BM, a query sequence was identified by the reference sequence by searching the best match having smallest genetic distance to the query. The BCM approach requires a threshold value which was calculated for each locus from pairwise summary. The threshold was a value below which 95% of all intraspecific distances were observed, leading to an upper bound value on the similarity of a barcode match (Meier et al. 2006). If both the query and subject sequences were from the same species, the identification was considered as successful. Whereas, if more than one query sequences from different species exhibited equally good match, then the samples were considered as ambiguous. Another character based analysis method, Barcoding with LOGic (BLOG), was also used (Weitschek et al. 2013). The method selects the unique nucleotide position of the sequence and derives the formula to differentiate among species. The advantage of this method is that it provides concise and meaningful classification rules (Bertolazzi et al. 2009).

S 1.2.2 Validation of DNA barcodes using NCBI sequences and unauthenticated samples

The DNA sequences of various *Dalbergia* species corresponding to the *matK, rbcL* and nrITS regions in addition to our own submission were downloaded from the NCBI database and used to validate the proposed regions (Table S1.6). Only those having multiple accessions of each species (minimum 2) and read length matching to the sequences generated for development of barcodes were downloaded. The analysis was carried out with all those parameters used for barcode development except WSR. Simultaneously, the few unauthenticated individuals of *Dalbergia* and a single accession of *D. sissoides* collected from the WG region and few more *D. latifolia* samples (authenticated) were also included in the analysis. These samples were later authenticated by BSI.

Table S1.6: Sample details of sequences downloaded from NCBI

Locus	Species for which sequences downloaded from NCBI
matK	**D. sissoo, *D. melanoxylon, D. candenatensis, D. paniculata, D. rubiginosa, D. volubilis, D. lanceolaria, D. latifolia, D. horrida, D. tamarindifolia, D. armata, D. nigrescens, D. cochinchinensis, D. thorelii, D. rimosa, D. retusa*
rbcL+matK	**D. sissoo, *D. melanoxylon, D. candenatensis, D. paniculata, D. rubiginosa, D. volubilis, D. lanceolaria, D. latifolia, D. horrida, D. tamarindifolia, D. thorelii, D. retusa*
nrITS	**D. sissoo, *D. melanoxylon, *D. candenatensis, D. paniculata, D. rubiginosa, D. volubilis, *D. lanceolaria, *D. latifolia, *D. horrida, D. tamarindifolia, D. pinnata, D. stipulacea, D. frutescens, D. hancei, D. tonkinensis, D. oliveri, D. nigrescens, D. cochinchinensis, D. assamica*

*- Species common to NCL and NCBI

S 1.2.3 Phylogenetic analysis

DNA sequences used for barcode development were also subjected to phylogenetic analysis (Table S1.7). The sequences were further fine-tuned using Aliview program v1.1 (Larsson 2014). Phylogenetic analysis was performed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003). Seaview drives the PhyML v3.1 to compute maximum likelihood (ML) trees (Gouy et al. 2010). The *matK, rbcL, trnH-psbA* and nrITS were

first analyzed separately and later combined analysis was performed. For Bayesian analysis the best fit nucleotide substitution model was found out using jmodel test v2.1.7 (Posada 2008). The best model was calculated under AIC (Akaike 1974) and used in the analysis. Bayesian posterior probabilities were obtained using Markov Chain Monte Carlo (MCMC) analysis. Two independent MCMC analyses with four chains were carried out for ten million generations and trees were sampled every 1000th generation. Trees were viewed in Treeview v1.6.6 (Page 2001).

Table S1.7: Sample details of phylogenetic analysis

Species name	**No. of accessions**	**Continent**
D. abbreviata	1	Asia
D. abrahami	1	Africa
D. acariiantha	1	Africa
D. adamii	1	Africa
D. afzeliana	1	Africa
D. arbutifolia	1	Africa
D. armata	1	Africa
D. assamica	6	Asia
D. aurea	1	Africa
D. balansae	1	Asia
D. baronii	1	Africa
D. benthamii	1	Asia
D. bignonae	1	Africa
D. bintuluensis	1	Asia
D. boehmii	1	Africa
D. bojeri	1	Africa
D. bracteolata	1	Africa
D. candenatensis	4	Asia, Australia
D. canescens	1	Asia
D. capuronii	1	Africa
D. cearensis	1	America
D. cochinchinensis	4	Asia
D. cultrata	1	Asia
D. dialoides	1	Asia

D. dongnaiensis	1	Asia
D. ecastaphyllim	1	America
D. entadoides	1	Asia
D. falcata	1	Asia
D. floribunda	1	Asia
D. frutescens var frutescens	1	America
D. frutescens var tomentosa	1	America
D. glomerata	1	America
D. godefroyi	1	Asia
D. greveana	1	Africa
D. hancei	2	Asia
D. havilandii	1	Asia
D. horrida	6	Asia
D. hostilis	1	Africa
D. humbertii	1	Africa
D. hupeana	2	Asia
D. inundata	1	America
D. junghuhnii	1	Asia
D. kurzii	1	Asia
D. lactea	1	Africa
D. lakhonensis	1	Asia
D. lanceolaria	3	Asia
D. lateriflora	1	America
D. latifolia	8	Asia
D. maritima	1	Africa
D. martinii	1	Africa
D. melanocardium	1	America
D. melanoxylon	8	Asia, Africa
D. miscolobium	1	America
D. monetaria	1	America
D. multijuga	1	Africa
D. nigrescens	3	Asia
D. obovata	1	Africa
D. odorifera	2	Asia

D. oliveri	4	Asia
D. parviflora	1	Asia
D. pinnata	2	Asia
D. revoluta	1	Asia
D. retusa	3	Not known
D. rimosa	1	Asia
D. rimosa var. foliacea	1	Asia
D. rostrata	1	Asia
D. sandakanensis	1	Asia
D. sericea	1	Asia
D. sissoo	7	Asia, Africa
D. stipulacea	1	Asia
D. spruceana	1	America
D. tamarindifolia	3	Asia
D. thorelii	2	Asia
D. tonkinensis	4	Asia
D. trichocarpa	1	Africa
D. velutina	1	Asia
D. paniculata	3	Africa
D. rubiginosa	3	Africa
D. volubilis	2	Africa

S 1.3 Results

S 1.3.1 Identifying a potential barcode for 10 *Dalbergia* species

Between 5 and 25 accessions of each species were collected from different geographical locations. Multiple accessions of each species were collected to understand the effect of geographical isolation on intraspecific variation in barcoding. The collections were planned based on the Botanical Survey of India records. However, some records were very old and no species were found to be distributed there.

Amplification success

The success rate for PCR amplification and sequencing of bidirectional reads was the highest for *rbcL* (97.6%), followed by *matK* (97.0%) and *trnH-psbA* (94.7%), while nrITS

exhibited the lowest rate (80.5%). Using BLAST analysis, all the loci correctly identified 100% of the samples at genus level; while at species level, nrITS had the highest identification rate i.e. 60% followed by *rbcL* (50%), *matK* (20%) and *trnH-psbA* (10%). The low rate of species level identification might be due to the absence of species records in NCBI database and high percentage of in-dels especially in the case of *trnH-psbA* sequences.

Nucleotide variation

The percentages of polymorphic informative (Pi) sites and variable sites were comparable for the respective loci. For nrITS, aligned length was 637 bp, with 29.83% variable sites and 28.89% polymorphic informative sites, which was the highest among all the loci (single locus as well as combination of loci). Based on the percentage of conserved sites, the most conserved loci were *rbcL* followed by *matK* and *matK*+*rbcL* (Table S1.8).

Table S1.8: Summary statistics for potential barcode loci from ten *Dalbergia* species

Locus	***matK***	***rbcL***	***trnH-psbA***	**nrITS**	***matK*+ *trnH-psbA***	***matK*+ *rbcL***	***rbcL*+ *trnH-psbA***
No. of sequences analysed	165	166	161	137	157	163	157
Total no. of sites	677	491	273	637	950	1168	764
Conserved sites	636 (93.94)	477 (97.15)	250 (91.58)	447 (70.17)	863 (90.84)	1113 (95.29)	724 (94.76)
Variable sites	41 (6.06)	14 (2.85)	23 (8.42)	190 (29.83)	87 (9.16)	55 (4.71)	40 (5.24)
Parsimony informative sites	40 (5.91)	14 (2.85)	15 (5.49)	184 (28.89)	55 (5.79)	54 (4.62)	29 (3.80)

The values in parentheses are expressed in percentage

Distance analysis and Wilcoxon signed rank test

The nrITS locus showed greater interspecific divergence than the plastid loci (*matK*, *rbcL* and *trnH-psbA* and their combinations) using both average inter specific distance and θ' parameters. However, in case of intraspecific divergence, nrITS and *rbcL* showed the highest and the lowest value, respectively. Thus, no single locus revealed the highest interspecific but the lowest intraspecific divergence (Table S1.9 and Figure S1.3). When the Wilcoxon signed rank test was used to compare the loci, nrITS exhibited the highest

interspecific divergence followed by *trnH-psbA*, whereas *rbcL* displayed the lowest intraspecific divergence (Tables S1.10A and S1.10B).

Table S1.9: Inter and intraspecific divergence values for potential barcode loci

Distance parameters	***matK***	***rbcL***	***trnH-psbA***	**nrITS**	***matK+ trnH-psbA***	***matK+ rbcL***	***trnH-psbA+ rbcL***
Average interspecific distance	0.014± 5.74E-05	0.007± 3.61E-05	0.017± 1.11E-4	0.140± 6.45E-4	0.015± 6.45E-05	0.011± 4.55E-05	0.010± 5.03E-05
Theta prime (θ')	0.015± 7.43E-4	0.008± 5.6E-4	0.018± 1.677E-3	0.114± 3.656E-3	0.016± 8.2E-4	0.012± 5.98E-4	0.011± 7.07E-4
Smallest inter specific distance	0.014± 7.12E-4	0.008± 5.43E-3	0.017± 1.627E-3	0.156± 7.922E-3	0.015± 8.17E-4	0.011± 5.66E-4	0.011± 6.83E-4
Average intraspecific distance	0.001± 4.50E-05	0.000± 1.52E-05	0.000± 3.52E-05	0.004± 5.48E-4	0.001± 3.77E-05	0.001± 2.78E-05	0.000± 1.50E-05
Theta (θ)	0.000± 2.96E-4	0.000± 1.03E-4	0.001± 3.85E-4	0.003± 2.112E-3	0.000± 2.61E-4	0.000± 1.77E-4	0.000± 1.29E-4

Barcode gap

Barcode gap represents the absence of overlapping regions between inter and intraspecific distances. The barcode gap was absent for all the marker loci used in the present study, indicating overlaps between inter and intraspecific distances (Figure S1.4). However, the mean interspecific divergence was significantly higher than that of the corresponding intraspecific divergence for each of the loci. This was further confirmed by analysis carried out using TaxonDNA.

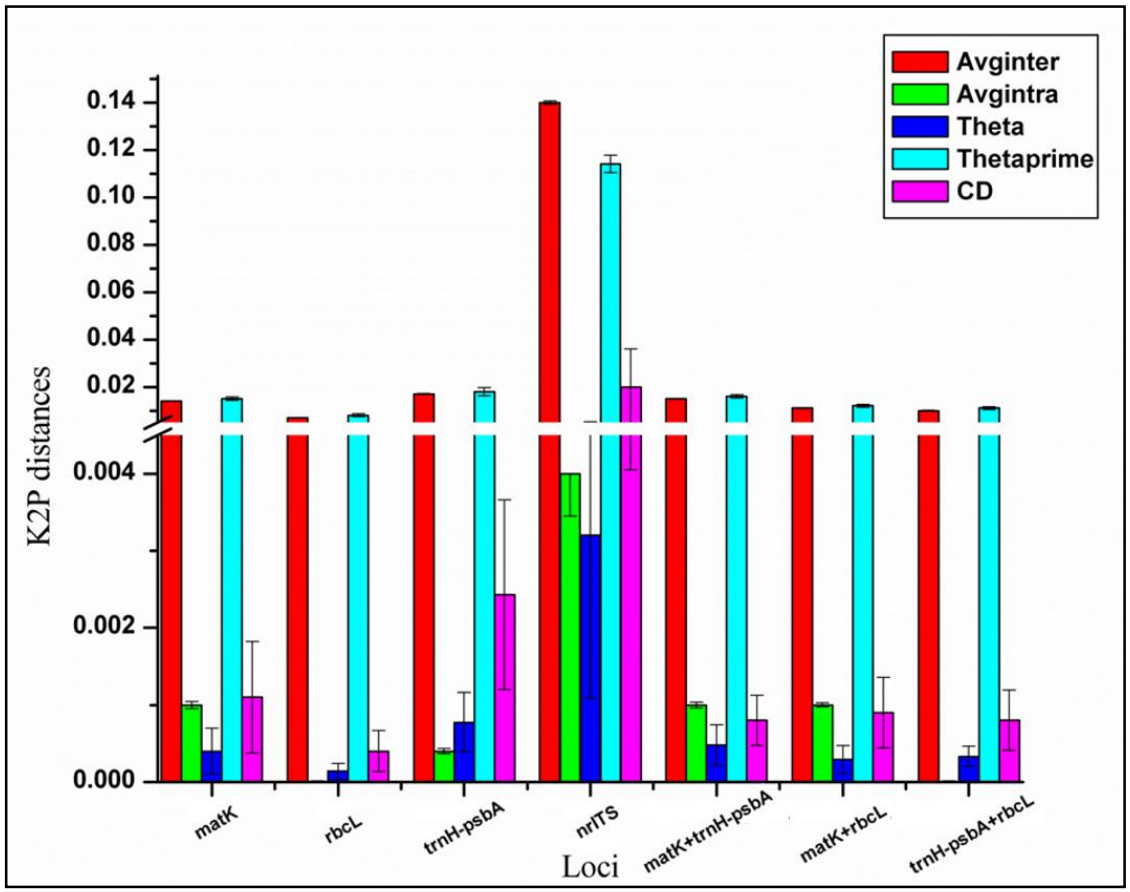


Figure S1.3: Distribution of inter and intraspecific divergence. The plot depicts inter and intraspecific divergence parameters for various loci. Avginter: Average inter specific distance, Avgintra: Average intraspecific distance, Theta, Theta prime, CD: coalescence depth.

Table S1.10A: Wilcoxon signed-rank tests results for interspecific divergence of the indicated loci

W+	W-	Inter relative ranks	Results
matK	*rbcL*	W+=85063, W-=15,n=412, p=0	*matK*>*rbcL*
matK	*matK + rbcL*	W+=73114, W-=39,n=412, p=0	*matK*>*matK+rbcL*
matK	*rbcL+trnH-psbA*	W+=68612, W-=6078, n=412, p=0	*matK*>*rbcL+trnH-psbA*
matK	*trnH-psbA*	W+=29484, W-=46370,n=412, p=0	*trnH-psbA*>*matK*
matK	nrITS	W+=1, W-=85077,n=412, p=0	nrITS>*matK*
matK	*matK+trnH-psbA*	W+=27245, W-=41019, n=412, p=0.001	*matK+trnH-psbA*>*matK*
rbcL	nrITS	W+=0, W-=85078,n=412, p=0	nrITS>*rbcL*
rbcL	*trnH-psbA*	W+=5996,W-=76625,n=412, p=0	*trnH-psbA*>*rbcL*
rbcL	*matK + rbcL*	W+=0, W-=84255, n=412, p=0	*matK+rbcL*>*rbcL*
rbcL	*matK+trnH-psbA*	W+=0, W-=84666, n=412, p=0	*matK+trnH-psbA*>*rbcL*
rbcL	*rbcL+trnH-psbA*	W+=1924, W-=60204, n=412, p=0	*rbcL+trnH-psbA*>*rbcL*
trnH-psbA	*matK+rbcL*	W+=63125, W-=17476.5, n=412, p=0	*trnH-psbA*>*matK+rbcL*
trnH-psbA	*matK+trnH-psbA*	W+=47083.5, W-=28771.50, n=412, p=0	*trnH-psbA*>*matK+trnH-psbA*
trnH-psbA	*rbcL+trnH-psbA*	W+=73380.5, W-=6020.5, n=412, p=0	*trnH-psbA*>*rbcL+trnH-psbA*
trnH-psbA	nrITS	W+=0, W-=85078,n=412, p=0	nrITS>*trnH-psbA*
nrITS	*matK+rbcL*	W+=85077, W-=1,n=412, p=0	nrITS>*matK+rbcL*
nrITS	*matK+trnH-psbA*	W+=85077, W-=1,n=412, p=0	nrITS>*matK+trnH-psbA*
nrITS	*rbcL+trnH-psbA*	W+=85078, W-=0,n=412, p=0	nrITS>*rbcL+trnH-psbA*
matK+ rbcL	*matK+trnH-psbA*	W+=37, W-=73116,n=412, p=0	*matK+trnH-psbA*>*matK+rbcL*
matK+ rbcL	*rbcL+trnH-psbA*	W+=49627, W-=22004,n=412, p=0	*matK+rbcL*>*rbcL +trnH-psbA*
matK+trnH-psbA	*rbcL+trnH-psbA*	W+=79759, W-=41,n=412, p=0	*matK+trnH-psbA*>*rbcL+trnH-psbA*

Table S1.10B: Wilcoxon signed-rank test results for intraspecific divergence of the indicated loci

W+	W-	Inter relative ranks	Results
matK	*rbcL*	W+=141, W-=12,n=53, p=0.002	*matK*>*rbcL*
matK	*rbcL+trnH-psbA*	W+=181.50, W-=49.50,n=53, p=0.020	*matK*>*rbcL+trnH-psbA*
matK	*matK + rbcL*	W+=18.50, W-=2.50,n=53, p=0.084	*matK*=*matK+rbcL*
matK	*matK+trnH-psbA*	W+=66, W-=39,n=53, p=0.369	*matK*=*matK+trnH-psbA*
matK	*trnH-psbA*	W+=135.5, W-=74.5,n=53, p=0.250	*matK= trnH-psbA*
matK	nrITS	W+=102.50, W-=932.50,n=53, p=0	nrITS>*matK*
rbcL	nrITS	W+=10, W-=1071,n=53, p=0	nrITS>*rbcL*
rbcL	*trnH-psbA*	W+=28, W-=63,n=53, p=0.212	*rbcL*=*trnH-psbA*
rbcL	*rbcL+trnH-psbA*	W+=42, W-=49,n=53, p=0.793	*rbcL+trnH-psbA=rbcL*
rbcL	*matK + rbcL*	W+=6.50, W-=146.50,n=53, p=0.001	*matK+rbcL*>*rbcL*
rbcL	*matK+trnH-psbA*	W+=8.5, W-=111.5,n=53, p=0.003	*matK+trnH-psbA*>*rbcL*
trnH-psbA	nrITS	W+=151.5, W-=1024.5,n=53, p=0	nrITS>*trnH-psbA*
trnH-psbA	*matK+rbcL*	W+=80.5, W-=129.5,n=18, p=0.356	*trnH-psbA=matK+rbcL*
trnH-psbA	*matK+trnH-psbA*	W+=74.5, W-=135.5,n=53, p=0.250	*trnH-psbA =matK+trnH-psbA*
trnH-psbA	*rbcL+trnH-psbA*	W+=63, W-=28,n=53, p=0.212	*trnH-psbA=rbcL+trnH-psbA*
nrITS	*matK+rbcL*	W+=1034, W-=47,n=53, p=0	nrITS>*matK+rbcL*
nrITS	*matK+trnH-psbA*	W+=1122.5, W-=102.5,n=53, p=0	nrITS>*matK+trnH-psbA*
nrITS	*rbcL+trnH-psbA*	W+=1209, W-=16,n=53, p=0	nrITS>*rbcL+trnH-psbA*
matK+rbcL	*matK+trnH-psbA*	W+=42, W-=63,n=53, p=0.485	*matK+trnH-psbA=matK+rbcL*
matK+rbcL	*rbcL+trnH-psbA*	W+=163, W-=47,n=53, p=0.028	*matK+rbcL>rbcL+trnH-psbA*
matK+trnH-psbA	*rbcL+trnH-psbA*	W+=167.50, W-=22.50,n=53, p=0.002	*matK+trnH-psbA>rbcL+trnH-psbA*

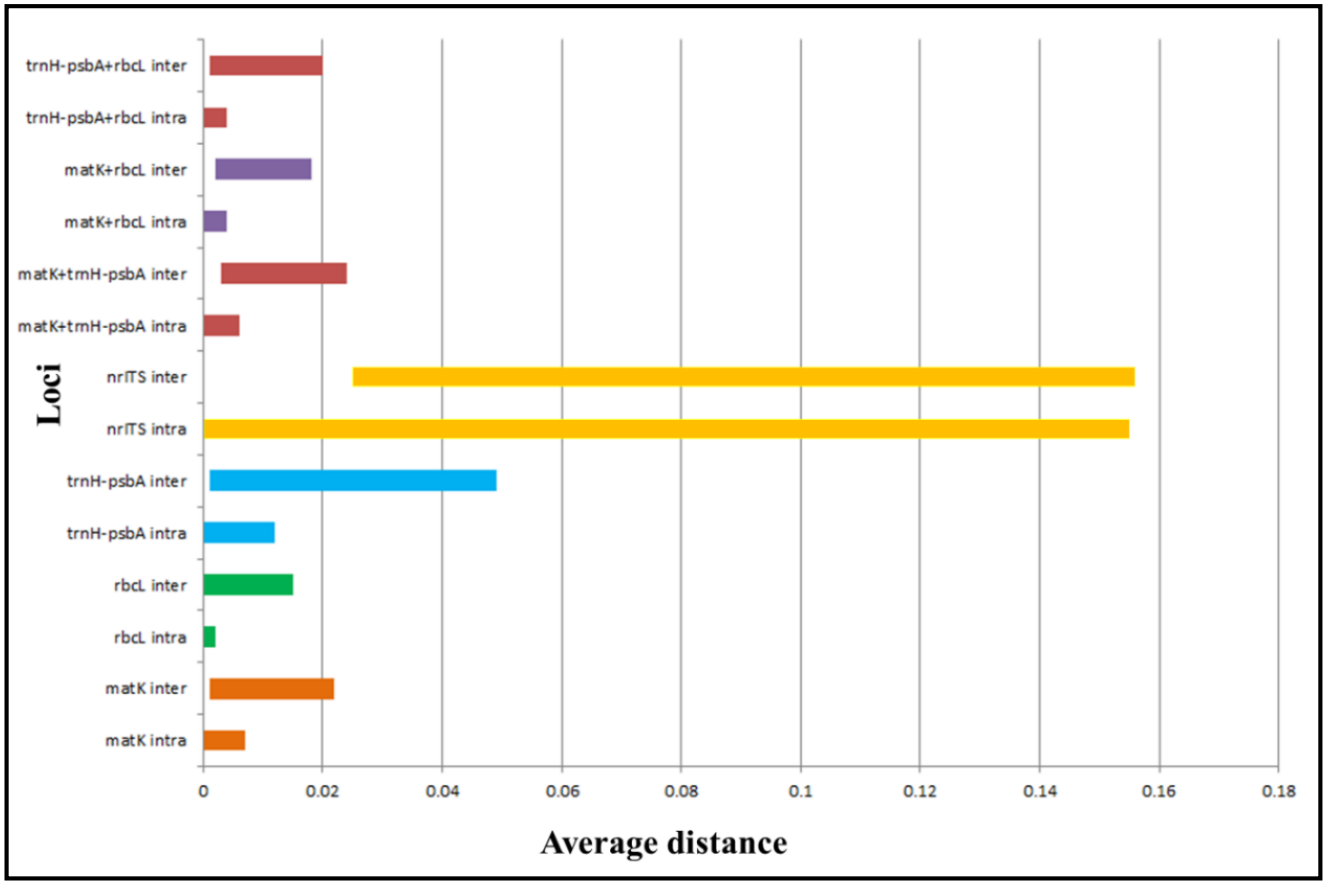


Figure S1.4: The barcoding gap. Graph of the smallest interspecific and the largest intraspecific distances highlighting the overlapping divergence

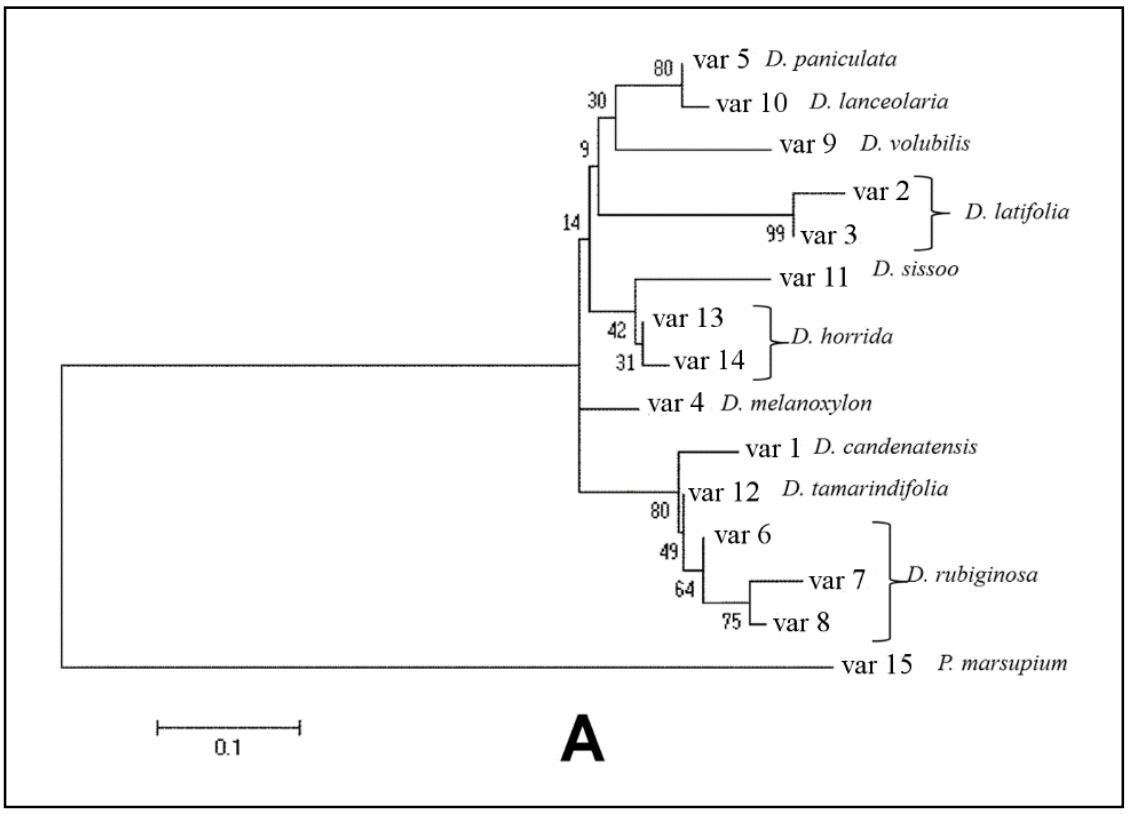
Tree based analyses

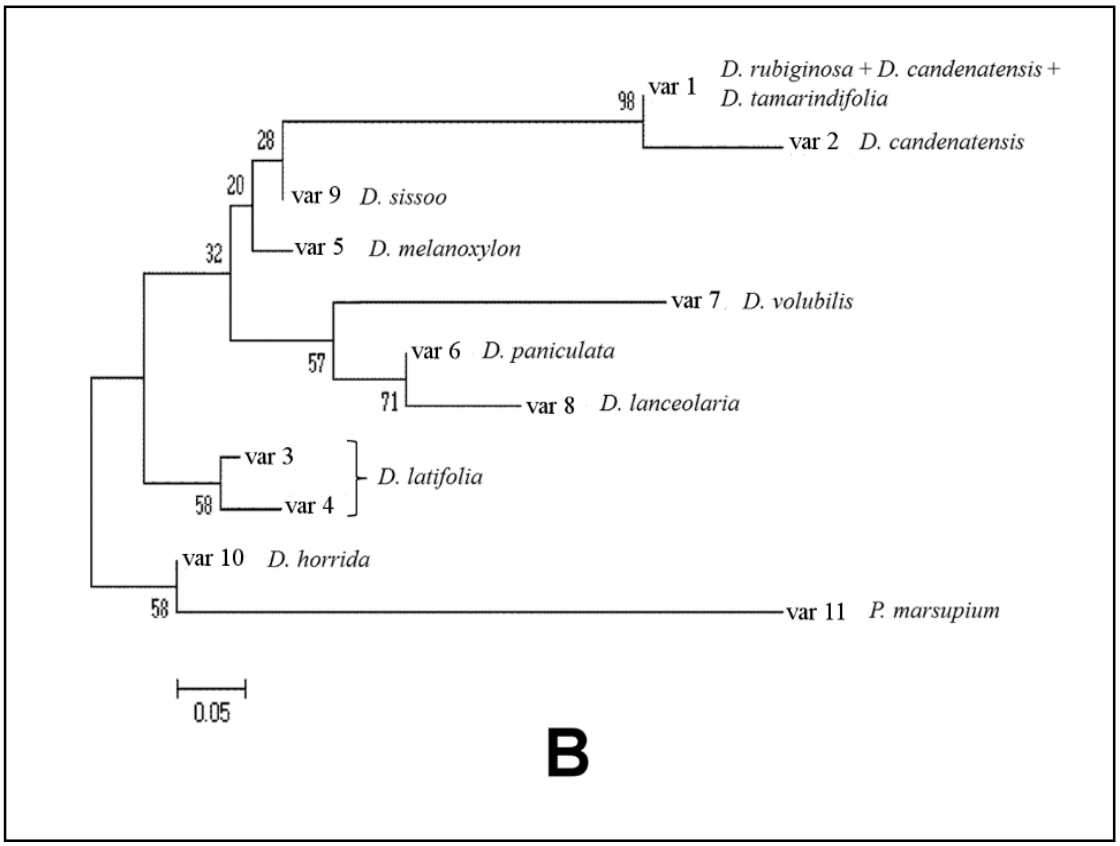
The sequence variants of each marker locus were determined using DnaSP 5.0 and MEGA 5.0 as mentioned previously. Among all loci, nrITS exhibited the maximum number of sequence variants (Table S1.11). By including all the sequence variants, seven NJ trees were constructed with *matK, rbcL, trnH-psbA* and nrITS either alone or in combinations (Figure S1.5). All of them except *rbcL* revealed a separate cluster for each species and *rbcL* could not differentiate between *D. rubiginosa*, *D. candenatensis* and *D. tamarindifolia*. Interestingly, except *trnH-psbA* all other loci (*matK*, *rbcL*, nrITS and *matK*+*rbcL*) either alone or in combination were capable of grouping together all three species-clusters from the section *Dalbergia* (*D. volubilis*, *D. lanceolaria* and *D. paniculata*). These observations indicated that *matK*, nrITS, *rbcL* and *matK*+*rbcL* could correctly identify the reported relationships among the *Dalbergia* species and hence, they could most likely be successful as barcodes for this genus.

