MOLECULAR MARKER BASED GENETIC DIVERSITY IN *SYMPLOCOS* SPP. FROM THE TWO BIODIVERSITY HOTSPOTS IN INDIA

THESIS SUBMITTED TO THE UNIVERSITY OF PUNE

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

BY

SOFIA BANU

RESEARCH GUIDE
Dr. VIDYA S. GUPTA

PLANT MOLECULAR BIOLOGY GROUP,
DIVISION OF BIOCHEMICAL SCIENCES,
NATIONAL CHEMICAL LABORATORY,
PUNE – 411008, INDIA

OCTOBER 2008

Molecular marker based genetic diversity in *Symplocos* spp. from the two biodiversity hotspots in India

A thesis submitted to the University of Pune

for the Degree of

DOCTOR OF PHILOSOPHY

in Biotechnology

SOFIA BANU
Plant Molecular Biology Group,
Division of Biochemical Sciences,
National Chemical Laboratory,
Pune – 411008, INDIA

OCTOBER 2008

CERTIFICATE

This is to certify that the work incorporated in the thesis, "**Molecular marker based genetic diversity in *Symplocos* spp. from the two biodiversity hotspots in India"** submitted by **Ms. Sofia Banu,** for the Degree of **Doctor of Philosophy,** was carried out by the candidate under my supervision in the Division of Biochemical Sciences, National Chemical Laboratory, Pune - 411 008, India. Such material as has been obtained from other sources has been duly acknowledged in the thesis.

Dr. Vidya S. Gupta
(Research Supervisor)
October 2008
Pune

DECLARATION

I hereby declare that the thesis **"Molecular marker based genetic diversity in *Symplocos* spp. from the two biodiversity hotspots in India"** submitted for the degree of Doctor of Philosophy to the University of Pune has not been submitted by me for a degree/diploma to any other University.

Sofia Banu

October 2008

Pune

Dedicated to my beloved parents

INDEX

ACKNOWLEDGEMENT

The journey in the path of research during last five years led me to associate with many people, their views, ideas, traits and criticism all that have influenced and contributed immensely in making of this thesis.

At the outset, I wish to express my sincere gratitude to my research guide, Dr. Vidya Gupta, National Chemical Laboratory (NCL), Pune, for accepting me as a student; I am previleged to have worked under her supervision. I am grateful for her constant support, encouragement, teachings, guidance and her godmotherly care during my trying times during the course of this work. Her enthusiastic attitude, ideas and scientific temperament have inspired me profoundly and have silently contributed to my understanding of the subject. Her views have given new diamensions to my opininion and my overall growth as a researcher. The concepts developed in the thesis are outcome of long sessions of pondering, framing-reframing of statements/views and brainstroming stimulating discussions with her. I wholeheartly thank her for everything.

My sincere thanks to Dr. Meena Lagu for always advicing me in planning the research work especially the collection tours. Her meticuolous ways have taught me a lot and has always inspired me to perform better. The constant help and encouragement rendered by her during my Ph.D. programme will always remain arched in my mind.

I want to express my heartfelt gratitude to Dr. Narendra Kadoo, without whom this thesis would not have had the present form. I especially thank him for his innovativeness in carrying out the sequencing reactions and for rendering all possible help in generating the data. He always inspired me to learn new techniques, especially has taught me a lot about computers and helped me to improve my technical skills. His good wishes and company will always be cherished.

I thank Dr. BG Kulkarni, Dr. PSN Rao, Dr. HS Suresh, Dr. Mao, Dr. A Baishya, Dr. PJ Handique for their generous help extended during field trips. I thank all the field personals from various forest offices who have helped me immensely during sample collection, whose names I might not recollect now but I am grateful to them.

I would like to thank the Director, National Chemical Laboratory for making available all the facilities for my research work. I also acknowledge financial support from the Council of Scientific and Industrial Research, New Delhi in the form of Research Fellowships and Department of Biotechnology, New Delhi in the form of research grant during the tenure of my work.

I am happy to thank sincerely to my collection team members Ram and Rasika, who stood behind all my decisions taken during field tours and had taken all the pains to make this project a success. I thank them for their critical views, stimulating ideas and mostly for supporting me unconditionally all these years.

My sincere thanks to Dr. Mohini Sainani, Dr. S. P Joshi and Dr. Ashok Giri for their help and advice.

I am grateful to Dr. Gauri Apte, Dr. Abhay, Dr. Savita Borse, Dr. Elangovan, Dr. Sirsha and Dr. Bhusan. They have immensely contributed to my intellectual growth during various phases of my research and have provided food for thought, encouraged me to analyze things in all possible ways and mostly inspired me to look beyond what is apparent.

I thank Anand, Bornali, Shraeddha, Lalitha, Sanjeeb Roy, Sharmili, Sujoy and Subhash for their constant support and friendship. They always raised my spirits and showed me the brighter side of life and inspired me to look forward.

I also wish to thank my seniors and more over my friends Ajay, Suhas, Ashwini, Rupali, Ramya, Neeta, Varsha, Sadhana, Suvarna, Rashmi, Trupti, Gayatri, Rohini, Charu, Gauri, Vaiju, Manasi, Radhika, Sagar, Tejas, Hemangi, Manasi M, Arun, Krishna, Shekhar, Vivek, Pavan, Prabhash, Yashwant, Rebecca, Abhijit, Medha, Pradeep, Manje, Richa, Aditi, Laxmi, Neha, Priya, Pankaj, Hemlata, Ajit, Jagtap ji, Nishant, Suman, Shweta, Poorva, Uma for their friendship and helping hands. Knowingly or unknowlingly all have contributed to my learning and growth, making this journey a memorable one.

The conception and progress of this research project has been positively gained due to valuable and critical inputs from many people and I remain forever indepted for their kind suggestions and critical views. I acknowledge all friends and well-wishers, whom I might have left out due to oversight, but whose best wishes have always encouraged me.

I would like to mention Carl Sagan, whose quote "There are many hypotheses in science which are wrong. That's perfectly all right; they're the aperture to finding out what's right" has influenced me immensely and kept me going all these years when I was struggling hard to find answers to many of the questions that boggled my mind.

Lastly, I would like to mention my parents for whom it is more than a dream and who have been eagerly looking forward to this milestone. I would like to thank my brothers for always asking me when I will finish my Ph. D work. I thank them for being there always, with their endless encouragement, support and most importantly for keeping faith in me.

Sofia

LIST OF ABBREVIATIONS

AFLP	Amplified fragment length polymorphism
AMOVA	Analysis of molecular variance
BHS	Biodiversity hotspots
bp	base pair
BP	Before Present
BSI	Botanical Survey of India
CAPS	Cleaved amplified polymorphic sequence
cpDNA	Chloroplast DNA
cpSSR	Chloroplast simple sequence repeat
CTAB	Cetyl-trimethyl-ammonium bromide
DNA	Deoxy ribonucleic acid
dNTPs	Deoxy ribonucleotide triphosphate
EDTA	Ethylene diamine tetra acetic acid
EG	Eastern Ghats
ESU	Evolutionary significant unit
g	gram
h	hour
ha	hectar
IAA	Iso-amyl alcohol
ISSR	Inter simple sequence repeat
kb	kilobase pair
kg	kilogram
km^2	square kilometer
LB	Luria Bertonii
LGM	Last glacial maxima
m	meter
M	molar
mg	milligram
min	minute
mL	milliliter
mm	millimeter

mM	millimolar
MoEF	Ministry of Environment and Forest
msl	mean sea level
MSN	Minimum spanning network
MST	Minimum spanning tree
mtDNA	Mitochondrial DNA
mya	million years ago
nDNA	nuclear DNA
NEI	Northeast India
ng	nanogram
°C	Degree celsius
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEC	Primer enzyme combination
RAPD	Randomly amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
rpm	revolution per minute
s	second
SI	South India
SSD	Sum of squared deviation
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TE	Tris-EDTA
WG	Western Ghats
µg	microgram
µL	microlitre
µM	micromolar

LIST OF TABLES

LIST OF FIGURES

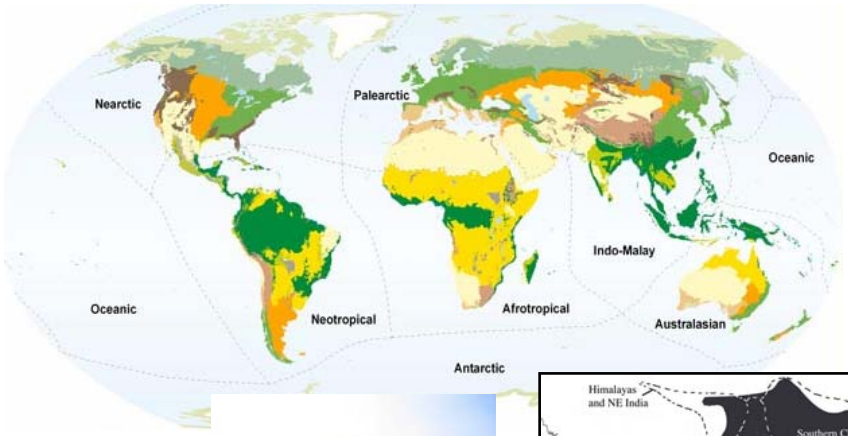
Chapter 1

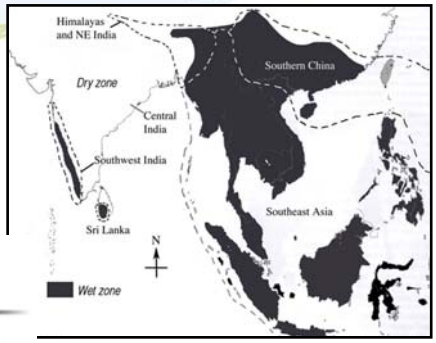
Chapter 2 - Section A

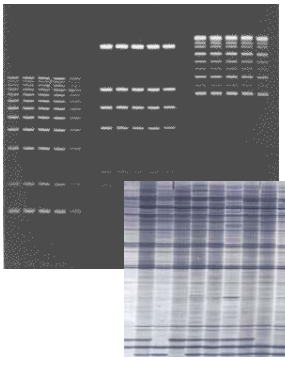
Chapter 4

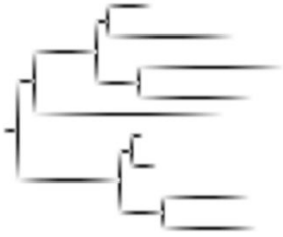
Chapter 1

Introduction and review of literature









OUTLINE

1.1 India, a biodiverse country

India is the seventh largest country in the world and Asia's second largest nation with an area of 3,287,263 km^2. India possesses a distinct identity, not only because of its geography, history and culture but also because of the great diversity of its natural ecosystems. The World Conservation Monitoring Centre, an agency of the United Nations Environment Programme, has identified 17 megadiverse countries across the globe (Fig. 1.1) and India is one of them.

Fig. 1.1 The megadiverse countries distributed across the globe
(Source:http://www.environment.gov.au/biodiversity/hotspots/images/megadiverse-countries.gif)

India lies at the junction of three major biogeographic realms, namely, the Indo-Malayan, the Eurasian and the Afro-tropical. It has a rich and varied heritage of biodiversity covering ten biogeographical zones, namely, the trans-Himalayan, the Himalayan, the Indian desert, the semi-arid zone(s), the Western Ghats, the Deccan Penninsula, the Gangetic Plain, Northeast India and the islands and coasts (Rodgers et al. 2000). The rich biodiversity is entrapped in these zones of which the forests form an important part. The forests cover an actual area of 63.73 million ha (19.39%) and consist of 37.74 million ha of dense forests, 25.51 million ha of open forests and 0.487 million ha of mangroves, apart from 5.19 million ha of scrub. The 16 major forest groups present in India account for species diversity; about 1,256 species belonging to 245 genera (MoEF 1999).

1.2 The biodiversity hotspots of India

The idea of biodiversity hotspots (BHSs) as a solution to prefer areas for conservation of biodiversity was first proposed by Myers (1988). He used species endemism and degree of threat as two basic criteria for defining BHSs. According to him, BHSs are the areas featuring exceptional loss of habitat. More precisely, to earn hotspot status, a region must harbour 1,500 or more endemic plant species, which are found in that particular area but nowhere else, and it must have lost at least 70% of its original habitat, primary vegetation. Based on criteria such as species richness, species endemism, unique higher taxa, unusual ecological or evolutionary phenomena and global rarity of major habitat types, three regions of India namely, Western Ghats (extending upto SriLanka), Himalaya and the Indo-Burma region have been included among the 34 global biodiversity hotspots spread over the globe (Conservation International) (Fig. 1.2). A broad comparison of existence of number of species in India vis-à-vis world is depicted in Table 1.1.

Table 1.1 Comparison between the number of species in India and the world

Group	Number of species in India (SI)	Number of species in the world (SW)	SI/SW (%)
Flowering plants	15,000	250,000	6.0
Mammals	350	4,629	7.6
Birds	1,224	9,702	12.6
Reptiles	408	6,550	6.2
Amphibians	197	4,522	4.4
Fishes	2,546	21,730	11.7

It has been reported that 325 families represented by 15,000 flowering plants are found in India, out of which 15% are believed to be under threat. Apart from the flowering plants, 64 gymnosperms, 1,200 pteridophytes, 2,850 bryophytes, 13,000 fungi and 12,500 algae are also observed in India (Hajra and Mudgal 1997). Together the three hotspots have 5,150 endemic plant species, or 20% of the world's recorded plant species, in only 746,400 km^2 or 0.5% of the earth's land surface.

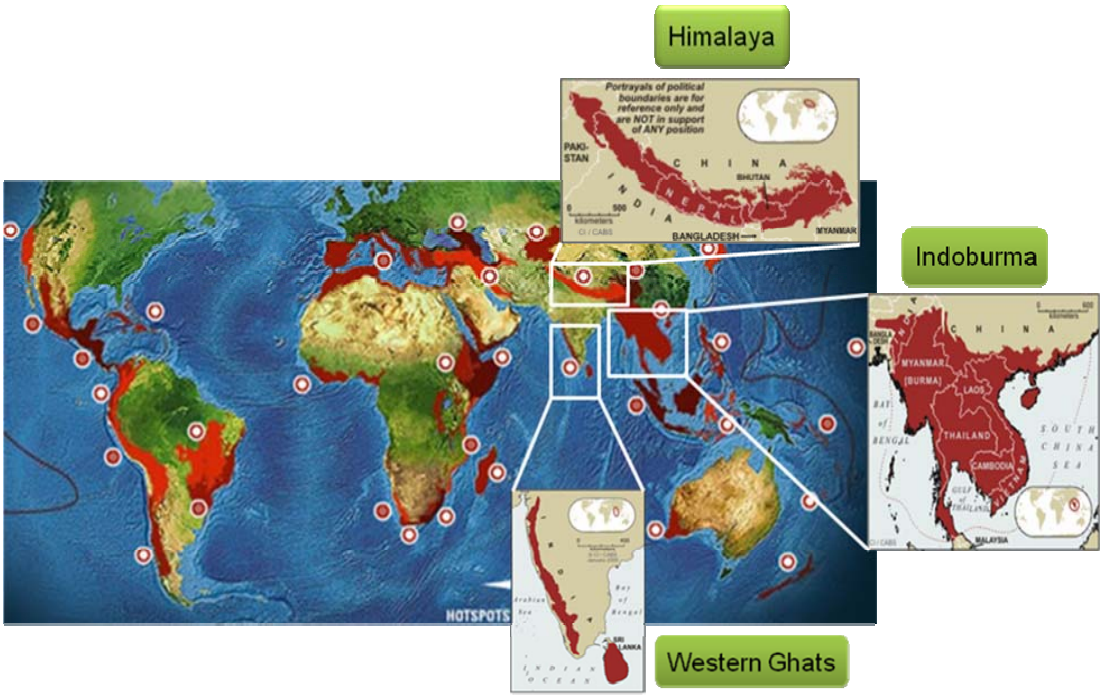


Fig. 1.2 The three biodiversity hotspots of India, in the background of 34 identified biodiversity hotspots in the globe

(Source: http://www.biodiversityhotspots.org/)

1.3 Western Ghats

The Western Ghats (WG), also known as the Sahyadri Hills, are well known for their rich and unique assemblage of flora and fauna and is among the eight hottest biodiversity hotspots. The hills north of the Krishna basin (largely Maharashtra and Gujarat) with fragile basaltic rocks are results of the same processes that gave rise to the Deccan trap. Arising abruptly from the narrow Konkan and Malabar coasts, these hills run 1,600 km north-south parallel to the west coast of India from 8°20'N to 21°40'N and 73°E to 77°E between the river Tapti in Gujarat and Kanyakumari in Tamil Nadu, covering an area approximately 160,000 km^2. In the east, they slope gently towards the Deccan Plateau. The northernmost segment that extends into Gujarat merges in the east with the Surat and Dangs. In the Nilgiris, Palanis and parts of Karnataka, the WG extend considerably eastwards, locally merging with the Eastern Ghats (EG). Towards the south, the hill chain is divided into two by the Palghat Gap (a mere 30 km gap) rendering a physically homogeneous high altitude plateau into two rather distinct biogeographic units *viz.*, the Nilgiris complex in the north and the Anamalai-Palnis complex in the south (Fig. 1.3).

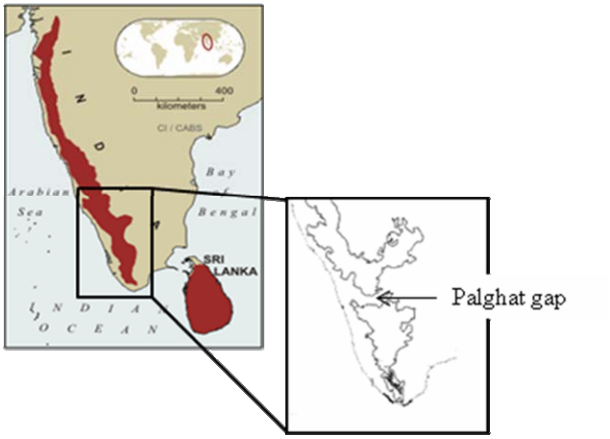


Fig. 1.3 The Western Ghats

(Source: http://www.biodiversityhotspots.org/)

Climatic conditions in the WG vary with the altitude and physical proximity to the Arabian Sea and the equator. Although the WG experience a tropical climate - being warm and humid during most of the year with mean temperature ranging from 20°C in the south to 24°C in the north, the higher elevations experience subtropical climates and on occasions frost. Further, it has been observed that the coldest periods in the southern WG coincide with the wettest.

WG is also known as the Malabar rain forest province and is one of the major tropical evergreen forest regions in India which exhibit enormous plant and animal diversity. As early as 1904, Hooker had drawn attention to the distinct flora of the WG, which he called the 'Malabar' floristic region. The major vegetation types include tropical evergreen forests, moist deciduous forests, dry deciduous forests, scrub jungles, sholas, savannas including high rainfall savannas, peat bogs and *Myristica* swamps. Amongst the lower plants, around 320 species of pteridophytes, 200 species of bryophytes, 300 species of algae and 800 species of lichens are known along with 600 species of fungi. The WG are estimated to harbor approximately 5,500 species of flowering plants (Nair and Daniel 1986). Fifty six genera of flowering plants are considered endemic to the WG with 2,180 endemic species, constituting 0.7% of global plant species (Myers et al. 2000). Sixty three percent of India's evergreen woody plants are endemic to the WG.

1.3.1 Origin and prehistory of WG

The peninsular India was a part of the Gondwana land till about 150 million years ago (mya), from which it split and started moving north. The northward drift which lasted about 100 million years (my) finally ended with the peninsula colliding with the Asian mainland 45 mya. Major geological transformations took place as the peninsula moved northwards. Soon after detachment from the Gondwana land, the Indian peninsula drifted over the Reunion Hotspots-localised volcanic centres in the earth's lithosphere, 200-300 km across, which have remained active for several my. This event happened some 120-130 mya and resulted in the uplift of the WG. Subsequently, there were a series of volcanic eruptions until around 65 mya giving rise to the extensive Deccan Traps. These volcanic episodes largely moulded the northern third of the WG. Since the WG are the result of domal uplift, the underlying rocks are ancient-around 2,000 million years old. The oldest of these rocks are found in the Nilgiris and the high ranges of the Western Ghats. The peninsular India broke along its line of weakness, and the western segment drifted westward into the sea (a process known as faulting), giving rise to the present day hill chain, the WG and the west coast. This happened during the Eocene (between 45 and 65 mya), even before India became part of the Asian mainland. It was only during the Pliocene (5-1.6 my BP) that WG came to be known more or less as they are today. Soon after this, during the Pleistocene (1.5 my BP) the peninsula experienced a marked eastward tilt permanently changing the pattern of drainage (Chapman and Reiss 1992). The western faulting led to 'river capture' and diversion of the easterly drainage to the west in many instances. The WG thus represent a tectonically active region with high rates of uplift, high summit altitudes, steep slopes, deep gorges and large potential energy for erosion and correspondingly high sediment yields.

Very little fossil evidence exists to reliably reconstruct the prehistoric biodiversity of the WG. The flora of the WG shares elements with Africa, Madagascar and South America (*e.g.* Family Bignonaceae; *Vinca rosea*, Family Balsaminaceae: *Impatiens* species etc.). Many species of invertebrates including a few species of butterflies are also shared with South America and Africa. Amongst freshwater fishes, few genera (*Notopterus, Barilius, Rasbora, Puntius, Labeo, Clarias*, *Aplocheilus, Mastacembelus, Garra, Aphanius*) that are common to India and Africa and are represented by one or more species in the WG.

1.3.2 Shola forests of WG and their flora and fauna

'Sholas' are the best examples of tropical evergreen forests or tropical rain forests (Malabar rain forest) and are some of the non-renewable natural resources (Hajra and Mudgal 1997). They are found mainly in the sheltered ravines, troughs, hollows and other depressions, where there is abundance of moisture and are occasionally seen 'flowing' to the valleys along with the streams. The montane shola-grassland complexes occur between 1,900 and 2,220 m. These consist of stunted montane forests surrounded by undulating grasslands. The upper story of the forests is characterized by *Pygeum gardneri, Schefflera racemosa, Linnociera ramiflora, Syzigium* spp., *Rhododenron nilgiricum, Mahonia nepalensis, Eleocarpus recurvatus, Ilex denticulata, Michaelia nilagirica, Actinodaphne bourdellonii* and *Litsea wightiana* (Karunakaran et al. 1998). A low, twisted second story of *Ilex wightiana, Rapanaea wightiana, Ternstroemia gymnanthera, Symplocos* spp. and *Microtropis* spp. and a dense shrub layer of saplings of *Strobilanthes, Psychotria* and *Lasianthus* spp. usually are present. The grasslands that surround the shola forests consist of several fire and frost-resistant grasses: *Chrysopogon zeylanicus, Cymbopogon flexuosus, Arundinella ciliata, Arundinella mesophylla, Arundinella tuberculata, Themeda tremula,* and *Sehima nervosum* (Karunakaran et al. 1998).

Frost, fire and wind are the three main factors determining the distribution of the sholas. When in the open grasslands, frosts occur during winter night and temperature goes subzero, it is invariably above 0°C inside the sholas (Legris and Blasco 1969). The nocturnal temperature as low as -16°C has been reported from the grasslands of Nilgiris (Ranganathan 1938). The plants are then subjected to intercellular crystallisation. In contrast to the climate that prevails in the temperate latitudes, a frosty night is followed by a hot day and the bright morning sun promotes rapid thawing of the crystallised water, which is quickly transpired by the plant. At the same time, the top soil is in a frozen state, so that the plants experience difficulty to absorb water to compensate the transpiratory losses. This results in water stress, analogous to physiological dryness, which in turn causes wilting and death of the plant. Seedlings are the most affected, since they have only shallow root system, but the trees once established manage to survive, with less damage. However, various levels of frost hardening are exhibited by the shola plants and so most of them, especially those found in the periphery have slightly re-curved, coriaceous leaves. At

the same time, grasses that have perennial rootstock and leaves with bulliform cells, can withstand these extremes of frost, by rolling their leaves to reduce transpiration rate and spreading quickly by vegetative means, even if badly damaged by frost.

Fires are the second factor; usually having an anthropogenic origin, have a vital role in shaping the spread of sholas to a certain extent. Fire once lit finds its way up, swallowing the grasslands, very often entering the sholas through the reed zones. In this process, it wipes out practically the last-seedling which had escaped the dangerous effect of frost and got established in the open grasslands, thereby preventing the expansion of the sholas. Shola trees always remain shallow rooted even in regions of considerable soil depth. Therefore, there are ample chances of heavy wind fall. This limits the expansion of the sholas into the grasslands where they are more prone to the effects of wind and hence found restricted only in regions protected from wind, such as hill-folds and depressions.

Palynological studies have revealed that these shola forest communities had been wide-spread in the past. They originated through gradual invasion of shrubs and undertrees into grassland, about 35,000 years BP, corresponding in time to the last glaciation in the north and were established about 24,000 years BP. The progressive recession of sholas had started around 7,000 years BP. The biotic and abiotic climatic factors (increase in aridity in central India) have reduced the shola to the present state and have created the condition under which these shola community plants have completely stopped regeneration outside the sholas in SI (Gupta 1990). Apart from these, owing to the growth in anthropogenic activities in these regions, the shola communities are receding as reported by Vishnu-Mittre and Gupta (1968). Their observations are based on pollen investigation in the shola community at Ooty in Nilgiris, and they further referred sholas as “living fossils”.

1.3.3 Human ecology and its impact on the biodiversity of the WG

The history of species extinctions in the WG was coincident with the climatic and human histories. Extended arid periods and human interference starting 12,000 years before present, led to slow but extensive transformation of habitats in and around the WG. The use of fire to clear forests for cultivation has had a major influence on the forests of the WG. Hill agroecosystems in the WG are today dominated by estates-chiefly of tea, coffee, rubber and monocultures of various tree species, including the

oil palm. It has also been highlighted that the Nilgiri district with a total area of 2,549 km^2 has around 1,000 km^2 under various forms of cultivation. In WG, smaller ranges and deforestation contribute to isolation of the shola forests. A loss of 25.6% of the total forest area has been noted in the WG region during the last two to three decades, whereas the dense forest cover and open forests have decreased by 19.5 and 33.2% of their total occupation, respectively, due to deforestation and land use (Jha et al. 2000).