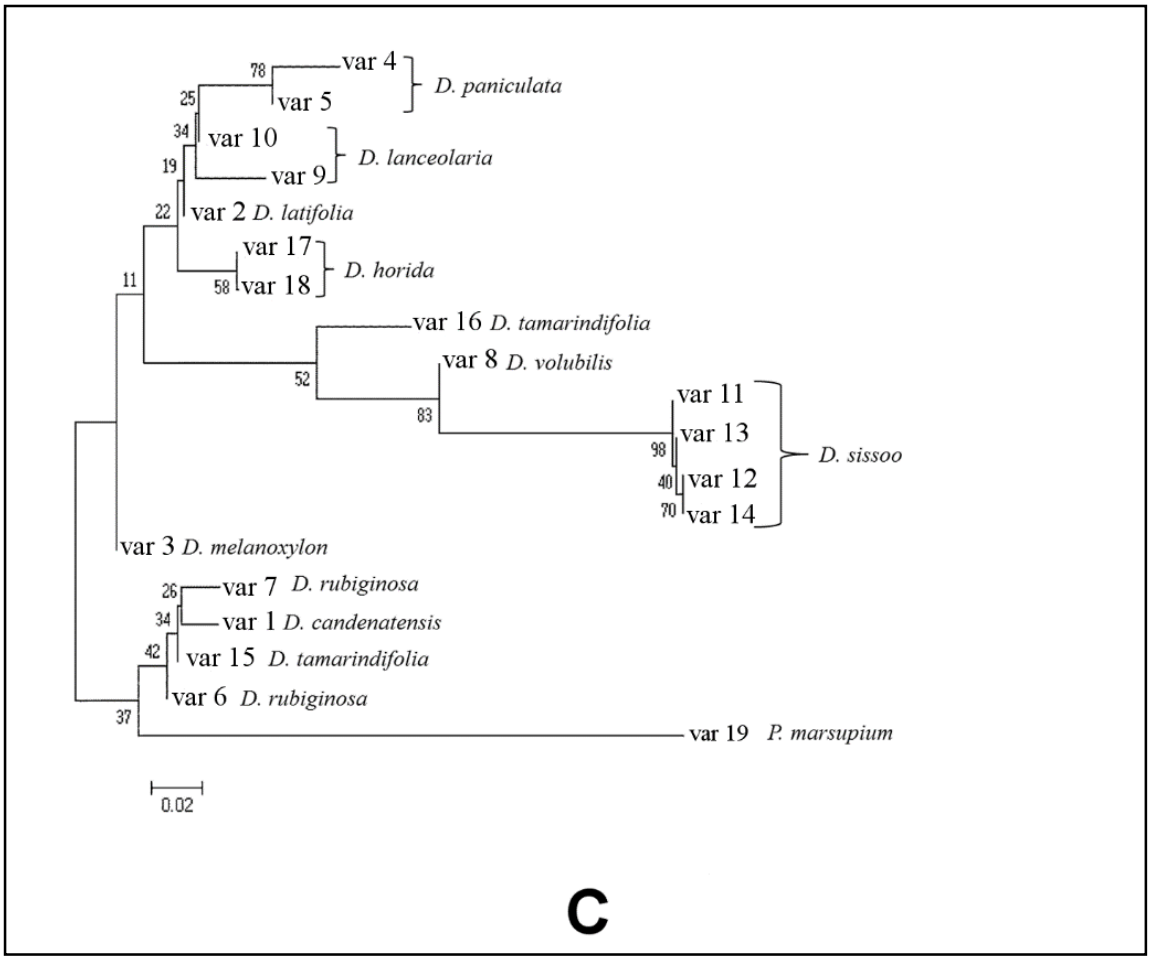
Table S1.11: Distribution of sequence variants among the ten *Dalbergia* species across all loci

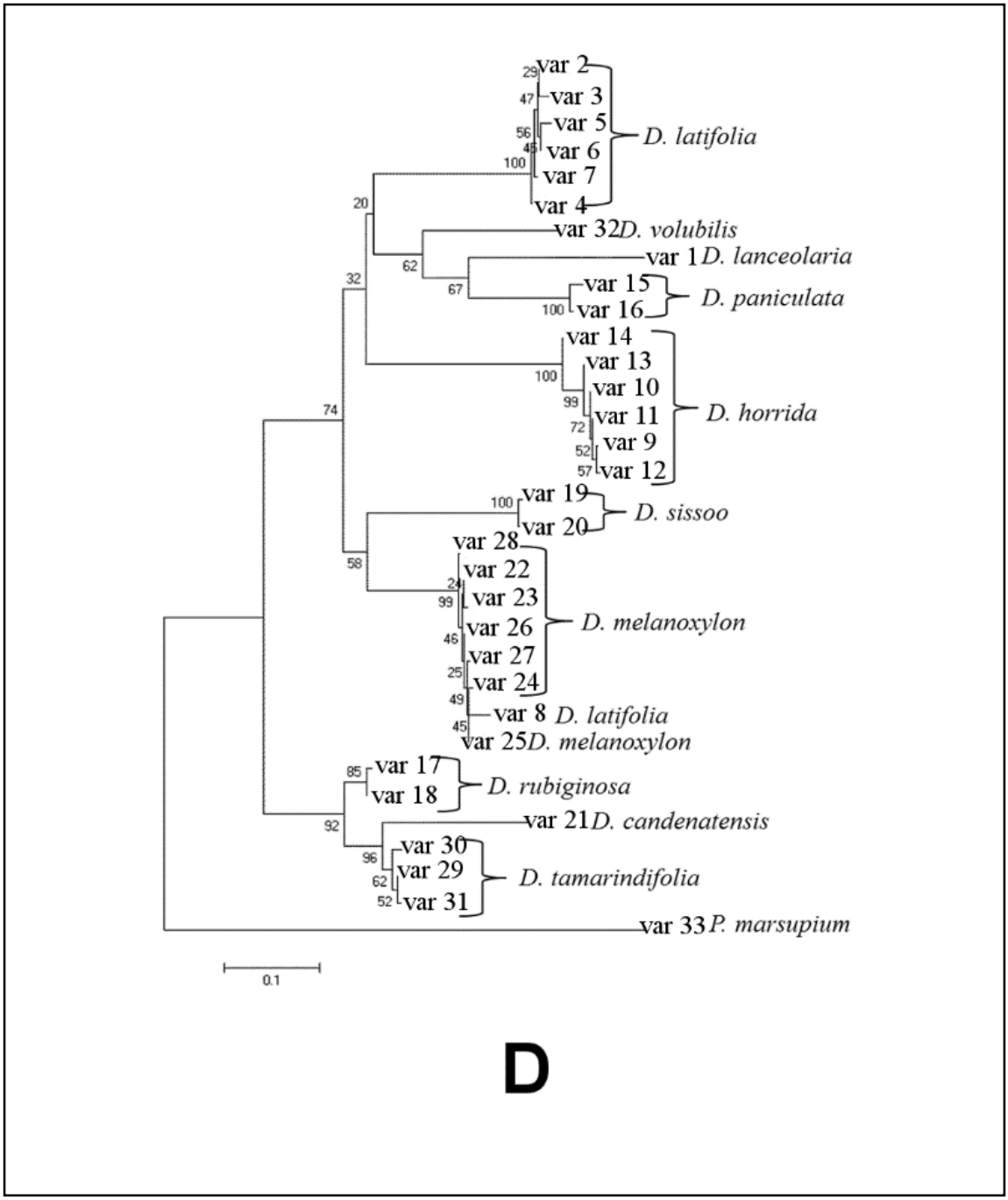
Locus / **Species code**	***matK***	***rbcL***	***trnH-psbA***	**nrITS**	***matK+rbcL***	***matK+trnH-psbA***	***rbcL+trnH-psbA***
	Number of sequence variants (var)						
Dc	1	2	1	1	2	1	2
Dlat	2	2	1	7	2	2	2
Dm	1	1	1	7	1	1	1
Dp	1	1	2	2	1	1	1
Dr	3	1	2	2	3	4	2
Dv	1	1	1	1	1	1	1
Dlan	1	1	2	1	1	2	2
Ds	1	1	4	2	1	4	4
Dt	1	1	2	3	1	2	2
Dh	2	1	2	6	2	2	2

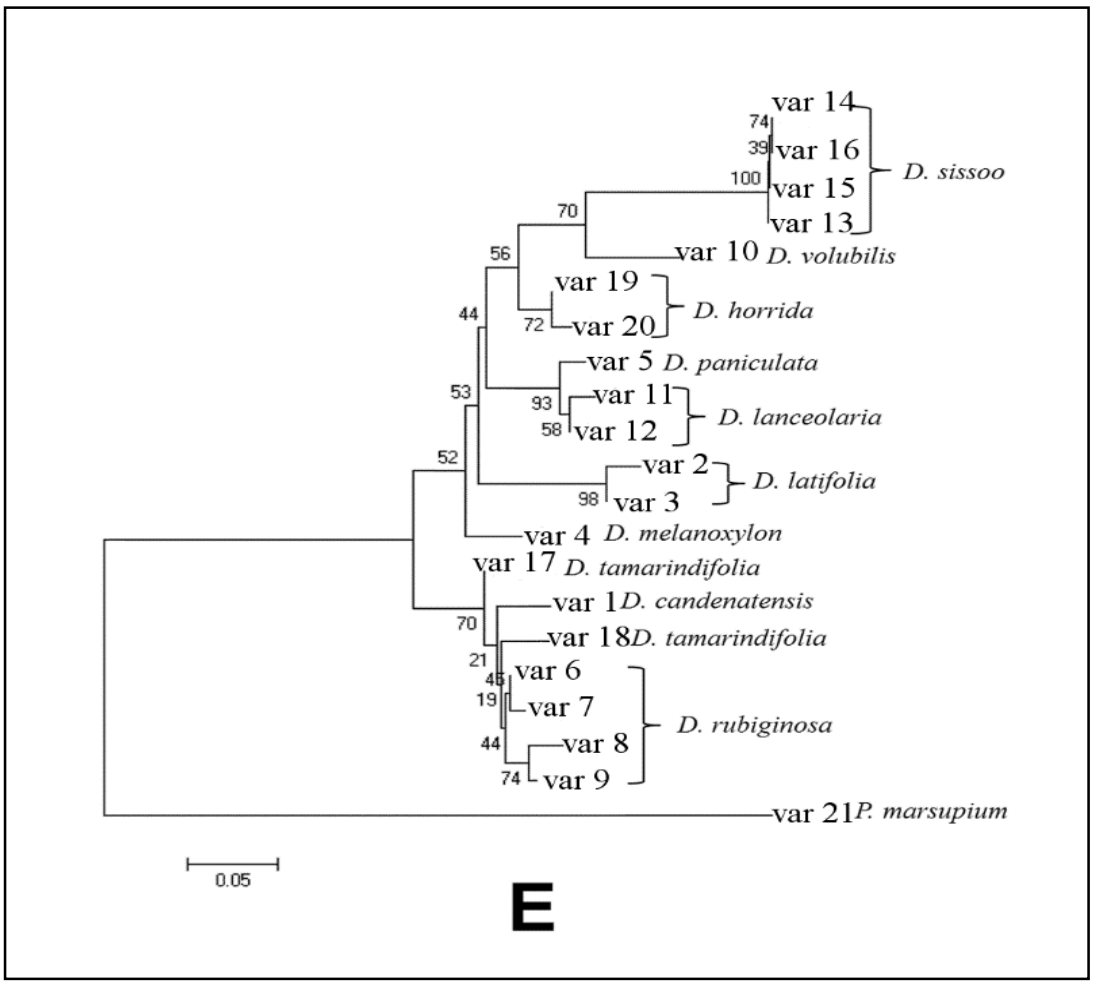
Dc: *D. candenatensis*, Dlat: *D. latifolia*, Dm: *D. melanoxylon*, Dp: *D. paniculata*, Dr: *D. rubiginosa*, Dv: *D. volubilis*, Dlan, *D. lanceolaria*, Ds: *D. sissoo*, Dt: *D. tamarindifolia*, Dh: *D. horrida*

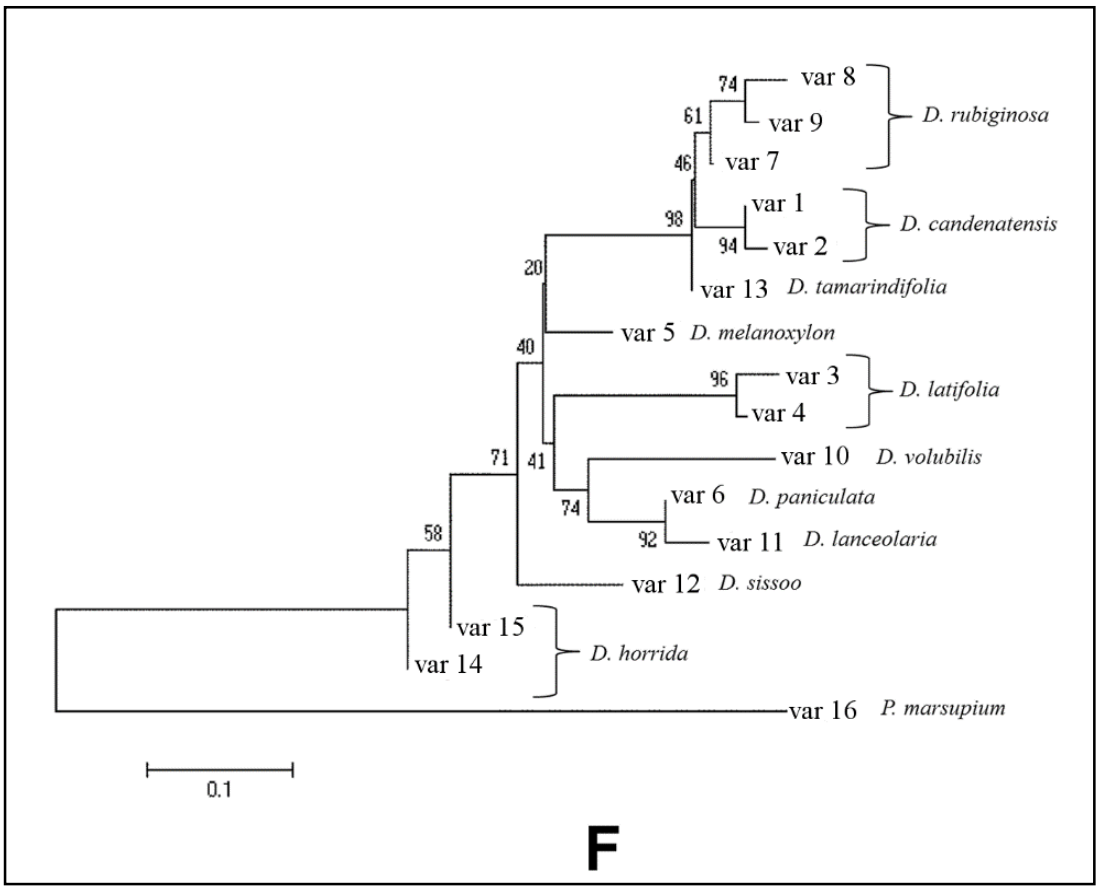

80
var 5 D. paniculata
30
var 10 D. lanceolaria
9
var 9 D. volubilis
var 2
D. latifolia
14
99
var 3
var 11 D. sissoo
42
var 13
D. horrida
31
var 14
var 4 D. melanoxylon
var 1 D. candenatensis
var 12 D. tamarindifolia
80
49
var 6
64
var 7
D. rubiginosa
75
var 8
var 15 P. marsupium
0.1
A


98
var 1
D. rubiginosa + D. candenatensis +
D. tamarindifolia
28
var 2 D. candenatensis
20
var 9 D. sissoo
32
var 5 D. melanoxylon
var 7 D. volubilis
57
var 6 D. paniculata
71
var 8 D. lanceolaria
var 3
D. latifolia
58
var 4
var 10 D. horrida
58
var 11 P. marsupium
0.05
B


78
var 4
var 5
D. paniculata
25
34
var 10
var 9
D. lanceolaria
19
22
var 2 D. latifolia
var 17
58
var 18
D. horida
11
var 16 D. tamarindifolia
var 8 D. volubilis
52
83
var 11
var 13
98
40
var 12
70
var 14
D. sissoo
var 3 D. melanoxylon
26
var 7 D. rubiginosa
34
var 1 D. candenatensis
42
var 15 D. tamarindifolia
var 6 D. rubiginosa
37
var 19 P. marsupium
0.02
C


var 2
var 3
var 5
var 6
var 7
var 4
D. latifolia
var 32 D. volubilis
var 1 D. lanceolaria
var 15
var 16
D. paniculata
var 14
var 13
var 10
var 11
var 9
var 12
D. horrida
var 19
var 20
D. sissoo
var 28
var 22
var 23
var 26
var 27
var 24
D. melanoxylon
var 8 D. latifolia
var 25 D. melanoxylon
var 17
var 18
D. rubiginosa
var 21 D. candenatensis
var 30
var 29
var 31
D. tamarindifolia
var 33 P. marsupium
0.1
D


var 14
74
39
var 16
100
var 15
var 13
D. sissoo
70
var 10 D. volubilis
56
var 19
72
var 20
D. horrida
44
var 5 D. paniculata
53
93
var 11
58
var 12
D. lanceolaria
52
var 2
98
var 3
D. latifolia
var 4 D. melanoxylon
var 17 D. tamarindifolia
var 1 D. candenatensis
70
var 18 D. tamarindifolia
21
46
var 6
19
var 7
44
var 8
74
var 9
D. rubiginosa
var 21 P. marsupium
0.05
E


74
var 8
61
var 9
D. rubiginosa
46
var 7
98
var 1
94
var 2
D. candenatensis
20
var 13 D. tamarindifolia
var 5 D. melanoxylon
40
96
var 3
var 4
D. latifolia
71
41
var 10 D. volubilis
74
var 6 D. paniculata
58
92
var 11 D. lanceolaria
var 12 D. sissoo
var 15
var 14
D. horrida
var 16 P. marsupium
0.1
F

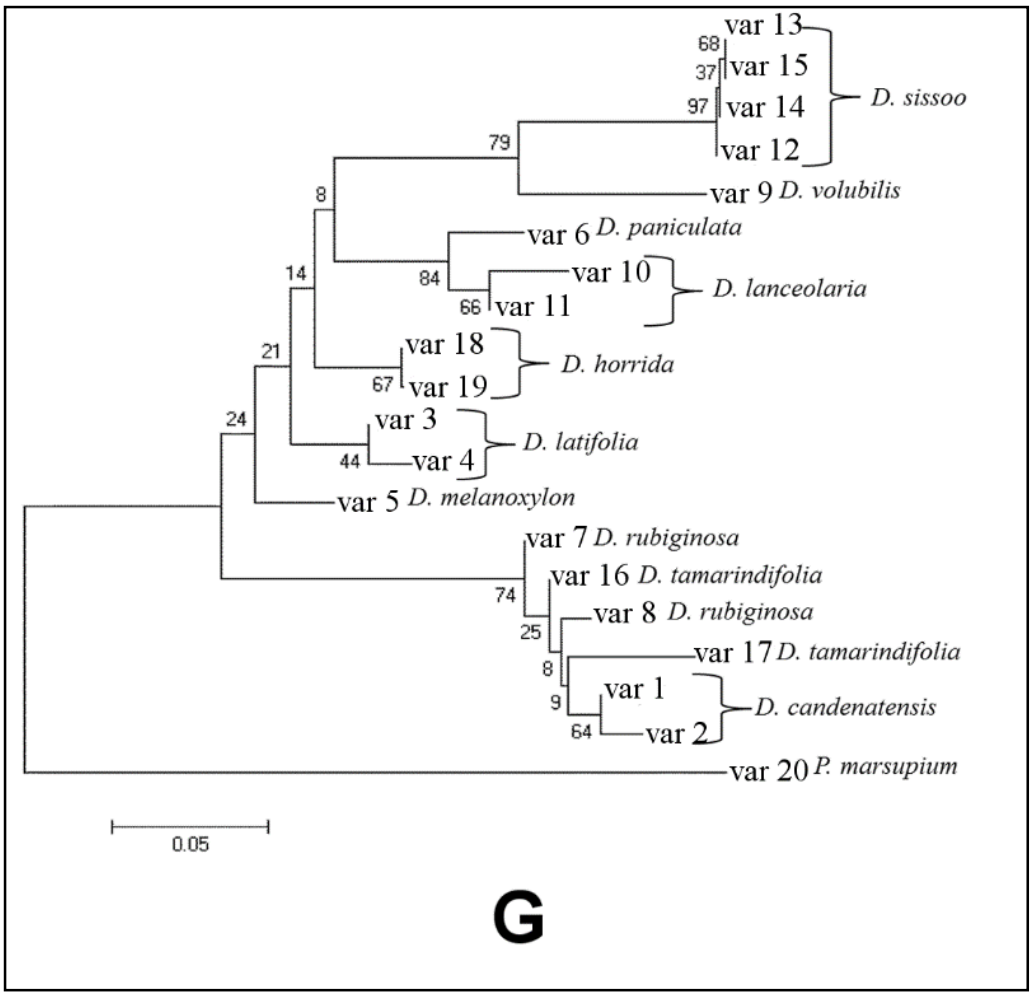


Figure S1.5: NJ trees. NJ trees were constructed using MEGA 5.0 based on K2P distance model– A, *matK*; B, *rbcL*; C: *trnH-psbA*, D, nrITS , E, *matK*+*rbcL*; F, *matK*+*trnH-psbA*; G, *rbcL*+ *trnH-psbA*

Similarity based approach

To evaluate the accuracy of these potential barcodes in species assignments, the BM and BCM parameters from TaxonDNA analysis were used (Table S1.12). Finding a standard threshold for BCM approach is difficult as there is a large variation in inter and intraspecific divergence across all loci in different plant systems (Pettengill and Neel 2010). Moreover, our approach to use multiple accessions of each species, as suggested by Pettengill and Neel (2010) has ensured that the basic requirement was fulfilled and therefore, we chose to use calculated thresholds. The calculated threshold value per locus varied from 0.12% in *rbcL+trnH-psbA* to 1.2% in nrITS. With the BM and BCM approaches, the success rate of correct identification was unambiguously 100% for *matK*, *matK+trnH-psbA* and *matK+rbcL* and 0% incorrect identification (Table S1.12).

Character based approach

The data analysis resulted into logic formulae as well as revealed information regarding correctly classified, wrongly classified and not classified species. Only the analysis done using *matK*, nrITS , *matK+rbcL* and *matK+trnH-psbA* loci could assign the characteristic nucleotide positions for all the species with 100% correct classification (Table S1.13).

Overall performance of the loci

Different parameters used for screening potential barcode loci were ranked based on their performance on a scale of 1-10. In case of NJ trees, the ranking was done based on clustering of the species. Those loci which separated all the species irrespective of intraspecific variation were given ten marks, while for the remaining loci, the scale was determined based on the number of species clubbed together. For inter- and intraspecific distances, the difference between the maximum and minimum distance was calculated to determine the scale for each locus. For BM and BCM methods, the percent values corresponding to correct, ambiguous and incorrect classification were used to rank the loci. A similar methodology was also applied for BLOG. Finally, for Wilcoxon signed rank test, the locus which performed the best in a pair in both, inter and intraspecific distance determinations, was ranked the highest (Table S1.14).

Table S1.12: Results from similarity based analysis using TaxonDNA

	Best Match			Best Close Match				
Regions	**Correct**	**Ambiguous**	**Incorrect**	**Correct**	**Ambiguous**	**Incorrect**	**Sequence without any match closer than threshold**	**Threshold**
matK	165 (100.0)	0 (0.0)	0 (0.0)	165 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.74
rbcL	126 (75.9)	40 (24.09)	0 (0.0)	126 (75.9)	40 (24.09)	0 (0.0)	0 (0.0)	0.2
trnH-psbA	109 (67.7)	51 (31.67)	1 (0.62)	108 (67.08)	51 (31.67)	0 (0.0)	2 (1.24)	0.37
nrITS	134 (99.3)	0 (0.0)	1 (0.73)	135 (99.26)	0 (0.0)	1 (0.73)	0 (0.0)	1.2
matK+ trnH-psbA	157 (100.0)	0 (0.0)	0 (0.0)	157 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.52
matK+ rbcL	163 (100.0)	0 (0.0)	0 (0.0)	163 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.25
rbcL+trnH-psbA	135 (85.98)	21 (13.37)	1 (0.63)	135 (85.98)	21 (13.37)	0 (0.0)	1 (0.63)	0.12

Values in parentheses and threshold are expressed in %

Table S1.13: Character based approach for species identification in *Dalbergia*

Locus	cc	wc	nc	Formula									
				Dc	**Dlat**	**Dm**	**Dp**	**Dr**	**Dv**	**Dlan**	**Ds**	**Dt**	**Dh**
matK	100	0	0	362=A	84=T	206=A	28=G, 166=T, 368=G	440=G	7=C	166=A	51=G	292=T, 362=T, 440=T	368=G, 422=A
rbcL	76.56	0	23.44	339=A	191=T	19=T	35=T, 186=G	-	485=T	35=C	19=A, 179=A,86=A,191=C,458=G,485=C	-	458=T
trnH-psbA	69.35	0	30.65	24=T	-	12=C	118=G	139=A	52=C, 114=A,228=A	26=T	114=G,228=A	52=A, 228=A	114=G,228=G
nrITS	100	0	0	107=C	122=C,456=T	83=C	621=A	132=T,231=G	43=C	132=T,456=C	128=A	107=G	539=C,637=C
matK +rbcL	100	0	0	1052=T	711=T	697=A	95=A, 657=T,711=G,1011=A	931=G OR 783=A	1011=G	657=A	1018=G	783=T,801=A,931=T,1052=C	922=T

Locus	**cc**	**wc**	**nc**	**Formula**									
				Dc	**Dlat**	**Dm**	**Dp**	**Dr**	**Dv**	**Dlan**	**Ds**	**Dt**	**Dh**
matK +trnH -psbA	100	0	0	362= A	84=T	206= A	795= G	440= G OR 292= A	422=C , 905= A	166= A	422= A,905 =A	292=T , 362=T , 440=T	422= A,905 =G
rbcL+ trnH-psbA	86.99	0	13.01	515=T	191=T	503=C	609= G	630= A	495=T ,734= C	95=A, 191=C ,734= A	495= G	-	191=C ,503= A,686 =A,73 4=A

cc: correctly classified, wc: wrongly classified, nc: not classified

Dc: *D. candenatensis*, Dlat: *D. latifolia*, Dm: *D. melanoxylon*, Dp: *D. paniculata*, Dr: *D. rubiginosa*, Dv: *D. volubilis*, Dlan, *D. lanceolaria*, Ds: *D. sissoo*, Dt: *D. tamarindifolia*, Dh: *D. horrida*

Table S1.14: Comparative ranking of loci used in DNA barcoding of *Dalbergia*

Parameters	***matK***	***rbcL***	***trnH-psbA***	**nrITS**	***matK+rbcL***	***matK+ trnH-psbA***	***rbcL+ trnH-psbA***
Barcode	10	7	8	8	10	8	8
Inter and intra specific distances	10	7	8	8	10	8	8
BM and BCM	10	7.2	4.9	6.6	10	10	3.3
BLOG	10	7	7	7.5	10	8.6	6.4
Wilcoxon Signed Rank test	4	5	9	5.5	3.5	5	6
Total	**44**	**33.2**	**36.9**	**35.6**	**43.5**	**39.6**	**31.7**

Larger values indicate better performance

S 1.3.2 Validation of loci using NCBI sequences

Inter and intraspecific divergence based analysis

The DNA sequences of various *Dalbergia* species having more than one accessions corresponding to the *matK*, *rbcL* and nrITS regions downloaded from the NCBI database were used to validate the proposed regions along with sequences from the present study. The analysis was carried out using all the parameters which have previously been used for the development of barcode. However, total number of sites for nrITS from the NCBI dataset included some variation in sequence length of nrITS loci (637 in our dataset as against 660 in the combined data). Summary statistics and within and between species divergence parameters are summarized in Table S1.15.

For nrITS , aligned length was 660 bp, with 36.36% sites variable and 32.12% polymorphic informative, which was the highest among the three loci under consideration; while, based on the percentage of conserved sites, the most conserved locus was a combination of *mat* and *rbcL* (Table S1.15). The nrITS showed greater inter specific divergence than *matK* and *matK+rbcL* using both the parameters (average inter specific distance and θ'). Similarly, even for intraspecific divergence, nrITS showed the highest value while, *rbcL+matK* showed the lowest (Table S1.15 and Figure S1.6). In case of barcode gap analysis, none of the loci used in the combined analysis showed any gap (Figure S1.7) supporting our previous results.

Table S1.15: Summary statistics and inter and intra specific divergence for samples used for validation

Locus / Parameters	*matK*	*matK+rbcL*	nrITS
Total number of sequences	183	172	181
Total no. of sites	677	1168	660
Conserved sites	598 (88.33)	1080 (92.47)	416 (63.03)
Variable sites	79 (11.67)	88 (7.43)	240 (36.36)
Parsimony informative sites	56 (8.27)	61 (5.22)	212 (32.12)
Average interspecific distance	0.017±4.11E-05	0.014±3.78E-05	0.172±3.257E-4
Theta prime (θ')	0.016±4.77518E-4	0.012±5.32E-4	0.159±2.814844E-3
Smallest interspecific distance	0.015±0.000469	0.011±0.000514	0.15±0.003576
Average intraspecific distance	0.001±4.25133E-05	0±2.7017E-05	0.005±5.76744E-4
Theta (θ)	0.000437±2.54947E-4	0.00025±1.54E-4	0.026±1.28E-2
Coalescent depth	0.00094±0.00054	0.00075±0.000392	0.044±0.016

Tree based analyses

The sequence variants were further determined using DnaSP 5.0 and MEGA 5.0. They revealed overall high intraspecific variation in case of nrITS supporting the previous results. Three NJ trees were constructed with *matK, rbcL+matK* and nrITS using MEGA 5.0 including all the sequence variants (Figure S1.8). All of them revealed separate cluster for each species.

Similarity based approach

Validity and accuracy of shortlisted potential barcodes were further checked with the BM

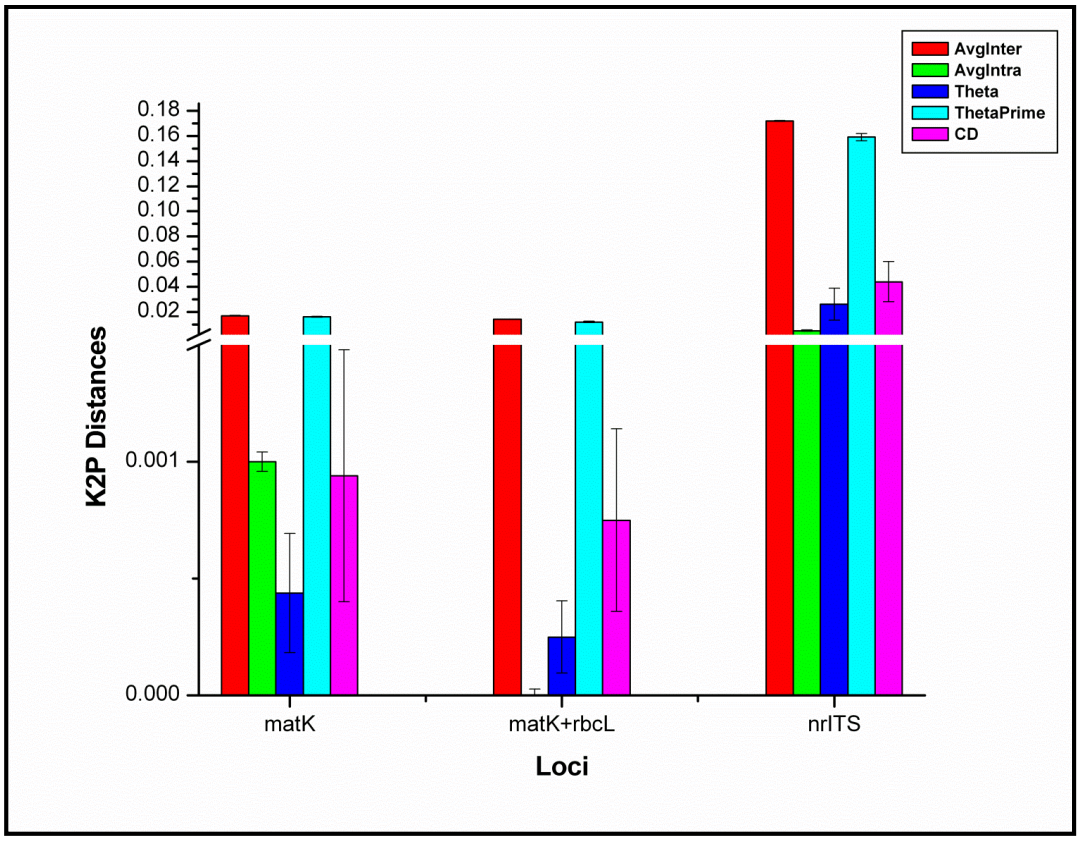


Figure S1.6: Plot of distribution of inter and intraspecific divergence

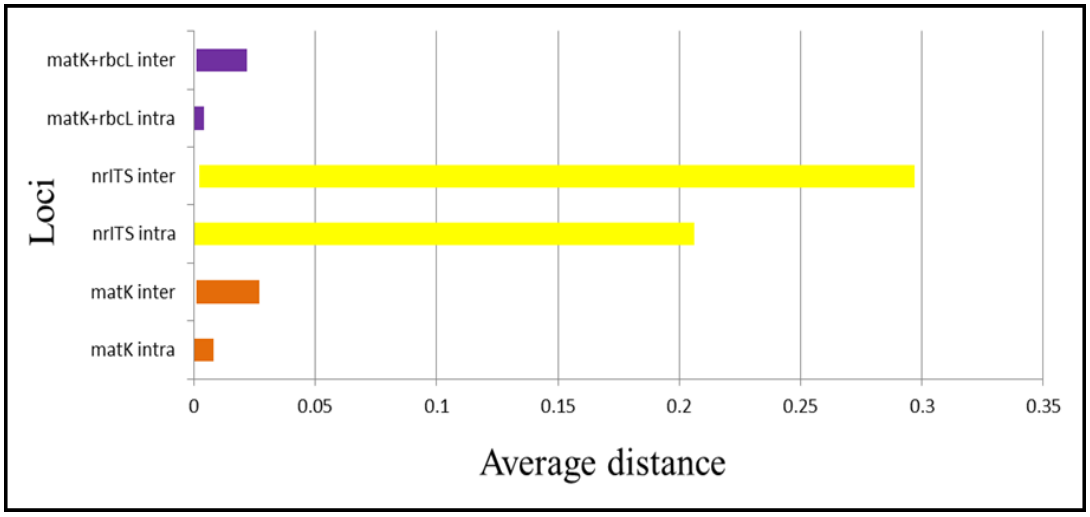


Figure S1.7: The barcoding gap

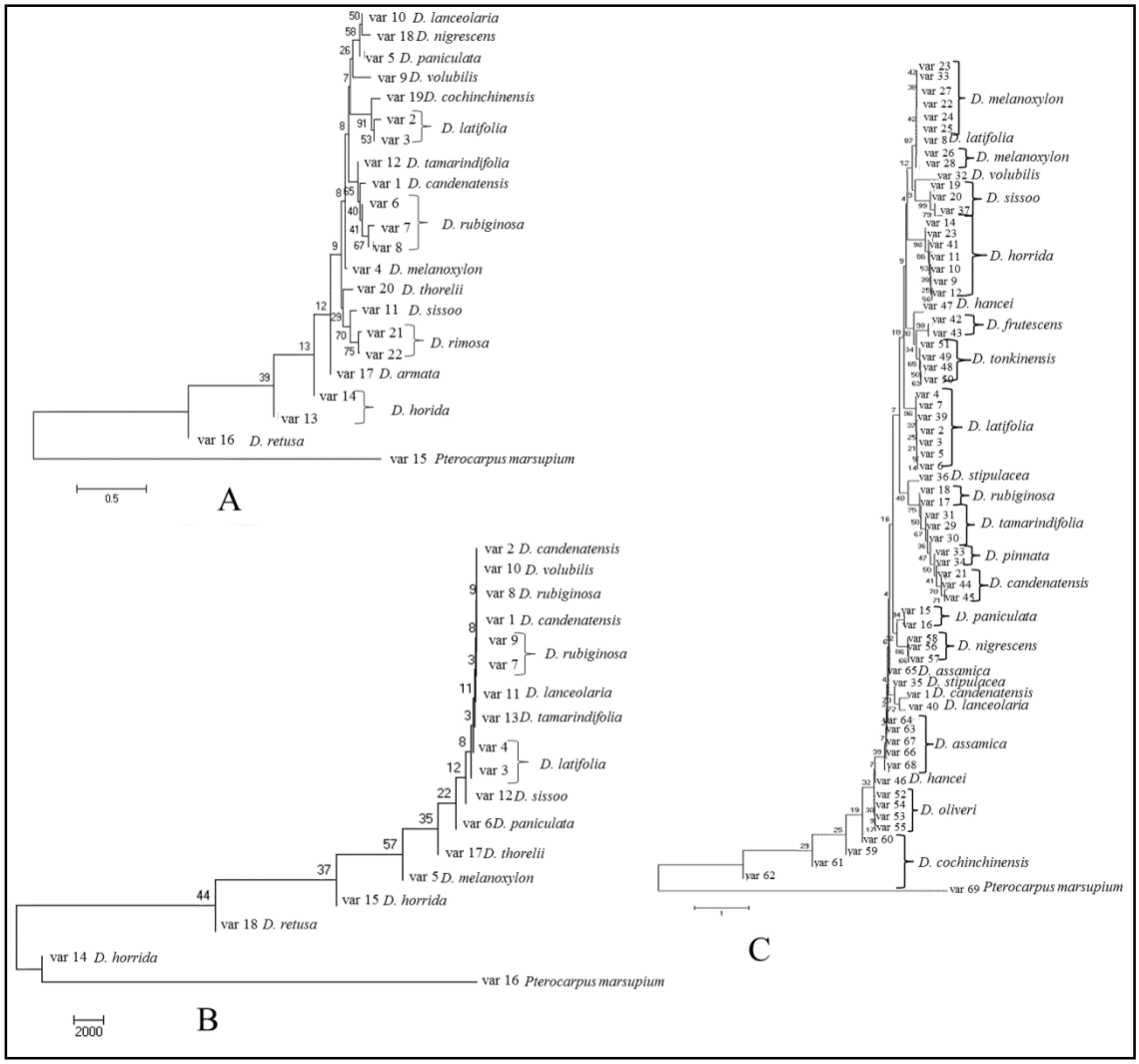


Figure S1.8: NJ trees constructed using MEGA 5.0 based on K2P distance model– A, *matK*; B, *matK+rbcL*; C, nrITS

and BCM parameters of TaxonDNA. With the BM and BCM approaches, the success rate of correct identification was unambiguously 100% in *matK* and *matK+rbcL*; however, nrITS exhibited 94.45% correct identification with 5.55% incorrect identification (Table S1.16). This has also supported the previous results where nrITS showed incorrect identification.