1.4 Indo-Burma biodiversity hotspot

The Indo-Burma hotspot encompasses 2,373,000 km^2 of tropical Asia east of the Ganges-Brahmaputra lowlands. Formerly including the Himalaya chain and the associated foothills in Nepal, Bhutan and India, the Indo-Burma hotspot has now been more narrowly redefined as the Indo-Chinese subregion (http://www.biodiversityhotspots.org/). It begins in eastern Bangladesh and then extends across north-eastern India, south of the Brahmaputra river, to encompass nearly all of Myanmar, part of southern and western Yunnan Province in China, all of the Lao People's Democratic Republic, Cambodia and Vietnam, the vast majority of Thailand and a small part of Peninsular Malaysia. In addition, the hotspot covers the coastal lowlands of southern China (in southern Guangxi and Guangdong), as well as several offshore islands, such as Hainan Island (of China) in the South China Sea and the Andaman Islands (of India) in the Andaman Sea (Fig. 1.4).

The hotspot encompasses 33 terrestrial ecoregions, which include tropical and subtropical moist broadleaf forests, tropical and subtropical dry broadleaf forests, tropical and subtropical coniferous forests, temperate broadleaf and mixed forests, montane forests and mangroves. There are also patches of shrublands and woodlands on karst limestone outcrops and, in some coastal areas, scattered heath forests. In addition, a wide variety of distinctive localized vegetation formations occurs in Indo-Burma, including lowland floodplain swamps, mangroves, and seasonally inundated grasslands. Throughout the hotspot, montane forests extend on humus-rich soils from about 800 m. These forests are lower in stature with fewer emergents; oaks (Fagaceae) dominate, while laurels (Lauraceae) and magnolias (Magnoliaceae) become notable constituents. Montane tree species composition is generally less diverse than that of lowland forests, but it contains proportionately more endemic species. Diverse edaphic, topographic and microclimatic conditions at higher

elevations give rise to a range of mixed coniferous and broadleaf evergreen forest formations. On dry hills and plateaus subjected to fire, conifer-dominated savanna forests occur (typically dominated by *Pinus merkusii* or *P. kesiya*). At the highest elevations, on ridgelines and ridge crests, stunted, xerophytic formations are characterized by the presence of *Rhododendron* spp.

Fig. 1.4 The Indo-Burma biodiversity hotspot region

(Source: http://www.biodiversityhotspots.org/)

An estimate of total plant diversity in this hotspot reveals presence of about 13,500 vascular plant species, of which about 7,000 (52%) are endemic (Dijk et al. 1999). Among the flora of the Indo-Burma Hotspot are a wide array of orchid and ginger species (there are more than 1,000 orchid species in Thailand alone) and many tropical hardwood trees, including commercially valuable dipterocarp species and teak (*Tectona grandis*).

The Northeastern region of India (NEI comprising the states of Assam, Manipur, Meghalaya, Mizoram, Nagaland and Tripura) is a part of the Indo-Burma biodiversity hotspot and falls in a unique transitional zone between the Indian, Indo-Malayan and Indo-Chinese biogeographical zones as well as the confluence of the Himalayan region (Rao 1993). The high rainfall, moist and cold climate coupled with factors like variable altitude, latitude and longitude have added to the multiplicity of habitats and thus provide varied microclimates and ecological niches both for plants and animals. The region is recognized as a centre of origin of several cultivated plants.

1.4.1 Origin and prehistory of Indo-Burma region

Indo-Burma has a complex geological and evolutionary history. They were created by the same tectonic processes that resulted in the formation of Himalaya. The Indian intrusion into the Asian continental landmass has been responsible for the formation of most of the hotspot's topography, including the general north-south orientation of the mountains and main rivers. The wide variation in land form, climate, and latitude within the hotspot has led to the development of diverse natural habitats that support a high diversity of plant and animal species. This diversity is enhanced by a significant endemic element, which may largely derive from habitat isolation caused by periods of high sea level and vegetation changes during the glacial episodes of the Pleistocene. Consequently, the hotspot contains many localized centers of endemism, particularly montane isolates, but also areas of lowland wet evergreen forests that were isolated at some stage and river basins. The ecoregion represents the subtropical forests of the Khasi and Garo hills that rise to more than 1,800 m in the eastern Indian states of Meghalaya and Assam.

1.4.2 Effect of anthropogenic factors on the Indo-Burma region

Indo-Burma is one of the most threatened of the Earth's 34 biodiversity hotspots. Only about 5% of its natural habitats remain in relatively pristine condition. Indo-Burma may have been one of the first places on the globe where agriculture developed (Solheim 1972; Diamond 1997), creating a long history of forest burning and clearance for shifting or permanent small-scale cultivation. In recent centuries, the steadily increasing trade in agricultural commodities and timber, combined with population growth, has led to widespread forest destruction. Tree plantations (teak, rubber, oil palm) have replaced large areas of lowland forest, while coffee, tea, vegetable crops and sugarcane plantations threaten montane and hill forests. Other threats to forests include logging, mining for gems and ore, firewood collection and charcoal production. India's ever increasing population has put tremendous pressure on these ecosystems. In order to increase food production, there is an increase in agricultural areas leading to reduction in the forests and grasslands (Upadhyay 1999). In the NEI region a 2,778 km^2 area was used for shifting cultivation from 1993 to 1997 and has caused extensive deforestation. In Meghalaya alone, 295 km^2 area has been used for similar purpose (Upadhyay 1999). Protected area systems are the

cornerstones of government conservation programs in the Indo-Burma hotspot. In total, 236,000 km^2 is officially protected, representing roughly 10% of the original extent of vegetation in the hotspot. However, only 132,000 km^2 (a little under six percent) is in IUCN protected area categories I to IV, the coverage of protected areas falls to around 6% of the hotspot.

1.5 Himalaya

Stretching in an arc over 3,000 km of northern Pakistan, Nepal, Bhutan and the Northwestern and Northeastern states of India, the Himalaya hotspot includes all of the world's mountain peaks higher than 8,000 m. This includes the world's highest mountain, Sagarmatha (Mt. Everest) as well as several of the world's deepest river gorges. This immense mountain range, which covers nearly 750,000 km^2, has been divided into two regions: the Eastern Himalaya, which covers parts of Nepal, Bhutan, the Northeast Indian states of West Bengal, Sikkim, Assam and Arunachal Pradesh, southeast Tibet (China), and Northern Myanmar; and the Western Himalaya, covering the Kumaon-Garhwal, Northwest Kashmir and Northern Pakistan (Fig. 1.5).

The Himalayan mountain range comprises three east-west-directed parallel zones. The southernmost outer Himalayas, or Siwaliks, lie alongside the Indo-Gangetic Plain and is composed of alluvial deposits that have washed down from the north. It is more recent in origin than the other ranges. The next is the Middle Himalayas, a highly folded system of ridges and valleys that rise to about 5,000 m. The third is the Inner Himalayas, which contain the tallest mountains in the world: Everest, Makalu, Dhaulagiri, and Jomalhari. Biogeographically, the Himalayan mountain range straddles a transition zone between the Palearctic and Indo-Malayan realms, with species from both contributing to its biodiversity. Understanding the distribution of biodiversity requires some knowledge of the genesis of the Himalaya and the complex geological and physical features that influence patterns of biodiversity (Molnar 1986). Paleobotanical evidence indicates that much of the forest vegetation in the Himalayan region is derived from the tropical evergreen forests of the Indian Peninsula, which was a part of Gondwanaland, while the alpine flora evolved after the Himalayan uplift (Ram and Singh 1994). Since this period, Pleistocene glaciation has influenced the flora and fauna, especially the southward migration of the species from the Central Asian, Alps, Mediterranean and Sino-

Japanese regions (Gupta 1994). At the same time, physical and climatic barriers along the altitudinal transect, temperature extremes, and high intensity of ultraviolet radiation have resulted in high endemism, and evolution of mutants, polyploids, and ecotypes among several taxa.

Two distinct ecological formations of broadleaf forests can be distinguished in this ecoregion depending on the geology and slope (moisture regime): the temperate evergreen forests of oaks (*Quercus* spp.), especially *Quercus lamellosa* in association with *Lithocarpus pachyphylla, Rhododendron arboreum, Rhododendron falconeri, Rhododendron thomsonii, Michelia excelsa, Michelia cathcartii, Bucklandia populnea, Symplocos cochinchinensis,* and other species of *Magnolia, Cinnamomum,* and *Machilus;* and temperate deciduous forests dominated by *Acer campbellii, Juglans regia, Alnus nepalensis, Betula alnoides, Betula utilis,* and *Echinocarpus dasycarpus* (Puri et al. 1989; Rao 1994; Shrestha and Joshi 1997). Of the estimated 10,000 species of plants in the Himalayan hotspot, about 3,160 are endemic, as are 71 genera. Furthermore, five plant families are endemic to the region, the Tetracentraceae, Hamamelidaceae, Circaesteraceae, Butomaceae and Stachyuraceae. The largest family of flowering plants in the hotspot is the Orchidaceae, with 750 species. The Eastern Himalaya is also a center of diversity for several widely distributed plant taxa, such as *Rhododendron*, *Primula* and *Pedicularis*. Although many of the other non-vascular taxonomic groups have yet to be adequately documented, nearly 13,000 species of fungi and around 1,100 species of lichens have been described.

The geological uplift and evolution of the Himalaya began when the Indian continent collided with mainland Asia at the beginning of the Cenozoic era (Valdiya 2002). The complex orogeny of the Himalaya, coupled with the ensuing climatic and edaphic changes facilitated its colonization by floral and faunal immigrants from the neighboring regions (Pandit et al. 2000). Exposed to novel selection pressures in these evolving ecosystems, the newly-arrived biotas underwent rapid evolutionary divergence and speciation (Kumar 1983). As a result, contemporary Himalaya is a repository of extremely rich and endemic biodiversity (Nayar 1996; Pandit et al. 2000). However, over the past few decades, the Himalaya has experienced unprecedented land-use changes driven by rapid human population growth and

intensifying human activities, such as agriculture and expanding human settlements (Nayar and Sastry 1987, 1988, 1990; Tikader 1983).

Fig. 1.5 The Himalayan biodiversity hotspot

(Source: http://multimedia.conservation.org/cabs/online_pubs/hotspots2/Himalaya.html)

1.6 Disjunct distribution of flora and fauna in India

The Indian subcontinent can be divided into two broad climatic zones. The first climatic zone also referred as the wet-zone, constitutes areas in the Northeast India (NEI) and South India (SI) as well as southwest SriLanka. These areas receive over 250 cm rainfall per year. The second climatic zone referred to as dry zone receives between 50 and 100 cm rainfall per year. The rest of the Indian subcontinent belongs to this climatic zone. An interesting outcome of this dichotomy in climate is the marked discontinuity in the distributions of many wet-zone species in the Indian subcontinent (Karanth 2003). For example, many plant species of temperate shrub savanna type such as *Gaultheria fragrantissima, Rhododendron arboretum, Mahonia leshenaultii* and others like *Ternstroemia japonica*, *Hypericum hookerianum*, *Thalictrum javanicum*, *Cotoneaster buxifolia*, *Parnassia wightiana*, *Lonicera ligustrina* are found in the wet evergreen montane (shola) forests of SI and SriLanka and are also found about 1500 km away in the wet evergreen forests of NEI and throughout Southeast Asia, although absent from the intervening dry zone (central India) (Khoshoo 1996). Such discontinuity in the distribution also called as disjunct distribution is particularly pronounced in the case of Indo-Chinese elements of the peninsular Indian biota (Fig. 1.6).

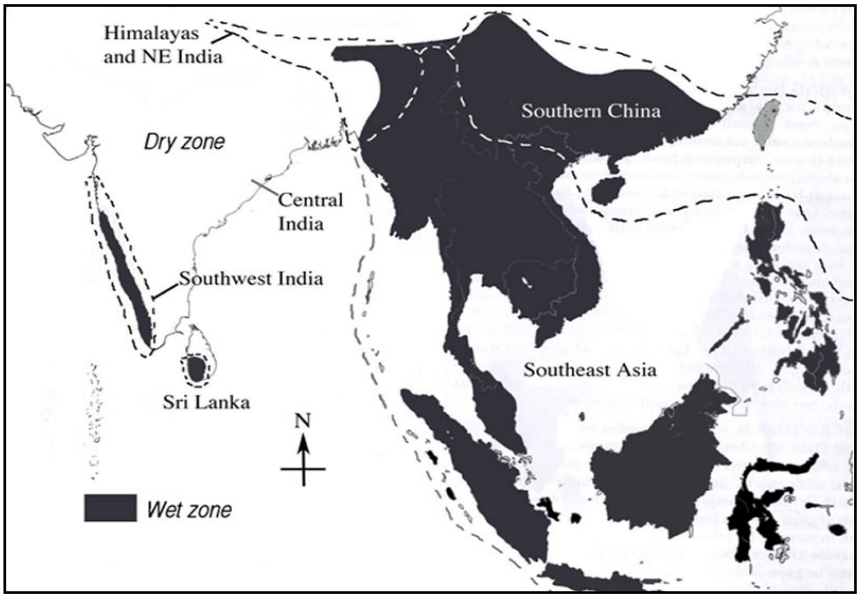


Fig. 1.6 The distribution spread of wet-zone species showing disjunct distribution (Source: Karanth 2003)

Moreover, the plant species showing the disjunct distribution were found to be part of montane forest community also known as "Shola". These 'Sholas' are the best examples of tropical evergreen forests or tropical rain forests (Malabar rain forest). The plants common to NEI and SI are either from the temperate stock or of a tropical stock and are confined to shrub savanna or the margins of the forests (Meher-Homji 1967). This type of disjunct distribution has also been observed in case of animals. For example, among birds of the genus *Dicaeum* (flowerpeckers) the species *Dicaeum concolor* is distributed in two widely separated areas *viz*., southwest India and NEI. The intermediate dry zone is inhabited by another species *Dicaeum erythrorhynchos*. In fact there are several cases of disjunct distributions among birds (Ripley and Beehler 1990), mammals (Morales and Melnick 1998), amphibians and reptiles (Daniels 2001; Das 2002), freshwater fishes (Dahanukar et al. 2004) and plants (Meher-Homji 1967, 1972, 1975).