Character based approach

The data analysis showed 100% correct classification with *matK* and *matK+rbcL*. These results supported the previous observations. In previous results, nrITS indicated 100% correct classification. However, when species other than ten *Dalbergia* species were added in the analysis, the correct classification reduced to 84.44%. The detailed results of character based approach are given in Table S1.17 (**Table is provided in a compact disc [CD] alongwith this dissertation**).

Overall results

The overall performance also concluded *matK* and *matK+rbcL* to be the best loci to identify *Dalbergia* species (Table S1.18).

S 1.3.3 Analysis using unauthenticated individuals

Individuals belonging to unauthenticated samples of *Dalbergia* were also collected from WG and were authenticated by BSI after the barcoding analysis. Hence, these were evaluated for their identity using *matK*, *matK+rbcL* and nrITS loci based NJ trees (Figure S1.9), as unknown samples. The results indicated that, in all the three loci, one sample clubbed with *D. melanoxylon*, while five with *D. horrida*. Secondly, only one accession of *D. sissoides* collected and authenticated from local person, actually clubbed with variant two of *D. latifolia* represented by five samples in *matK*, while with variant 17 containing the same five samples of *D. latifolia* in *matK+rbcL*. However, in nrITS *D. sissoides* combined with variant 33 containing one of the above five samples along with variant 34 (containing remaining four samples) which was closest than rest *D. latifolia* samples. Similarly, another accession of *D. paniculata* (DpA) collected and authenticated from Mysore region by a local person, clustered with *D. lanceolaria* individuals. Moreover, all the unauthenticated samples formed more haplotypes only in case of nrITS.

Table S1.16: Similarity based analysis using TaxonDNA for samples used for validation

	Best Match			Best Close Match				
Regions	**Correct**	**Ambiguous**	**Incorrect**	**Correct**	**Ambiguous**	**Incorrect**	**Sequence without any match closer than threshold**	**Threshold**
matK	183 (100)	0 (0)	0 (0)	180 (98.36)	0 (0)	0 (0)	3 (1.63)	0.28
nrITS	170 (94.45)	0 (0)	10 (5.55)	159 (88.33)	0 (0)	0 (0)	21 (11.66)	0.63
matK*+ *rbcL	172 (100)	0 (0)	0 (0)	170 (98.83)	0 (0)	0 (0)	2 (1.16)	0

Table S1.18: Comparative ranking loci used in validation

Parameters	***matK***	***rbcL+matK***	**nrITS**
Barcode	10	10	10
Inter and intra specific distances	4.31	5.215	5
Best match & best close match	10	10	8.14
BLOG	10	6.67	3.33
Total score	34.31	31.89	26.47

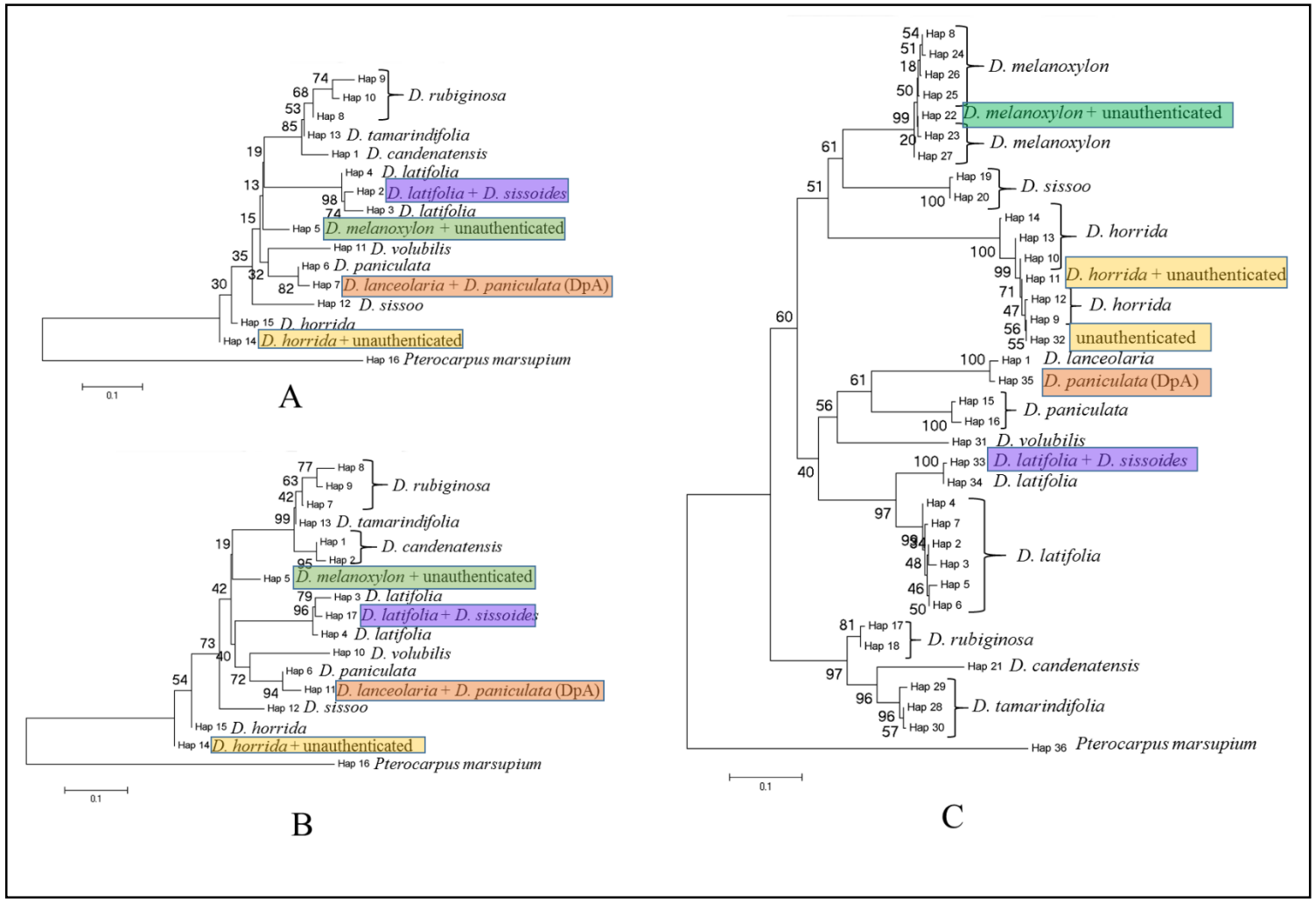


Figure S1.9: NJ trees with unknown samples, A, *matK*; B, *matK*+*rbcL*; C, nrITS

S 1.3.4 Phylogenetic analysis

The combined (*matK*+*rbcL*+nrITS+*trnH-psbA*) phylogenetic analysis was carried out to study the relationship among different *Dalbergia* species (Figure S1.10). The outgroup species such as (*Pterocarpus marsupium, P. acapulcensis, Aeschynomene virginica, A. pfundii, Soemmeringia sempeiflorens, Picteria marginata, Diphysa ormocarpoides* and *Ormocarpum muricatum*) were well separated from the rest of the species. The phylogenetic tree was divided into three groups, group I, group II and group III, while group III was further divided into three subgroups such as IIIa, IIIb and IIIc. Group I consisted both old and new world species. Both the species of section *Ecastaphyllum* (*D. ecastaphyllum* and *D. monetaria*) were clustered in group I. Group II had only New World species and all of them were from section *Triptolemea* (*D. inundata*, *D. lateriflora* and *D. revoluta*). All the subgroups in third group consisted of both Old and New World species. In group IIIa species from Asia, Africa and America were found to be distributed. In IIIb the distribution was found to be dominated by Asian species (14 species) with some of them from section *Dalbergia* (*D. paniculata, D. lanceolaria, D. assamica, D. balansae, D. godefroyi, D. volubilis, D. dongnaiensis, D. lakhonensis, D. oliveri, D. falcata* and *D. kurzii*), few African (three species) and one American species while in IIIc three species were Asian and three species were American. Group I clades showed both the highest posterior probability value (pp=1) as well as the lowest value (pp=0.5). The clades of *D. candenatensis, D. rubiginosa*, section *Ecastaphylla*, *D. abbreviata*, *D. sericea* and *D. havilandii* were to be evolved with high pp values (>0.8). Group II were observed to be evolved with high pp values (pp=0.97 and 0.89). In group III many including *D. horrida, D. frutescens, D. assamica, D. lanceolaria, D. hupeana, D. paniculata* and *D. cochinchinensis* having multiple accessions were found to be evolved with high posterior probabilitie values (pp>0.9).

The current analysis involved species with single accessions as well as multiple accessions. Multiple accessions of *D. latifolia, D. cochinchinensis, D. lanceolaria, D. hupeana, D. nigrescens, D. assamica, D. oliveri, D. tonkinensis, D. thorelii, D. horrida, D. melanoxylon, D. sissoo,* and *D. candenatensis* clubbed according to their species. However, multiple accessions of *D. odorifera* (group I and group IIIc), *D. hancei* (group IIIa and IIIb) and *D. pinnata* (group I and group IIIc) were placed far away from each other.

As mentioned above, the species with multiple accessions clustered irrespective of their geographical distribution. Individuals of *D. latifolia, D. cochinchinensis* (group

IIIc), *D. lanceolaria, D. hupeana, D. nigrescens, D. assamica, D. hancei, D. oliveri* (group IIIb), *D. tonkinensis, D. thorelii, D. horrida* included in this analysis were found to be distributed across the entire Asia, however, individuals of *D. melanoxylon, D. sissoo* (group IIIc), and *D. candenatensis* (group I) analysed here were from different continents i.e. Asia, Africa and America. Thus, the overall distribution pattern was found to be assorted with no differentiation between Old World and New World species.

S 1.4 Discussion

Paul Hebert's research in 2003 on species identification using short stretches of DNA from a well characterized region of the genome, gave birth to the concept of DNA barcoding (Hebert et al. 2003a). Initial efforts proved the reliability of mitochondrial cytochrome c oxidase 1 (*cox1*) gene as an impressive barcode in animals (Hebert et al. 2003b). However, initial research on plant DNA barcoding suggested that species discrimination in plants with a single universal locus is difficult. This is primarily due to various phenomena such as polyploidy, hybridization, heteroplasy etc., which results in the formation of continuous range of variable characters and making delineation a difficult task. Alternatively, sufficient time is often required to accumulate mutations in organisms which are responsible for separation of closely related species. However, the lack of adequate genetic variation hampers species level discrimination of plants by DNA barcoding (Hollingsworth et al. 2011). This problem is exaggerated in woody plants because of longer generation time and lower mutation rate. It is also difficult to differentiate species in taxonomically complex groups where species are narrowly defined. Additionally, large ancestral population sizes and low levels of within species gene flow for plastid markers create difficulty in barcode based identification (Fazekas et al. 2009; Hollingsworth et al. 2011). In order to resolve these problems, several attempts have been made to establish DNA barcodes using multiple genes from different plant genomes for specific families such as Myristicaceae (Newmaster et al. 2008), Lemnaceae (Wang et al. 2010b), Zingiberaceae (Shi et al. 2011), Podocarpaceae (Little et al. 2013a) or genera such as *Paeonia* (Zhang et al. 2009), *Acacia* (Newmaster and Ragupathy 2009), *Paphiopedilum* (Parveen et al. 2012), *Parnassia* (Yang et al. 2012) and *Gossypium* (Ashfaq et al. 2013). However, from different studies, it appears that finding a universal barcode or even a barcode at family level is difficult and it may be possible to establish a discriminating barcode only at genus level (Tiayyba Riaz 2011).

There are few reports on DNA barcoding of tropical tree species (Gonzalez et al. 2009;

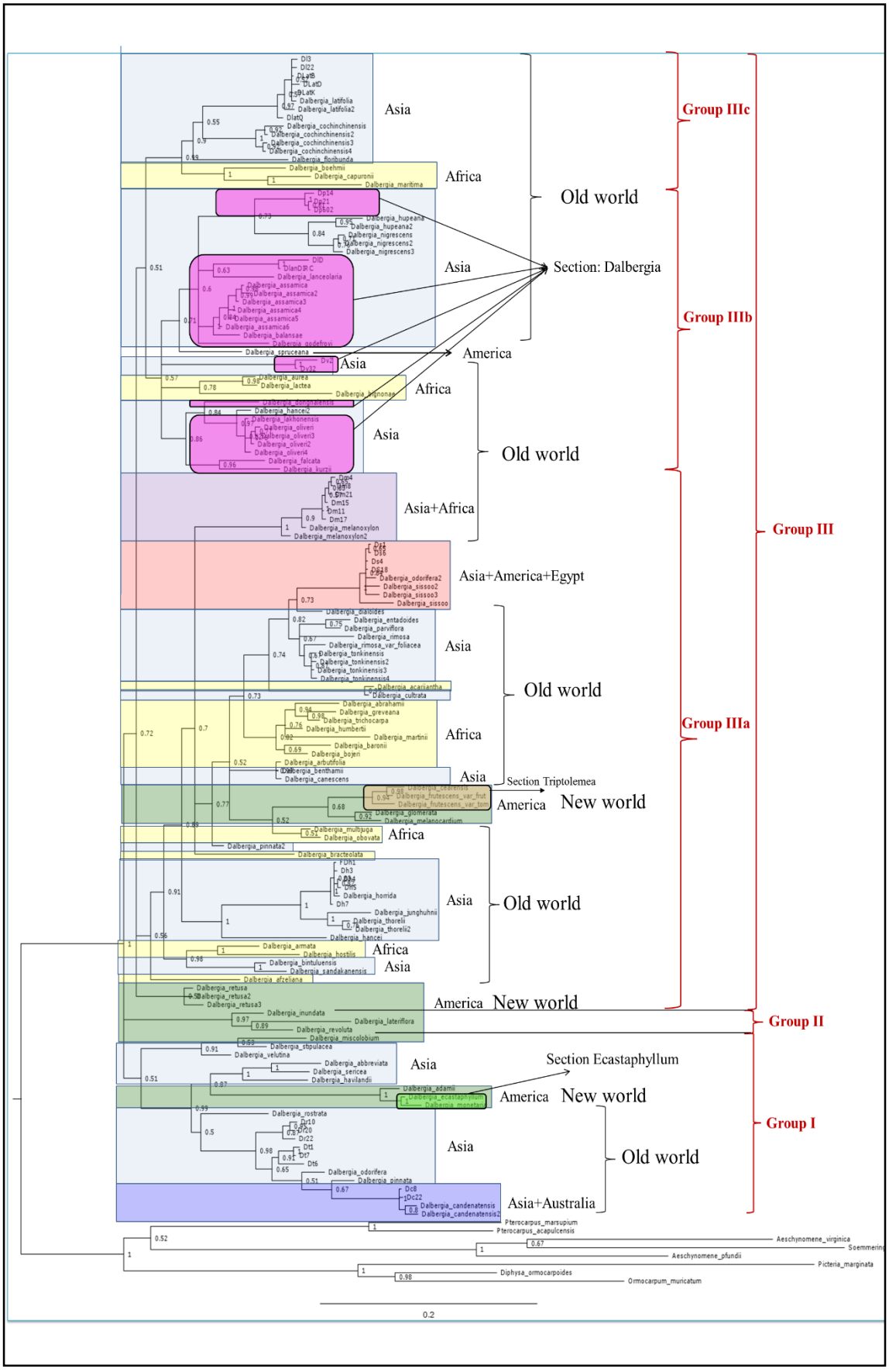


Figure S1.10: Phylogenetic tree

Nithaniyal et al. 2014; Tripathi et al. 2013) which include Amazonian as well as Indian forest trees. These studies have used nrITS, *matK*, *rbcL* and *trnH-psbA* loci. However, there are scanty reports on DNA barcoding of trees exclusively from WG of India. A study on 143 tree species from tropical dry evergreen forests in India covering 114 genera and 42 families revealed that combination of *matK* and *rbcL* loci gave the highest success in accurate identification (Nithaniyal et al. 2014). Similarly, DNA barcoding of medicinal plants from the family Fabaceae revealed 80% and 96% success at species and genus level, respectively using *matK* locus, while the ITS2 locus gave more than 80% success at species level and 100% success at genus level (Gao et al. 2010a). However, none of the above mentioned studies included *Dalbergia*. A recent study on tropical tree species from India (149 species from 82 genera and 38 families) included three *Dalbergia* species and suggested that ITS and *trnH-psbA* might not be highly successful (Tripathi et al. 2013). Efforts to resolve the sister species complex of *Acacia* from Fabaceae using *rbcL*, *trnH-psbA* (same primer sequence as we have used in our study) and *matK* recommended all the three regions for barcoding (Newmaster and Ragupathy 2009). On the contrary, studies on *Aspalathus* using ITS (different primers than the ones used in our study), *psbA-trnH* and *trnT-trnL* concluded that all the three loci were unable to resolve the species (Edwards et al. 2008). It was observed that the output from *matK* analysis was variable based on the plant systems as well as on the combination of primers used for analysis. However, the Consortium for the Barcode of Life (CBOL) proposed 90% success with *matK* for plants. Our study also identified *matK* as one of the potential loci for DNA barcoding. Thus, *matK*, nrITS and *rbcL* individually or in their combinations could be explored as the potential DNA barcodes in various plant genera (Gonzalez et al. 2009).

Assessment of the four candidate barcodes in *Dalbergia* genus

In the present study, the amplification and sequencing success rate in *Dalbergia* ranged from 80.5% (for nrITS) to 97.6% (for *rbcL*). While the *rbcL* locus has been reported to be easy to amplify and sequence across a broad range of plant taxa, but offers low species resolution, the rapidly evolving *matK*, locus, is known for its high discriminatory power with low universality (Hollingsworth et al. 2009). Hence, the *matK* locus is popular for species discrimination in case of angiosperms (Fazekas et al. 2009). However, mixed results ranging from high success rate (Lahaye et al. 2008; Hollingsworth et al. 2009) to poor discrimination (Kress and Erickson 2007; Fazekas et al. 2009) have been reported for *matK*. Even in the present study, *matK* showed good resolving power and although *trnH-psbA* showed good universality and higher discrimination, it also has variable

length, presence of homopolymers, inversions and insertion of *rps19* gene (Starr et al. 2009; Whitlock et al. 2010; Pang et al. 2012). Similarly, while the nrITS locus is a commonly used nuclear marker for phylogenetic studies (Alvarez and Wendel 2003), it was, however, not preferred for barcoding studies initially because of fungal contamination, paralogous gene copies and problems in recovery (Hollingsworth et al. 2011b). In our study, similarity search using BLAST did not reveal any problem of fungal contamination in nrITS sequences; however, the sequencing success was low (80%), which might be due to the presence of divergent gene copies as reported earlier (Alvarez and Wendel 2003). In case of *trnH-psbA* which gave 94.7% sequencing success, our data revealed the presence of T and A repeats, without any insertion of *rps19* gene when checked by BLAST.

The overall interspecific distances were high compared to intraspecific distances and no significant barcode gap was observed in the present study. Usually in the closely related plant species, plastid regions such as *rbcL* and *matK* do not generate a barcode gap (Lahaye et al. 2008). Several studies have also revealed the absence of barcode gap in different plant systems such as *Agalinis* (Pettengill and Neel 2010), medicinal plants (Chen et al. 2010), *Dioscorea* (Sun et al. 2012), *Parnassia* (Yang et al. 2012) and *Gossypium* (Ashfaq et al. 2013). Furthermore in the NJ tree based analysis all the loci except *trnH-psbA* could cluster the *Dalbergia* species as per their sections. This agrees with a previous report on genome size variation and evolution of *Dalbergia* species which found that *D. lanceolaria* and *D. paniculata* were closely related (Hiremath and Nagasampige 2004). NrITS, *matK* and *trnH-psbA* and their combinations formed separate clusters for each species. However, *rbcL* could not differentiate *D. rubiginosa*, *D. candenatensis* and *D. tamarindifolia*, which could be because of the conserved nature of the gene (Albert et al. 1994). Similar behavior of *rbcL* was also reported in *Carex* (Starr et al. 2009). Together this suggested that individually *rbcL* might not serve as a good barcode but can be utilized in combination with other loci.

A recent report on DNA barcoding of eight *Dalbergia* species from Vietnam recommended ITS locus as a potential barcode based on UPGMA analysis and nucleotide diversity (Phong et al. 2014). It has been reported that being a multigene family, 18s-26s rDNA is subjected to concerted evolution. In certain cases, ITS1 (Campbell et al. 2005; Blaalid et al. 2013) and ITS2 (Pang et al. 2012; Blaalid et al. 2013; Han et al. 2012; Chen et al. 2013; Chen et al. 2010) have been used as separate loci for DNA barcoding. However, point mutations displayed by ITS1 and ITS2 also contribute to high

intraspecific variations (Baldwin 1992). We used the complete ITS region (ITS1-5.8S-ITS2) as a single barcoding locus. In our study, nrITS showed high intraspecific variation with high species discrimination, leading to incorrect identification with BM and BCM. However, DNA barcoding of eight *Dalbergia* species from Vietnam (Phong et al. 2014), did not use the species from our study. A reanalysis of the data from NCBI for the species used in the Vietnam study along with dataset from this study revealed a high number of sequence variants for most of the species. Moreover, from the available sequence data in NCBI for the Vietnam study (Phong et al. 2014), we could find only one nrITS sequence each for *D. dialoides*, *D. entadoides* and *D. hancei* making it difficult to assay the intraspecific variation. It was therefore, not possible to comment on either the intraspecific diversity of these species, which is an important factor for DNA barcoding or the suitability of nrITS as the potential barcode for *Dalbergia* species. It is essential to sample enough number of accessions for each of these species, ideally from different geographical locations, to sample the intraspecific variation from the entire distributional range (Gonzalez et al. 2009).

matK and _mastK+rbcL_: most suitable barcode loci for _Dalbergia_ genus

In the validation of DNA barcodes *Dalbergia* species sequences downloaded from NCBI were analysed along with the present study. For the validation purpose the *matK* and *matK+rbcL* (proposed loci for *Dalbergia* genus) and nrITS were used. Total 16, 12 and 19 species were analysed using *matK*, *matK+rbcL* and nrITS, respectively. The nrITS was found to be highly variable and revealed the highest average interspecific distance while *matK+rbcL* was less variable and revealed the lowest average intraspecific distance. No barcode gap was observed with any of the loci. When the data was analysed using NJ trees all the three loci showed separate clusters for each species however, nrITS displayed higher number of sequence variants. In similarity based approach also *matK* and *matK+rbcL* revealed 100% correct identification. All these results correspond well to the previous conclusions based on our data alone. The overall performance concluded that *matK* and *matK+rbcL* are the suitable loci to identify *Dalbergia* species. There are few reports in which *in silico* analysis was carried out to develop DNA barcodes, for medicinal plants (http://dnabarcodes2015.org/wp-content/uploads/2015/09/SaloniMalik.pdf.), using ITS2 region for 50,790 plants and 12,221 animals (Yao et al. 2010), for diptera (Meier et al. 2006) and to check the utility of *trnH-psbA* intergenic spacer (Pang et al. 2012) etc. On the other hand, large number of studies developed DNA barcodes through wet lab validations (Kress and Erickson 2007; Yan et al. 2011; Giudicelli et al. 2015). In

the analysis by de Groot et al. (2011) the *in silico* analysis for ferns was performed for those species which could not be collected for wet lab experimentation for barcode determination. However, there are no such reports to validate the available wet lab data with other complementary tools. In our case, the established barcode loci for *Dalbergia* were clearly shown to be the most suitable loci even after *in silico* analysis using NCBI data. Thus, our data indicated good merge of *in silico* and wet lab analysis.

Further, the unauthenticated samples of *Dalbergia* collected from WG were checked for their identity using NJ trees by analysing them along with the previously collected samples. All the three loci (*matK, matK+rbcL* and nrITS) revealed separate cluster for each species. Out of eight such samples, six could cluster with *D. melanoxylon* and *D. horrida* as detailed in results section. Remaining two samples authenticated by local collectors as *D. sissoides* and *D. paniculata*, respectively matched with *D. latifolia* and *D. lanceolaria* using barcoding loci. *D. sissoides* and D. *latifolia* share the section *Sissoa*, while, *D. lanceolaria* and *D. paniculata* are from section *Dalbergia*. *D. sissoides* is morphologically similar to *D. latifolia* and appears to be evolved from it (Hiremath and Nagasampige 2004). It has also been considered as a variety of *D. latifolia* (Panda 2004). Similarly, *D. lanceolaria* and *D. paniculata* are morphologically similar and difficult to separate in the fruiting condition. Apparently *D. paniculata* appears to be evolved from *D. lanceolaria (*Hiremath and Nagasampige 2004). Thothathri (1987) has rather proposed that *D. paniculata* may be merged with primitive taxon *D. lanceolaria* and given a subspecies status. Thus, the confusion in identification of *D. sissoides* and *D. paniculata* by morphological characters was revealed by our study. Hence more number of accessions of *D. sissoides* and *D. paniculata* from the respective regions need to be collected and studied using DNA barcoding approach to eliminate this misidentification. Secondly, the NJ tree by nrITS marker loci indicated that the unauthenticated samples of *D. horrida* and *D. melanoxylon* revealed more variations leading to additional variants than the existing ones. This was not the case by *matK* and *matK+rbcL*, thus, suggesting *matK* and *matK+rbcL* as more robust DNA barcoding loci than nrITS for genus *Dalbergia*.

Relationship between different *Dalbergia* species

Dalbergia has a pantropical distribution with high species diversity in Asia (ca. 100 species) Neotropics and Africa. Long distance dispersal and vicariance are the two processes which have been considered to explain the wide distribution in Africa, Asia and South America (Bartish et al. 2011). Also, *Dalbergia* species grow in varied habitats right

from moist and dry tropical forests, to savannas, coastal dunes and rocky outcrops including shrubs, trees and woody climbers. This diversity in habitats and life forms may have contributed to the ability of *Dalbergia* species to inflate the distribution range. Eighty six South East Asian *Dalbergia* species were classified into two subgenera, five sections and 24 series by Prain (1904). Thothathri (1987) classified 46 *Dalbergia* species from Indian subcontinent into four sections and seven series. The present study included 79 species from all the sections and various geographic regions. Using Bayesian phylogenetic analysis, our results indicated mixed distribution of species supporting the monophyletic origin of *Dalbergia* species as suggested by Vatanparast et al. (2013). The species when studied based on the sections; it was revealed that section *Dalbergia* (low support value), section *Ecastaphyllum* (high support value), section *Triptolemea* (high support value), section *Selenolobium* (high support value) and section *Sissoa* (low support value except *D. rubiginosa*) were monophyletic in nature. The monophyly of section *Triptolemea* and *Ecastaphylla* was in accordance with the results of Vatanparast et al. (2013) (based on ITS sequences) and Ribeiro et al. (2007) (based on ITS and *trnL* sequences). Further, according to Vatanparast et al. (2013), section *Selenolobia* and section *Dalbergia* are considered as non-monophyletic in nature however, our results indicated the monophyly of these two sections. Ribeiro et al. (2007) and Vatanparast et al. (2013) et al. 2013 suggested the inflorescence and fruit type as a sources of synapomorphies for classifications of *Dalbergia*.

Our preliminary analysis of *Dalbergia* agrees with the monophyletic nature of the genus. However, more work needs to be done with more number of samples and different markers covering broader genomic region to comment on the relationship between species and biogeography.

Section 2: Genetic diversity studies in *Symplocos* species

Publications:

- **Evaluation of genetic variability in *Symplocos laurina* Wall. from two biodiversity hotspots of India (2014) Plant Systematics and Evolution (DOI 10.1007/s00606-014-1046-4)**
- **Effect of Anthropogenic Activities on Patterns of Genetic Diversity in *Symplocos racemosa* Roxb. From Western Ghats, India (2014) International Journal of Biodiversity (http://dx.doi.org/10.1155/2014/296891)**

S 2.1 India: A Megadiverse Country

The concept of megadiverse countries was proposed by Russell Mittermeier in 1988, to prioritise the conservation activities. India is among the 17 megadiverse countries of the world (as identified by United Nations Environment Programme). It is rich in biodiversity and accounts for 7.8% of the world's species. It is also diverse in terms of languages, religions and demographics. Figure S2.1 illustrates the megadiverse countries of the world.

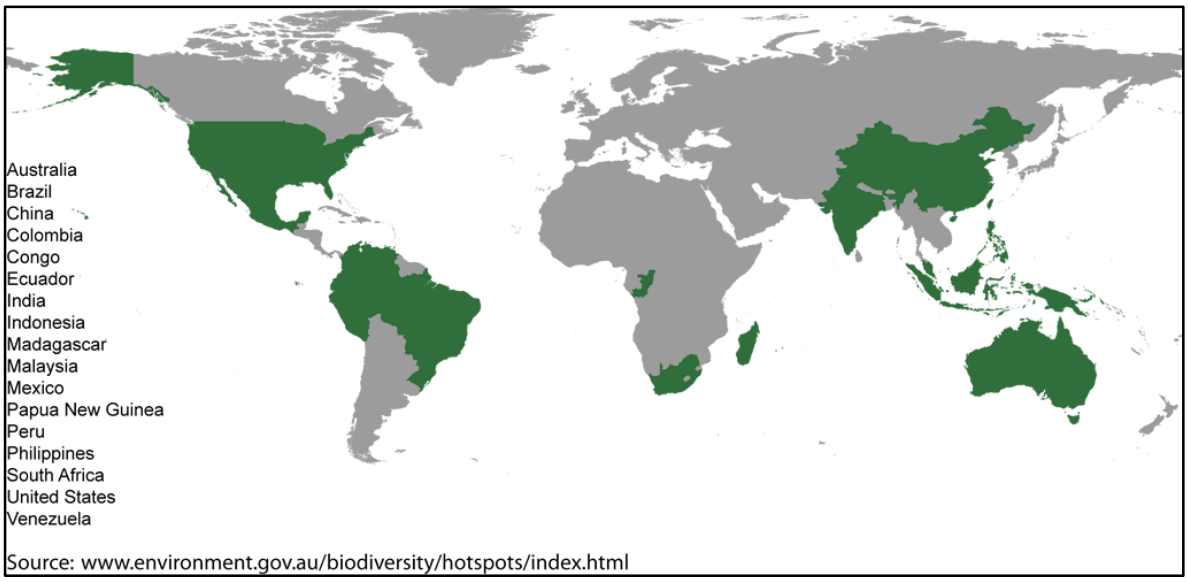


Figure S2.1: 17 megadiverse countries on a world map

India is the world's seventh and Asia's second largest country with an area of 3,287,263 km^2 surrounded by 7,516 km of coastline. The mainland lies between 8°4' to 37°6' N latitudinally and 68°7' to 97°25' E longitudinally. The climate of India ranges from tropical in southern India to temperate in the Himalayas. Rains usually fall between June and October from the South-West, with drier winds from the North between December and February followed by dry and hot climate from March to May (http://ces.iisc.ernet.in/hpg/cesmg/indiabio.html). India is divided into four relatively well defined regions - the Himalayan Mountains, the Gangetic river plains, the southern (Deccan) plateau, and the islands of Lakshadweep, Andaman and Nicobar (http://ces.iisc.ernet.in/hpg/cesmg/indiabio.html). Biogeographically, India is situated at the junction of the Afro-tropical, Indo-Malayan and Palearctic domains (http://nbaindia.org/undp/) describing 10 biogeographic zones and 25 provinces (Rodgers

and Panwar 1988). However, the revised version described 26 provinces (Table S2.1) (Rodgers et al. 2002).

Table S2.1: Biogeographic zones in India

Biogeographic zones	**Biotic provinces**
Trans-Himalaya	Ladakh mountains, Tibetan plateau
Himalaya	Northwest, West, Central and East Himalayas
Desert	Thar, Kutch
Semi-arid	Punjab plains, Gujarat Rajputana
Western Ghats	Malabar plains, Western Ghats
Deccan Peninsula	Central highlands, Chhota-Nagpur, Eastern highlands, Central Plateau, Deccan South
Gangetic plains	Upper and lower Gangetic plains
Coast	West and East coast, Lakshadweep
North-East	Brahmaputra valley, Northeast hills
Islands	Andaman and Nicobar

This classification was carried out on the basis of altitude, moisture, topography, rainfall, etc. and the wildlife protected areas are planned on the basis of these biogeographic zones (http://thewesternghats.in/node/28). Human activities causing exploitation of resources restricted in some of these areas to conserve the biodiversity therein. These protected areas of India occupy 160499.31 km^2 area (4.88% of total geographic area) (as on 06 October, 2015) (http://www.wiienvis.nic.in/Database/Protected_Area_854.aspx). The forest cover of India is 6,97,898 km^2 (21.23% of total geographic area) which is further divided into very dense forest (83,502 km^2), moderately dense forest (3,18,745 km^2), open forest (2,95,651 km^2) and mangroves (4,628 km^2) (State of Forest Report, 2013) (http://fsi.nic.in/details.php?pgID=mn_93).

S 2.2 Biodiversity Profile of India

The term "Biodiversity hotspot" was coined by the British biologist Norman Myers (Myers 1988) to define a biogeographic region with a significant reservoir of biodiversity, containing a remarkable level of plant endemism that is under threat from humans. To qualify as a biodiversity hotspot, the prerequisites are: the region must contain at least 1,500 species (>0.5% of the world's total) of vascular plants as endemics, and it must have lost at least 70% of its primary vegetation. Conservation International in 1999 identified 25 biodiversity hotspots all over the world which were raised to 34 presently. The Asia-Pacific region harbours eight hotspots, of which three regions exist in India and are described below.

S 2.2.1 The Eastern Himalayas region

The Himalayas have originated due to the movement of the Indian tectonic plate into the Asian plate and have complex topography. The Eastern Himalayas (EH) region ranges from central Nepal to Myanmar encompassing Bhutan, Tibet (China), Sikkim, North Bengal, and North-Eastern India (NEI) (Figure S2.2). It also includes the world's highest mountain, Mt. Everest. The region is geologically young with high altitudinal variation from less than 300 m (tropical lowlands) to more than 8,000 m (high mountains) which has led to a variety of vegetation patterns. Largely, vegetation types in the EH have been categorized as: a) tropical, b) sub-tropical, c) warm temperate, d) cool temperate, e) sub-alpine and f) alpine types (Chettri et al. 2010). There are approximately 7,500 flowering plants, 700 orchids (12 endangered species, 16 vulnerable species and 31 near threatened species), 58 bamboo species [25 species extremely rare (Thomas et al. 1998)], 64 citrus species, 28 conifers, 500 mosses, 700 ferns and 728 lichens recorded from this region. The EH biodiversity hotspot is rich in endemic flora and many species are used for medicinal and edible purposes and have commercial value. It is also the centre of diversity for *Rhododendron*, *Primula* and *Pedicularis* (Chettri et al. 2010; Banu 2008). However, evaluating the consequences of climate change in the EH region is a big challenge. Various methods are set to throw light on the potential impacts and vulnerabilities of biological diversity due to climate change.

S 2.2.2 Indo-Burma region

The Indo-Burma region was also formed by the tectonic processes, which were responsible for the formation of Himalayas. The topography of the Indo-Burma hotspot

is complex. The region occupies over 2 million km^2 of tropical Asia, ranges from Eastern Bangladesh to Malaysia and comprises NEI (Assam, Manipur, Meghalaya, Mizoram, Nagaland and Tripura), Myanmar, the southern part of China's Yunnan province, Lao People's Democratic Republic, Cambodia, Vietnam and Thailand (Figure S2.2).

Figure 2.2: The map of India with Himalaya and Indo-Burma biodiversity hotspot (Source: http://www.cepf.net/SiteCollectionImages/Maps/482_Himalaya_Indo_Burma_map.jpg)

Because such a huge area is spread across different countries, a diverse climate and habitat pattern has been recorded for this region. It is a home to many primate species and some freshwater turtle species are endemic to this region. About 13,500 plant species are reported to be present in the Indo-Burma hotspot with at least half of them being endemic (Dijk et al. 1999). Zingiber officinale, many orchid species, as well as commercially valuable teak species are native to this area (Banu 2008); (http://www.biodiversityofindia.org/index.php?title=Biodiversity_hotspots_in_India#cite_note-24). Among the 34 hotspots, the Indo-Burma region has the largest human population (over 315 million) (Mittermier et al. 2004), creating a tremendous pressure on the natural ecosystems of the region. To satisfy the needs of the increasing population, expansion of agriculture (for plantation of crops such as rice, rubber and oil palm),

infrastructure development (especially roads and hydropower dams), timber extraction by destructing forest areas, as well as illegal trade in wildlife have made it the most threatened biodiversity hotspot (Tordoff et al. 2012).

S 2.2.3 The Western Ghats and Sri Lanka region

The WG is among the first 18 biodiversity hotspots identified by Myers et al. (2000). The WG are much older than Himalayas in terms of geological age. The Indian peninsula was a part of the Gondwana supercontinent, which later split from it and drifted over the Reunion Hotspots localised volcanic centres and then resulted in uplift of the WG. This event took place some 120-130 mya (Banu 2008). The WG comprises southwestern region of India (chain of unbroken hills except Palghat gap for around 1600 km that run along the western edge of peninsular India, parallel to Arabian Sea coast) and the highlands of southwestern Sri Lanka (Figure S2.3). These two regions are separated by 400 km; however, they are very similar in terms of geology, climate and evolutionary history (http://bsienvis.nic.in/files/Biodiversity%20Hotspots%20in%20India.pdf). The WG encompasses six states in India namely, Gujarat, Maharashtra, Goa, Karnataka, Kerala and Tamil Nadu (MoEF et al. 2011-2012) and possesses various climatic conditions supporting diverse vegetation types (Gunawardene et al. 2007). The region shows high species diversity with high levels of endemism (Daniels 2001). The vegetation of WG is classified into four major types:

1. Moist tropical forest (tropical evergreen and semi evergreen forests, tropical moist deciduous forests and littoral and swamp forests)
2. Dry tropical forests (tropical dry deciduous forests, tropical thorn forests)
3. Montane subtropical forests (subtropical broad leaved hill forests)
4. Montane temperate forests (montane wet temperate forests)

Along with the above mentioned forests, tropical dry deciduous forests, scrubs, jungles, sholas, montane grasslands, savannas, peat bogs and *Myristica* swamps also occur along the WG (Rodgers and Panwar 1988; Champion and Seth 1968) http://shodhganga.inflibnet.ac.in/bitstream/10603/12580/7/07_chapter%203.pdf).

Approximately 4000 species of flowering plants (27% of the country's total species) are known from the WG of which 56% are endemic (Nair and Daniel 1986; Rao et al. 2013). However, WG and Sri Lanka are facing tremendous pressure on the forests because of large population. Further, the intense impact has been observed on forests as

they have been lost drastically (from original area of 1,82,500 km^2 to only 12,450 km^2) due to demands for timber and agricultural land (http://www.cnrs.fr/inee/recherche/fichiers/Biodiversite_hotspots.pdf); (Gunawardene et al. 2007).

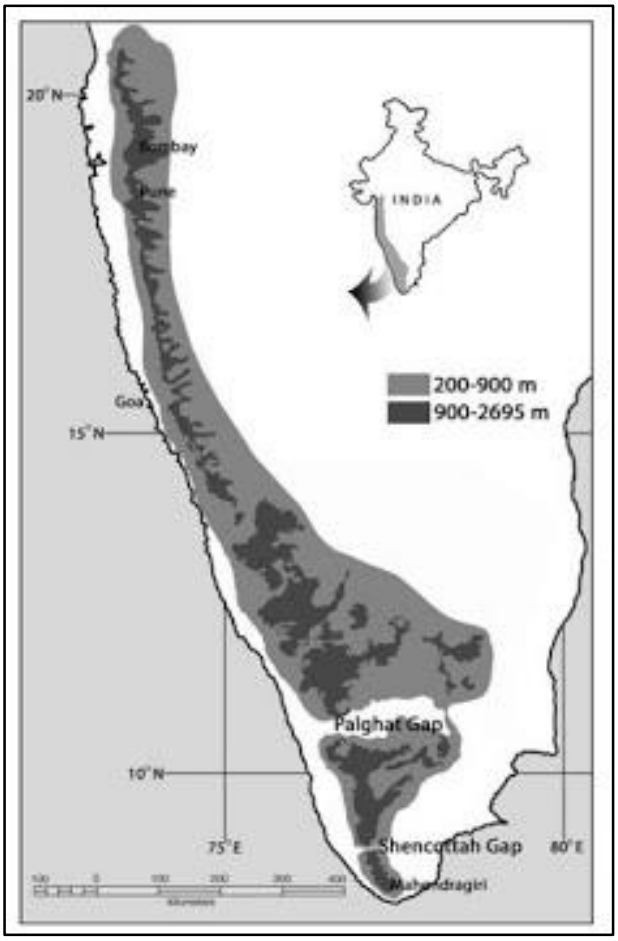


Figure S2.3: Map of India with Western Ghats and Sri Lanka biodiversity hotspot

(Source:http://novataxa.blogspot.in/2011/07/herpetology-2009-10-new-aorchestes.html)

S 2.3 Disjunct Distribution of Some Plant Species in India

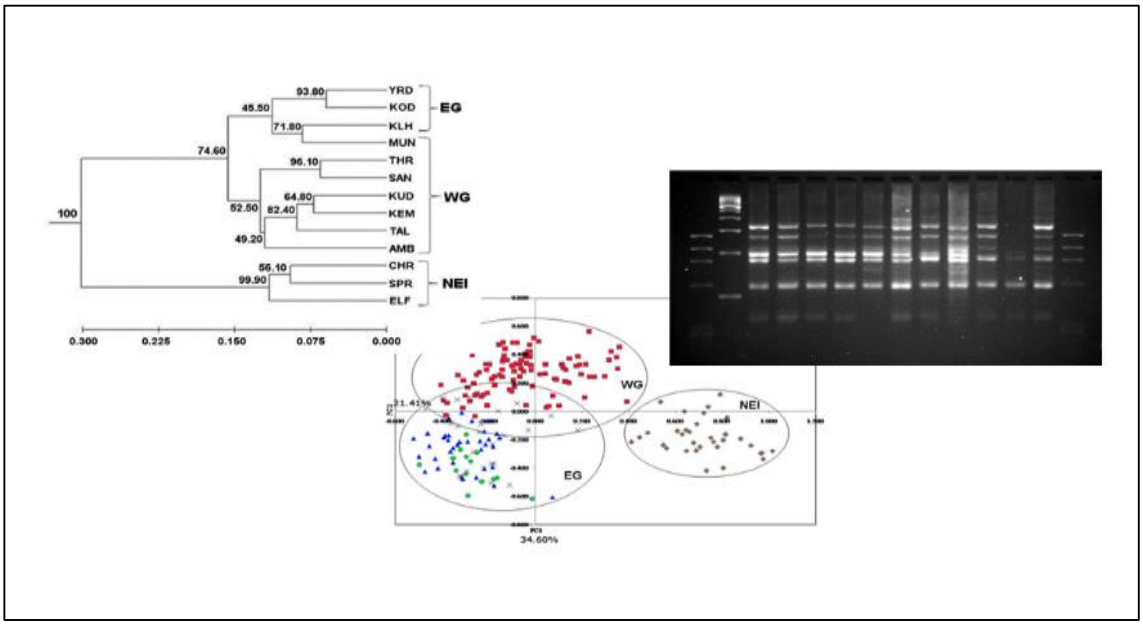
The Indian subcontinent is divided into two climatic zones, namely, wet zone [comprises north east India (NEI), South India (SI)] and dry zone (Central India). It was observed that some plant species (*Gaultheria fragrantissima, Rhododendron arboretum, Mahonia leshenaultii, Ternstroemia japonica, Hepercum hookerianum, Thalictrum javanicum, Cotoneaster buxifolia, Parnassia wightiana, Lonicera ligustrina* etc.) are found to be distributed in wet zone (wet evergreen forest of NEI and SI though they are geographically far) but absent in dry zone (Central India) (Banu 2008). Along with plants, mammals, birds, freshwater fishes, amphibians and reptiles, insects are also found to be

disjunctly distributed in the Indian subcontinent (Karanth 2003). This disjunct distribution pattern is explained by Satpura hypothesis (Hora 1949) and other alternative routes such as Eastern Ghats (Abdulali 1949), the Brij area (Dilger 1952) and the Aravalli range (Mani 1974).

Among various plant species, *Symplocos laurina* is the one which is also disjunctly distributed in India (Banu 2008) and hence, the studies were conducted in our laboratory to throw light on its phylogeography. Previous studies were conducted to understand the geographical pattern of distribution (Banu et al. 2010b) and to describe the effect of glaciation and to analyse the population structure (Banu et al. 2010a). In continuation of the previous work, it was planned to undertake DNA barcoding of *Symplocos* species from the WG. However, as detailed in the Genesis of the Thesis, out of 68 species of *Symplocos* reported in India only two species, namely, *S. laurina* and *S. racemosa* could be collected in different regions. In spite of serious efforts to collect other species, it was not possible to collect other reported species. Hence, the studies were restricted to genetic diversity analysis.

Family Symplocaceae is monogeneric containing the genus *Symplocos* (Cronquist 1981; Holub 1999; Thorne 2000) which is a woody, mostly evergreen angiosperm and relatively a large genus with approximately 290 species reported in the world and 68 from India (Ahmad et al. 2005). It is mainly found in humid tropical montane forests (Fritsch and Al 2015). Genetic diversity studies of *S. laurina* and *S. racemosa* have been described in subsection 2A and 2B.

Section 2A: Genetic diversity studies in *Symplocos laurina* using ISSR markers



S 2A.1 Introduction

Symplocos laurina is one of the important members of biodiversity hotspots in India and represented in the shola forests of Western Ghats (WG), Eastern Ghats (EG) and Northeast India (NEI) (Meher-Homji 1975; Banu et al. 2010a; Banu et al. 2010b). *S. laurina* is a small tree reaching up to 7 m in height. The leaves are lanceolate, elliptic or oblong with serrate margins. The inflorescence is branched axillary spike with yellowish white fragrant flowers. Fruits are drupaceous, purple in colour, globose and ribbed (Figure S2A.1). The flowering season of *S. laurina* is from September to December (Almeida 1990; Banu et al. 2010a). The seed dispersal of this species is endozoic by birds and bats (Meher-Homji 1975). *S. laurina* propagates by clonal and sexual reproduction (Zhang et al. 2005) and inhabits a new environment first by invasion with grown-up plantlets and then quickly occupies the space through ramets, thus making it one of the pioneering plants in the establishment of shola ecosystem (Meher-Homji 1975). However, regeneration of *S. laurina* is very difficult in open areas due to its susceptibility to winter frost, droughts and fires. These plants typically occur near water source and are always associated with the species of *Syzygium* (Deshpande et al. 2001). *S. laurina* has many medicinal properties and the yellow dye prepared from its leaves is used as a mordant in textile industry, while the wood is used for fuel (Hore 1990; Warrier et al. 1996).

As forests are being replaced with crop plantations at many places to satisfy the needs of increasing population; forests areas are diminishing or shrinking at an alarming rate in India as well as other parts of the world (Puyravaud et al. 2010). As a result, the shola ecosystems with their constituent plant species such as *Symplocos* are rapidly disappearing from these regions. Various molecular markers have been utilized to screen, characterize and evaluate genetic diversity of several forest plant species. Inter simple sequence repeat (ISSR) markers have emerged as the markers of choice for such studies, as they are highly sensitive, reproducible and cost effective compared to other PCR-based markers (Reddy et al. 2002). As prior information of DNA sequence is not required; ISSR markers are technically simpler than other markers such as simple sequence repeats (SSRs). Further, ISSR markers can detect very low levels of genetic variation effectively, therefore they are ideal for population genetic studies of plant species (Zietkiewicz et al. 1994). These techniques have been successfully used in genetic diversity studies in many forest plants such as *Primula obconica* (Nan et al. 2003), *Gmelina arborea* (Naik et al.

2009), economically important agarwood plant (Banu et al. 2015) as well as medicinal plants such as *Glycyrrhiza uralensis* (Yao et al. 2008), *Salvia miltiorrhiza* (Song et al. 2010), *Cunila* species (Agostini et al. 2008) and *Pinus nigra* (Rubio-Moraga et al. 2012). In the present study, we used the ISSR markers to study the genetic diversity of 13 *S. laurina* populations from three ecological regions in India namely, WG, EG and NEI for broader genome coverage, as they are neutral and randomly distributed in the genome.

S 2A.2 Materials and Methods

Plant tissue collection

For this study, we collected a total of 208 accessions of *S. laurina* from 13 populations. Among these, seven populations were collected from WG, three from EG and three from NEI (Table S2A.1; Figure S2A.2). The collection sites were authenticated based on herbarium collections of Botanical Survey of India (BSI) (Western Circle, Pune, India). Among population, distances were more than 10 Km as suggested by Nybom and Bartish (2000) whereas a distance among the individuals in each population was maintained around 10 m to avoid sampling from the same clonal individuals.

DNA isolation and PCR amplification

The DNA was isolated from leaf tissues as explained in materials and methods of section one (Richards et al. 1994). The extraction method is detailed in materials and methods of section 1. A set of 100 ISSR primers, procured from the University of British Columbia (UBC) (Vancouver, Canada), was screened for amplification using a small subset of 10 representative accessions. The primers generating well-resolved, non-ambiguous and reproducible patterns with these representative samples in two independent experiments were only used for further studies with all the accessions. PCR was carried out in 25 µL reaction volume, using 20 ng template DNA, 10X PCR buffer (10 mM Tris–HCl, pH 8.3; 50 mM KCl, 1.5 mM $MgCl_2$), 0.4 mM spermidine, 0.1 mM dNTPs, 0.3 µM primer and 0.48U *Taq* DNA polymerase (Bangalore Genei, India). Initial denaturation was carried out at 94°C for 5 min followed by 45 cycles of denaturation at 94°C for 30 sec, annealing at 50°C for 45 sec and extension at 72°C for 2 min with a final extension of 72°C for 5 min. The amplified products were resolved on 1.5% agarose gel in 0.5X Tris acetate EDTA (TAE) buffer. The amplicons were detected by staining with GelRed (Biotium, USA) and photographed using the Dyversity gel documentation system (Syngene, UK).

Figure S2A.1: Photograph of *S. laurina* displaying habit, flowers and fruits

Photo credit: Dr. Sofia banu

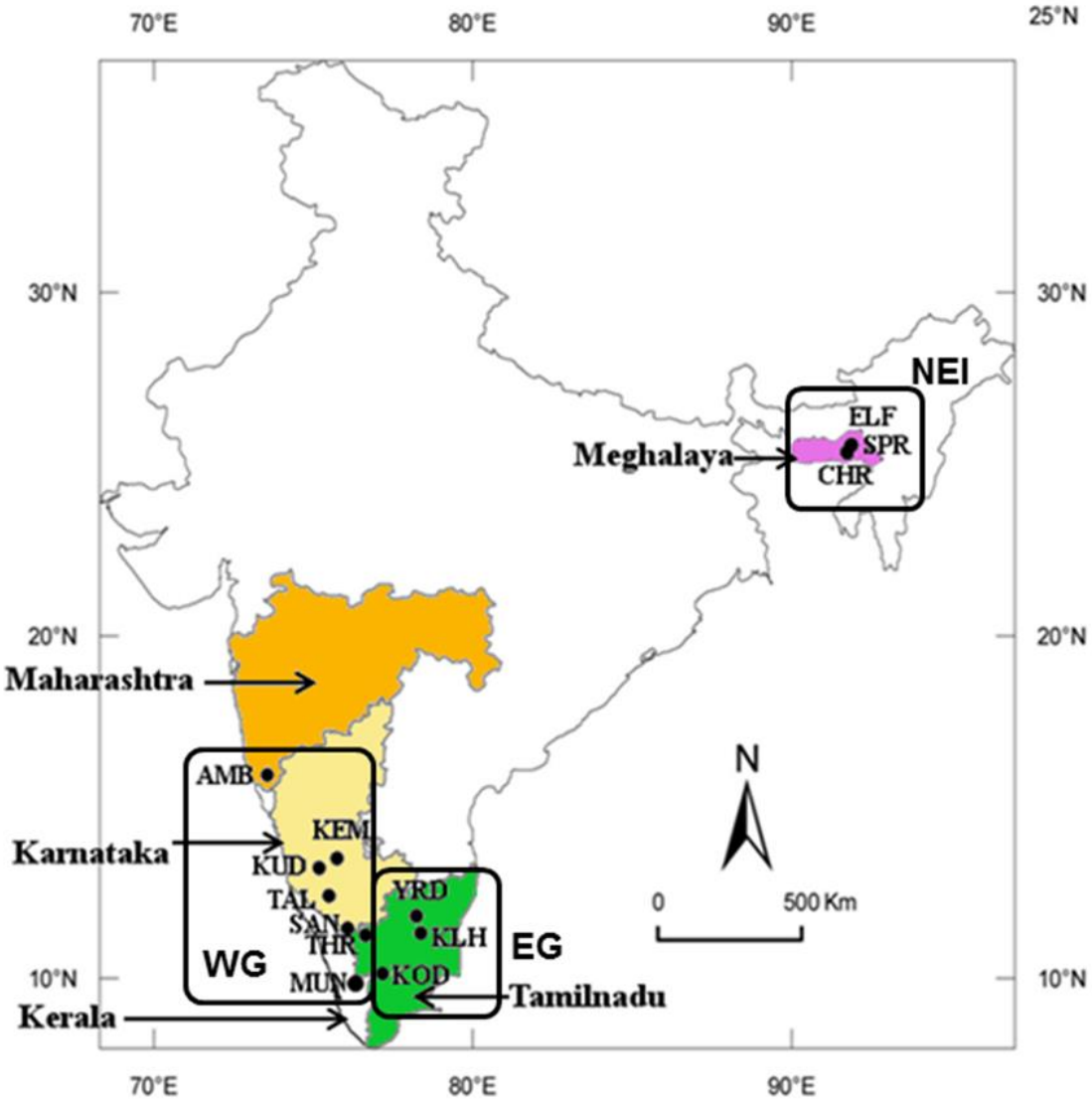


Figure S2A.2: Locations of the 13 populations of *S. laurina* from Western Ghats, Eastern Ghats and North Eastern India on geographical map of India

Table S2A.1: Sampling details of *S. laurina* populations

Region	Population site	Population code	Individuals sampled	Latitude	Longitude	Altitude (m)
WG	Amboli	AMB	29	E°74.06654	N° 15.99816	825
	Kudremukh	KUD	18	E° 75.18670	N° 13.20770	836
	Talakaveri	TAL	14	E° 75.48790	N° 12.38866	1243
	Kemmangundi	KEM	13	E° 75.74579	N° 13.47867	1369
	Sandynallah	SAN	11	E° 76.06321	N° 11.43310	2227
	Thaishola Road	THR	14	E° 76.62558	N° 11.24087	1877
	Munnar	MUN	14	E° 77.15291	N° 11.81318	1606
EG	Kodaikanal	KOD	10	E° 77.47947	N° 10.23795	2131
	Kolli hills	KLH	24	E° 78.36096	N° 11.30846	1303
	Yercaud	YRD	27	E° 78.22367	N° 11.81318	1293
NEI	Shillong Peak Road	SPR	13	E° 91.85528	N° 25.53548	1903
	Cherrapunji Road	CHR	13	E° 91.73666	N° 25.33120	1661
	Elephanta Falls	ELF	8	E° 91.82544	N° 25.53368	1754
	Total / Average		**208**			

Data analysis

The ISSR amplicons were treated as dominant markers and each locus was considered as bi-allelic with one amplifiable and the other null allele. Only the bands that could be unambiguously scored as present (1) or absent (0) were considered for the analysis. The bands amplified in <15% of the accessions were considered as rare bands, those amplified in up to 70% of accessions were considered as shared bands, while the bands amplified in more than 70% of accessions were considered as similar bands (Grativol et al. 2011). GenAlEx ver. 6.4 (Peakall and Smouse 2006) was used to plot principal co-ordinate analysis (PCA) and calculate analysis of molecular variance (AMOVA) to assess the overall distribution of diversity among regions and among populations within regions. The genetic diversity within and among the populations was estimated in terms of *h* (Nei's genetic diversity) , H_T (total heterozygosity), H_S (heterozygosity within-population), G_{ST} (degree of population differentiation), D_{ST} (heterozygosity among-populations) (Nei 1973) and *Nm* (number of migrants per generation or gene flow) (Slatkin and Barton 1989). All the above parameters were calculated using POPGENE ver. 1.31 (Yeh et al. 2000). The comparison among the populations was made using the distance matrix based on Nei's genetic distance (Nei 1972), which was then used to build an unweighted pair group method with arithmetic mean (UPGMA) dendrogram after bootstrapping 1,000 times using the TFPGA (Tools For Population Genetic Analyses) software (Miller 1997). A similarity matrix within-population was constructed using the WINDIST package of WINBOOT software (Yap and Nelson 1996) with Jaccard's coefficients.

The polymorphism information content (PIC), which measures the variability at specific loci, was calculated using the formula 1 - p2 - q2, where p is the frequency of present bands and q is the frequency of no bands (Ghislain et al. 1999; Rajwade et al. 2010). ISSR primer index (SPI) was then calculated by summing up the PIC values of all the loci amplified by the same primer. The probability of identical match (PI) of amplification profiles for two randomly selected genotypes was also calculated for each primer using the formula $PI = (XD)^n$, where XD represents the average similarity index for all pairwise comparisons and n is the average number of fragments amplified by a single primer (Wetton et al. 1987). Mantel test was performed to calculate the correlation between genetic distances (D) and geographical distances (in Km) among the populations

using XLSTAT ver. 2012.2.03 (Addinosoft; http://www.xlstat.com/en/) with 10,000 permutations.

S 2A.3 Results

ISSR profiling

From the set of 100 ISSR primers screened, 19 primers produced amplifications. On the basis of their clear amplification profiles and reproducibility, eight primers were selected for further study. Of these, seven primers were with dinucleotide repeats and one primer had tetranucleotide repeats (Table S2A.2). These eight primers together generated 103 loci resulting in 8,111 amplicons with molecular size range of 500–2,100 bp from 208 accessions. The primers UBC835 and UBC880 amplified the highest number of polymorphic loci (16). All the bands produced by all the primers were polymorphic (100% polymorphism). On an average, each primer amplified 13 loci. The number of rare bands (NRB) ranged from 30 (UBC845) to 135 (UBC880), while the number of shared bands (NSB) ranged from 343 (UBC815) to 726 (UBC845). Likewise the number of similar bands (NSiB) ranged from 0 in UBC824 to the highest of 857 in UBC859 (Table S2A.3). A maximum of 1,556 bands were amplified by UBC859, while UBC824 amplified only 693 bands. On an average, the eight ISSR primers produced 1,014 bands, of which 6.58% were rare bands, 55.96% were shared bands and 37.46% were similar bands. A representative amplification pattern with ISSR primer UBC852 has been depicted in Figure S2A.3.

Genetic polymorphism and distance

Of the13 populations of *S. laurina*, CHR, ELF, AMB, KUD, KLH and YRD showed very high polymorphism [percentage of polymorphic bands (PPB) >75%]; SPR, TAL, KEM, THR and MUN showed intermediate polymorphism (61–71% PPB), while the remaining two populations, SAN and KOD showed less polymorphism (52–59% PPB) (Table 2A.3). Interestingly, the altitudes of the first two groups were in a range of 800–1,900 m, while for the third group (SAN and KOD) it was above 2,100 m. To check the effect of altitudinal variation on genetic diversity, Pearson's correlation coefficient was calculated between altitude and genetic diversity (*h*), which revealed significant negative correlation (r =-0.59).