1.6.1 Exploring the links between the Western Ghats and Indo-Burma biodiversity hotspots

The present day disjunct distribution of plants and animals in the Indian subcontinent was first explained by Medlicott and Blanford (1879). Of the many schools of thought, the most influential "Satpura hypothesis", put forward by Hora (1949), postulates that wet-zone species colonized SI by way of once continuous corridor of

tropical evergreen forests from Eastern Himalayas across Vindhya-Satpura ranges to the WG (*via* Rajmahal hills-Chhota Nagpur plateau-Vindhya-Satpura) and also suggests alternative corridors for dispersal such as the Eastern Ghats (*via* Chhota Nagpur hills through Orissa and Jeypore hills) (Abdulali 1949), Brij country (Dilger 1952) and the Aravalli range (Mani 1974). The migration routes suggested for the dispersal of organisms in Indian subcontinent is depicted in Fig. 1.7 (Karanth 2003). However, geological studies done by Auden (1949) do not support the Satpura hypothesis.

Fig. 1.7 Different dispersal routes that wet-zone species from NEI could have taken to reach SI and SriLanka according to dispersal hypothesis put forth by various authors

The routes include the Aravalli range (1), Brij region (2), Vindhya-Satpura ranges (3) and Eastern Ghats (4). Arrows indicate the direction of dispersal (Karanth 2003).

Alternative hypothesis for migration from South to NEI has also been suggested. According to Hooker and Thompson (1855) and Razi (1954), the hills of Bihar and Orissa served as a route of migration from Southeast to Northeast or vice versa. Randhwa (1945) and Dilger (1952), on the basis of various evidences, suggested that a much wetter climate than the present existed in the central India,

which they termed “Brij Country” that was covered with much evergreen tropical forests about 2,000 years ago and that had affinities with the wet tropical forests of the Eastern Himalayas, Assam and Burma as well as with the WG of SI. However, according to the second school of thought the present disjunct distributions for some species probably represent relict populations of once widely distributed species. According to Medlicott and Blanford (1879), Burkill (1924), Meher-Homji (1975) and lately by Daniels (2001), the Pleistocene glaciation, which pushed the Himalayan species southwards to SriLanka, and subsequent post-Pleistocene changes towards warmer climate might have resulted in expansion of the montane flora and fauna on the hilltops of WG and EG where the conditions were favourable for them to persist.

1.6.2. Models suggested to understand the evolution of disjunct distributions among wet-zone species of India

Though the schools of thought described above have existed from a long, Ripley and Beehler (1990) were the first to define models to illustrate the disjunct distribution in India. Of the four models suggested by them, the dispersal model and the vicariance model were found to be more effective in Indian biogeographical context as pointed out by Karanth (2003). If the wet-zone species arrived on the Indian subcontinent after establishment of the dry zone forests as a stepping-stone to reach the WG (SI), then this would suggest a dispersal event. On the other hand, if the wet-zone forests arrived on the Indian subcontinent before establishment of the present day arid zone and were isolated from their counterparts of NEI and SI, then this would suggest a vicariance event.

The current discontinuity of some species (mainly faunal population) may represent relict of former continuous distribution (Ali 1935; Mani 1974; Ripley et al. 1986; Das 1996; Daniel 2002). This relict fauna scenario supposes that much of India and Southeast Asia was covered with extensive and continuous humid forest that has recently been broken into isolated patches mainly due to climatic changes (Fig. 1.8). These isolated patches are home for many relict populations which were once distributed in humid forests. The existence of rainforest taxa in Northeast India has been suggested by Ashton and Gunatilleke (1987) based on fossil flora of mid-miocene (18 to 11 mya). This area is now covered with semi arid to desert habitat.

Additionally, the fossil evidence, point towards existence of humid climate in Indian peninsula also during most of the Miocene period (Meher-Homji 1977).

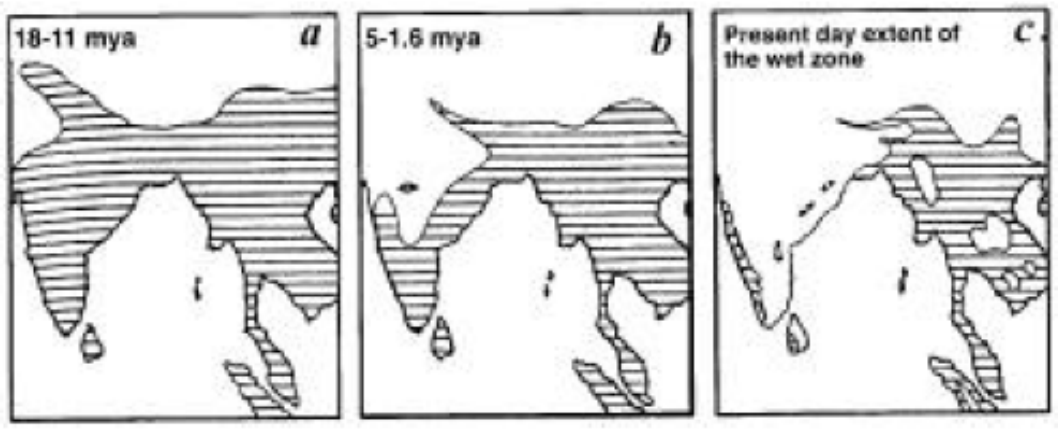


Fig. 1.8 The regression of the wet-zone in the Indian subcontinent

a: Fossil evidence suggests that during mid-Miocene times, much of India was covered with humid forest that was continuous with the forests of South East Asia Much of India received high rainfall that could support this humid forest ***b:*** Onset of drier climate resulted in regression of the wet-zone ***c***: further regression of the wet-zone. (Source: Karanth 2003)

The aridity is linked to the geological as well as the climatic history of the region. The northward movement of the Indian plate, which resulted in the formation of Himalayas and the WG triggered the onset of drier climate and the replacement of the tropical evergreen forests by deciduous ones over the majority of the Indian peninsula (Meher-Homji 1973). The climatic history of these regions has been held as one of the most influensive factors determining the distribution of the organisms in these areas. The subcontinent has experienced multiple cycles of dry and wet periods. The $\delta^{13}C$ records from the late Quaternary period in southern India suggest an arid period 6,000 to 3,500 years ago and a short wet phase about 600 years ago (Sukumar 1993). Based on the archaeological and ancient literature, Randhwa (1945) suggested that certain parts of northern India were wetter than today, till 2,000 years ago. Studies based on pollen from Donger-sambar swamps in Madhya Pradesh by Chauhan (2002) also indicated that a moist and warm climate condition prevailed in this region around 1,800 years ago. All these evidences indicate that the large tracts of humid forest cover progressively declined and retreated to wetter parts of India (wet-zone) during late Miocene and Pliocene times. This was probably followed by expansion and shrinking of forest cover due to multiple cycles of wet and dry periods

during the Quaternary. As the wet-zone species of peninsular India were separated or cut off from those in the NEI by the dry zone, the establishment of the dry zone could form a vicariant event (Fig 1.8). According to Daniels (1997), the alternating arid-cold and humid-warm climates that the entire tropics experienced during the Pleistocene period is believed to have favoured periodic invasions of both forests and montane birds into southern and western India. One of the arguments is that the most species that fit the dispersal model might have arrived in peninsular India recently (during Pleistocene times), whereas the species that fit the vicariance model might have arrived in the peninsular India from late Miocene to Holocene times (Das 2002). This has been supported by studies by Daniels (1997) on birds, he observed that the montane and grassland birds were isolated in the WG at a much earlier period than the birds of the lower elevation rainforests and more open habitats.

Other hypotheses, such as long distance dispersal by wind or birds (Blasco 1970, 1971), have been proposed. However, Meher-Homji (1972) has revealed problems associated with long distance dispersal hypothesis based on the presence of different species of *Rhododendron* and *Mahonia* on WG and NEI regions ***e.g.*** *Rhododendron arboreum* in the Himalaya and *R. nilagiricum* in the Ceylon highlands; *Berberis*- syn. *Mahonia nepalensis* in the Himalayas and *M. leschnaultii* in SI. He pointed out that not more than one species per genus was observed to be common to the Himalayan and Southeastern Ghats (WG and EG) suggesting that migration between these regions has taken place over a long period and not by wind or bird dispersal.

Many authors have invoked variations of either the dispersal or the vicariance model to explain disjunct distribution. While, Karanth (2003) and Ripley and Beehler (1990) pointed that Satpura or Brij hypothesis is nothing more than a special case of dispersal model, Meher-Homji (1975) assumed these plants to be relict flora suggesting vicariance model. However, it has been noted that incorrect taxonomy, due to convergence in morphological characters, can erroneously suggest disjunct distribution. The first step in studying disjunct distribution is to determine if the observed pattern is real (true disjunct) and not due to convergence *i.e.,* an artifact of incorrect taxonomy (false disjunct).

Recently, Karanth (2003) has posed an important question as how to distinguish between the dispersal model and the vicariance model. The relative ages

of populations and the barriers between them need to be resolved first in order to distinguish between the models. However, based on the current faunal and floral distribution and geological data, it is difficult to determine the age of the population versus their barriers. Therefore, to test these models, it is imperative to know the phylogeny, distribution and population structure of the species, which are common to NEI and SI.

1.7 Biogeography

Biogeography deals with geographic patterns of species distribution and the processes that result in such patterns. It is defined as the "synthetic science, related to geography, biology, soil science, geology, climatology, ecology and evolution". It also tries to explain in part the causes of biodiversity and is also concerned with the factors responsible for the variations in distributions. It, therefore, answers the questions as to where do species occur, how the species got to be where they are, and where are the greatest concentrations (Jønsson and Fjeldså 2006). Biotic factors such as interactions among competing species, coevolutionary influences, and the reproductive and nutritional requirements of populations and species are also studied. The studies done by Alfred Russel Wallace on the distribution of flora and fauna in the Malay Archipelago in the 19th century was the first recorded document on biogeography.

Based on relatively complete compilations of species within well-studied groups, such as birds and mammals, biogeographers identified eight different ecozone or biogeographic realms within which species tend to be closely related and between which turnovers in major groups of species are observed. Eight terrestrial biogeographic realms are typically recognized, corresponding roughly to continents (Table 1.2). Although similar ecosystems (such as tropical moist forests) share similar processes and major vegetation types wherever they are found, their species composition varies markedly depending on the biogeographic realm in which they are found. Assessing biodiversity at the level of biogeographic realms is important because the realms display substantial variation in the extent of change, they face different drivers of change, and there may be differences in the options for mitigating or managing the drivers.

Table 1.2 The eight terrestrial biogeographic realms recognized all over the world

Realm	Continental area (million km^2)	Continents included
Nearctic	22.9	Most of North America
Palearctic	54.1	The bulk of Eurasia and North Africa
Afrotropic	22.1	Sub-Saharan Africa
Indomalaya	7.5	Afghanistan and Pakistan, the South Asian subcontinent and Southeast Asia
Australasia	7.7	Australia, New Guinea, and neighbouring islands. The northern boundary of this zone is known as the Wallace line
Neotropic	19.0	South America and the Caribbean
Oceania	1.0	Polynesia, Fiji and Micronesia
Antarctic	0.3	Antarctica

Explaining the geographical deployment of biodiversity through time in these realms forms the science of historical biogeography. During the past 30 years, this science has endured several paradigm shifts. For example, prior to 1970, most workers assumed that the species' ancestors of a species, when dispersed across preexisting barriers, became isolated and evolved into new species (Udvardy 1969). This, the vicariance school (Nelson and Platnick 1981) proposed, was the major way in which biodiversity was generated through the fragmentation of wide spread ancestors by vicariant (isolating) events, such as uplifting of mountain ranges. Hence, dispersal occurred prior to the presence of the isolating barrier, not afterwards. Critics of the vicariance school argued that dispersal must play a more prominent role in the evolution of diversity, especially for highly vagile organisms such as birds, whereas vicariance advocates (Rosen 1978) that dispersal should be invoked only after falsification of vicariance model. Lately, methods of reconstructing biogeographical history have been proposed in which both dispersal and vicariance are allowed (Ronquist 1997). Hence, there is a growing plurality in the theoretical and methodological tools of biogeography. Nonetheless, few empirical studies document the relative roles of vicariance and dispersal (Lynch 1989). In the vicariance paradigm, rare but extensive dispersal (range expansion) is followed by a series of allopatric isolation events (Bush 1975; Endler 1977), interrupted by occasional random dispersals. If the isolation events affect many organisms simultaneously, this

process will generate congruent tree topologies. Dispersalists consider range expansion to be a more common and regularly occurring phenomenon. Both these views laid the foundation for the expansion of classical biogeography, the development of molecular systematics, resulting in creation of a new discipline known as phylogeography. This development allowed scientists to test theories about the origin and dispersal of populations.