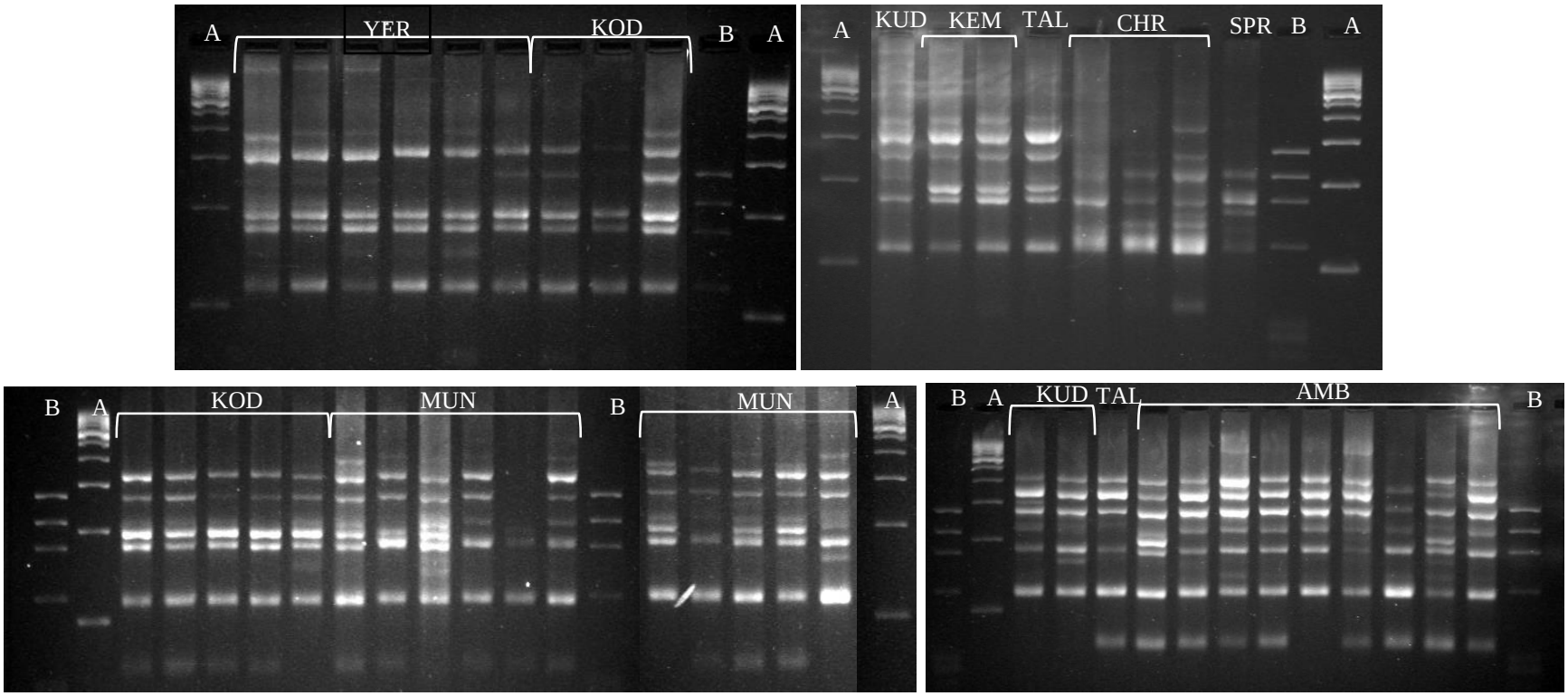


Figure S2A.3: Gel picture represents the profile obtained by UBC852 with few representative samples. YER- Yeracaud, KOD- Kodaikanal, KUD- Kudremukh, KEM- Kemmangundi, TAL- Talkaveri, MUN- Munnar, AMB- Amboli, CHR- Cherapunji, SPR- Shilong Peak Road, A- mixture of 100 bp and 500 bp ladder, B- ØX ladder

Table S2A.2: Primers utilized for ISSR amplification and genotype construction

Primer	Sequence (5'→3') R= (A, G); Y= (C,T)	No. of Polymo-rphic bands	SPI	PI	ISSR genotypes detected												
					AMB (29)	**KUD (18)**	**TAL (14)**	**KEM (13)**	**SAN (11)**	**THR (14)**	**MUN (14)**	**KOD (10)**	**KLH (24)**	**YRD (27)**	**SPR (13)**	**CHR (13)**	**ELF (8)**
UBC815	$(CA)_8T$	10	3.37	3.3×10^{-2}	11	12	11	10	8	12	7	3	8	10	11	11	**8**
UBC824	$(TC)_8G$	10	3.68	2.1×10^{-2}	11	11	9	10	6	9	7	8	21	19	4	6	5
UBC835	$(AG)_8YC$	16	5.40	4.0×10^{-3}	27	11	6	9	7	13	6	3	19	19	12	11	**8**
UBC845	$(CT)_8RG$	14	5.02	5.6×10^{-3}	14	14	10	10	10	12	7	8	17	13	9	11	3
UBC852	$(TC)_8RA$	9	3.16	4.4×10^{-2}	13	10	9	6	7	6	11	5	15	6	10	8	**8**
UBC854	$(TC)_8RG$	13	3.98	2.2×10^{-2}	12	12	9	7	3	8	5	7	15	11	9	9	7
UBC859	$(TG)_8RC$	15	4.99	6.9×10^{-3}	23	17	12	9	9	12	**14**	9	22	19	12	**13**	**8**
UBC880	$(GGAT)_4$	16	5.08	5.3×10^{-3}	19	16	12	9	5	13	13	9	17	21	5	6	7
Average		12.88	4.34	$1.8\text{x}10^{-2}$	16.25	12.88	9.75	8.75	6.88	10.63	8.75	6.50	16.75	14.75	9.00	9.38	6.75
				PPB (71.40)	81.55	83.50	68.93	61.17	58.25	67.96	64.08	52.43	85.44	78.64	70.87	79.61	75.73
				h (0.226)	0.263	0.235	0.227	0.205	0.177	0.224	0.199	0.170	0.271	0.231	0.230	0.253	0.256

SPI, ISSR primer index; PI, Probability of identical match; Sample sizes in parentheses; PPB, Percent polymorphic bands; *h*, genetic diversity. Unique banding pattern of each individual in a particular population is indicated in bold type. The average values of PPB and *h* in parentheses.

Based on PPB, genetic distances were calculated by pairwise comparisons between populations and corresponding geographic distances were used for correlation. Furthermore, Mantel test showed significant positive correlation [r(AB) = 0.925, $P < 0.0001$ (two-tailed), a = 0.01, 10,000 permutations] between genetic distance and geographic distance of the two data matrices. The SPI, which is indicative of the efficiency of the marker system, ranged from 3.16 (UBC852) to 5.40 (UBC835). Similarly, PI values for each primer ranged from 6.9 X 10^{-3} (UBC859) to 4.4 X 10^{-2} (UBC852) (Table S2A.3).

Cluster and principal co-ordinate analysis

Genetic distances among these 13 populations were used to build a UPGMA dendrogram using TFPGA software and the individuals were grouped according to the populations (Figure S2A.4). The NEI populations showed clear separation from the WG and EG populations. All the WG populations (KUD, KEM, TAL, AMB, SAN and THR) except MUN formed one cluster, while EG populations (YRD, KOD and KLH) formed another cluster along with MUN of WG. Within the EG cluster, YRD and KOD were closer and MUN and KLH formed a separate cluster. To assess the level of intra-population genetic structuring, multivariate ordering was performed based on pairwise genetic distances between all the individuals of 13 populations. The first two co-ordinates explained 34.60 and 21.41% of total variability (Figure S2A.5). The PCA showed similar pattern of clustering as in the dendrogram, which further confirmed the grouping of individuals. All the genotypes from NEI region and WG region (except MUN) formed two independent clusters. However, the EG cluster also included all the genotypes from MUN population (WG). Interestingly, the KLH population from EG showed distribution in WG as well as in EG clusters.

Genetic structure

The total heterozygosity (H_T) was 0.3407 when all the three regions were considered together, while it was in the range of 0.2690–0.2934 when considered separately. The average heterozygosity (H_S) values for all the three regions combined as well as separately were in the range of 0.2186–0.2465 and were lower than the corresponding H_T values (Table S2A.4), which indicated less population variability.

Table S2A.3: Information of amplification profiles of the ISSR primers

Primer	TNB	NPB (%)	NRB (%)	NSB (%)	NSiB (%)
UBC815	892	892 (100.00)	90 (10.09)	343 (38.45)	459 (51.46)
UBC824	693	693 (100.00)	33 (4.76)	660 (95.24)	0 (0.00)
UBC835	1166	1166 (100.00)	86 (7.38)	580 (49.74)	500 (42.88)
UBC845	1111	1111 (100.00)	30 (2.70)	726 (65.35)	355 (31.95)
UBC852	715	715 (100.00)	63 (8.81)	479 (66.99)	173 (24.20)
UBC854	901	901 (100.00)	44 (4.88)	522 (57.94)	335 (37.18)
UBC859	1556	1556 (100.00)	53 (3.41)	646 (41.52)	857 (55.08)
UBC880	1077	1077 (100.00)	135 (12.53)	583 (54.13)	359 (33.33)
Total	8111	8111 (100.00)	534 (6.58)	4539 (55.96)	3038 (37.46)
Minimum	693	693 (100.00)	30 (4.33)	343 (49.49)	0 (0.00)
Maximum	1556	1556 (100.00)	135 (8.68)	726 (46.66)	857 (55.08)
Average	1014	1014 (100.00)	67 (6.58)	567 (55.96)	380 (37.46)

TNB, Total number of bands; NPB, Number of polymorphic bands; NRB, Number of rare bands; NSB, Number of shared bands; NSiB, Number of similar bands

The populations from EG and NEI showed lower G_{ST} values indicating less differentiation in these two regions as compared to WG. The estimated gene flow within WG, EG and NEI was 1.3149, 2.3927 and 2.6241, respectively; while it was comparatively low (0.9890) among these three regions.

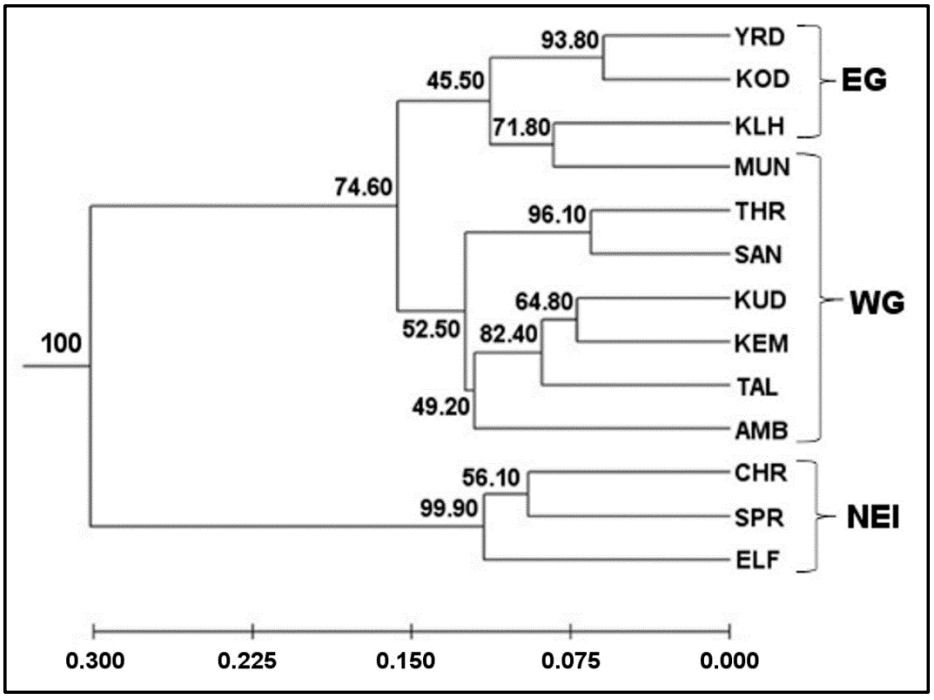


Figure S2A.4 UPGMA dendrogram of genetic distances (Nei 1972) between populations of *S. laurina*. Numbers above the branches represent bootstrap values from 1000 replicates

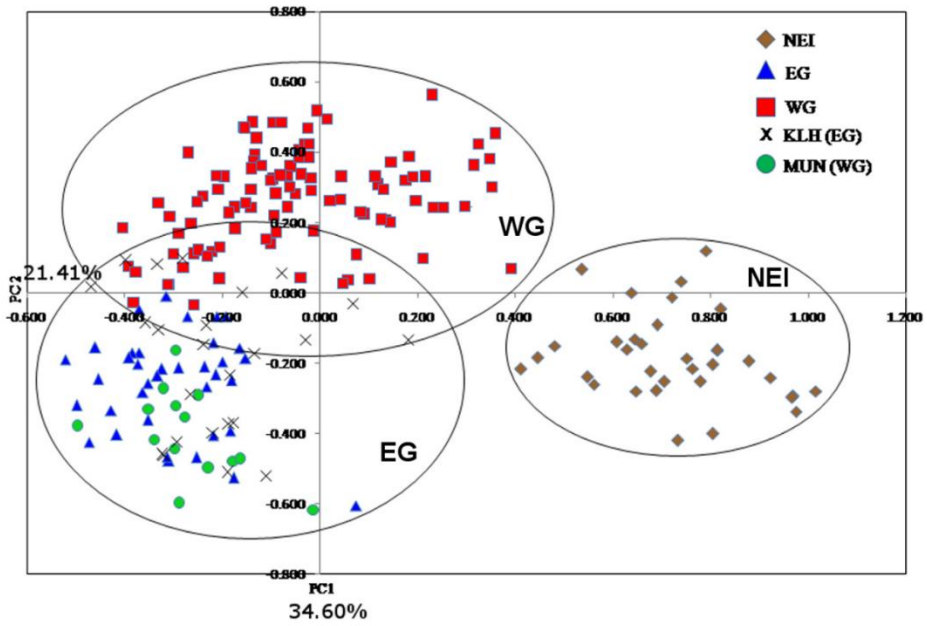


Figure S2A.5 Principal co-ordinate analysis among 208 individuals from 13 populations of *S. laurina*

Table S2A.4: Genetic diversity analysis in 13 populations of *S. laurina*

Region	h	H_T	H_S	G_{ST}	D_{ST}	Nm
	Mean±SD	Mean±SD	Mean±SD			
WG	0.2990 ±0.1544	0.2932 ±0.0240	0.2186 ±0.0128	0.2546	0.0746	1.3149
EG	0.2785 ±0.1545	0.2690 ±0.0243	0.2241 ±0.0174	0.1671	0.0449	2.3927
NEI	0.2900 ±0.1587	0.2934 ±0.0246	0.2465 ±0.0185	0.1600	0.0469	2.6241
Overall	0.3367 ±0.1288	0.3407 ±0.0161	0.2263 ±0.0080	0.3358	0.1144	0.9890

h, Nei's genetic diversity; H_T, total heterozygosity; H_S, heterozygosity within population; G_{ST}, degree of population differentiation; D_{ST}, heterozygosity between populations; *Nm*, number of migrants per generation or gene flow

Partitioning of genetic diversity

The total genetic variation was partitioned into variation among groups, among-populations within groups and within-populations (P = 0.001). AMOVA was carried out considering three separate regions as WG, EG and NEI as well as only two regions as South India (SI) (WG + EG) and NEI. This analysis showed more genetic variation in comparison of two regions (23%) than three regions (15%) (Table S2A.5).

S 2A.4 Discussion

ISSR markers with unique profiles

Symplocos laurina is among the Pleistocene relict plant pool with pioneering characteristics involved in the establishment of new shola forest system or habitats in WG, making it an important plant species from the ecological perspective (Banu et al. 2010a; Banu et al. 2010b). In the present study, we employed ISSR markers to analyse the genetic diversity and population structure of *S. laurina* and also determined the efficiency of ISSR markers by evaluating the rare, shared and similar bands.

Table S2A.5: Analysis of molecular variance (AMOVA)

A – Three regions – WG, EG, NEI

Source	**df**	**SS**	**MS**	**VC**	**%**
Among Regions	2	483.687	241.843	2.810	15%
Among Populations	10	632.363	63.236	3.254	17%
Within Populations	195	2490.398	12.771	12.771	68%
Total	207	3606.447		18.836	100%

B – Two regions – SI (WG+EG), NEI

Source	**df**	**SS**	**MS**	**VC**	**%**
Among Regions	1	338.353	338.353	4.890	23%
Among Populations	11	777.697	70.700	3.613	17%
Within Populations	195	2490.398	12.771	12.771	60%
Total	207	3606.447		21.274	100%

df, degrees of freedom; SS, sum of Squares; MS, mean square; VC, variance components; %, total variance (percentage of total variance contributed by each component)

The detection of rare polymorphic bands was very useful in identifying a particular accession, while the shared and similar polymorphic bands described the resemblance among the accessions from different geographic regions (Tatikonda et al. 2009; Grativol et al. 2011). As UBC880 produced the maximum number of rare bands, it can be considered as the most suitable primer to identify a particular accession. Likewise, based on high SPI value, UBC835 was highly efficient in detecting polymorphism in these *S. laurina* populations. Similarly, UBC859 detected the highest number of unique genotypes from these populations compared to other ISSR markers and had the lowest PI value.

S. laurina indicates good differentiation within populations and within SI and NEI regions

In our study, 17% variation was detected among the populations of *S. laurina*, which was in the range of 10–20% of among-population genetic variation reported for outcrossing species (Hamrick and Godt 1989). Genetic differentiation among the populations was low as most of the variation (68%) was conserved within the populations. This level of variation is usually found in the long-lived, perennial and outcrossing species of late successional stages (Loveless and Hamrick 1984) and could be applied to *S. laurina*. Buckley et al. (2010) and Kremer et al. (2012) stated that increase in genetic variance resulted in improved adaptation to changes in local environment with improved gene flow, which was also true in our study. Slatkin (1987) suggested that low gene flow ($Nm \leq 1$) was responsible for substantial local differentiation leading to genetic drift, whereas higher gene flow ($Nm \geq 1$) was accountable for fitness of the populations. To improve the effect caused by inbreeding and to maintain the population fitness, 1–10 migrants per generation are required and this can restore the effects of drift and selection on genetic variation (Lopez et al. 2009; Blanquart and Gandon 2011). In *S. laurina*, the overall gene flow was slightly low (0.9890) compared to the threshold value. This could be due to the long distance between NEI and SI, which might be acting as a barrier to gene flow. However, the gene flow or the number of migrants per generation within WG, EG and NEI populations was higher (1.3149, 2.3927 and 2.6241, respectively), indicating more frequent exchange of genes within the regions, which is good to prevent genetic drift.

The coefficient of genetic differentiation among-populations (G_{ST}) is a statistical measure widely applied in studies of genetic differentiation. The G_{ST} value ≥0.25 is regarded as a threshold beyond which significant population differentiation occurs (Han et al. 2007). G_{ST} value considering all the *S. laurina* populations together was 0.3358, which indicated good genetic differentiation. The overall high G_{ST} values obtained could be mainly due to the total differentiation of NEI from SI populations (WG+EG combined) and partly due to the differentiation of EG populations (along with MUN) from the rest of *S. laurina* populations in SI. Similar G_{ST} value of 0.25 was reported in *Hagenia abyssinica* and ancient tea (*Camellia sinensis* var. *Assamica*) ($G_{ST} = 0.3911$) with ISSR analysis (Feyissa et al. 2007; Ji et al. 2011).

When the three regions (WG, EG and NEI) were considered separately (Table 2A.5), lower G_{ST} values were obtained in NEI and EG compared to WG, indicating less population

differentiation within NEI and EG populations. This suggested, seed dispersal mechanism and gene flow probably more efficient within these two regions as compared to WG. Alternatively, the lower G_{ST} values in NEI and EG populations could be attributed to the sampling locations within the three regions. Among the three regions, WG had seven populations, while EG and NEI had three populations each. Moreover, a large sample collection area (up to 732 Km) could be covered in WG with the broad range of latitude, longitude and altitude (Table 2A.2) as compared to narrow range of collection in EG (up to 192 Km) and NEI (up to 253 Km). More populations need to be collected and analysed from the EG and NEI regions.

Influence of geographic and topographic factors on genetic diversity in *S. laurina*

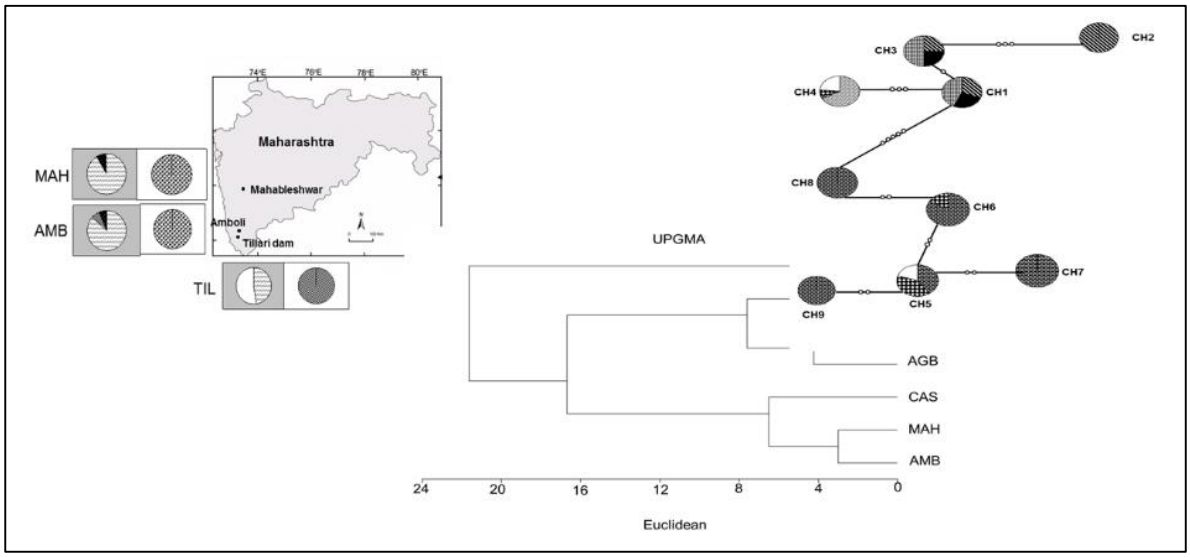
S. laurina is known to be an altitude indicator as it survives efficiently above an altitude of 1,000 m above mean sea level (msl) (Meher-Homji 1975). The altitude of the 13 collection sites ranged from 825 to 2,227 m (Table 2A.2). Interestingly, the negative Pearson's correlation suggested more genetic diversity (h = 0.205–0.271) in lower altitudinal populations compared to the populations at higher altitudes (h = 0.170–0.256) (Table 2A.2). This indicated the prevalence of natural selection of some genotypes over time, which was better adapted to higher altitudes. Natural selection operates by sieving of genetic variation found within-populations where the genetic diversity within-population is the driving force for evolutionary change.

Estimation of genetic diversity within and among-populations is the primary step in evolutionary studies. The genetic diversity within-populations of *S. laurina* (H_S = 0.2263) was close to the average values for long-lived species (H_S = 0.25) and outcrossing species (H_S = 0.27) as reported by (Nybom 2004) based on RAPD analysis. In WG, most of the populations showed clustering according to their geographical location except the MUN population, which clustered with the EG populations. The PCA displayed a similar pattern of clustering. The first axis separated the NEI populations from SI (WG and EG), thus explaining separation of different geographical regions, while the second axis explained the separation of EG and WG with the exception of MUN population. The grouping of MUN population with EG rather than WG populations, could be due to the presence of Palghat gap (a 30 Km discontinuity in WG), which separates MUN from other WG populations. The Palghat gap and the surrounding region have been referred as one of the main centres of

endemism in India (Ramesh et al. 1997). This discontinuity would have probably created a marked hindrance in gene flow in WG. The Palghat gap is of considerable interest as it has been found to disrupt species continuity and acts as a barrier to gene flow as observed in population genetics studies of fresh water fishes in WG (Dahanukar et al. 2004), Asian elephants in southern India (Vidya et al. 2005), Ichthyophis in WG (Gower et al. 2007) and endemic birds in sky islands of southern India (Robin et al. 2010).

Another interesting observation was with respect to the KLH population, which is geographically located in EG. Though in PCA, the individuals of this population were mainly clustered with EG, they were distributed close to the axis separating WG and EG clusters, indicating that the KLH population could be a link between WG and EG. Probably, the less xeric conditions across WG and EG allowed gene flow among them via the Biligirirangan hills, which links WG and EG near Nilgiri hills. In case of *S. laurina*, seed dispersal can occur through birds, bats (Meher-Homji 1975) and wind, which could be important for long-distance gene flow (Banu et al. 2010b). This differentiation could be due to an ineffective exchange of seed dispersal or gene flow between the two regions owing to the loss of evergreen forests linking them and increase in the aridity post-glaciation in the central Indian region.

Section 2B Effect of anthropogenic activities on patterns of genetic diversity in *S. racemosa*



S 2B.1 Introduction

Symplocos racemosa is a medicinally treasured species of genus *Symplocos* found in forests of WG with moderate sized tree. It is a height indicator inhabiting an altitude range of 400–1000 m. The leaves are, orbicular to elliptic, crowded at the end of branches, dark green and glabrous above, lanceolate and subcoriaceous (Kulkarni 1988; Yadav and Sardesai 2002). Male flowers are in pubescent spikes and fragrant with a foliaceous bract to each flower. Stamens are in bundles. Style is with a ring of hairs and orange red coloured glands at base. Fruits are ovoid-oblong in shape and dark blue coloured (Figure 2B.1). The flowering and fruiting season is from November to January. The bark of *S. racemosa* is cooling, mildly astringent, light, and useful in treating dropsy, elephantiasis, filaria, liver complaints, bowel complaints, eye diseases, ulcers, menorrhagia, and leucorrhoea (Ahmad et al. 2005). The species is reported with anti-fibrinolytic activity and inhibitory activity against snake-venom phosphodiesterase I (Ahmad et al. 2003). Phenolic compounds present in it are potential candidates for arthritis (Choudhary et al. 2004). Due to such high medicinal importance, *S. racemosa* is exploited from its native habitats. Hence, we undertook this study to analyse and examine systematically the impact of habitat fragmentation on the genetic diversity of this species. Depending on the extent of anthropogenic disturbances, the seven collection sites in WG were classified as protected, semi-protected and disturbed. We used chloroplast and mitochondrial DNA as they are conserved and carry more information than nuclear markers. Also, maternally inherited markers are better indicators of within population diversity (Skuza et al. 2013) and chloroplast-specific markers help to understand the post-glacial histories of plant species (King and Ferris 1998). Therefore, we evaluated the genetic variation existing in intergenic sequences of chloroplast and mitochondrial DNA of *S. racemosa* and documented the patterns of genetic diversity within and among the sites in the context of effect of habitat disturbance on its genetic diversity and need for conservation.

S 2B.2 Materials and Methods

Study areas

The study area comprised seven sites from WG (Figure S2B.2 and Table S2B.1) covering two states of India *viz.* Karnataka and Maharashtra, and divided in three groups. Group I characterized by the sites from protected areas like Agumbe (AGB), Bhagmandala (BHM) and Madikeri (MAD). Group II, defined as semi-protected areas, comprised protected forests

disturbed to different extent and included Castle Rock (CAS) and Amboli (AMB). Similarly, Group III characterized by sites from disturbed areas, Mahabaleshwar (MAH) and Tillari dam (TIL). The sites were grouped on the basis of semi-quantitative field observations. The rate of disturbance at each site was measured using different drivers of disturbances such as fire, grazing, paths construction, cutting trees and tourism/collection (Table S2B.2). Though all these areas were parts of protected forest zones, mostly classified as forest reserves, human intervention as well as conservation methods varied necessitating an elaborate analysis of the anthropogenic activities on these forest regions.

Figure S2B.1: Photograph of *S. laurina* displaying habit, flowers and fruits

Photo credit: Dr. Sofia Banu

Table S2B.1: Sampling locations with global position and number of sampled individuals of *Symplocos racemosa*

States	Collection sites	Latitude	Longitude	Altitude (m)	Individuals sampled	Protection regime	Group
Karnataka	Agumbe (AGB)	N 13.50031	E 75.08308	640	11	RF	I
	Bhagmandala (BHM)	N 12.38837	E 75.48798	1273	8	RF	I
	Madikeri (MAD)	N 12.43467	E 75.71656	1109	5	RF	I
	Castle Rock (CAS)	N 15.41788	E 74.33014	600	20	RF	II
Maharashtra	Amboli (AMB)	N 15.96496	E 73.99225	912	15	RF	II
	Mahabaleshwar (MAH)	N 17.91670	E 73.66670	1352	12	ESA	III
	Tillari Dam (TIL)	N 15.80165	E 74.19120	790	20	RF	III

RF- Reserved forest, ESA: Ecosensitive area. Group I: Protected area; Group II: Semi-protected area; Group III: Disturbed area

Table S2B.2: Sampling locations with drivers of disturbances and their weighted score

Location	Fire	Grazing	Paths	Cutting	Tourism/ collection	Weighted score	Group
AGB	0	0	1	1	1	3	I
BHM	1	0	1	1	1	4	I
MAD	0	1	0	1	2	4	I
CAS	0	1	2	2	1	6	II
AMB	1	1	3	2	2	9	II
MAH	0	2	3	2	3	10	III
TIL	0	2	3	3	2	10	III

score 0-3 indicates 0- no activity, 1- low activity, 2- moderate activity and 3- higher activity. For total weighted score 0-4- no disturbance; 5-9- moderate disturbance and 10- >10 high disturbance. Group I: Protected area; Group II: Semi-protected area; Group III: Disturbed area

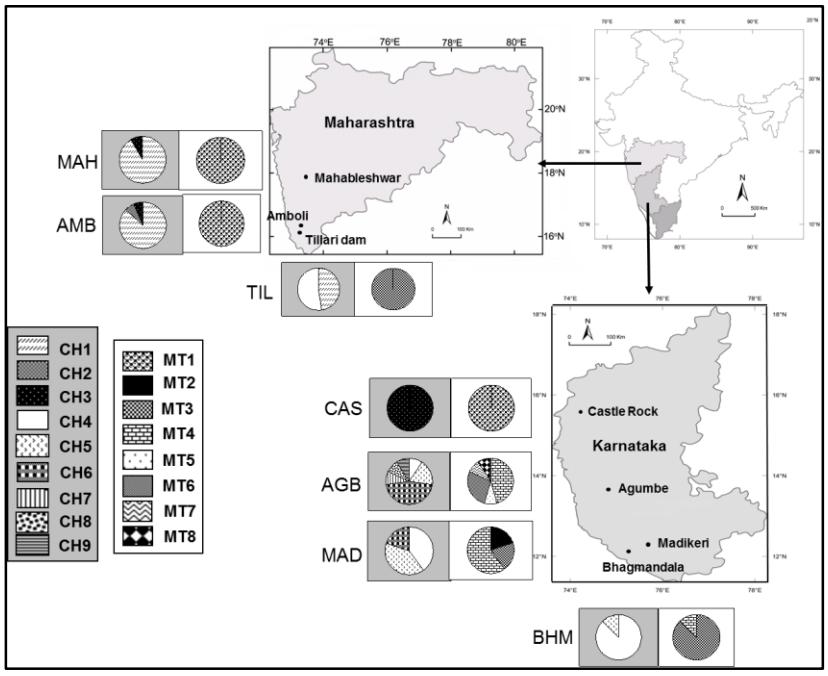


Figure S2B.2: Sampling sites of *Symplocos racemosa* populations in Western Ghats within the states of Maharashtra and Karnataka projecting out distribution of nine chlorotypes (CH1 to CH9) (shaded boxes) and eight mitotypes (MT1 to MT8) (unshaded boxes)

Sampling

S. racemosa populations were collected from seven natural sites. The minimum and maximum aerial distances among the sites were 25 Km and 648 Km, respectively. The number of plants collected per site ranged from five to twenty. Leaf samples were collected, cleaned and frozen immediately in liquid nitrogen and stored at -80 °C till further processing.

DNA extraction and PCR amplification

Total genomic DNA was extracted using a modified Cetyl trimethyl ammonium bromide (CTAB) method (Richards et al. 1994) as mentioned in materials and methods of section one and quantified on 0.8 % agarose gel stained with ethidium bromide. PCR conditions were standardized for twenty pairs of universal primers representing various intergenic regions in cp and mt DNA as described by Demesure et al. (1995), Dumolin-Lapegue et al. (1997) and Chiang et al. (1998). Each 25 µL reaction consisted of 60 ng template DNA, 200 µM of each dNTPs, 0.2 µM of each primer, 1.5 mM $MgCl_2$, 0.8 U of *Taq* DNA polymerase (Banglore Genei, India) and 10X buffer provided by the manufacturer. The PCR amplifications were carried out in PTC 200 thermal cycler (MJ Research, USA) at initial denaturation 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 30 sec at 50-63°C (depending upon the average annealing temperature of the primers used) and 2 min at 72 °C with a final extension of 5 min at 72°C. The amplicons were resolved on 1 % agarose gels in 0.5X TAE buffer.

In all, ten cp and six mt DNA primers giving good amplicons during initial screening were considered for further study. For the present study, PCR-RFLP approach was used. Five µL of PCR product from the selected primers was digested separately with six different tetranucleotide specific restriction endonucleases (REs) (*Alu*I, *Hae*III, *Hha*I, *Msp*I, *Rsa*I and *Taq*I) and four hexanucleotide specific REs (*Eco*RI, *Hin*fI, *Hin*dIII and *Xba*I) (Promega, USA) using 1-2 units of REs in 1X restriction buffer (provided by the supplier) and incubating the reaction at 37°C for 12-14 h for all the enzymes except *Taq*I. In case of *Taq*I the reaction was incubated for 65°C for the same time period as mentioned above. The restriction digests were resolved on 1.5% agarose gels (for fragments above 1.3 kb) and on 2-3.5% MetaPhor agarose gels (Cambrex Bio Science, USA) (for fragments having size range of 75 bp to 1.3 kb) in 0.5X tris borate EDTA (TBE) buffer. The agarose and MetaPhor

agarose gels were stained with ethidium bromide and photographed using ImageMaster VDS gel documentation system (Amersham Biosciences, USA) under UV light.

Data analysis

Definitions of chlorotypes and mitotypes were generated by combining the alleles observed using four and three primer pair-restriction endonuclease combinations (PECs) specific for cp and mt DNA, respectively. The diversity parameters (H_T – total diversity, H_S – average intrapopulation diversity, G_{ST} – the level of population subdivision of diversity using unordered alleles) and the frequency of haplotypes were analysed using the HAPLODIV program (Pons and Petit 1995). Similarly, the HAPLONST program (Pons and Petit 1996) was used to calculate N_{ST} (the level of population subdivision for ordered alleles) and v_T and v_S (analogues of H_T and H_S). Likewise, PERMUT (Pons and Petit 1996) was used to evaluate the differences between G_{ST} and N_{ST} values for statistical significance to identify the existence of geographical structure. G_{ST} is defined as the proportion of genetic diversity that exists among populations (Nei 1973) and is calculated from the total genetic diversity in the pooled population (H_T) and the mean diversity within each population (H_S) (Nei 1973, 1977). A G_{ST} value of '1' indicates that all populations are completely differentiated and all differences in haplotypic composition lie among populations rather than within populations. A G_{ST} value of '0' indicates that all populations are the same and all differences in haplotypic composition lie within rather than among populations.

The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrograms was constructed with coefficient of distances using MVSP for Windows (http://www.kovcomp.co.uk/mvsp/) with 1000 permutations to analyse the relationships among the sites, for cp and mt data each. A minimum spanning network (MSN) among the haplotypes was constructed using ARLEQUIN version 3.11 (Excoffier et al. 2005) and visualized using TREEVIEW version 1.6.6 (Page 1996). The influence of spatial separation on the degree of differentiation among the sites was investigated by calculating pairwise G_{ST} and N_{ST} values using the program DISTON (http://www.pierroton.inra.fr/genetics/labo/Software/Diston/index.html) and by plotting the means of these parameters against geographical distance classes. ARLEQUIN was used to partition the variance among the regions, sites and within sites based on analysis of molecular variance (AMOVA). The AMOVA was performed in two ways: (i) the whole dataset with

all the sites taken together, and (ii) partitioning the sites in terms of their ecological status (protected, semi-protected and disturbed).

S 2B.3 Results

The seven *S. racemosa* sites were represented by nine chlorotypes (Table S2B.3) and eight mitotypes (Table S2B.4). In case of cpDNA, the noncoding region between the *psa*A and *trn*S genes revealed polymorphism upon digestion with *Hae*III, *Hha*I, *Hin*fI and *Rsa*I REs. From the initial screening, seven primer enzyme combinations (PEC) yielding unambiguous amplicons and digestion patterns were selected to analyse the genetic diversity among the 91 individuals from seven *S. racemosa* sites. Four of these patterns were cp specific, while three were mt specific. The size of the polymorphic fragments ranged from 75 - 872 bp for cpDNA and 170 - 740 bp for mtDNA (Tables S2B.3 and S2B.4). The PEC *psa*A-*trn*S:*Hae*III for cpDNA and *nad*7/2-*nad*7/3r:*Hha*I for mtDNA showed the highest polymorphism of 57.14%. In all, alleles from 14 loci defined the chlorotypes, while the alleles from 11 loci contributed to the definition of mitotypes.

Chlorotypes, mitotypes and their diversity indices

The distribution of the haplotypes with respect to geographical locations is presented in Figure S2B.2 and Tables S2B.5 and 2B.6. No haplotype common to all the sites was detected for both cp and mt DNA but the CH1 and MT1 haplotypes were dominant haplotypes. AGB represented the highest number of haplotypes for both the genomes while in contrast, TIL was represented by only one cp and mt DNA haplotypes each. The number of chlorotypes detected per individual was also the lowest (0.05) for TIL and highest for AGB (0.55) and MAD (0.60) (Table S2B.5). The scenario was similar for the number of mitotypes as well. As clear from the Table S2B.6, the diversity was high in the protected group and low in the disturbed group in case of both the genomes.

The diversity indices estimated using HAPLODIV and HAPLONST revealed high total diversity (H_T) for cp (0.786) and mt (0.778) genomes, although the within sites diversity (H_S) was low (cp = 0.352, mt = 0.245). The trends were similar for the ordered counterparts. V_T for cp genome is 0.787 and for mt 0.769 as well as v_S for cp is 0.352 and for mt 0.312. Permutation analysis of chlorotypes and mitotypes with 1000 permutations gave 5% of the

permuted values greater than the observed value of N_{ST}, indicating that the difference between G_{ST} and N_{ST} was significant.

Table S2B.3: Definitions of the chlorotypes with respect to the polymorphic bands obtained with the PECs (primer pair-restriction endonuclease combination)

PEC % polymorphism	Size of the poly-morphic fragment (bp)	Chlorotypes								
		CH1	CH2	CH3	CH4	CH5	CH6	CH7	CH8	CH9
***psa*A- *trn*S *Hae*III (57.14)**	872	0	1	0	0	1	0	1	0	1
	850	1	1	1	1	0	1	0	1	0
	490	1	1	1	1	1	1	1	1	0
	470	0	0	0	0	0	0	0	0	1
***psa*A- *trn*S *Hha*I (50.00)**	570	0	0	0	0	1	1	1	1	1
	450	1	1	1	1	0	0	0	0	0
***psa*A- *trn*S *Hin*fI (25.00)**	440	1	1	1	1	0	0	1	0	0
	420	1	1	1	1	1	1	0	1	1
	350	1	1	1	0	0	0	0	1	0
	310	0	1	0	1	1	1	1	0	1
	281	1	1	1	0	1	1	1	1	1
	75	0	1	1	0	0	0	0	0	0
***psa*A- *trn*S *Rsa*I (40.00)**	550	0	1	0	0	1	1	1	1	1
	370	1	1	1	1	0	0	0	0	0

Relationships among the haplotypes

The dendrograms drawn using the UPGMA algorithm considering the frequencies of individual chlorotypes and mitotypes in each population are represented in Figures 2B.3 and 2B.4.Surprisingly, both the dendrograms were almost similar and showed clustering of the sites as per the geographical locations; the two Maharashtra sites (AMB and MAH) along with the north Karnataka site, CAS, formed one cluster; while the southern Karnataka sites (AGB, MAD and BHM) formed another cluster. However, the TIL (Maharashtra) separated out from both the clusters in both the dendrograms.

Table S2B.4: Definitions of the mitotypes with respect to the polymorphic bands obtained with the PECs (primer pair-restriction endonuclease combination)

PEC (% polymorphism)	Size of the polymorphic fragment (bp)	Mitotypes							
		MT 1	MT 2	MT 3	MT 4	MT 5	MT 6	MT 7	MT 8
***nad*7/2 – *nad*7/3r *Alu*I (33.33)**	281	1	1	1	1	1	1	0	1
	267	0	1	0	0	0	1	1	0
	220	0	0	0	1	1	0	1	1
	180	1	0	1	1	1	0	1	1
nad7/2 – *nad*7/3r *Hha*I (57.14)	350	0	1	1	0	0	0	0	1
	310	1	0	0	1	1	1	1	0
	210	0	0	0	1	1	0	1	0
***nad*7/2 – *nad*7/3r *Rsa*I (50.00)**	740	0	1	0	0	0	0	0	0
	281	0	1	0	1	1	1	1	0
	220	1	0	1	0	1	0	0	1
	170	0	0	0	1	1	0	0	0

Table S2B.5: Distribution of the chlorotypes in seven populations of *S. racemosa*

Populations	Karnataka				Maharashtra			Total (N)	Frequency
	AGB	**BHM**	**MAD**	**CAS**	**AMB**	**MAH**	**TIL**		
Group*	**I**	**I**	**I**	**II**	**II**	**III**	**III**		
Chlorotypes									
CH1	-	-	-	18	13	11	-	42	0.462
CH2	-	-	-	-	**1**	-	-	**1**	0.011
CH3	-	-	-	2	1	1	-	4	0.044
CH4	1	7	2	-	-	-	20	30	0.330
CH5	2	1	2	-	-	-	-	5	0.055
CH6	5	-	1	-	-	-	-	6	0.066
CH7	**1**	-	-	-	-	-	-	**1**	0.011
CH8	**1**	-	-	-	-	-	-	**1**	0.011
CH9	**1**	-	-	-	-	-	-	**1**	0.011
No. of different chlorotypes	6	2	3	2	3	2	1	9	
Haplotypes /individual	0.55	0.25	0.60	0.10	0.20	0.17	0.05	0.10	
Haplotype diversity	0.80 ± 0.11	0.25 ± 0.18	0.80 ± 0.16	0.19 ± 0.11	0.26 ± 0.14	0.17 ± 0.13	0.00 ± 0.00		

Table S2B.56: Distribution of the mitotypes in seven populations of *S. racemosa*

Populations	Karnataka				Maharashtra			Total (N)	Frequency
	AGB	BHM	MAD	CAS	AMB	MAH	TIL		
Group*	**I**	**I**	**I**	**II**	**II**	**III**	**III**		
Chlorotypes									
MT1	-	-	-	20	15	12	-	47	0.516
MT2	-	-	**1**	-	-	-	-	**1**	0.011
MT3	-	7	1	-	-	-	20	28	0.044
MT4	5	1	3	-	-	-	-	9	0.099
MT5	**1**	-	-	-	-	-	-	**1**	0.011
MT6	**3**	-	-	-	-	-	-	**3**	0.033
MT7	**1**	-	-	-	-	-	-	**1**	0.011
MT8	**1**	-	-	-	-	-	-	**1**	0.011
No. of different mitotypes	5	2	3	1	1	1	1	8	
haplotypes /individual	0.45	0.25	0.60	0.05	0.07	0.08	0.05	0.09	
Haplotype diversity	0.76 ± 0.11	0.25 ± 0.18	0.70 ± 0.22	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00		

The MSN depicting the phylogenetic relationships among the chlorotypes and mitotypes (Figure S2B.5) showed the presence of some of the unique haplotypes (CH2, CH7, CH9 and MT2, MT8) in the edges of the tree; while the dominant ones occupied anterior position in the network. In the cpDNA MSN, a maximum of five mutations were observed between CH1 and CH8, while in the mtDNA network, two alternative connections were observed between MT1 and MT5, and MT6 and MT7 (Figure S2B.5).

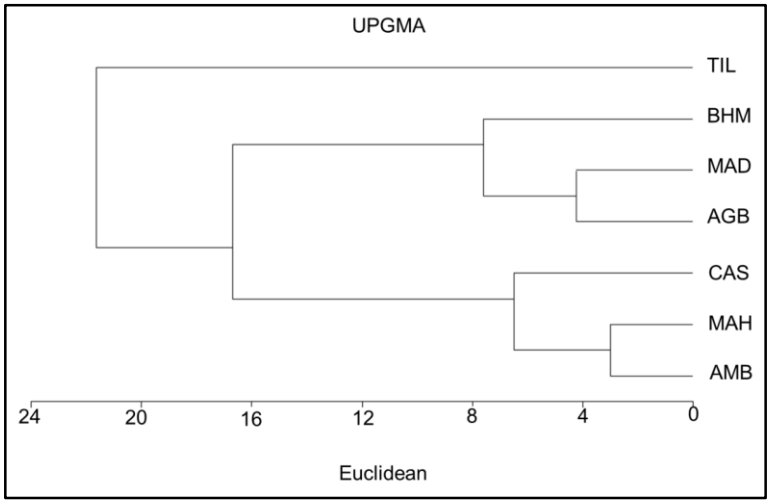


Figure S2B.3: Dendrogram drawn with frequencies of chlorotypes using coefficient of Euclidean distances and UPGMA algorithm

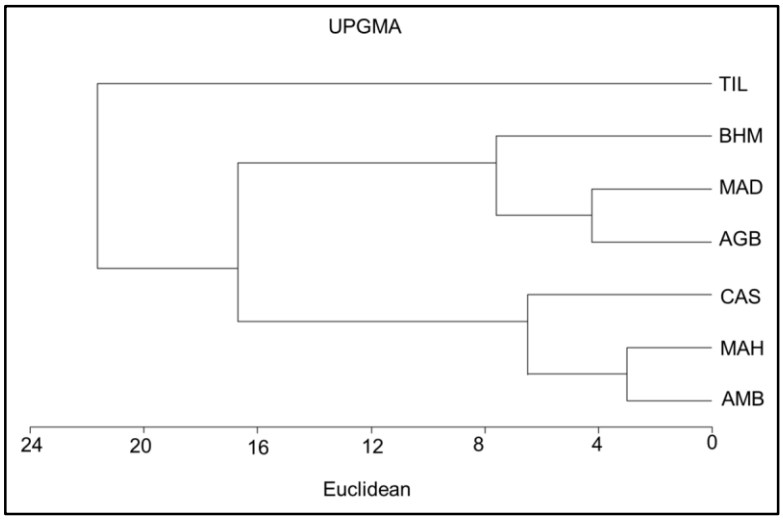


Figure S2B.4: Dendrogram drawn with frequencies of mitotypes using coefficient of Euclidean distances and UPGMA algorithm

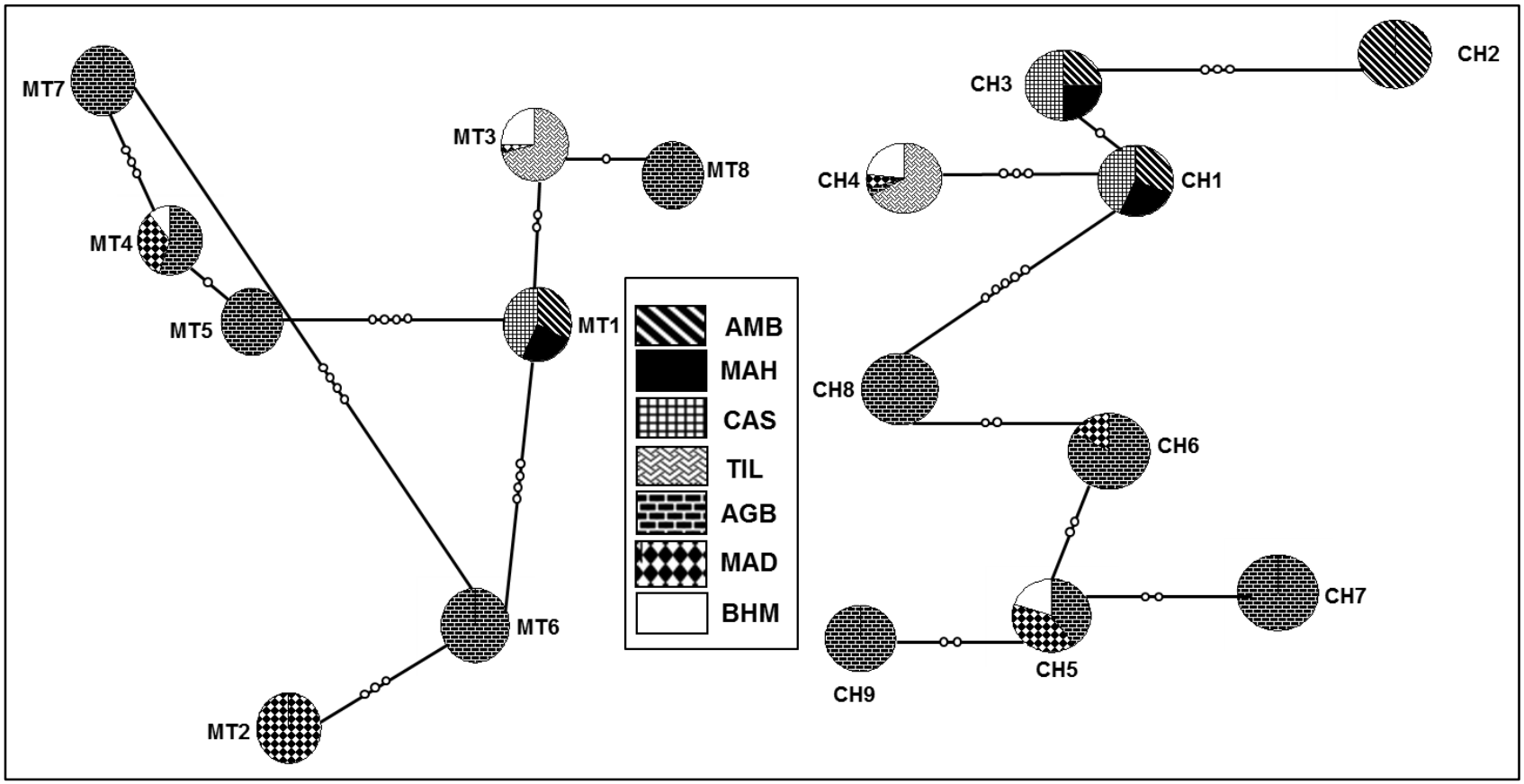


Figure S2B.5: Minimum Spanning Network of *S. racemosa* showing relationships among the chlorotypes (CH) and mitotypes (MT). The circles in the figure indicate the type of chlorotype and mitotype. The shades within each circle indicate the contribution of each population to the chlorotype and mitotype. The small circles within the lines connecting the circles indicate the number of mutations between the haplotypes

Analysis of molecular variance

The AMOVA performed with all the sites taken together showed that 74.32% and 69.26% variation was partitioned among the seven sites for cp and mt DNA, respectively, while the rest of the variation was within the populations (Table S2B.6). However, when the AMOVA was performed with the populations grouped according to their ecological status (protected, semi-protected and disturbed), a majority of the variation was partitioned among populations within groups; while the rest of the variation was found within populations and among groups (Table S2B.6).

Table S2B.6: Analysis of molecular variance (AMOVA) of *S. racemosa* populations

	Source of variation	**d.f.**	**SS**	**Variance components**	**Percentage variation**	
a) AMOVA with all the populations taken together						
cpDNA	Among populations	6	106.508	1.367	74.32	F_{ST} = 0.743
	Within populations	84	39.668	0.472	25.68	
mtDNA	Among populations	6	68.802	0.877	69.26	F_{ST} = 0.693
	Within populations	84	32.670	0.389	30.74	
b) AMOVA with populations divided into three ecological groups (protected, semi-protected and disturbed)						
cpDNA	Among groups	2	10.362	0.08294	21.05	F_{ST} =0.668
	Among populations within groups	4	9.064	0.18021	45.74	
	Within populations	84	10.992	0.13085	33.21	
mtDNA	Among groups	2	12.330	0.10485	27.26	F_{ST} =0.811
	Among populations within groups	4	10.115	0.20731	53.89	
	Within populations	84	6.093	0.07254	18.86	

S 2B.4 Discussion

Distribution of *S. racemosa* haplotypes *vis-à-vis* their geographical distance

In the present study, the haplotypes were distributed *vis-à-vis* the geographical distances among them and the genetic diversity was relatively low within sites, similar to that reported for *Oryza glumaepatula* (Buso et al. 1998) and *Orobanche Cumana* (Gagne et al. 1998). The low within-sites diversity might be because of fixation of most of the populations (except the AGB) for one or two types of haplotypes, while the high total diversity observed in this study was probably contributed by the unique haplotypes. Unique haplotypes are characteristic of expanding population and are not known to be found in population where the anthropogenic activities are the highest (Bajc et al. 2011). The Karnataka sites (AGB, MAD and BHM) were more diverse, being represented by maximum number of haplotypes, which was apparently contributed by the unique haplotypes, especially in AGB. The TIL site shared one chlorotype and one mitotype with these populations, which are more than 400 Km away from it; indicating that the genetic affinities among the sites were irrespective of the geographical distances. However, after this, TIL site apparently got fixed only for these haplotypes and no other haplotypes were detected even after sampling a maximum number of individuals from this site.

The patchy or fragmented landscape of TIL seems to contribute to isolation of the population, which might have resulted in reduction of genetic diversity in this area. Long lived woody perennials in particular, are expected to be resilient to changes in genetic diversity due to ample gene flow especially if they are cross-pollinating and have long generation times. However, habitat fragmentation might lead to genetic changes in remnant plant populations, which include the erosion of genetic variability and acceleration of genetic divergence among the populations by means of two mechanisms: reduced gene flow and an increase in random genetic drift, which might be true for the TIL. Due to this, the CAS and TIL sites, though separated by a distance of only 45 Km, did not share any common haplotype. This indicated the lack of gene flow between them and underlined the existence of possible barriers in gene flow among these two populations.

High genetic differentiation and low gene flow in *S. racemosa*

Population differentiation is relatively high in both gymnosperms and angiosperms for cp and mt DNA, since the genomes are generally maternally inherited (Demesure et al. 1996). This has been found to be true in the present study as well. Low values of H_S (in this study, for cp, $H_S = 0.352$ and for mt $H_S = 0.245$) indicate either low mutation rate or a consequence of genetic drift. Significant genetic differentiation in populations of tree species is observed when there is restricted seed and pollen dispersal resulting in low gene flow (Cavers et al. 2005; Hardy et al. 2006) as observed in this study. The lack of haplotype sharing along with high differentiation values obtained here implied an ancient nature of the sites. The species, which experience long-term biogeographic barriers to gene flow, are known to be composed of geographical populations that belong to different genealogies and are correlated to geographical boundaries (Avise 1994). Most of the sites of *S. racemosa* were restricted to mountains in small size and were strongly isolated from each other by mountains and valleys, which might have enhanced fragmentation in their habitat. Past fragmentation leading to geographical isolation and limited pollen/seed dispersal among the populations might have resulted in much lower diversity within populations and elevated genetic differentiation among the sites of *S. racemosa*.

Partitioning of genetic variation among the populations

In general, for out-crossing, wind-pollinated and long-lived tree species, genetic variation is expected to be maintained within populations, while predominantly selfing, short-lived species harbour comparatively higher variation among the populations. Although *S. racemosa* is a perennial out-crossing woody angiosperm, most of the genetic variation in this study was partitioned among the sites. This could be because these sites might have become susceptible to stochastic effects of genetic drift leading to reduction in heterozygosity and loss of alleles. Such lower genetic diversity has also been reported previously in many other relic species of East Asia, e.g., *Alsophila spinulosa* (Wang et al. 2004), *Ammopiptanthus mongolicus* and *A. nanus* (Ge et al. 2005). The genetic distances among the populations analyzed in this study did not correspond to the geographic distances among them. The pattern obtained in the genetic differentiation *vs*. geographical distances indicated the same. Similar observations have been reported by Cronberg (2002) and Galeuchet et al. (2005) as well.

The high haplotype diversity observed in the protected sites highlights the fact that the genetic variability is higher in protected areas, suggesting a relation between genetic variation and habitat disturbance. The occurrence of only one or two haplotypes in disturbed areas despite the size of the samples being large; might imply erosion of genetic variation due to various anthropogenic activities in these areas. In contrast, the presence of relatively high number of haplotypes, and as a result, high diversity in protected and semi-protected areas emphasizes the roles played by various natural factors in maintaining the diversity. The study areas AMB and MAH, which are under high demographic pressures due to various anthropogenic activities and other factors, have already been identified as reserve forests of high conservation value.

The lack of genetic diversity observed in disturbed areas, resulting in reduced fitness of the populations, has been observed in many studies (Fischer and Matthies 1998). Such genetic impoverishment is usually assumed to be the effect of genetic erosion within isolated populations in fragmented environments. A variety of conservation approaches ranging from strict protection to alteration of existing land use patterns to targeted management interventions for a particular species would be necessary across these areas for sustainable management of resources in these forests (Fischer and Matthies 1998). As the protected areas harboured higher genetic diversity than semi-protected and disturbed areas, these areas need to be strictly protected from human exploitation to conserve the inherent biodiversity. However, human intervention would be necessary to augment the genetic diversity in semi-protected and disturbed areas by introducing individuals with different haplotypes from protected areas into these areas. In view of this, the AGB population would be a good source of different haplotypes of *S. racemosa* to be introduced in other areas.

Summary and future prospects

Summary and future prospects

DNA barcoding of *Dalbergia* species collected from Western Ghats, India

Identifying a universal barcode locus is difficult for plants as mentioned in previous chapters of the thesis. Hence, the efforts were directed to develop the DNA barcodes at family or genus level. The present study was initiated to develop DNA barcodes for discrimination of *Dalbergia* species. The study included 7-26 accessions of ten *Dalbergia* species each collected from different geographic locations in WG region of India and were screened using 37 primer pairs from several nuclear and plastid genes. Of these, four loci (*rbcL*, *matK*, *trnH-psbA* and nrITS) and their combinations were further evaluated with five different analyses such as inter and intra specific distances, Neighbor Joining (NJ) trees, Best Match (BM) and Best Close Match, (BCM) character based approach and Wilcoxon signed rank (WSR) test. The results generated by all these parameters were ranked based on their performance to select the best barcode loci that could discriminate the *Dalbergia* species. The three criteria (inter and intraspecific distances, NJ trees, and BLOG) revealed *matK* and *matK*+*rbcL* as the best performers. While, BM and BCM revealed *matK*, *matK*+*rbcL* and *matK*+*trnH-psbA* as the best loci with 100% correct identification. The Wilcoxon Signed Rank test indicated *rbcL*+*trnH-psbA* followed by nrITS as the suitable loci. The overall performance after ranking all the results concluded *matK*, followed by *matK*+*rbcL* as the most suitable barcode loci to discriminate the *Dalbergia* species. The validation of these two loci along with nrITS was further carried out using sequences downloaded from NCBI database. The DNA sequences of various *Dalbergia* species (only those having minimum 2 accessions of each species) corresponding to the *matK*, *rbcL* and nrITS regions were downloaded from the NCBI database. All the above mentioned parameters except WSR were used for further analysis. In this analysis too, the ranking revealed highest scores for *matK* followed by *matK*+*rbcL*. In addition, few unauthenticated samples were also used and analyzed based on the NJ trees. The results showed that, in all the three loci, one unauthenticated sample clustered with *D. melanoxylon* while five clustered with *D. horrida*. One accession of *D. sissoides* collected and authenticated from local person clustered with one of the variants (containing same five samples in *matK* and *matK*+*rbcL* while split into two in nrITS) of *D. latifolia* in all the three loci. Similarly, another accession of *D. paniculata* (DpA) collected and authenticated from Mysore region clustered with *D. lanceolaria* group. Phylogenetic analysis was carried out by combining all the above four primers with

sequences used in the development of barcodes and downloaded from NCBI database. The same species were found to cluster together irrespective of their geographical locations. The study revealed mixed pattern of distribution of the species irrespective of Old World and New World origin, indicating monophyletic nature of the genus *Dalbergia*.

Genetic diversity studies in *S. laurina* with ISSR markers

We analysed 208 accessions of *S. laurina* from 13 populations belonging to three ecoregions in India namely, Western Ghats (WG), Eastern Ghats (EG) and Northeast India (NEI). A total of 100 inter simple sequence repeat primers were screened, of which eight were selected based on their clear and reproducible amplification and polymorphism patterns. These eight ISSR primers produced on an average 1,014 bands from 208 accessions, of which 6.58% were rare bands, 55.96% were shared bands and 37.46% were similar bands. The primers, UBC835 and UBC880, amplified the highest number of polymorphic loci, while the primer UBC852 amplified the least number of amplicons. UBC880 produced the maximum number of rare bands and hence can be considered as the most suitable primer to identify a particular accession. Similarly, UBC835 with high SPI value (5.40), was highly efficient in detecting polymorphism in *S. laurina* populations. Likewise, UBC859 detected the highest number of unique genotypes and had the lowest PI value ($6.9\ X10^{-3}$).

S. laurina is an indicator of altitude and is found above 800 m up to 2100 m. To evaluate the effect of altitudinal variation on genetic diversity, Pearson's correlation coefficient was calculated between altitude and genetic diversity (*h*). The results revealed significant negative correlation (r =-0.59) indicating prevalence of natural selection of some genotypes over time, which are better adapted to higher altitudes over time. The overall gene flow was slightly low (0.9890) compared to threshold value; however, within region it showed more than one migrant per generation indicating good differentiation within populations and within regions. A UPGMA dendrogram was constructed based on genetic distances, which revealed clear separation of the NEI populations from the WG and EG populations. All the WG populations except Munnar formed one cluster, while EG formed another cluster along with Munnar of WG. The PCA showed similar pattern of clustering. The grouping of Munnar population with EG rather than WG populations, could be due to the presence of Palghat gap (a 30 Km discontinuity in WG), which separates Munnar from other WG populations. Thus, the present study revealed the importance of the Palghat gap in shaping the distribution of genetic diversity in *S. laurina.*

Further, based on PCA analysis, individuals of Kolli hills population from EG were found to be distributed close to the axis separating WG and EG clusters, indicating that the Kolli hills population could be a link between WG and EG. The total genetic diversity (H_T) and the average heterozygosity (H_S) values were 0.3407 and 0.2263, respectively; while the coefficient of genetic differentiation (G_{ST}) was observed to be 0.3358 and the genetic diversity within-populations were found to be 0.2263 which was in similar lines with the long-lived, perennial and outcrossing species of late successional stages. Within-population variance was higher (68%) than among-population variance (17%) indicating high fitness and adaptability to changing environmental conditions.

Effect of anthropogenic activities on patterns of genetic diversity in *S. racemosa*

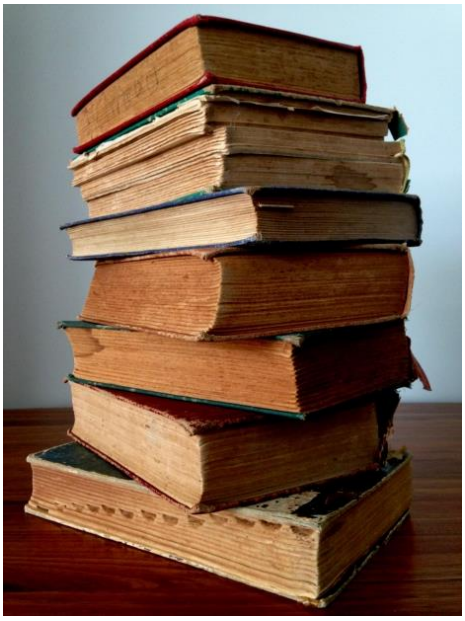
To study the effect of anthropogenic activities on genetic diversity in *S. racemosa,* seven populations from protected, semi-protected and disturbed regions of WG were analysed. The analysis was carried out using PCR-RFLP approach in intergenic sequences of chloroplast (cp) and mitochondrial (mt) DNA. The nine chlorotypes and eight mitotypes were identified by combining the alleles generated by primer pairs and restriction endonucleases. The Agumbe population represented the highest number of haplotypes while the Tillari population represented the lowest number of haplotypes. The diversity parameters (H_T – total diversity, H_S – average intrapopulation diversity, G_{ST} – the level of population subdivision of diversity using unordered alleles) and the frequency of haplotypes were calculated. High total diversity was revealed for both the genomes. UPGMA dendrograms were drawn to study the relationship among populations. Both the dendrograms indicated that the genetic affinities among the sites were irrespective of the geographical distances. The genetic variation was expected to be maintained within populations; however, in this study, it was partitioned among the sites. This could be because these sites might have become susceptible to stochastic effects of genetic drift leading to reduction in heterozygosity and loss of alleles. The study identified *S. racemosa* populations with low diversity as well as with unique haplotypes that could be used for formulating conservation strategies for the species. Among all the populations, the populations from protected areas exhibited high diversity while the disturbed areas showed less diversity. Since the Agumbe population depicted the highest number of haplotypes for both cp and mt genomes, it would be a good source of different haplotypes to be introduced in the other areas. However, more such studies entailing other important plants from these biodiversity hotspots are needed to identify the target sites for effective maintenance and conservation of biodiversity of WG.

Future prospects

1. The current study analyzed ten species of *Dalbergia* from WG region of India for development of DNA barcodes for the genus. Validation of the barcodes was carried out using NCBI sequences as well as few unauthenticated samples. The confusion in identification of *D. sissoides* and *D. paniculata* by morphological characters was also revealed in our study. In this study, we have included only one sample of *D. sissoides* and 26 accessions of *D. latifolia* while 14 accessions of *D. paniculata* and seven accessions of *D. lanceolaria*. Hence, more number of accessions of *D. sissoides* and *D. paniculata* as well as *D. lanceolaria* covering the broader geographical range need to be collected and studied using DNA barcoding approach to eliminate this misidentification.
2. Globally, more than 120 species of *Dalbergia* are reported; while in India, total 30 and in WG approximately 10-15 species are reported. To develop robust DNA barcodes, Indian and world collections of various *Dalbergia* species need to be evaluated for both, morphological characters and DNA barcodes. Also, there are few new additions in genus *Dalbergia* (Jongkind, 2007; Pradeep and Balan, 2010) which need to be analysed.
3. More work needs to be done with more number of samples and populations across the wide region and different markers covering broader genomic region to comment on the genetic diversity, relationship between species and biogeography of *Dalbergia* genus.
4. Similarly, more genera from family Fabaceae need to be analysed.
5. Many times, *S. laurina* is confused with *S. cochinchinensis*. It has been always identified as one of the four subspecies under a circumscribed species *S. cochinchinensis* (Nooteboom, 2004). *S. atlantica* described and illustrated as a new species from the Atlantic Rain Forest of Brazil, is morphologically similar to *S. glandulosomarginata* and *S. glaziovii*. It can be distinguished from these two by young leaves with an eglandular margin or rarely with 1–3 early caducous glands per cm, a pilose and dome-like to short-cylindrical disc in flower, and fertile ovules 0.2–0.5(-0.6) mm long (Filho et al. 2009). To solve such confusions and also to develop powerful DNA barcodes, more number of *Symplocos* species need to be collected from India as well as outside India. The specimen numbers can be increased either by explorations or acquisition through different herbaria deposits.

6. The diversity of *S. laurina* was analysed using various markers and a comment was made on their population structure. Similar studies need to be carried out on other high altitude plant species such as *Gaultheria* as well as associates of *S. laurina* such as *Memecylon* and *Syzygium* to check the effect of altitude.
7. *S. racemosa* is a medicinally important tree species and commonly used in a treatment of ulcers, bowel and liver complaints, eye diseases etc. Our studies showed the negative effects of anthropogenic activities on genetic diversity of *S. racemosa* and identified the populations with unique haplotypes. These populations can be used to formulate conservation strategies for the species. The conservation efforts can be put along with the forest department and Botanical Survey of India (BSI).
8. Such studies need to be performed with associated species of *S. racemosa* to analyse the effect of anthropogenic activities on their diversity. This will help to understand the scenario in terms of biodiversity of a particular locality.