1.8 Phylogeography

Phylogeography is the study of the processes controlling the geographic distributions of lineages by constructing the genealogies of populations and genes (Avise 2000). The term phylogeography was coined in 1987 (Avise 2000) and was introduced to describe geographically structured genetic signals within and among species. Phylogeography till recently was a domain of historical biogeography and addressed question such as how historical, geological, climatic and ecological conditions influenced the current distribution of species. As part of historical biogeography, researchers have been evaluating the geographical and evolutionary relationships of organisms years before. Paleogeography, geology and paleoecology are all important fields that supply information that is integrated into phylogeographic analyses (Fig 1.9). Two developments during the 1960s and 1970s were particularly important in laying the platform for modern phylogeography; the first was the spread of cladistic thought and the second was the development of plate tectonics theory (De Queiroz 2005).

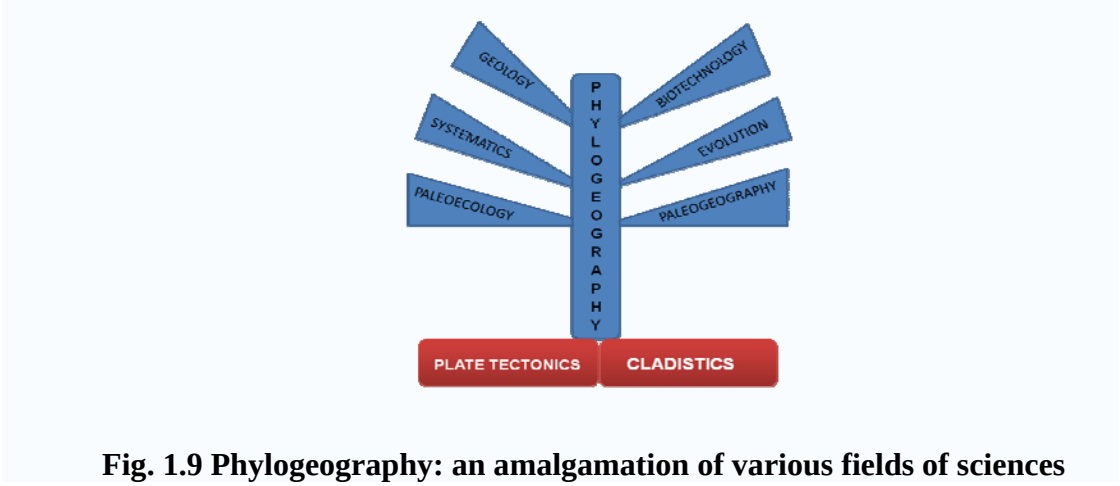


Fig. 1.9 Phylogeography: an amalgamation of various fields of sciences

Gene genealogies offer a great promise for furthering our understanding of plant evolution (Schaal et al. 1998), and when genetic variation is organised into a

genealogy, with subsequent overlay of geography, the resulting analysis has been referred to as “intraspecific phylogeography” (Templeton et al. 1995). An explicit focus on a species' biogeographical past sets phylogeography apart from classical population genetics (Knowles and Maddison 2002). One of the goals of phylogeographic analyses is to evaluate the relative role of history in shaping the genetic structure of populations relative to important ongoing processes (Cruzan and Templeton 2000). Phylogeography thus provides insights into historical processes that have shaped the current geographic distribution of genetic variation, including environmental and ecological events, and population processes such as isolation events and bottlenecks (Avise 1998; Zink 1996, 2002). Revealing the distribution of genetic variation and the mechanisms responsible are common goals in population genetics. Phylogeographic approaches can be highly informative about these aspects of population biology. By superimposing phylogenies over geography, phylogeography can reveal the phylogenetic relationship among populations (Pongratz et al. 2003; Zink 2002). The validity of phylogeographic inferences can be increased by including the integration of data from external fields such as ecology, palaeopalynology, archaeology and geology (Avise 1998; Bermingham and Moritz 1998). Input from external fields may facilitate the conception of *a priori* hypotheses to be tested, provide insight into past selective forces and may provide support for results. For instance, inferred locations and size of past refugia may be substantiated by input from palaeomodelling (Hugall et al. 2002), plate tectonic theories could be used to generate hypotheses based on vicariance events (Trewick 2000) or fossil records could be used as an indication of past distribution patterns (Taberlet et al. 1998; Hugall et al. 2002).

Phylogeographic studies in plants have lagged behind those in animals, primarily because of difficulties in finding ordered, neutral intraspecific variation required for constructing gene trees (Olsen and Schaal 1999; Schaal and Olsen 2000). Detection of “phylogenetically informative” intraspecific variation is a prerequisite for plant population biologists aiming to reconstruct population history (Schaal et al. 1998). The ability of phylogeographic methods to detect geographic associations depends upon there being resolution in the haplotype tree, and historical events not being older than the coalescence time for the gene region being investigated (Templeton et al. 1995). However, in order to construct gene trees (for

phylogeographic use), significant genetic variation must occur at the appropriate level; *i.e.* among the populations or taxonomic units being investigated (Schaal et al. 1998). In the search for highly informative markers within a single species, genetic markers with recurrent mutation rates high enough to yield multiple mutations over the time-frame are of interest (Smouse 1998).

1.9 Tools for phylogeographic analysis

Biodiversity encompasses variety of morphology, physiology and biochemistry in living organisms. Underlying this phenotypic diversity is a diversity of genetic blueprints (nucleic acids) that specify phenotypes and direct their development. Genetic variation is in fact the raw material for evolutionary change. Using genetic markers, the patterns of genetic variation can be studied directly, that is, without the influence of environment. If the markers are neutral and not subject to selection, the pattern of genetic variation can be used to study the genetic relationship among populations. The history of a species or a population can have a great impact on its current genetic structure and therefore, it can also be used to deduce its past. There are however, several forces that can act to obscure the historical pattern, such as selection, mutation and current migration, but nonetheless, the genetic structure of current populations can be used to trace past migration patterns (Hewitt 1996) and also past and present hybridization events.

There is an apparently bewildering array of genetic markers available for phylogeographic studies that have evolved gradually over the years. The studies started with usage of morphological markers and with technological advancement new methodologies have been incorporated.

1.9.1 Morphological markers

The inheritance of these markers can be monitored visually without specialized biochemical or molecular techniques. Morphological traits that are controlled by a single/multiple locus can be used as genetic markers provided their expression is reproducible over a range of environments. Besides environment, the expression of such markers is also altered by epistatic and pleiotropic interactions. The number of morphological markers is very limited; their alleles interact in a dominant-recessive

manner, thereby making it impossible to distinguish the heterozygous individuals from homozygous individuals.

1.9.2 Molecular markers

1.9.2.1 Isozyme markers

Lewontin (1974) was the pioneer in using protein electrophoresis as an invaluable tool for measuring genetic variation and inferring evolutionary processes. This technique has allowed a remarkable degree of understanding of population-level processes (Brown and Allard 1970; Hamrick and Holden 1979; Hamrick and Godt 1989, 1996) and still continues to be a powerful tool to identify and analyze genetic structure in plants (*e.g*. Cruse Sanders and Hamrick 2004; Koch and Bernhardt 2004; Cheng et al. 2005b, Albaladejo and Aparicio 2007; López-Pujol et al. 2007; Quiroga and Premoli 2007; Rao et al. 2007; Torres-Diaz et al 2007; Fady et al. 2008). The problems associated with the isozymes include less informativeness, susceptibility to error, difficulty in inferring homology and polarizing of the variants. However, they have a great deal of power to measure population substructuring within species and to discriminate between populations of closely related taxa.

1.9.2.2 DNA based markers

Technological advances that took place after mid 1960's gave rise to an array of new methods for studying genetic variation. This is evident from the growth in a number of studies reporting population genetic variation in a wide variety of organisms and blooming of new areas like molecular systematics and phylogeography. Of these, DNA based markers have become an indispensable tool for analyzing genetic variation in natural populations after the invention of PCR (Saiki et al. 1988) and its technical simplicity made it a very popular technique among researchers. They represent nuclear as well as organellar genome (chloroplast and mitochondria) and reveal sequence polymorphisms that can be used as highly informative markers for the structure and dynamics of genomes at the level of populations and individuals. The choice of marker for population genetic studies depends on the amount of genetic variation needed to elucidate a particular phenomenon. Analysis of genetic variation has the potential to contribute materially to studies related to elucidation of history of population migration from the present day plant population (Hewitt 1996). DNA

based markers have substantially overtaken those based on protein and enzymes. One reason for this is that when analyzing variation within the genetic material itself, it is possible to avoid difficulties with differentiating gene expression related to the developmental state of material or environmental influences. Other reason is that there are potentially an enormous number of DNA markers that can be used and they are also constantly been modified to enhance their utility and to bring about automation in the process of genome analysis.

Random marker systems

The last decade has seen an increasing application of PCR based techniques such as random amplified polymorphic DNA (RAPD, Williams et al. 1990) or arbitrarily primed PCR (AP-PCR, Welsh and McClelland 1990), Inter simple sequence repeats (ISSRs, Zietkiewicz et al. 1994) and Amplified fragment length polymorphism (AFLP, Vos et al. 1995) in population genetic studies of plants as they can detect very low levels of genetic variation and donot require prior sequence information. However, the lack of reproducibility has been a major weakness of RAPD and ISSR markers (Riedy et al. 1992; Ellsworth et al. 1993). The dominant nature of these markers further reduces the information content compared to co-dominant markers such as isozymes. Inspite of these flaws RAPD and ISSR were the most common marker systems in applied research, *e.g.* plant breeding (review by Weising et al. 1995) and also in studies of natural plant populations (reviews by Bartish et al. 1999; Bussell 1999). Among the random marker systems, AFLP has gained popularity among population biologists owing to its capacity to detect thousands of independent loci with minimal cost and time, reproducibility and easy cross transferability. AFLP is now widely used to analyze the historical events that influenced distribution of organisms (Dixon et al. 2007; Nettel and Dodd 2007; Meudt and Bayly 2008; Ronikier et al. 2008).

Cytoplasmic markers: Noncoding regions of chloroplast and mitochondrial DNA

The extension of DNA analyses to population biology studies started with the application of RFLP using mitochondrial DNA for faunal diversity analysis. The last two decades has seen an immense boom in the usage of chloroplast DNA (cpDNA) and mitochondrial DNA (mtDNA) for a wide range of systematic studies. The pioneering efforts of various authors like Schnabel and Asmussen (1989) showed that

cpDNA variation can provide a great deal of useful information to plant population biologists. The chloroplast and mitochondrial genomes of plants are circular DNA molecules that are predominantly uniparently inherited (Harris and Ingram 1991). In the absence of heteroplasmy (presence of more than one organelle genotype per cell) the organelle genomes can be treated as effectively haploid (Birky et al. 1989). Further, for the purpose of population genetic analysis, the organelle genomes may be regarded as a single haploid gene. However, both the cytoplasmic genomes have separate mechanisms of replication and also have different rates of mutations and accumulation potentials (Clegg et al. 1986). Availability of intergenic chloroplast (cp) and mitochondrial (mt) DNA markers (Taberlet 1991; Demesure et al. 1995; Dumolin-Lapègue et al. 1997a) and the use of PCR-RFLP (Polymerase Chain Reaction-Restriction Fragment Length Polymorphism) or CAPS technique (Cleaved Amplified Polymorphic Sequence) and DNA sequencing have helped researchers to proficiently study genetic diversity in natural populations.

As the cytoplasmic genomes are maternally inherited they are more likely to record events in population history, firstly because the effective population size of haploid organelle genes in a hermaphrodite, outbreeding plant species is half of that of diploid nuclear genes (Birky et al. 1989). Secondly, when the recolonizing populations meet, gene flow for the maternally inherited organelle marker is far less substantial than that of nuclear genome. The organelle markers thus provide a good source of species-specific markers because of low levels of interspecific variation and are also useful for studying of intraspecific variation when the highly mutable portions of the organelle genomes are targeted.

The cytoplasmic genome are widely used to reconstruct the glacial and postglacial history of the plants. The much greater differentiation for a maternally inherited organelle marker than for biparentally inherited nuclear markers, even after a possible 50 generations of contact between recolonizing populations, is the feature of these markers that has been harvested to construct the past population structure of many plants. As many plant communities either possessing widespread distribution or disjunct distribution are known to be greatly affected by the climatic and environmental changes during the last glacial maxima (LGM, 20,000 years ago), these markers can be used to provide insight into the plant population history and the effect of glaciation on these plants. This information can also be used to design management

and conservation strategies for the populations. In plants, chloroplast DNA has been successfully used for the deduction of past colonisation routes in many species (Ferris et al. 1993; Dumolin-Lapègue et al. 1997b; Jøhnk and Siegismund 1997; Petit et al. 1997; King and Ferris 1998; Chiang et al. 2001; Gielly et al. 2001; Huang et al. 2002; Mohanty et al. 2002; Trewick et al. 2002; Johansen and Latta 2003; Palmé et al. 2003; Paula and Leonardo 2006; Afzal-Rafii and Dodd 2007; López de Heredia et al. 2007; Ikeda and Setoguchi 2007; Markwith and Parker 2007; Speranza et al. 2007) and also for the identification of hybridisation events (Terry et al. 2000; Bleeker and Hurka 2001; Petit et al. 2002; Lorenz-Lemke et al. 2006; Lihová et al. 2007).

The organeller DNA has many advantages making it a preferred marker for tracing the evolutionary historyof plants. However; there are few limitations in using this marker system. The data obtained on population genetic structure from an organelle genome are unreplicated as the organelle genome is effectively only a single gene. Hence, caution is needed while interpreting intraspecific chloroplast DNA phylogenies. Additionally, one of the limitations of using organelle markers pertaining to studies related to hybridization is that, though they provide unequivocal parentage for trees in the first generation, analysis of codominant or dominant nuclear markers is required for identifying the nature of genotypes in later generations.