Bibliography



References

Abdulali H Some peculiarities of avifaunal distribution in peninsular India. In: Proceedings of the National Institute of Sciences, India, 1949. pp 387-393

Agostini G, Echeverrigaray S, Souza-Chies TT (2008) Genetic relationships among South American species of *Cunila* D. Royen ex L. based on ISSR. Plant Systematics and Evolution 274:135-141

Ahmad VU, Abbasi MA, Hussain H, Akhtar MN, Farooq U, Fatima N, Choudhary MI (2003) Phenolic glycosides from *Symplocos racemosa*: natural inhibitors of phosphodiesterase I. Phytochemistry 63 (2):217-220

Ahmad VU, Zubair M, Athar Abbasi M, Abid Rashid M, Rasool N, Nahar Khan S, Iqbal Choudhary M, Kousar F (2005) Structure determination of bioactive galloyl derivatives by NMR spectroscopy. Magnetic Resonance in Chemistry 43 (6):486-488

Akaike H (1974) A new look at the statistical model identification. Automatic Control, IEEE Transactions on 19 (6):716-723

Aladele SE, Ariyo OJ, Lapena Rd (2008) Genetic relationships among West African okra (*Abelmoschus caillei*) and Asian genotypes (*Abelmoschus esculentus*) using RAPD. African Journal of Biotechnology 7 (10):1426-1431

Albert VA, Backlund A, Bremer K, Chase MW, Manhart JR, Mishler BD, Nixon KC (1994) Functional constraints and *rbcL* evidence for land plant phylogeny. Annals of the Missouri Botanical Garden 81 (3):534-567. doi:10.2307/2399902

Allen O, Allen E (1981) The Leguminosae. Univer-sity of Wisconsin Press. Macmillan Publishing Company, Madison, WI,

Almeida S (1990) The flora of Savantwadi: Maharashtra, India. Jodhpur: Scientific Publishers 304p-col illus En Keys Plant records Geog 6

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29 (3):417-434

Alves S, Tiago L, Chauveau O, Eggers L, Souza-Chies TT (2014) Species discrimination in *Sisyrinchium* (Iridaceae): assessment of DNA barcodes in a taxonomically challenging genus. Molecular Ecology Rresources 14 2):324-335

Arif M, Zaidi NW, Singh YP, Haq QMR, Singh US (2009) A comparative analysis of ISSR and RAPD markers for study of genetic diversity in Shisham (*Dalbergia sissoo*). Plant Molecular Biology Reporter 27 (4):488-495. doi:10.1007/s11105-009-0097-0

Ashfaq M, Asif M, Anjum ZI, Zafar Y (2013) Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). Molecular Ecology Resources 13 (4):573-582. doi:10.1111/1755-0998.12089

Atchison E (1951) Studies in the Leguminosae. VI. Chromosome Numbers Among Tropical Woody Species. American Journal of Botany:538-546

Avise JC (1994) Molecular markers, natural history and evolution. Published by Chapman and Hall.

Bahulikar RA, Lagu MD, Kulkarni BG, Pandit SS, Suresh HS, Rao MKV, Ranjekar PK, Gupta VS (2004) Genetic diversity among spatially isolated populations of *Eurya nitida* Korth. (Theaceae) based on inter-simple sequence repeats. Current Science India 86 (6):824-831

Bajc M, Čas M, Ballian D, Kunovac S, Zubić G, Grubešić M, Zhelev P, Paule L, Grebenc T, Kraigher H (2011) Genetic differentiation of the western capercaillie highlights the importance of south-eastern Europe for understanding the species phylogeography. PloS one 6 (8)

Bakshi M, Sharma A (2011) Assessment of genetic diversity in *Dalbergia sissoo* clones through RAPD profiling. Journal of Forestry Research 22 (3):393-397. doi:10.1007/s11676-011-0183-7

Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. Molecular Phylogenetics and Evolution 1 (1):3-16. doi:10.1016/1055-7903(92)90030-K

Banu S (2008) Molecular marker based genetic diversity in *Symplocos* spp. from the the two biodiversity hotspots in India. Pune, CSIR- National Chemical Laboratory

Banu S, Baruah D, Bhagwat RM, Sarkar P, Bhowmick A, Kadoo NY (2015) Analysis of genetic variability in *Aquilaria malaccensis* from Bramhaputra valley, Assam, India using ISSR markers. Flora-Morphology, Distribution, Functional Ecology of Plants 217:24-32

Banu S, Bhagwat RM, Kadoo NY, Lagu MD, Gupta VS (2010a) Understanding the genetic structure of *Symplocos laurina* Wall. populations using nuclear gene markers. Genetica 138 (2):197-210

Banu S, Lagu MD, Gupta VS (2010b) Phylogeographical studies in disjunct populations of *Symplocos laurina* Wall. using cytoplasmic PCR-RFLP approach. Tree Genetics & Genomes 6 (1):13-23

Bartish IV, Antonelli A, Richardson JE, Swenson U (2011) Vicariance or long distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). Journal of Biogeography 38 (1):177-190

Bentham G (1860) Synopsis of Dalbergieae, a tribe of the Leguminosae. Synopsis of Dalbergieae, a Tribe of the Leguminosae Journal of the Proceedings of the Linnean Society IV (Supplement):1-134

Bertolazzi P, Felici G, Weitschek E (2009) Learning to classify species with barcodes. BMC Bioinformatics 10. doi:10.1186/1471-2105-10-S14-S7

Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H (2013) ITS1 versus ITS2 as DNA metabarcodes for fungi. Molecular Ecology Resources 13 (2):218-224. doi:10.1111/1755-0998.12065

Blanquart F, Gandon S (2011) Evolution of migration in a periodically changing environment. American Naturalist 177 (2):188-201. doi:10.1086/657953

Bleeker W, Klausmeyer S, Peintinger M, Dienst M (2008) DNA sequences identify invasive alien Cardamine at Lake Constance. Biological Conservation 141 (3):692-698. doi:http://dx.doi.org/10.1016/j.biocon.2007.12.015

Botstein D, White R, Skolnick M, Davis R (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. The American Journal of Human Genetics 32:314-331

Bradford J, Barnes R (2001) Phylogenetics and classifi cation of Cunoniaceae (Oxalidales) using chloroplast DNA sequences and morphology. Systematic Botany 26:354-385

Buckley J, Bridle JR, Pomiankowski A (2010) Novel variation associated with species range expansion. BMC Evolutionary Biology 10. doi:Artn 382 Doi 10.1186/1471-2148-10-382

Buso G, Rangel P, Ferreira M (1998) Analysis of genetic variability of South American wild rice populations (*Oryza glumaepatula*) with isozymes and RAPD markers. Molecular Ecology 7 (1):107-117

Butcher PA, Moran GF, Perkins HD (1998) RFLP diversity in the nuclear genome of *Acacia mangium*. Heredity 81 (2):205-213

Campbell CS, Wright WA, Cox M, Vining TF, Major CS, Arsenault MP (2005) Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): sequence divergence and structure. Molecular Phylogenetics and Evolution 35 (1):165-185. doi:10.1016/j.ympev.2004.11.010

Carvalho A (1997) A synopsis of the genus *Dalbergia* (Fabaceae: Dalbergieae) in Brazil. Brittonia 49 (1):87-109. doi:10.2307/2807701

Carvalho Ad (1989) Systematic studies in the genus *Dalbergia* L. f. in Brazil. University of Reading,

Cavers S, Degen B, Caron H, Lemes M, Margis R, Salgueiro F, Lowe A (2005) Optimal sampling strategy for estimation of spatial genetic structure in tree populations. Heredity 95 (4):281-289

Champion SH, Seth SK (1968) A revised survey of the forest types of India. Published by Delhi, Manager of Publications.

Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC (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

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madrinan S, Petersen G, Seberg O, Jorgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M (2007) A proposal for a standardised protocol to barcode all land plants. Taxon 56 (2):295-299

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society of London B Biological Science 360 (1462):1889-1895. doi:10.1098/rstb.2005.1720

Chawla H (2009) Introduction to Plant Biotehnology. Oxford and IBH publishing Co. Pvt. Ltd., New Delhi

Chen R, Jiang LY, Liu L, Liu QH, Wen J, Zhang RL, Li XY, Wang Y, Lei FM, Qiao GX (2013) The gnd gene of Buchnera as a new, effective DNA barcode for aphid identification. Systematic Entomology 38 (3):615-625. doi:10.1111/Syen.12018

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C (2010) Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. Plos One 5 (1):e8613. doi:10.1371/journal.pone.0008613

Chester M, Cowan RS, Fay MF, Rich TCG (2007) Parentage of endemic *Sorbus* L. (Rosaceae) species in the British Isles: evidence from plastid DNA. Botanical Journal of the Linnean Society 154 (3):291-304. doi:10.1111/j.1095-8339.2007.00669.x

Chettri B, Bhupathy S, Acharya BK (2010) Distribution pattern of reptiles along an eastern Himalayan elevation gradient, India. Acta Oecologica 36 (1):16-22

Chiang T-Y, Schaal BA, Peng C-I (1998) Universal primers for amplification and sequencing a noncoding spacer between the *atpB* and *rbcL* genes of chloroplast DNA. Botanical Bulletin of Academia Sinica 39

Choudhary MI, Fatima N, Abbasi MA, Jalil S, Ahmad VU (2004) Phenolic glycosides, a new class of human recombinant nucleotide pyrophosphatase phosphodiesterase-1 inhibitors. Bioorganic & Medicinal Chemistry 12 (22):5793-5798

Clarkson J, Knapp S, Garcia V, Olmstead R, Leitch A, Chase M (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Molecular Phylogenetics and Evolution 33:75-90

Clement WL, Donoghue MJ (2012) Barcoding success as a function of phylogenetic relatedness in *Viburnum*, a clade of woody angiosperms. BMC Evolutionary Biology 12:73. doi:10.1186/1471-2148-12-73

Cole TCH (2015) Angiosperm Phylogeny Group (APG) in jeopardy- Where have the flowers gone? PeerJ PrePrints

Coletta Filho HD, Machado MA, Targon M, Moreira M, Pompeu Jr J (1998) Analysis of the genetic diversity among mandarins (*Citrus* spp.) using RAPD markers. Euphytica 102 (1):133-139

Cooke T (1901) The Flora of the Bombay Presidency. London,

Cowan RS, Chase MW, Kress WJ, Savolainen V (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon 55 (3):611-616

Cronberg N (2002) Colonization dynamics of the clonal moss *Hylocomium splendens* on islands in a Baltic land uplift area: reproduction, genet distribution and genetic variation. Journal of Ecology 90 (6):925-935

Cronquist A (1981) An integrated system of classification of flowering plants. Columbia University Press,

Dahanukar N, Raut R, Bhat A (2004) Distribution, endemism and threat status of freshwater fishes in the Western Ghats of India (vol 31, pg 123, 2004). Journal of Biogeography 31 (8):1379-1379

da-Mata TL, Segeren MI, Fonseca AS, Colombo CA (2009) Genetic divergence among gerbera accessions evaluated by RAPD. Scientia Horticulture 121:92-96

Daniels RR (2001) Endemic fishes of the Western Ghats and the Satpura hypothesis. Current Science-Bangalore- 81 (3):240-244

Darlington CD, Wylie AP (1956) Chromosome atlas of flowering plants. Published by George Allen and Unwin.

de Groot GA, During HJ, Maas JW, Schneider H, Vogel JC, Erkens RHJ (2011) Use of *rbcL* and *trnL-F* as a Two-Locus DNA Barcode for Identification of NW-European Ferns: An Ecological Perspective. Plos One 6 (1):e16371. doi:10.1371/journal.pone.0016371

Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. Molecular Ecology 18 (9):2022-2038

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. Evolution:2515-2520

Demesure B, Sodzi N, Petit R (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Molecular ecology 4 (1):129-134

Deshpande AU, Apte GS, Bahulikar RA, Lagu MD, Kulkarni BG, Suresh HS, Singh NP, Rao MKV, Gupta VS, Pant A, Ranjekar PK (2001) Genetic diversity across natural populations of three montane plant species from the Western Ghats, India revealed by intersimple sequence repeats. Molecular Ecology 10 (10):2397-2408

Dezaho C, Dianxiang Z, Kai L (2010) Flora of China, vol 10.

Dijk P, Aston P, Ma J (1999) Indo-Burma. In: Mittermeier RA, Myers N, Mittermeier CG, Robles Gil P (eds) Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions 1999 pp. 431 pp. . CEMEX, S.A., Agrupación Sierra Madre, S.C. , Mexico City,

Dilger WC (1952) The Brij Hypothesis as an explanation for the tropical faunal similarities between the Western Ghats and the eastern Himalayas, Assam, Burma, and Malaya. Evolution 6 (1):125-127

Doebley J, Durbin M, Golenberg EM, Clegg MT, Ma DP (1990) Evolutionary Analysis of the Large Subunit of Carboxylase (rbcL) Nucleotide Sequence Among the Grasses (Gramineae). Evolution 44 (4):1097-1108. doi:10.2307/2409569

Dong L-N, Wortley AH, Wang H, Li D-Z, Lu L (2011) Efficiency of DNA barcodes for species delimitation: A case in *Pterygiella* Oliv. (Orobanchaceae). Journal of Systematics and Evolution 49 (3):189-202. doi:10.1111/j.1759-6831.2011.00124.x

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

Du Z-Y, Qimike A, Yang C-F, Chen J-M, Wang Q-F (2011) Testing four barcoding markers for species identification of Potamogetonaceae. Journal of Systematics and Evolution 49 (3):246-251. doi:10.1111/j.1759-6831.2011.00131.x

Duminil J, Di Michele M (2009) Plant species delimitation: a comparison of morphological and molecular markers. Plant Biosystems 143 (3):528-542

Dumolin-Lapegue S, Pemonge M, Petit R (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. Molecular ecology 6 (4):393-397

Edwards D, Horn A, Taylor D, Savolainen V, Hawkins J (2008) DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). Taxon 57 (4):1317–1327

Elameen A, Klemsdal SS, Dragland S, Fjellheim S, Rognli OA (2008) Genetic diversity in a germplasm collection of roseroot (*Rhodiola rosea*) in Norway studied by AFLP. Biochemical Systematics and Ecology 36 (9):706-715

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47

Fay MF, Krauss SL (2003) Orchid conservation genetics in the molecular age. Orchid conservation Kota Kinabalu: Natural History Publications (Borneo):91-112

Fay MF, Smith R, Zuiderduin K, Hooper E, Samuel R, Bateman R, Chase M (2007) How does hybridization influence the decision making process in conservation? The genus *Orchis* (Orchidaceae) as a case history. Lankesteriana 7:135-137

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SC (2008) Multiple multilocus DNA barcodes from the plastid

genome discriminate plant species equally well. PloS One 3 (7):e2802. doi:10.1371/journal.pone.0002802

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SCH, Newmaster SG, Hajibabaei M, Husband BC (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Molecular Ecology Resources 9:130-139. doi:10.1111/j.1755-0998.2009.02652.x

Feyissa T, Nybom H, Bartish IV, Welander M (2007) Analysis of genetic diversity in the endangered tropical tree species *Hagenia abyssinica* using ISSR markers. Genetic Resources and Crop Evolution 54 (5):947-958

Fischer M, Matthies D (1998) RAPD variation in relation to population size and plant fitness in the rare *Gentianella germanica* (Gentianaceae). American Journal of Botany 85 (6):811-811

Fritsch PW, Al F (2015) A taxonomic revision of Antillean *Symplocos* (Symplocaceae). Phytotaxa 194 (1):1-67

Gagne G, Roeckel-Drevet P, Grezes-Besset B, Shindrova P, Ivanov P, Grand-Ravel C, Vear F, De Labrouhe DT, Charmet G, Nicolas P (1998) Study of the variability and evolution of *Orobanche cumana* populations infesting sunflower in different European countries. Theoretical and Applied Genetics 96 (8):1216-1222

Galeuchet D, Perret C, Fischer M (2005) Microsatellite variation and structure of 28 populations of the common wetland plant, *Lychnis floscuculi* L., in a fragmented landscape. Molecular Ecology 14 (4):991-1000

Gao T, Sun Z, Yao H, Song J, Zhu Y, Ma X, Chen S (2011) Identification of Fabaceae plants using the DNA barcode *matK*. Planta Medica 77 (1):92-94. doi:10.1055/s-0030-1250050

Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, Pang X, Xu H, Chen S (2010a) Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. Journal of Ethnopharmacology 130 (1):116-121. doi:10.1016/j.jep.2010.04.026

Gao T, Yao H, Song J, Zhu Y, Liu C, Chen S (2010b) Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. BMC Evolutionary Biology 10:324. doi:10.1186/1471-2148-10-324

Ge X-J, Liu M, Wang W-K, Schall BA, Chiang T-Y (2005) Population structure of wild bananas, *Musa balbisiana*, in China determined by SSR fingerprinting and cpDNA PCR-RFLP. Molecular Ecology 14 (4):933-944

Geng Q-F, Liu J, Sun L, Liu H, Ou-Yang Y, Cai Y, Tang X-S, Zhang H-W, Wang Z-S, An S-Q (2015) Development and Characterization of Polymorphic Microsatellite Markers (SSRs) for an Endemic Plant, *Pseudolarix amabilis* (Nelson) Rehd.(Pinaceae). Molecules 20 (2):2685-2692

Gere J, Kowiyou Y, Daru B, Mankga L, Maurin O, van der Bank M (2013) Incorporating *trnH-psbA* to the core DNA barcodes improves significantly species discrimination within southern African Combretaceae. ZooKeys 365: 127-147. Doi: 10.3897/zookeys.365.5728

Ghislain M, Zhang D, Fajardo D, Huaman Z, Hijmans RJ (1999) Marker-assisted sampling of the cultivated Andean potato *Solanum phureja* collection using RAPD markers. Genetic Resources and Crop Evolution 46 (6):547-555

Ghneim HT, Posso DD, Pérez A, I., Torrealba NG, Pieters AJ, Martinez CP, Tohme JM (2008) Assessment of genetic diversity in Venezuelan rice cultivars using simple sequence repeats markers. Electronic Journal of Biotechnology 11 (5):3-4

Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. Molecular Biology and Evolution 11 (5):769-777

Giudicelli GC, Mäder G, Brandão de Freitas L (2015) Efficiency of ITS Sequences for DNA Barcoding in *Passiflora* (Passifloraceae). International Journal of Molecular Sciences 16 (4):7289-7303

Goldblatt P (1981) Chromosome numbers in legumes II. Annals of the Missouri Botanical Garden:551-557

Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, Roger A, Thébaud C, Chave J (2009) Identification of Amazonian Trees with DNA Barcodes. Plos One 4 (10):e7483. doi:10.1371/journal.pone.0007483

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution 27 (2):221-224

Gower DJ, Dharne M, Bhatta G, Giri V, Vyas R, Govindappa V, Oommen OV, George J, Shouche Y, Wilkinson M (2007) Remarkable genetic homogeneity in unstriped, long-tailed Ichthyophis along 1500 km of the Western Ghats, India. Journal of Zoology 272 (3):266-275. doi:10.1111/j.1469-7998.2006.00266.x

Grativol C, Lira-Medeiros CD, Hemerly AS, Ferreira PCG (2011) High efficiency and reliability of inter-simple sequence repeats (ISSR) markers for evaluation of genetic diversity in Brazilian cultivated *Jatropha curcas* L. accessions. Molecular Biology Reporter 38 (7):4245-4256. doi:DOI 10.1007/s11033-010-0547-7

Gu J, Su JX, Lin RZ, Li RQ, Xiao PG (2011) Testing four proposed barcoding markers for the identification of species within *Ligustrum* L. (Oleaceae). Journal of Systematics and Evolution 49 (3):213-224. doi:10.1111/j.1759-6831.2011.00136.x

Gunawardene NR, Daniels DA, Gunatilleke I, Gunatilleke C, Karunakaran P, Nayak GK, Prasad S, Puyravaud P, Ramesh B, Subramanian K (2007) A brief overview of the Western Ghats–Sri Lanka biodiversity hotspot. Current Science 93 (11):1567-1572

Guo X, Simmons MP, But PP-H, Shaw P-C, Wang R-J (2011) Application of DNA barcodes in *Hedyotis* L. (Spermacoceae, Rubiaceae). Journal of Systematics and Evolution 49 (3):203-212. doi:10.1111/j.1759-6831.2011.00130.x

Gupta R (2011) Plant taxonomy: past, present and future. Published by The Energy and Resources Institute, TERI

Haider N, Allainguillaume J, Wilkinson MJ (2009) Spontaneous capture of oilseed rape (*Brassica napus*) chloroplasts by wild B. rapa: implications for the use of chloroplast transformation for biocontainment. Current Genetics 55 (2):139-150. doi:10.1007/s00294-009-0230-5

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer, Sunderland, Massachusetts, USA, pp 43-63

Han J, Shi L, Chen X, LIN Y (2012) Comparison of four DNA barcodes in identifying certain medicinal plants of Lamiaceae. Journal of Systematics and Evolution 50 (3):227–234. doi:10.1111/j.1759-6831.2012.00184.x

Han YC, Teng CZ, Zhong S, Zhou MQ, Hu ZL, Song YC (2007) Genetic variation and clonal diversity in populations of *Nelumbo nucifera* (Nelumbonaceae) in central China detected by ISSR markers. Aquatic Botany 86 (1):69-75

Hansen LB, Siegismund HR, Jorgensen RB (2003) Progressive introgression between *Brassica napus* (oilseed rape) and *B. rapa*. Heredity (Edinb) 91 (3):276-283. doi:10.1038/sj.hdy.6800335

Hardy OJ, Maggia L, Bandou E, Breyne P, Caron H, CHEVALLIER MH, Doligez A, Dutech C, Kremer A, Latouche-Hallé C (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. Molecular Ecology 15 (2):559-571

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. Proc Biol Sci 270 (1512):313-321. doi:10.1098/rspb.2002.2218

Hebert PD, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Systematic Biology 54 (5):852-859

Hebert PD, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society B Biological Sciences 270 Suppl 1:S96-99. doi:10.1098/rsbl.2003.0025

Hiremath SC, Nagasampige MH (2004) Genome size variation and evolution in some species of *Dalbergia* Linn.f. (Fabaceae). Caryologia 57 (4):367-372

Hoban SM, McCleary TS, Schlarbaum SE, Romero-Severson J (2009) Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. Biology Letters 5 (3):324-327. doi:10.1098/rsbl.2009.0031

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim KJ, Kress WJ, Schneider H, van AlphenStahl J, Barrett SCH, van den Berg C, Bogarin D, Burgess KS, Cameron KM, Carine M, Chacon J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TAJ, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR, Kim JS, Kim YD, Lahaye R, Lee HL, Long DG, Madrinan S, Maurin O, Meusnier I, Newmaster SG, Park CW, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi DK, Little DP, Grp CPW (2009) A DNA barcode for land plants. Proceedings of National Academy of Sciences USA 106 (31):12794-12797. doi:10.1073/pnas.0905845106

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and Using a Plant DNA Barcode. Plos One 6 (5):e19254. doi:10.1371/journal.pone.0019254

Holub J (1999). Folia Geobotanica 34 (4):497-499

Hora SL (1949) Climates as affecting the Satpura Hypothesis. Proceedings of National Institute of Sciences, India (8):15

Hore DK (1990) Symplocaceae. Fascicles of flora of India, vol 21. Botanical Survey of India,

Jaakola L, Suokas M, Häggman H (2010) Novel approaches based on DNA barcoding and high-resolution melting of amplicons for authenticity analyses of berry species. Food Chemistry 123 (2):494-500. doi:http://dx.doi.org/10.1016/j.foodchem.2010.04.069

Jeanson ML, Labat J-N, Little DP (2011) DNA barcoding: a new tool for palm taxonomists? Annals of Botany-London. doi:10.1093/aob/mcr158

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable'minisatellite' regions in human DNA. Nature 314 (6006):67-73

Ji PZ, Li H, Gao LZ, Zhang J, Cheng ZQ, Huang XQ (2011) ISSR diversity and genetic differentiation of ancient tea (*Camellia sinensis* var. *assamica*) plantations from China: Implications for precious tea germplasm conservation. Pakistan Journal of Botany 43 (1):281-291

Jiang Y, Ding C-b, Zhang L, Yang R, Zhou Y, Tang L (2011) Identification of the genus *Epimedium* with DNA barcodes. Journal of Medicinal Plants Research 5 (28):6413-6417

Jones C, Edwards K, Castaglione S, Winfield M, Sala F, Van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular breeding 3 (5):381-390

Jones FA, Erickson DL, Bernal MA, Bermingham E, Kress WJ, Herre EA, Muller-Landau HC, 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

Joshi SP, Ranjekar PK, Gupta VS (1999) Molecular markers in plant genome analysis. Current Science India 77 (2)

Karanth KP (2003) Evolution of disjunct distributions among wet-zone species of the Indian subcontinent: Testing various hypotheses using a phylogenetic approach. Current Science 85 (9):1276-1283

Kardolus JP, van Eck HJ, van den Berg RG (1998) The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (Solanaceae). Plant Systematics and Evolution 210 (1-2):87-103

Kelly LJ, Ameka GK, Chase MW (2010) DNA Barcoding of African Podostemaceae (River-Weeds): A Test of Proposed Barcode Regions. Taxon 59 (1):251-260

Kesanakurti PR, Fazekas AJ, Burgess KS, Percy DM, Newmaster SG, Graham SW, Barrett SC, Hajibabaei M, Husband BC (2011) Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. Molecular Ecology 20 (6):1289-1302. doi:10.1111/j.1365-294X.2010.04989.x

Khan SA, K.J. M, Tayaab M, Abdin MZ (2009) RAPD Profile for Authentication of Medicinal Plant *Glycyrrhiza glabra* Linn. Journal of Food Safety 11: 24-28

Kharkwal M, Jain H (2004) Plant breeding: Mendelian to molecular approaches. Published by Kluwer/ Springer

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. doi:10.1007/Bf01731581

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. Molecular Ecology 7 (9):1151-1161

Kochzius M, Seidel C, Antoniou A, Botla SK, Campo D, Cariani A, Vazquez EG, Hauschild J, Hervet C, Hjörleifsdottir S (2010) Identifying fishes through DNA barcodes and microarrays. Plos One 5 (9):e12620

Kothari MJ, Moorthy S (1993) Flora of Raigad District, Maharashtra State.

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology Letters 15 (4):378-392

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). doi: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

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 USA 102 (23):8369-8374. doi:10.1073/pnas.0503123102

Kulkarni BG (1988) Flora of Sindhudurg.

Kumar P, Gupta VK, Misra AK, Modi DR, Pandey BK (2009) Potential of Molecular Markers in Plant Biotechnology. Plant Omics Journal 2 (4):141-162

La Salle J, Wheeler Q, Jackway P, Winterton S, Hobern D, Lovell D (2009) Accelerating taxonomic discovery through automated character extraction. Zootaxa 2217:43-55

Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences USA 105 (8):2923-2928. doi:10.1073/pnas.0709936105

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30 (22):3276-3278

Lavin M, Pennington RT, Klitgaard BB, Sprent JI, de Lima HC, Gasson PE (2001) The dalbergioid legumes (Fabaceae): Delimitation of a pantropical monophyletic clade. American Journal of Botany 88 (3):503-533. doi:10.2307/2657116

Lee G, Park SY, Yum S, Woo S, Lee Y-H, Hwang SY, Park H-S, Moh SH, Lee S, Lee T-K (2012) Development of DNA chip for verification of 25 microalgae collected from southern coastal region in Korea. BioChip Journal 6 (4):325-334

Levin D, Funderburg S (1979) Genome size in angiosperms: temperate versus tropical species. American Naturalist:784-795

Li F-W, Kuo L-Y, Rothfels CJ, Ebihara A, Chiou W-L, Windham MD, Pryer KM (2011) *rbcL* and *matK*: Earn Two Thumbs Up as the Core DNA Barcode for Ferns. Plos One 6 (10):e26597. doi:10.1371/journal.pone.0026597

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* (Orchidaceae) based on AFLP. Genetica 133 (2):159-166

Little DP, Knopf P, Schulz C (2013) DNA barcode identification of Podocarpaceae-the second largest Conifer family. Plos One 8 (11). doi:10.1371/journal.pone.0081008

Liu J, Moller M, Gao LM, Zhang DQ, Li DZ (2011) DNA barcoding for the discrimination of *Eurasian yews* (Taxus L., Taxaceae) and the discovery of cryptic species. Molecular Ecology Resources 11 (1):89-100. doi:10.1111/j.1755-0998.2010.02907.x

Liu P, Yang Y, Hao C, Guo W (2007) Ecological risk assessment using RAPD and distribution pattern of a rare and endangered species. Chemosphere 68 (8):1497-1505

Liu Z, Zeng X, Yang D, Chu G, Yuan Z, Chen S (2012) Applying DNA barcodes for identification of plant species in the family Araliaceae. Gene 499 (1):76-80. doi:10.1016/j.gene.2012.02.016

Long D, Paton J, Squirrell J, Woodhead M, Hollingsworth P (2006) Morphological, ecological and genetic evidence for distinguishing *Anastrophyllum joergensenii* Schiffn. and *A. alpinum* Steph.(Jungermanniopsida: Lophoziaceae). Journal of Bryology 28 (2):108-117

Lopez S, Rousset F, Shaw FH, Shaw RG, Ronce O (2009) Joint effects of inbreeding and local adaptation on the evolution of genetic load after fragmentation. Conservation Biology 23 (6):1618-1627. doi:10.1111/j.1523-1739.2009.01326.x

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. Annual Review of Ecology and Systematics 15:65-95

Lu H-P, Cai Y-W, Chen X-Y, Zhang X, Gu Y-J, Zhang G-F (2006) High RAPD but no cpDNA sequence variation in the endemic and endangered plant, *Heptacodium miconioides* Rehd.(Caprifoliaceae). Genetica 128 (1-3):409-417

Lu J, Knox MR, Ambrose MJ, Brown JK, Ellis TH (1990) Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. Nucleic Acids Research 18 (24):7213-7218

Ma X-Y, Xie C-X, Liu C, Song J-Y, Yao H, Luo K, Zhu Y-J, Gao T, Pang X-H, Qian J (2010) Species identification of medicinal pteridophytes by a DNA barcode marker, the chloroplast *psbA-trnH* intergenic region. Biological and Pharmaceutical Bulletin 33 (11):1919-1924

Mani M (1974) Biogeographical evolution in India. In: Ecology and biogeography in India. Springer, pp 698-724

Manktelow M (2010) History of taxonomy. Lecture from Dept. of Systematic Biology, Uppsala University, 2010.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437 (7057):376-380

Meher-Homji VM (1975) On the montane species of Kodaikanal, South India. Phytocoenologia 2:28-39

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. doi:10.1080/10635150600969864

Meusnier I, Singer GA, Landry J-F, Hickey DA, Hebert PD, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9 (1):214

Miller MP (1997) Tools for population genetic analysis (TFPGA) Version 13. Department of Biological Science, Northern Arizona University, Arizona, USA

Mittermier R, Gil PR, Hoffmann M, Pilgrim J, Brooks T, Mittermeier C, Lamoreaux J, Da Fonseca G (2004) Hotspots revisited. CEMEX, SA de CV

Miwa H, Odrzykoski IJ, Matsui A, Hasegawa M, Akiyama H, Jia Y, Sabirov R, Takahashi H, Boufford DE, Murakami N (2009) Adaptive evolution of *rbcL* in *Conocephalum* (Hepaticae, bryophytes). Gene 441 (1-2):169-175. doi:10.1016/j.gene.2008.11.020

MoEF, Forests MoE, India. Go (2011-2012) ANNUAL REPORT 2011-12.