Nuclear gene based markers: low copy nuclear genes

Nuclear introns and genes represent a largely untapped source of genetic variation for population genetics and phylogeography. Many protein coding gene sequences are available from the GenBank (EMBL database etc) but as the nuclear genes have a higher rate of evolution (broad surveys have revealed that synonymous substitution rates of nuclear genes are up to five times greater than those of chloroplast genes and 20 times greater than those of mitochondrial genes: Wolfe et al. 1987; Gaut 1998), it is in general more difficult to design universal primers for nuclear genes. However, studies have proved the utility of single copy nuclear genes for understanding population structure and phylogeny (*e.g.* Strand 1997). The nuclear genes employed to analyze phylogeographic structure in plants include, *glyceraldehyde-3-phosphate dehydrogenase* gene (Olsen and Schaal 1999; Tani et al. 2003, Koehler-Santos et al. 2006a) and vacuolar invertase gene (Caicedo and Schaal 2004). Studies with these traditional phylogeographical goals often use nuclear data to corroborate initial results

based on cytoplasmic loci (Hare 2001). Even though the genealogical tree constructed using nuclear genes may contain errors introduced by recombination (Schierup and Hein 2000), selection, and allelic discrimination (Zhang and Hewitt 2003), nuclear genes can more completely reveal the mosaic of genealogical patterns evolving in genomes in response to historical and environmental conditions (Hare 2001). Hence, special consideration needs to be given while using these markers. Studies using nDNA intron variation to analyze study plant phylogeography indicate that the anticipated technical and biological hurdles can be overcome (Olsen and Schaal 1999; Gaskin and Schaal 2002; Olsen 2002; Caicedo and Schaal 2004). Despite these difficulties, nuclear low-copy sequences are beginning to find use in phylogeny and population structure analysis. The most prudent way of utilizing nuclear markers is in combination with organelle markers. Nuclear markers can provide replicated data to support or refute population history scenarios suggested by the geographic distribution of organelle markers. The comparison between organelle and nuclear population genetic structure can also yield estimates of seed and pollen flow.

Microsatellites markers

Microsatellites, or simple sequence repeats (SSRs), are the sequences of repetitive DNA where a single motif consisting of one to six base pairs is repeated tandemly a number of times. Microsatellites are rapidly becoming established as an extremely useful tool in population genetics due to their high levels of variability and co-dominance, which have allowed the analysis of natural populations to be carried out at a higher degree of resolution than had previously been possible using isozymes (Jarne and Lagoda 1996; Powell et al. 1996a). Microsatellite sequences have been identified in all the three eukaryote genomes: nuclear, chloroplast (Powell et al. 1995) and mitochondrial (Soranzo et al. 1999) (Table 1.3). With respect to the organelle genomes, mitochondrial microsatellites have had little impact so far, but the chloroplast microsatellites (cpSSRs) and nuclear microsatellites have been increasingly used in population genetics since their discovery.

Chloroplast microsatellites (cpSSRs)

Chloroplast simple sequence repeats (SSRs or microsatellites; cpSSRs) were developed for such genetic analyses in the 1990s (Powell et al. 1995; Vendramin et al. 1996; Vendramin and Ziegenhagen 1997; Provan et al. 2001). The development of

such cpSSR markers has been achieved through the use of chloroplast DNA sequence information from GenBank (http://www.ncbi.nlm.nih.gov).

Table 1.3 Mutation rates in plant nuclear and organellar genomes

Genome	Type	Rate	Reference
Nuclear	Substitution	3×10^{-9} - 5×10^{-8}	Wolfe et al. 1987
Mitochondrial	Substitution	2×10^{-10} - 5×10^{-8}	Wolfe et al. 1987
Chloroplast	Substitution SSR length polymorphism	1×10^{-9} - 3×10^{-9} 3.2 - 7.9×10^{-5}	Wolfe et al. 1987 Provan et al. 1999b

Chloroplast microsatellites present higher levels of polymorphism and are easily genotyped. Chloroplast-specific markers are considered good indicators of historical bottlenecks, founder effects and genetic drift. The use of chloroplast microsatellites has allowed the examination of these events at a finer level of detail than it was previously possible (Li et al. 2007). Because microsatellites are mostly derived from noncoding regions, they are more likely than allozymes to evolve neutrally; this, together with their abundance in most eukaryotic genomes and typically high level of polymorphism, makes microsatellites the marker of choice for many population genetic studies (Jarne and Lagoda 1996; Bryan et al. 1999; Ishii and McCouch 2000; Selkoe and Toonen 2006) and for studying seed dispersal (Lian et al. 2003). Weising and Gardner (1999) developed and used ten consensus chloroplast microsatellite primers (ccmp) and concluded that such primers had utility to discriminate closely related genotypes in a broad array of plant species. Chloroplast SSRs have provided information on the post-glacial migration routes and the location of refugia in several species (Vendramin et al. 2000; Walter and Epperson 2001; Marshall et al. 2002; Grassi et al. 2006; Bucci et al. 2007; Edh et al. 2007).

Nuclear microsatellites

The development of codominant, highly polymorphic nuclear microsatellite markers (Morgante and Olivieri 1993; Powell et al. 1996b) have contributed immensely to the field of phylogeography. The high resolution obtained with these markers in estimation of relationships between individuals (Blouin et al. 1996; Gerber et al. 2000), have produced several studies of gene dispersal in plant species (Chase et al.

1996; Dow and Ashley 1998; Sampson 1998; Streiff et al. 1999). Microsatellites are commonly known to be highly variable and therefore, are a powerful tool for detecting genetic variation within species; this property has been harvested to study genetic variation in many plant species (Ouborg et al. 1999; Bakker et al. 2001; Balloux and Lugon-Moulin 2002; Nagamitsu et al. 2004; Breton et al. 2006; Truong et al. 2007).

1.10 *Symplocos laurina*: the plant under study

Symplocos laurina Wall. (Symplocaceae) is a montane plant species distributed in tropical and subtropical Asia and has a disjunct distribution in India. Itis a small evergreen tree growing up to ~7 m in height with thin, smooth, light grey bark and white wood, leaves simple alternate, very thick, lanceolate, elliptic or oblong, shortly acuminate at the apex, irregularly crenate or serrate; flowers yellow white fragrant in close cluster, in axillary spikes; fruits globose, purple ribbed drupes and seeds 1-3, testa thin (Fig. 1.10). Plants flower during September to December followed by fruiting (Almeida 1990). Seed dispersal is endozoic by birds and bats (Meher-Homji 1975). The bark is grayish green to slight gray with patches of crustose lichens. The outer bark is thin and the inner greenish or light brown, the cut surface of the thick bark of a mature tree is reddish brown on drying, and when broken short thin fiber tips are seen. *S. laurina* is distributed typically above an altitude of 800 m in Western Ghats, Eastern Ghats and North Eastern Himalayas, sparsely in Western Himalayas extending up to Burma, China, Japan and Korea. *S. laurina* is among the tropical stock that forms the montane forest proper. The plants typically occur near a water source and are always associated with *Syzygium* species (personal observations and Deshpande et al. 2001). A chromosomal number of $n = 11$ is reported for plants from Kodaishola in Tamil Nadu (Singh and Gill 1984). It is a species of shola fringes and not a true shola species (Meher-Homji 1967). Regeneration of these species is very difficult in open areas subjected to winter frost, occasional droughts and fires. However, its distribution shows its ability to withstand cold and is often among the first species to invade the shrubby formation in shola forest (Meher-Homji 1967). But even then in its beginning stages, it thrives as long as the shrubby layer protects it against frost (Agarwal et al. 1961).

1.10.1 Molecular studies of *Symplocos* species

Caris et al. (2002) proposed phylogenetic relationship of the monogeneric Symplocaceae on the basis of floral morphological characters. The study, on the basis of molecular as well as morphological results showed that taxa Symplocaceae is possibly related to the members of *Ericales sensu lato*. The most recent taxonomic revision of *Symplocos* genus by Soejima and Nagamasu (2004) using ITS region and two intergenic spacers between *trn*L and *trn*F and *trn*H and *psb*A of cpDNA has ascertained the monophyletic origin of the *Symplocos* species in Japan and have provided insight into the classification (which until then was uncertain). Wang et al. (2004) studied the phylogeny and intrageneric classification of *Symplocos*, using 111 species and inferred an Eastern Asian origin with subsequent dispersal to the America, which is contrary to North American origin inferred from the occurrence of earliest fossil from North America. They have placed *Symplocos laurina* (with only East Asian distribution) under the subgenus Hopea, section Bobu after phylogenetic revision using ITS, *mat*K, *rpl*16 and *trn*L-*trn* F sequences.

1.10.2 Classification of *Symplocos laurina* Wall.

Kingdom: Plantae – Plants
Subkingdom: Tracheobionta – Vascular plants
Superdivision: Spermatophyta – Seed plants
Division: Magnoliophyta – Flowering plants
Class: Magnoliopsida – Dicotyledons
Subclass: Dilleniidae
Order: Ericales
Family: Symplocaceae – Sweetleaf family
Genus: *Symplocos*
Species: *laurina*

1.10.3 Vernacular names of *S. laurina*

English: Chunga
Hindi: Bholiya
Assamese: Bhomroti
Kannada: Lodha / Chunga
Malayalam: Poccotti
Marathi: Lodhra
Sanskrit: Lodhra
Tamil: Kambli- vetti

The tree lodhra, rodhra or srimata means “propitious" and "Tilaka” because it is used in making the tilaka mark on the forehead. In Europe, it was looked upon as a cinchona bark substitute and was known at various times as ‘Ecarce de Lau tour’, ‘china nova’, ‘china calafornica’, ‘china brasilarsis’ and ‘ china paraquatan’.

1.10.4 Ethnopharmacology and other uses

S. laurina has many uses in indigenous system of medicine. The bark is astringent, acrid, refrigerant, ophthalmic, expectorant, anti-inflammatory, depurative, febrifuge, haemostatic, stomachic and constipating. According to the Ayurveda system of medicine it is useful in vitiated conditions of pitta and kapha, asthma, bronchitis dropsy, arthritis, ulcers, leprosy, skin disease, ulemorrhagia, haemorrhage, haemoptysis, dyspepsia, leucorrhoea, diarrhea, dyssentry and gonorrhoea. Its bark is described as bitter and pungent which is used as aphrodisiac and in menorrhagia, the diseases of ‘*raktpitta’* and the disease of the eyes (Warrier et al. 1996). The leaves impart a yellow dye used as a mordant. The decotion of leaves is valued in Indian medicines. The wood is used for fuel (Hore 1990).

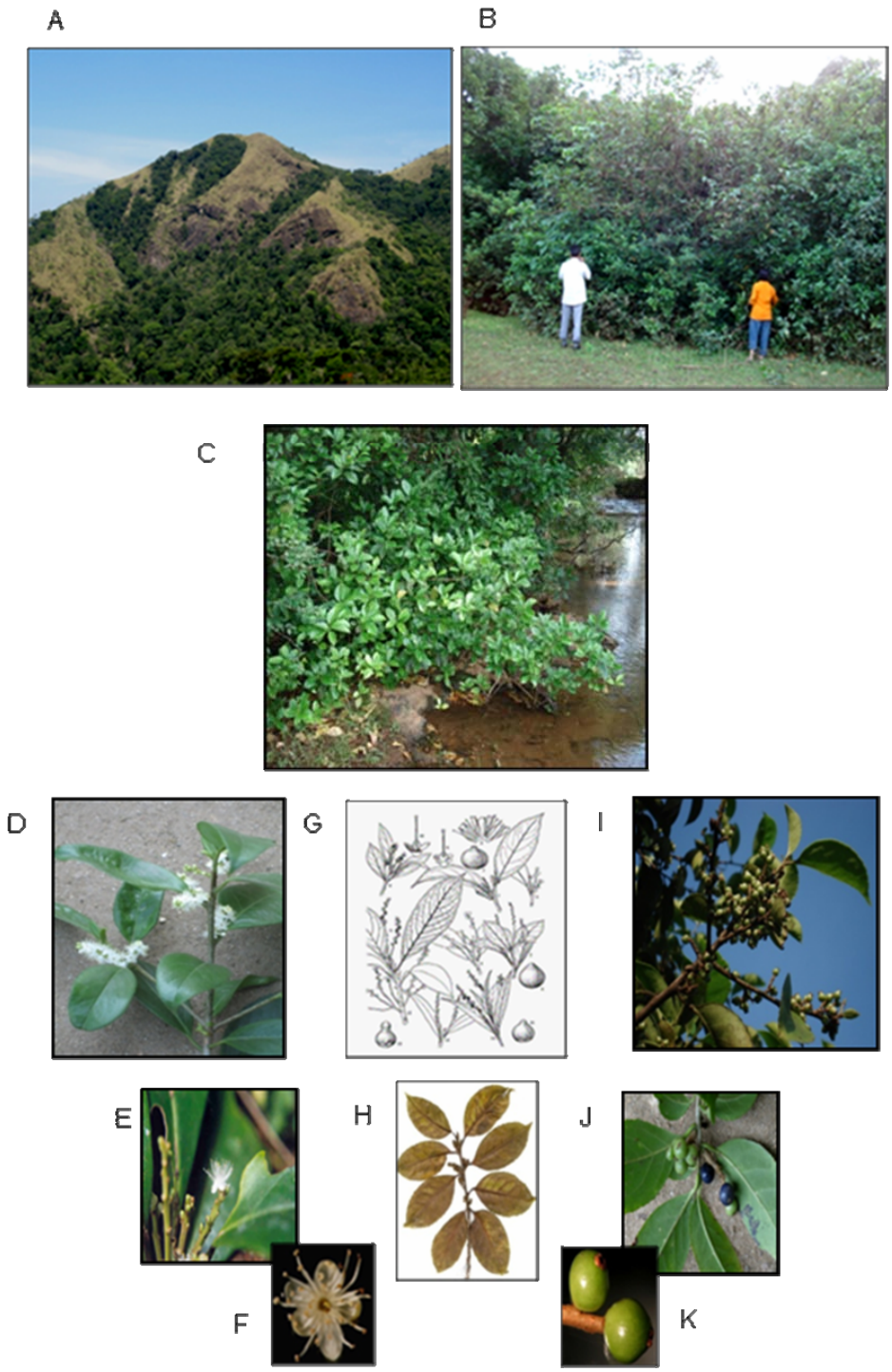


Fig. 1.10 A: Shola forest: habitat of montane plants, B: Collection of plant material from its habitat, C: *S. laurina* in its natural habitat, D: A flowering twig; E: Solitary flower, F: Solitary flower showing numerous stamens (2X), G: Line diagram of *S. laurina*, H: Herbarium voucher of *S. laurina,* I: Fruiting twig of *S. laurina*, J: Fruits ripe and unripe, K: Fruits (2X)

1.11 Genesis of the thesis

The existence of plants showing disjunct distribution in the Indian subcontinent has been a subject of discussion for over a century. However, attempts to capture the picture with the aid of modern technological advancements are almost negligible. The present thesis comprises studies carried out to understand the genetic structure of montane plants with disjunct distribution using *Symplocos laurina* as a model plant system. This work is an initial effort to understand the effect of last glaciation, climatic dynamics on the genetic structure and present day population distribution pattern of montane plants in India.

1.12 Objectives

1. To study the genetic diversity present in *S. laurina* chloroplast, mitochondrial and nuclear genomes
2. To analyze geographical association and relationships among haplotypes/ genotypes to infer the historical processes that contributed to the present state of distribution of *S. laurina* in India
3. To characterize events that led to contemporary distribution of *S. laurina* in India
4. To identify the most tenable hypothesis (vicariance vs dispersal) for the distribution of *S. laurina* in India

1.13 Organization of thesis

The thesis is presented in five chapters and the contents of each chapter are as follows

Chapter 1: Introduction and review of literature (the current chapter)

Chapter 2: Genetic diversity in *S. laurina* based on chloroplast and mitochondrial DNA markers

This chapter describes the genetic variation obtained with chloroplast and mitochondria based markers and provides an insight into the geographical pattern of distribution of haplotypes. The study identifies the factors, which influenced the genetic structure and provides evidences in support of the vicariance model. This chapter is divided into two sections:

- **Section A**: Analysis of populations based on chloroplast and mitochondrial universal markers using PCR-RFLP technique
- **Section B**: Analysis of populations with chloroplast microsatellite markers

Chapter 3: Understanding the distribution of genetic diversity in *S. laurina* based on chloroplast genome sequence to elucidate the phylogeographical structure

This chapter describes the effects of glaciation on the genetic pattern obtained from sequencing of three noncoding chloroplast DNA regions from *S. laurina* populations in India. The study provides evidences supporting the vicariance model and identifies existence of glacial refugia.

Chapter 4: Assessing the phylogeographic pattern of genetic structure in *S. laurina* using nuclear markers

This chapter deals with identification of polymorphic nuclear markers for analysis of population structure. The nuclear DNA diversity in *S. laurina* populations using *GapC* gene is presented, which identifies haplotypes that help to understand the evolutionary history of the plant.

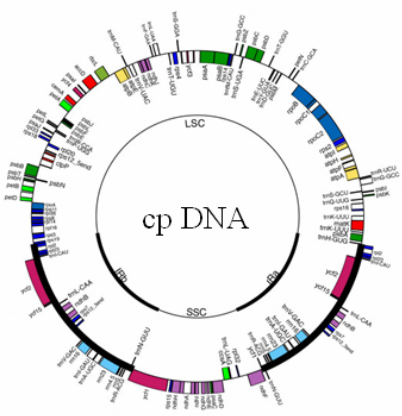
Chapter 5: Thesis summary and future directions

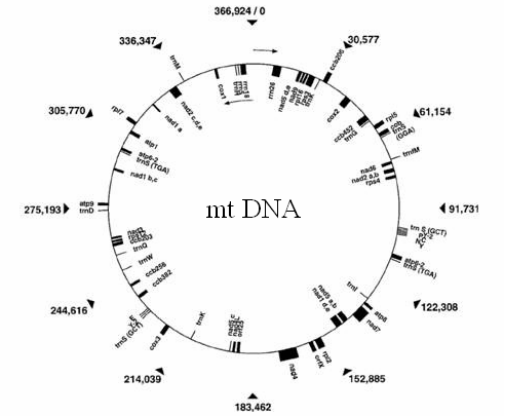
This chapter summarizes the important findings of the present research work. A chronicle of historical events contributing to the present day disjunct distribution of *S. laurina*, which were uncovered during the course of the study, is being synthesized.

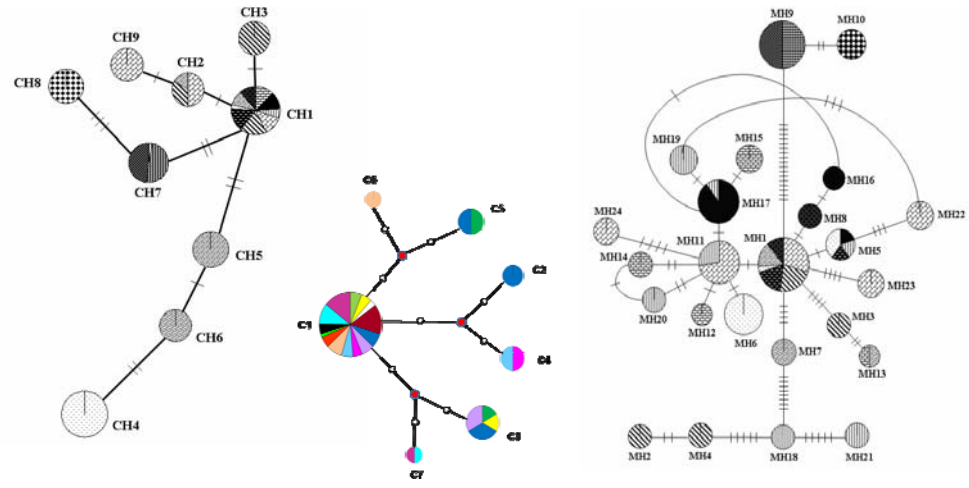
This is followed by bibliography.

Chapter 2

Genetic diversity in *S. laurina* based on chloroplast and mitochondrial DNA







Section A: Analysis of populations based on chloroplast and mitochondrial universal markers using PCR-RFLP technique

OUTLINE

2.A.1 Introduction

In recent years a great deal of interest has surfaced in the quantification and evaluation of biological diversity globally. The biodiversity of India has been implicated to be influenced by an array of factors, of which climatic changes of Quaternary have played a crucial role in defining population structure of plant species especially of montane zone. During late Miocene and Pliocene times, large tracts of humid forest cover progressively declined and retreated to wetter parts of India (wet-zone) but, during more recent (Quaternary) climate cycles there has been repeated expansion and shrinkage of forest cover (Gupta et al. 2003). Thus, the multiple cycles of dry and wet climate of Quaternary greatly influenced distribution of vegetation and vertebrate fauna, creating a distribution pattern that is vastly different from what was present before the last glacial maxima (LGM) (Chauhan 2002; Gupta et al. 2003). The present disjunct distributions observed in some species, therefore, represent a complex scenario influenced by a large number of ecological, climatological, geographical and genetic factors.

Analyses of spatial genetic patterns in contemporary populations have been useful in inferring postglacial histories of a number of species and in identifying refugia and migration routes in many European species. Several phylogeographical data sets have been produced during the last decade that focused mainly on European tree species for which substantial palaeobotanical data are available (*e.g.* Webb and Bartlein 1992; Konnert and Bergmann 1995; Demesure et al. 1996; King and Ferris 1998; Abbott et al. 2000; Mardulyn 2001; Despres et al. 2002; Petit et al. 2002; Petit and Grivet 2002; Burban and Petit 2003; Grivet and Petit 2003; Hampe et al. 2003; Jaramillo-Correa and Bousquet 2003; Kropf et al. 2003; Palmé et al. 2003; Petit et al. 2003; Abbott and Comes 2004; Heuertz et al. 2004; Hewitt 2004; Jaramillo-Correa et al. 2004; Petit et al. 2005 Anderson et al. 2006), and on plant species of high mountains (Schönswetter et al. 2005) or northern latitudes (Abbott and Brochmann 2003; Brochmann et al. 2003). However, the phylogeographical pattern of plants of Asian region is still in infant stage; the reported studies are mainly confined to Japan, Tibet, Taiwan, China and few other countries (*e.g. Abies mariesii* - Tsumura et al. 1994; *Pedicularis chamissonis* and *Primula cuneifolia* - Fujii et al. 1997; *Stachyurus macrocarpus* - Ohi and Murata 2000; *Cycas taitungensis* - Huang et al. 2001; *Fagus crenata* - Fujii et al. 2002; *Fagus crenata* - Okaura and Harada 2002; Stone oaks -

Cannon and Manos 2003; *Primula sieboldii* - Honjo et al. 2004; *Trochodendron aralioides* - Huang et al. 2004; *Alsophila spinulosa* - Su et al. 2005; *Euphrasia transmorrisonensis* - Wu et al. 2005; *Potentilla matsumurae* - Ikeda et al. 2006; *Hibiscus tiliaceus* – Takayama et al. 2006; *Machilus thunbergii* and *Machilus kusanoi* - Wu et al. 2006; *Picea jezoensis* - Aizawa 2007; *Pedicularis* ser. *Gloriosae* - Fujii 2007; *Taxus wallichiana* - Gao et al. 2007; *Phyllodoce nipponica* - Ikeda and Setoguchi 2007; *Picea crassifolia* - Meng et al. 2007; *Quercus mongolica* var. crispula - Okaura et al. 2007; *Metagentiana striata* - Chen et al. 2008). Till date, no reports are available on the phylogeographical patterns of plants belonging to Indian subcontinent except for studies done on the genetic structure ***e.g.*** Jain et al. 2000; Deshpande et al. 2001; Bahulikar et al. 2004 and Apte et al. 2006. Therefore, in order to obtain a comprehensive picture for the effect of glaciation on the genetic structure of flora of Indian subcontinent and the postglacial recolonization processes, it may be particularly useful to study species with disjunct distribution spanning over large region with specific climatic preferences (ecological niches are specific ***e.g.*** montane species). The montane plants of India occurring in a number of discrete populations in habitats that are naturally fragmented form the best candidates for phylogeographical analysis. *Symplocos laurina* a montane plant species, is among the species showing disjunct distribution as described in Chapter 1. This species assumes further importance in view of the fact that it belongs to the two biodiversity hotspots of India. An understanding of how biodiversity is distributed and maintained, particularly, within the species rich tropical forest region that is being rapidly depleted (Daniels 1992) is important and needs to be analyzed. Sholas provide such a narrow ecological niche to their members that they are considered living fossils or relict communities, which may vanish or eventually become extinct (Vishnu-Mittre and Gupta 1968). In order to conserve the germplasm pool of the two hotspots, it is important to elucidate the evolutionary history and to study diversity inherent within the organism studied. The evolutionary studies can be used to understand the past bottlenecks, founder effects and other genetic selection pressures that have shaped the present day genetic structure in *S. laurina*. This, in conjunction with biogeography of the subcontinent will help us to understand the role played by environmental factors in determining the distribution of shola species.

In recent times, molecular analyses of cytoplasmic genomes have become powerful tools for testing phylogeographical hypotheses and for studying the historical processes that have led to geographic variation within species (*e.g.* Dumolin-Lapègue et al. 1997a; Desplanque et al. 2000; Palmé et al. 2003). Chloroplast (cp) and mitochondrial (mt) DNA are nonrecombinant (no recombination occurs and are inherited as a single gene), homoplasmic (copies of their DNA within an individual are identical) and highly conserved molecules (Ouborg et al. 1999). However, mtDNA shows intramolecular recombinations occasionally, due to the long hairpin like secondary structure (Dumolin-Lapègue et al. 1998). In most plants, cytoplasmic genomes such as chloroplast and mitochondria are dispersed through seed; hence they are more suitable for the reconstruction of past processes and patterns than nuclear genes (McCauley 1995). In angiosperms both cp and mt DNA show maternal inheritance in most species (Reboud and Zeyl 1994). As a consequence, the two organellar genomes are expected to remain associated, that is, they should behave as if they are completely linked (Schnabel and Asmussen 1989) and are expected to give similar information on dispersal and gene flow by seeds. However, there are reports where cpDNA is biparentally inherited or shows 'paternal leakage' in plants (Reboud and Zeyl 1994). Additionally, cpDNA evolves four times slower than nuclear genome and mtDNA evolves three times slower than cpDNA (Wolfe et al. 1987). The availability of universal intergenic cp and mt DNA markers (Taberlet et al. 1991; Demesure et al. 1995; Dumolin-Lapègue et al. 1997b) and use of the PCR-RFLP (or CAPS) technique have helped researchers to study genetic structure of various plant species to identify the past processes involved in shaping the contemporary distribution of plants. In this study, I have used the variation in intergenic regions of cp and mt DNA to analyze the effect of glaciation on the present day population structure of *S. laurina*, and to elucidate the historical events involved. Through the present study I intend to a) analyse geographical association and relationships among cp and mt DNA haplotypes to infer the historical processes that contributed to the present state of distribution of *S. laurina* in India, b) to assess the level of association between cp and mt DNA haplotypes, c) to characterize events that led to contemporary distribution of *S. laurina* in India and d) to examine whether the molecular analysis of data corroborate or dispute inferences made from biogeographical analysis and fossil data.