Mohana GS, Shaanker RU, Ganeshaiah KN, Dayanandan S (2001) Genetic relatedness among developing seeds and intra fruit seed abortion in *Dalbergia sissoo* (Fabaceae). American Journal of Botany 88 (7):1181-1188. doi:10.2307/3558328

Mower JP, Stefanovic S, Young GJ, Palmer JD (2004) Plant genetics: Gene transfer from parasitic to host plants. Nature 432 (7014):165-166. doi:http://www.nature.com/nature/journal/v432/n7014/suppinfo/432165b_S1.html

Muellner AN, Schaefer H, Lahaye R (2011) Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). Molecular Ecology Resources 11 (3):450-460. doi:10.1111/j.1755-0998.2011.02984.x

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology 155:335

Myers N (1988) Threatened biotas:" hot spots" in tropical forests. Environmentalist 8 (3):187-208

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403 (6772):853-858

Nagoshi RN, Brambila J, Meagher RL (2011) Use of DNA barcodes to identify invasive armyworm *Spodoptera* species in Florida. Journal of Insect Science 11 (1):154

Naik D, Singh D, Vartak V, Paranjpe S, Bhargava S (2009) Assessment of morphological and genetic diversity in *Gmelina arborea* Roxb. New Forest 38 (1):99-115

Nair K (1986) Preservation of *Dalbergia* Lf in Kerala by establishment of a germplasm bank. KFRI Research Report (43)

Nair N, Daniel P (1986) floristic diversity of the Western Ghats and its conservation: a review. Proceedings Animal sciences-Indian Academy of Sciences

Nan P, Shi S, Peng SL, Tian CJ, Zhong Y (2003) Genetic diversity in *Primula obconica* (Primulaceae) from central and south-west China as revealed by ISSR markers. Annals of Botany-London 91 (3):329-333

Neale DB, Devey ME, Jermstad ISD, Ahuja MR, Alosi MC, Marshall KA (1992) Use of DNA markers in forest tree improvement research New Forest 6:391-407

Nei M (1972) Genetic distance between populations. American Naturalist 106 (949):283-287

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences 70 (12):3321-3323

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics 41 (2):225-233

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

Newmaster SG, Ragupathy S (2009) Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). Molecular Ecology Resources 9:172-180. doi:10.1111/j.1755-0998.2009.02642.x

Nithaniyal S, Newmaster SG, Ragupathy S, Krishnamoorthy D, Vassou SL, Parani M (2014) DNA barcode authentication of wood samples of threatened and commercial timber trees within the tropical dry evergreen forest of India. Plos One 9 (9):e107669. doi:10.1371/journal.pone.0107669

Niyomdham C (2002) An account of *Dalbergia* (Leguminosae-Papillionoideae) in Thailand. Thailand Forest Bulletin (BOT) 30:124-166

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13 (5):1143-1155

Nybom H, Bartish IV (2000) Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. Perspectives in Plant Ecology, Evolution and Systematics 3 (2):93-114

Ogden R, McGough HN, Cowan RS, Chua L, Groves M, McEwing R (2009) SNP-based method for the genetic identification of ramin *Gonystylus* spp. timber and products: applied research meeting CITES enforcement needs. Endangered Species Research 9 (3):255-261

Ojeda DI, Santos-Guerra A, Oliva-Tejera F, Jaen-Molina R, Caujape-Castells J, Marrero-Rodriguez A, Cronk Q (2014) DNA barcodes successfully identified Macaronesian *Lotus* (Leguminosae) species within early diverged lineages of Cape Verde and mainland Africa. AoB Plants 6. doi:10.1093/aobpla/plu050

Orgiazzi A, Dunbar MB, Panagos P, de Groot GA, Lemanceau P (2015) Soil biodiversity and DNA barcodes: opportunities and challenges. Soil Biology and Biochemistry 80:244-250. doi:http://dx.doi.org/10.1016/j.soilbio.2014.10.014

Pacey-Miller T, Henry R (2003) Single-nucleotide polymorphism detection in plants using a single-stranded pyrosequencing protocol with a universal biotinylated primer. Analytical Biochemistry 317 (2):166-170

Page RDM (1996) TreeView: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12: 357-358

Page RDM (2001) TreeView. Glasgow University, Glasgow, UK

Palmer JD (1985) Comparative organization of chloroplast genomes. Annual Review of Genetics 19 (1):325-354

Panda H (2004) Handbook on Ayurvedic Medicines with formulae, processes and their uses. National Institute of Industrial Re,

Pandit SS, Mitra S, Giri AP, Pujari KH, Patil BP, Jambhale ND, Gupta VS (2007) Genetic diversity analysis of mango cultivars using inter simple sequence repeat markers. Current Science (00113891) 93 (8)

Pang X, Song J, Zhu Y, Xie C, Chen S (2010) Using DNA barcoding to identify species within Euphorbiaceae. Planta Medica 76 (15):1784-1786. doi:10.1055/s-0030-1249806

Pang XH, Liu C, Shi LC, Liu R, Liang D, Li H, Cherny SS, Chen SL (2012) Utility of the *trnH-psbA* intergenic spacer region and Its combinations as plant DNA barcodes: A Meta-Analysis. Plos One 7 (11). doi:10.1371/journal.pone.0048833

Parveen I, Singh HK, Raghuvanshi S, Pradhan UC, Babbar SB (2012) DNA barcoding of endangered Indian Paphiopedilum species. Molecular Ecology Resources 12 (1):82-90. doi:10.1111/j.1755-0998.2011.03071.x

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6 (1):288-295

Pennisi E (2007) Taxonomy. Wanted: a barcode for plants. Science (New York, NY) 318 (5848):190

Pettengill JB, Neel MC (2010b) An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus *Agalinis* (Orobanchaceae). American Journal of Botany 97 (8):1391-1406. doi:10.3732/ajb.0900176

Phong DT, Tang DV, Hien VTT, Ton ND, Van HN (2014) Nucleotide diversity of a nuclear and four chloroplast DNA regions in rare tropical wood species of *Dalbergia* in Vietnam: a DNA barcode identifying utility. Asian Journal of Applied Sciences 02 (02):116-125

Pires AC, Marinoni L (2010) DNA barcoding and traditional taxonomy unified through Integrative Taxonomy: a view that challenges the debate questioning both methodologies. Biota Neotropica 10 (2):339-346

Pons O, Petit R (1995) Estimation, variance and optimal sampling of gene diversity. Theoretical and Applied Genetics 90 (3-4):462-470

Pons O, Petit R (1996) Measwring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144 (3):1237-1245

Posada D (2008) jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25 (7):1253-1256

Pourmohammad A (2013) Application of molecular markers in medicinal plant studies. Agriculture and Environment 5:80-90

Prain D (1904) The species of *Dalbergia* of South-eastern Asia. Annals of the Royal Botanical Gardens (10):1-114

Puyravaud JP, Davidar P, Laurance WF (2010) Cryptic destruction of India's native forests. Conservation Letters 3 (6):390-394

Pyle RL, Earle JL, Greene BD (2008) Five new species of the damselfish genus *Chromis* (Perciformes: Labroidei: Pomacentridae) from deep coral reefs in the tropical western Pacific. Zootaxa 1671:3-31

Rajwade AV, Arora RS, Kadoo NY, Harsulkar AM, Ghorpade PB, Gupta VS (2010) Relatedness of Indian Flax Genotypes (*Linum usitatissimum* L.): An Inter-Simple Sequence Repeat (ISSR) Primer Assay. Molecular Biotechnology 45 (2):161-170

Ramesh BR, Menon S, Bawa KS (1997) A vegetation based approach to biodiversity gap analysis in the Agastyamalai region, Western Ghats, India Ambio 26 (8):529-536

Rao G, Krishnakumar G, Dudani SN, Chandran MS, Ramachandra T (2013) Vegetation Changes along Altitudinal Gradients in Human Disturbed Forests of Uttara Kannada, Central Western Ghats. Journal of Biodiversity 4:61-68

Rasolomampianina R, Bailly X, Fetiarison R, Rabevohitra R, Bena G, Ramaroson L, Raherimandimby M, Moulin L, De Lajudie P, Dreyfus B, Avarre JC (2005) Nitrogen-fixing nodules from rose wood legume trees (*Dalbergia* spp.) endemic to Madagascar host seven different genera belonging to alpha- and beta-Proteobacteria. Molecular Ecology 14 (13):4135-4146. doi:10.1111/j.1365-294X.2005.02730.x

Reddy MP, Sarla N, Siddiq EA (2002) Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica 128 (1):9-17

Ren BQ, Xiang XG, Chen ZD (2010) Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. Molecular Ecology Resources 10 (4):594-605. doi:10.1111/j.1755-0998.2009.02815.x

Ribeiro RA, Lavin M, Lemos-Filho JP, Filho CVM, Dos Santos FR, Lovato MB (2007) The genus *Machaerium* (Leguminosae) is more closely related to *Aeschynomene* Sect. *Ochopodium* than to *Dalbergia*: Inferences from combined sequence data. Systematic Botany 32 (4):762-771

Ribeiro RA, Ramos ACS, Filho JPD, Lovato MB (2005) Genetic variation in remnant populations of *Dalbergia nigra* (Papilionoideae), an endangered tree from the Brazilian Atlantic Forest. Annals of Botany-London 95 (7):1171-1177. doi:10.1093/Aob/Mci128

Richards E, Reichardt M, Rogers S (1994) Preparation of genomic DNA from plant tissue. Current Protocols in Molecular Biology 1:2.3.1 - 2.3.7

Richardson JE, Fay MF, Cronk QC, Chase MW (2003) Species delimitation and the origin of populations in island representatives of *Phylica* (Rhamnaceae). Evolution 57 (4):816-827

Rieseberg LH, Wood TE, Baack EJ (2006) The nature of plant species. Nature 440 (7083):524-527.

Robin VV, Sinha A, Ramakrishnan U (2010) Ancient geographical gaps and paleo-climate shape the phylogeography of an endemic bird in the sky islands of southern India. PloS One 5 (10). doi:10.1371/journal.pone.0013321

Rodgers W, Panwar H (1988) Planning a wildlife protected area network in India2 vols. Project FO: IND/82/003. FAO, Dehradun. 339, 267 pp.

Rodgers W, Panwar H, Mathur V (2002) Wildlife Protected Area Network in India: A Review (Executive Summary), Wildlife Institute of India. Dehradun 44p

Ronikier M (2001) The use of AFLP markers in conservation genetics--a case study on *Pulsatilla vernalis* in the Polish lowlands. Cellular & Molecular Biology Letters 7 (2B):677-684

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19 (12):1572-1574

Rouhan G, Gaudeul M (2014) Plant Taxonomy: A Historical Perspective, Current Challenges, and Perspectives. In: Molecular Plant Taxonomy. Springer, pp 1-37

Rout GR, Bhatacharya D, Nanda RM, Nayak S, Das P (2003) Evaluation of genetic relationships in *Dalbergia* species using RAPD markers. Biodiversity and Conservation 12 (2):197-206. doi:10.1023/A:1021996020947

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19 (18):2496-2497. doi:10.1093/bioinformatics/btg359

Rubio-Moraga A, Candel-Perez D, Lucas-Borja ME, Tiscar PA, Vinegla B, Linares JC, Gomez-Gomez L, Ahrazem O (2012) Genetic diversity of *Pinus nigra* Arn. populations in southern Spain and northern Morocco revealed by inter-simple sequence repeat profiles. International Journal of Molecular Sciences 13 (5):5645-5658. doi:10.3390/Ijms13055645

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, vol 2. Cold spring harbor laboratory press New York,

Sandral G, Degtjareva G, Kramina T, Sokoloff D, Samigullin T, Hughes S, Valiejo-Roman C (2010) Are Lotus creticus and *Lotus cytisoides* (Leguminosae) closely related species? Evidence from nuclear ribosomal ITS sequence data. Genetic Resources and Crop Evolution 57 (4):501-514. doi:10.1007/s10722-009-9486-3

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, USA 74 (12):5463-5467

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

Schneider H, Schuettpelz E (2006) Identifying fern gametophytes using DNA sequences. Molecular Ecology Notes 6 (4):989-991. doi:10.1111/j.1471-8286.2006.01424.x

Selvaraj D, Sarma RK, Sathishkumar R (2008) Phylogenetic analysis of chloroplast matK gene from Zingiberaceae for plant DNA barcoding. Bioinformation 3 (1):24-27

Shaw AJ (2001) Biogeographic Patterns and Cryptic Speciation in Bryophytes. Journal of Biogeography 28 (2):253-261

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92 (1):142-166

Shi LC, Zhang J, Han JP, Song JY, Yao H, Zhu YJ, Li JC, Wang ZZ, Xiao W, Lin YL, Xie CX, Qian ZZ, Chen SL (2011) Testing the potential of proposed DNA barcodes for species identification of Zingiberaceae. Journal of Systematics and Evolution 49 (3):261-266. doi:10.1111/j.1759-6831.2011.00133.x

Simeone MC, Piredda R, Papini A, Vessella F, Schirone B (2013) Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (*Quercus*, Fagaceae): problems, prospects and phylogenetic implications. Botanical Journal of the Linnean Society 172 (4):478-499. doi:10.1111/boj.12059

Singh NP, Lal RK, Shasany AK (2009) Phenotypic and RAPD diversity among 80 germplasm accessions of the medicinal plant isabgol (*Plantago ovata*), Plantaginaceae. Genetics and Molecular Research 8:1273-1284

Skuza L, Filip E, Szućko I (2013) Use of organelle markers to study genetic diversity in soybean. INTECH Open Access Publisher,

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236 (4803):787-792

Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43:1349–1368

Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada 37 (6):977-983

Smith MA, Fisher BL, Hebert PD (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philosophical Transactions of the Royal Society London B Biological Sciences 360 (1462):1825-1834. doi:10.1098/rstb.2005.1714

Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH, Ims RA, Yoccoz NG (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Frontiers in Zoology 6 (1):16

Sokal RR, Sneath PHA (1963) Principles of numerical taxonomy. A Series of books in biology. W. H. Freeman, San Francisco,

Soltis DE, Kuzoff RK (1995) Discordance between Nuclear and Chloroplast Phylogenies in the *Heuchera* Group (Saxifragaceae). Evolution; International Journal of Organic Evolution 49 (4):727-742. doi:10.2307/2410326

Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS (2007) Autopolyploidy in Angiosperms: Have We Grossly Underestimated the Number of Species? Taxon 56 (1):13-30. doi:10.2307/25065732

Song J, Yao H, Li Y, Li X, Lin Y, Liu C, Han J, Xie C, Chen S (2009) Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. Journal of Ethnopharmacology 124 (3):434-439. doi:10.1016/j.jep.2009.05.042

Song ZQ, Li XF, Wang HG, Wang JH (2010) Genetic diversity and population structure of Salvia miltiorrhiza Bge in China revealed by ISSR and SRAP. Genetica 138 (2):241-249. doi:10.1007/s10709-009-9416-5

Srirama R, Senthilkumar U, Sreejayan N, Ravikanth G, Gurumurthy BR, Shivanna MB, Sanjappa M, Ganeshaiah KN, Shaanker RU (2010) Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepato-protective plant using molecular tools. Journal of Ethnopharmacology 130 (2):208-215. doi:10.1016/j.jep.2010.04.042

Starr JR, Naczi RFC, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). Molecular Ecology Resources 9:151-163. doi:10.1111/j.1755-0998.2009.02640.x

Staudacher K, Wallinger C, Schallhart N, Traugott M (2011) Detecting ingested plant DNA in soil-living insect larvae. Soil Biology Biochemistry 43 (2):346-350. doi:10.1016/j.soilbio.2010.10.022

Stebbins Jr C (1950) Variation and evolution in plants. In: Variation and evolution in plants. 1950 pp. 643 pp.

Stech M, Quandt D (2014) 20,000 species and five key markers: the status of molecular bryophyte phylogenetics. Phytotaxa 9 (1):196-228

Stoeckle M (2003) Taxonomy, DNA, and the bar code of life. BioScience 53 (9):796-797

Sun CQ, Wang XK, Li ZC, Yoshimura A, Iwata N (2001) Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. Theoretical and Applied Genetics 102 (1):157-162

Sun XQ, Zhu YJ, Guo JL, Peng B, Bai MM, Hang YY (2012) DNA barcoding the *Dioscorea* in China, a vital group in the evolution of monocotyledon: Use of *matK* gene for species discrimination. Plos One 7 (2). doi:10.1371/journal.pone.0032057

Sunarno B, Ohashi H (1997) *Dalbergia* (Leguminosae) of Borneo. Journal of Japanese Botany 72 (4):198-220

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast *trnL* (*UAA*) intron for plant DNA barcoding. Nucleic Acids Research 35 (3):e14-e14

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. doi:10.1093/molbev/msr121

Tatikonda L, Wani SP, Kannan S, Beerelli N, Sreedevi TK, Hoisington DA, Devi P, Varshney RK (2009) AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L., a biofuel plant. Plant Science 176 (4):505-513

Theerakulpisut P, Kanawapee N, Maensiri D, Bunnag S, Chantaranothai P (2008) Development of species-specific SCAR markers for identification of three medicinal species of *Phyllanthus*. Journal of Systematics and Evolution 46 (4):614-621

Thomas S, Haridasan K, Borthakur S (1998) Floristic study on rattans and its relevance in forestry of Arunachal Pradesh. Arunachal Forest News 16 (1-2):19-24

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. doi:10.1093/nar/22.22.4673

Thorne RF (2000) The classification and geography of the flowering plants: dicotyledons of the class Angiospermae. The Botanical Review 66 (4):441-647

Thothathri K (1983) Critical notes on the genus *Dalbergia* Linn. f. Bulletin of the Botanical Survey of India 25 (1-4):169-173

Thothathri K (1987) Taxonomic revision of the tribe Dalbergieae in the Indian subcontinent. Botanical Survey of India (Calcutta),

Tiayyba Riaz WS, Alain Viari, François Pompanon, Pierre Taberlet and Eric Coissac (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research:1-11. doi:10.1093/nar/gkr732

Tordoff A, Baltzer M, Fellowes J, Pilgrim J, Langhammer P (2012) Key biodiversity areas in the Indo-Burma Hotspot: process, progress and future directions. Journal of Threatened Taxa 4 (8):2779-2787

Tripathi AM, Tyagi A, Kumar A, Singh A, Singh S, Chaudhary LB, Roy S (2013) The internal transcribed spacer (ITS) region and *trnH-psbA* are suitable candidate loci for DNA barcoding of tropical tree species of India. Plos One 8 (2). doi:10.1371/journal.pone.0057934

Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F, Gielly L, Cruaud C, Nascetti G, Wincker P, Swenson JE, Taberlet P (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol Ecol Resour 9 (1):51-60. doi:10.1111/j.1755-0998.2008.02352.x

Valentini A, Miquel C, Taberlet P (2010) DNA Barcoding for Honey Biodiversity. Diversity 2 (4):610

Van De Wiel C, Van Der Schoot J, Van Valkenburg J, Duistermaat H, Smulders M (2009) DNA barcoding discriminates the noxious invasive plant species, floating pennywort (*Hydrocotyle ranunculoides* Lf), from non-invasive relatives. Molecular Ecology Resources 9 (4):1086-1091

Van E, Benjamin W, Jelinski N, Berry PE, Hipp AL (2006) Phylogeny and biogeography of *Croton alabamensis* (Euphorbiaceae), a rare shrub from Texas and Alabama, using DNA sequence and AFLP data. Molecular Ecology 15 (10):2735-2751

Vanderpoorten A, Goffinet B (2006) Mapping uncertainty and phylogenetic uncertainty in ancestral character state reconstruction: an example in the moss genus *Brachytheciastrum*. Systematic Biology 55 (6):957-971

Vatanparast M, Klitgard BB, Adema FACB, Pennington RT, Yahara T, Kajita T (2013) First molecular phylogeny of the pantropical genus *Dalbergia*: implications for infrageneric circumscription and biogeography. South African Journal of Botany 89:143-149. doi:10.1016/j.sajb.2013.07.001

Vidya TNC, Fernando P, Melnick DJ, Sukumar R (2005) Population differentiation within and among Asian elephant (*Elephas maximus*) populations in southern India. Heredity 94 (1):71-80. doi:10.1038/sj.hdy.6800568

Vijayan K, Tsou C (2010) DNA barcoding in plants: taxonomy in a new perspective. Current Science (Bangalore) 99 (11):1530-1541

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23 (21):4407-4414

Wang J-m, Yang J-m, Zhu J-h, Jia Q-j, Tao Y-z (2010a) Assessment of genetic diversity by simple sequence repeat markers among forty elite varieties in the germplasm for malting barley breeding. Journal of Zhejiang University Science B 11 (10):792-800

Wang T, SU YJ, LI XY, Zheng B, CHEN GP, ZENG QL (2004) Genetic structure and variation in the relict populations of *Alsophila spinulosa* from southern China based on RAPD markers and cpDNA atpB-rbcL sequence data. Hereditas 140 (1):8-17

Wang WQ, Wu YR, Yan YH, Ermakova M, Kerstetter R, Messing J (2010b) DNA barcoding of the Lemnaceae, a family of aquatic monocots. BMC Plant Biology 10. doi:10.1186/1471-2229-10-205

Wang Z-S, An S-Q, Liu H, Leng X, Zheng J-W, Liu Y-H (2005) Genetic structure of the endangered plant *Neolitsea sericea* (Lauraceae) from the Zhoushan Archipelago using RAPD markers. Annals of Botany-London 95 (2):305-313

Warrier PK, Nambiar VPK, Ramankutty C (1996) Indian Medicinal Plants, vol 5. Orient Longman, India

Weitschek E, Van Velzen R, Felici G, Bertolazzi P (2013) BLOG 2.0: a software system for character-based species classification with DNA Barcode sequences. What it does, how to use it. Molecular Ecology Resources 13 (6):1043-1046. doi:10.1111/1755-0998.12073

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18:7213-7218

Wetton JH, Carter RE, Parkin DT, Walters D (1987) Demographic study of a wild house sparrow population by DNA fingerprinting. Nature 327 (6118):147-149. doi:10.1038/327147a0

Wheeler QD, Raven PH, Wilson EO (2004) Taxonomy: impediment or expedient? Science (New York, NY) 303 (5656):285

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

Widmer A, Cozzolino S, Pellegrino G, Soliva M, Dafni A (2000) Molecular analysis of orchid pollinaria and pollinaria-remains found on insects. Molecular Ecology 9:1911–1914

Yadav S, Sardesai M (2002) Flora of Kolhapur district.

Yamaguchi A, Kawamura H, Horiguchi T (2006) A further phylogenetic study of the heterotrophic dinoflagellate genus, *Protoperidinium* (Dinophyceae) based on small and large subunit ribosomal RNA gene sequences. Phycological Research 54 (4):317-329

Yan HF, Hao G, Hu CM, Ge XJ (2011) DNA barcoding in closely related species: A case study of *Primula* L. sect. *Proliferae Pax* (Primulaceae) in China. Journal of Systematics and Evolution 49 (3):225-236

Yang JB, Wang YP, Moller M, Gao LM, Wu D (2012) Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae). Molecular Ecology Resources 12 (2):267-275. doi:10.1111/j.1755-0998.2011.03095.x

Yang L, Fu S, Khan MA, Zeng W, Fu J (2013) Molecular cloning and development of RAPD-SCAR markers for *Dimocarpus longan* variety authentication. Springerplus 2 (1):501

Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Xu H, Zhu Y, Xiao P (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. Plos One 5 (10):e13102

Yao H, Song JY, Ma XY, Liu C, Li Y, Xu HX, Han JP, Duan LS, Chen SL (2009) Identification of *Dendrobium* species by a candidate DNA barcode sequence: the chloroplast *psbA-trnH* intergenic region. Planta Medica 75 (6):667-669. doi:10.1055/s-0029-1185385

Yao H, Zhao Y, Chen DF, Chen JK, Zhou TS (2008) ISSR primer screening and preliminary evaluation of genetic diversity in wild populations of *Gycyrrhiza uralensis*. Biologia Plantarum 52 (1):117-120. doi:10.1007/s10535-008-0022-7

Yap IV, Nelson RJ (1996) Winboot: A program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms IRRI Discussion Paper Series, vol 14. IRRI, Manila, Philippines

Yeh F, Yang RC, Boyle T (2000) PopGene (v1.32) Microsoft windows-based freeware for population genetic analysis.

Yesson C, Barcenas RT, Hernandez HM, De La LUZ RM, Prado A, Rodriguez VM, Hawkins JA (2011) DNA barcodes for Mexican Cactaceae, plants under pressure from wild collecting. Molecular Ecology Resources 11 (5):775-783

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European patent application 92402629 (7)

Zawko G, Krauss S, Dixon K, Sivasithamparam K (2001) Conservation genetics of the rare and endangered *Leucopogon obtectus* (Ericaceae). Molecular Ecology 10 (10):2389-2396

Zhang JM, Wang JX, Xia T, Zhou SL (2009) DNA barcoding: species delimitation in tree peonies. Science in China Series C-Life Sciences 52 (6):568-578. doi:10.1007/s11427-009-0069-5

Zhang Y, Du X, Zhang Q, Gao X, Su Z (2005) Fitness analysis of seed and vegetative reproduction of clonal tree *Symplocos laurina*. Ying Yong Sheng Tai Xue Bao 16 (9):1687-1692

Zheng W, Wang L, Meng L, Liu J (2008) Genetic variation in the endangered *Anisodus tanguticus* (Solanaceae), an alpine perennial endemic to the Qinghai-Tibetan Plateau. Genetica 132 (2):123-129. doi:10.1007/s10709-007-9154-5

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20 (2):176-183. doi:10.1006/geno.1994.1151

Rasika M. Bhagwat

8 Shree Society, cross road no.2
Ashok nagar, Kandivali east
Mumbai-400101

Email: rasikamb@gmail.com
Tel: +919881046747

EDUCATION:

- **Ph.D**. Biotechnology, pursuing, CSIR-National Chemical Laboratory, Pune, India

- **Master of Science** (M. Sc.) 2005, Botany (GPA 6.95 with 'A' grade), University of Pune, India.

- **Bachelor of Science** (B. Sc.) 2002, Botany (74 % First class with distinction), University of Pune, India.

RESEARCH EXPERIENCE:

- **Senior Research Fellow** at NCL. Apr 2011-till date
  I am working on DNA barcoding of *Dalbergia* species from Western Ghats (one of the biodiversity hotspots) of India with an objective to design a barcode from either chloroplast, or nuclear region to identify the different *Dalbergia* species.

- **Project Assistant** at NCL Dec 2007- Mar 2011
  DNA barcoding of *Dalbergia* species from Western Ghats

- **Project Assistant** at NCL Sep 2005- Nov 2007
  Genetic diversity studies of *Symplocos* using low copy nuclear gene and ISSR markers. The populations of *Symplocos racemosa* collected from the biodiversity hotspot were analyzed using chloroplast and mitochondrial markers with PCR-RFLP technique.

ACADEMIC PROJECTS

- **Drought Tolerance Studies in Wheat** Jan 2005- May 2005

- **Building Database of Orchids in Western Ghats India** May 2004-July 2004
  Centre for Ecological Sciences, Indian Institute of Science (IISc), Bangalore (http://wgbis.ces.iisc.ernet.in/biodiversity/sahyadri_database/database1/index.htm)

- **Protocol for the Micropropagation of Different Orchids** Jun 2002-Apr 2003

Skills:

- Extraction of DNA
- Gel electrophoresis and PAGE
- PCR technology based molecular marker analysis (PCR-RFLP, ISSR)
- Cloning of gene of interest using various vector systems and sequencing
- DNA sequencing with automated sequencer (MegaBACE 1000)
- Sequence analysis using tools like MEGA, CLUSTAL X, Dnasp and BIOEDIT
- Bioinformatics tools for search, blast and alignment of sequences and analysis

Rasika M. Bhagwat

8 Shree Society, cross road no.2
Ashok nagar, Kandivali east
Mumbai-400101

Email: rasikamb@gmail.com
Tel: +919881046747

- Phylogenetic and biostatistical analysis of data using various software packages such as MVSP, HAPLODIV, HAPLONST, Arlequin, GenAlEx, PAUP AliView, seaview, MrBays, POPGENE, TaxonDNA, etc.
- Tissue culture techniques
- Field experience for tissue collection

Peer-reviewed Publications:

1. S. Banu, **R.M. Bhagwat**, N.Y. Kadoo, M.D. Lagu and V.S. Gupta (2010) Understanding the genetic structure of *Symplocos laurina* Wall. populations using nuclear gene markers **Genetica 138:197–210**
2. **R.M. Bhagwat**, S. Banu, B.B. Dholakia, N.Y Kadoo, M.D. Lagu, and V.S. Gupta (2014) Evaluation of genetic variability in *Symplocos laurina* Wall. from two biodiversity hotspots of India **Plant systematics and evolution 300 (10): 2239-2247**
3. Sofia Banu, **Rasika M. Bhagwat**, Meena D. Lagu, Narendra kadoo, B. G. Kulkarni and Vidya S. Gupta (2014) Effect of anthropogenic activities on patterns of genetic diversity in *Symplocos racemosa* Roxb. from the Western Ghats, India **International journal of Biodiversity vol. 2014, Article ID 296891**
4. **Rasika M. Bhagwat**, Bhushan B. Dholakia, Narendra Y. Kadoo, M. Balasundaran and Vidya S. Gupta (2015) Two new potential barcodes to discriminate *Dalbergia* species PLOS ONE DOI: 10.1371/journal.pone.0142965
5. Banu, S., Baruah, D., **Bhagwat**, **R**. **M**., Sarkar, P., Bhowmick, A., & Kadoo, N. Y. (2015). Analysis of genetic variability in Aquilaria malaccensis from Bramhaputra valley, Assam, India using ISSR markers. Flora-Morphology, Distribution, Functional Ecology of Plants, 217, 24-32.
6. Roy Subarna, Rathore Poonam, **Bhagwat Rasika**, Hegde Harsha, Hegde Satisha, Madihalli Samreen, Kholkute Sanjiva, Gupta Vidya, Jha Timir (2015) Assessment of Genetic diversity of Gymnema sylvestre (Retz.) R.Br. from Western Ghats and Eastern Himalayas, India (communicated)

Oral Presentations:

1. Darshana Baruah, **Rasika Bhagwat**, Ananya Bhowmick, Purabi Sarkar, Narendra Y. Kadoo, Sofia Banu (2014) Genetic diversity assessment of economically important and endangered plant *Aquilaria malaccensis* from Assam. In International Conference on Biodiversity, Bioresources and Biotechnology Organized by Association for the Advancement of Biodiversity Science Society for Applied Biotechnology Imperial Scientific Publishing from 30-31 January, 2014
2. **R. M. Bhagwat**, B.B. Dholakia, M. Balasundaran, N.Y. Kadoo, V.S. Gupta (2011) DNA barcoding: A new tool for delimitation of species boundaries in *Dalbergia* spp. In International Conference on Biodiversity and its Conservation at Pune, Maharashtra, India from 28-30 January, 2011

Rasika M. Bhagwat

8 Shree Society, cross road no.2
Ashok nagar, Kandivali east
Mumbai-400101

Email: rasikamb@gmail.com
Tel: +919881046747

3. **R. M. Bhagwat**, B.B. Dholakia, M. Balasundaran, N.Y. Kadoo, V.S. Gupta (2010) DNA barcoding of *Dalbergia* spp. From Western Ghats in India. In National Conference on Biodiversity of Medicinal and Aromatic plants: Collection, Characterization and Utilization at Anand, Gujarat, from 24-25 November, 2010.

Poster Presentations:

1. Darshana Baruah, Karishma Kashyap, **Rasika Bhagwat**, Narendra Y Kadoo and Sofia Banu (2013) How genetically fit is endangered populations of *Aquilaria agallocha* in Assam.
   In7th Annual Convention of Association of Biotechnology and Pharmacy – International Conference on Plant Biotechnology, Molecular Medicine and Human Health under the theme Recent Advances in Plant Biotechnology at University of Delhi South Campus, New Delhi from 18-20 October, 2013

2. Banu S, **Bhagwat RM**, Lagu MD, Kadoo NY, Kulkarni BG and Gupta VS (2012) Deciphering genetic structure of *Symplocos laurina*, a montane plant: a glimpse into its past and present. In 4th International Ecosummit , ecological sustainability, restoring the planet's ecosystem services at Columbus, Ohio, USA from 30 September-5 October, 2012

3. **Bhagwat RM**, Dholakia BB, Balasundaran M, Kadoo NY, Gupta VS (2012) Applying DNA barcoding to discriminate *Dalbergia* spp. from Western Ghats. In National conference on biodiversity assessment, conservation and utilizations, at Abasaheb Garware College, Pune, from 9-11February, 2012

4. **Bhagwat RM**, Dholakia BB, Balasundaran M, Kadoo NY, Gupta VS. (2011) Improving species identification of *Dalbergia* spp. with DNA barcoding. In Forth International Barcode of Life Conference at University of Adelaide, Adelaide, Australia, from 30 November. - 3 December, 2011
5. **Bhagwat RM**, Kadoo NY, Dholakia BB, and Gupta VS. (2010) DNA barcoding for delimitation of species boundaries in *Dalbergia* spp. from Western Ghats at National Chemical Laboratory, for national science day celebration
6. **RM Bhagwat**, Sofia Banu, Meena D Lagu and Vidya S Gupta (2008) *Symplocos racemosa*: a case study to analyze effect of habitat fragmentation in Western Ghats at National Chemical Laboratory, for national science day celebration
7. Sofia Banu, **RM Bhagwat**, Meena D Lagu and Vidya S Gupta (2007) *Symplocos laurina:* A glimpse into its present and past analyzed using molecular markers at National Chemical Laboratory, for national science day celebration

Achievements & Honors:

- **Best poster award** for the poster entitled '*Symplocos laurina:* A glimpse into its present and past analyzed using molecular markers'(2007)
- Completed **German language certificate course** from Pune University

Rasika M. Bhagwat

8 Shree Society, cross road no.2
Ashok nagar, Kandivali east
Mumbai-400101

Email: rasikamb@gmail.com
Tel: +919881046747

References:

1. Dr. Vidya S. Gupta, Former Chief Scientist and Chair, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune 411008, India. Tel: +91-20-25902237; Fax: +91-20-25902648
   E-mail: vs.gupta@ncl.res.in

2. Dr. Narendra Y. Kadoo, Senior Scientist, Plant Molecular Biology Group, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune 411008, India.
   Tel: +91-20-25902724; Fax: +91-20-25902648; E-mail: ny.kadoo@ncl.res.in

3. Dr. Ashok P. Giri, Senior Scientist, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune 411008, India. Tel: +91-20-25902710; Fax: +91-20-25902648
   E-mail: ap.giri@ncl.res.in