2.A.2 Materials and methods

2.A.2.1 Sample collection

A total of 218 plants were sampled from 12 locations (subsequently referred to as populations) from the two biodiversity hotspots, Northeast India (NEI) and Western Ghats (WG) along with Eastern Ghats (EG) (referred hereafter as regions). The details of sampling are given in Table 2.A.1 and Fig. 2.A.1. Collection sites were identified using herbarium collections of Botanical Survey of India (BSI, Western Circle, Pune, India) and field observations that were conducted over a period of two years in collaboration with BSI. An effort was made to provide a wide representation over the entire natural habitat range to avoid artifactual patterns caused by gaps in the distribution of samples (Templeton 1995). Even after extensive exploration in Sikkim, Darjeeling area and in Arunachal Pradesh, which are part of northeastern Himalayas, samples could not be collected as they were found to be absent from the places reported previously in literature. The SPR, CHR and ELF populations represent the Khasi hill range of Northeast India. The other populations include AMB (Sahayadri ranges), KUD (Gangamoola-Aroli-Gangrikal range), KEM (Baba Budangiri hill range) and TAL (Brahmagiri hills) whereas the populations SAN and THR represent Niligiri hills and KOD the Palani Hills. The remaining populations KLH and YRD represent the EG, while the KLH population is from the Kollimalai and the population YRD represents the Shevaroy hills. In the present study, the KOD population was included in EG regions for all the analyses carried out. The scale of sampling ranged from 10 km to 2247 km for the most distant populations (SPR and KOD) except for two populations from NEI where interpopulation distance was less (3.02 km), but separated by physical barriers (Khasi Hills). Frequency of plants in the study area varied from 10%-60% per sq. km. Depending on accessibility, more than 50% of the available plants were sampled from each population. At some of the sites, even after extensive exploration, only a limited number of plants could be sampled. In each of these populations, leaves were collected from nine to thirty individuals separated by minimum distance of 50 m to avoid clone sampling (Van Rossum et al. 2004). For all the collected samples the global position was noted with Global Positioning System (Garmin ver. 2.11). The vouchers of the samples were submitted to the herbarium of BSI.

Table 2.A.1 Sampling locations of 12 populations of *S. laurina* in India

Population sites	Population codes	Regions	No. of individuals sampled	Latitude	Longitude	Altitude (m)
Shillong Peak Road	SPR	NEI	13	E 91.85528	N 25.53548	1903
Cherrapunji Road	CHR	NEI	13	E 91.73666	N 25.33120	1661
Elephanta Falls	ELF	NEI	9	E 91.82544	N 25.53368	1754
Amboli	AMB	WG	29	E 74.06654	N 15.99816	825*
Kudremukh	KUD	WG	23	E 75.18670	N 13.20770	836
Kemmangundi	KEM	WG	17	E 75.74579	N 13.47867	1369
Talakaveri	TAL	WG	17	E 75.48790	N 12.38866	1243
Thaishola Road	THR	WG	15	E 76.62558	N 11.24087	1877
Sandynallah	SAN	WG	12	E 76.06321	N 11.43310	2227#
Kodaikanal	KOD	EG	10	E 77.47947	N 10.23795	2131
Kolli hills	KLH	EG	30	E 78.36096	N 11.30846	1303
Yercaud	YRD	EG	30	E 78.22367	N 11.81318	1293

* Population showing the lowest altitude from where the plants were sampled,
# Population showing the highest altitude from where the plants were sampled

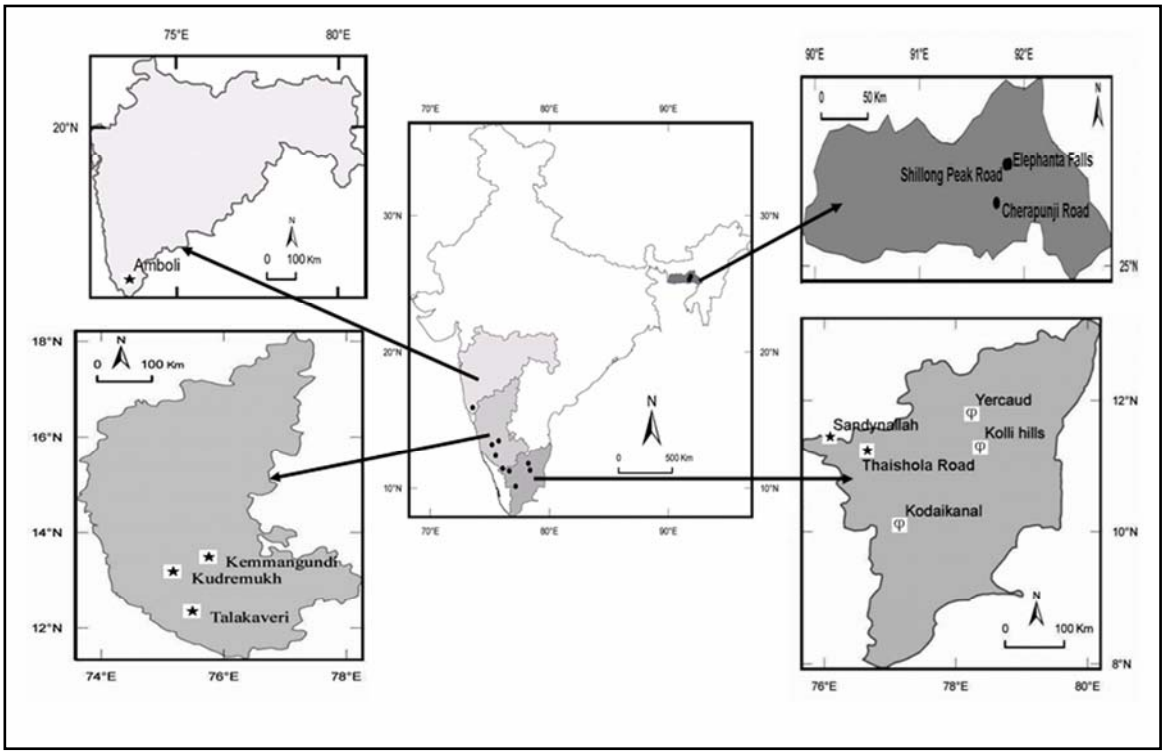


Fig. 2.A.1 Map of India depicting sampling sites in NEI, WG and EG, projected out state wise

The symbols represent ●-populations from NEI, □-populations from WG, φ-populations from EG.

2.A.2.2 DNA isolation

After collection, leaves were cleaned and immediately frozen in liquid nitrogen and stored at -70ºC. Isolation of DNA was carried out using modified hexadecyl-trimethyl-ammonium bromide (CTAB) method (Richards et al. 1994). In the first step, 1 g of leaf tissue was ground to fine powder in liquid nitrogen, which was quickly transferred in a polypropylene tube containing 10 mL of pre-warmed extraction buffer (2% CTAB, 1.4 M NaCl, 100 mM Tris, 20 mM EDTA, 100 mM Sodium metabisulphite, pH 8.0). The mixture was swirled gently and incubated at 65°C in water bath for 60 min with intermittent shaking for efficient lyses of cells. After incubation, the tube was cooled, equal volume of chloroform: isoamyl alcohol (24:1) was added and shaken gently it was then centrifuged at 10,000 rpm, for 10 min at room temperature. The aqueous layer was removed carefully with Pasteur pipette, collected in fresh polypropylene tube, $1/10^{th}$ volume of 10% CTAB was added followed by equal volume of chloroform: isoamyl alcohol. The mixture was swirled gently and the tube was centrifuged at 10,000 rpm, for 10 min at room temperature. The aqueous layer was removed carefully and collected in fresh polypropylene tube to which ¼ volume of precipitation buffer (2% CTAB, 50 mM Tris, 10 mM EDTA, 1% β-mercaptoethanol, pH-8.0) was added followed by equal volume of sterile deionized water. The tube was centrifuged at 10,000 rpm, for 10 min at room temperature and the solution was decanted carefully. As the plant was found to contain high amount of polysaccharides the pellet was dissolved in HSTE buffer (High salt Tris EDTA - 10 mM Tris, 0.5 mM EDTA, 1M NaCl, pH 8.0) and kept at room temperature for some time till it dissolved. This solution was then precipitated with twice the volume of chilled ethanol. Centrifugation was done at 10,000 rpm, for 10 min at 4°C followed by two 70% ethanol washes. The pellet was dried at room temperature. Sufficient volume of TE buffer was added to dissolve the pellet following which the RNase treatment was given. The DNA was quantified spectrophotometrically at 260/280 nm as well as visually by ethidium bromide staining on 0.8% agarose gel and comparing it with commercial DNA samples with known concentrations (Sambrook et al. 1989).

2.A.2.3 PCR-RFLP analysis of chloroplast and mitochondrial intergenic regions

To assess cp and mt DNA polymorphisms in *S. laurina*, a subset of 60 individuals was screened using PCR-RFLP. Five individuals per population were chosen

randomly as representative of sample range. Twelve cpDNA primers (*trn*H-*trn*K, *trn*C-*trn*D, *trn*D-*trn*T, *psb*C-*trn*C, *trn*S-*trn*fM, *psb*A-*trn*S, *trn*S-*trn*T, *trn*M-rbcL, *trn*T-*trn*F, *trn*H(GUG)-psbA, *trn*K2-*trn*Qr and *trn*Q-*trn*Rr) and ten mtDNA primers (*nad*1/B-*nad*1/C, *nad*4/1-*nad*4/2, *nad*4/2-*nad*4/4, *rps*14-*cob*, *cox*2/1-*cox*2/2r, *nad*1/4-*nad*1/5r, *nad*4/3-*nad*4/4r, *nad*7/1-*nad*7/2r, *nad*7/2-*nad*7/3r and *nad*4/2-*nad*4/3r) as described by Taberlet et al. (1991), Demesure et al. (1995), Dumolin-Lapègue et al. (1997b) and Hamilton (1999) were used for screening. These primers represent various intergenic regions of cpDNA and mtDNA, respectively.

Polymerase chain reaction was carried out using total DNA as template employing the primers described above. Each 25 µL consisted of 50 ng template DNA, 200 µM of dNTPs, 0.2 µM of each primer, 1.5 mM $MgCl_2$, and 0.8 U of *Taq* DNA Polymerase (Bangalore Genei, Bangalore, India) in the buffer provided by the manufacturer. The PCR amplifications were carried out in PTC 200 thermal cycler (MJ Research, Waltham, Massachusetts, USA) using initial denaturation at 94°C for 5 min, followed by 35 cycles of 1 min at 94°C, 30 s - 1 min at 50°C-63°C (depending upon the average annealing temperature of the primers used) and 2 min at 72°C with a final extension of 5-7 min at 72°C. The PCR amplicons were separated by electrophoresis on 1% agarose gels in 0.5 X TAE (Tris-Acetate-EDTA) buffer. In all ten cpDNA and eight mtDNA primers giving good amplicon were digested with six different tetracutter restriction endonucleases (*Alu*I, *Hae*III, *Hha*I, *Msp*I, *Rsa*I and *Taq*I) and four hexacutter restriction endonucleases (*Eco*RI, *Hinf*I, *Hind*III and *Xba*I) (Promega, Madison, Wisconsin, USA). Restriction digestions were performed in a total volume of 20 µL consisting of 5 µL of PCR product, 1 X restriction buffer (provided by the supplier) and 1 - 2 units of restriction enzyme and was incubated at 37°C (in case of *Taq*I at 65°C) for 12 - 14 h. The restriction digests were resolved on 1.5% agarose gels (fragments above 1.3 kb), on 2% - 3.5% MetaPhor agarose gel (Cambrex Bio Science, Rockland, ME, USA) (fragments having size range 300 bp - 1.3 kb) and on 4% - 6% native polyacrylamide gels (fragments having size range 50 bp - 400 bp) in 0.5 X TBE (Tris-Borate-EDTA) buffer. The agarose and MetaPhor agarose gels were stained with ethidium bromide and photographed using ImageMaster VDS gel documentation system (Amersham Pharmacia Biotech, Denver, USA) under UV light, whereas the polyacrylamide gels were silver stained (Creste et al. 2001). Restriction fragments were labeled in decreasing order of

molecular weight as described by Dumolin-Lapègue et al. (1998). The sizes of the digested products were analysed using the software SEQAID ver. 3.81 (Rhoads and Roufa 1991). Overall 80 cp and 64 mt primer-enzyme combinations (PEC) were tested. On the basis of initial screening, three cp and eight mt PEC that gave good amplification and non-ambiguous polymorphism were selected and were used for all the populations (Table 2.A.2).

2.A.2.4 Statistical analysis

To identify chlorotypes (cpDNA haplotypes) and mitotypes (mtDNA haplotypes), alleles observed for the three and eight PEC specific for chloroplast and mitochondrial DNA, respectively were combined. Various parameters of diversity (h_T-total diversity, h_S-average intrapopulation diversity, G_{ST}-level of population subdivision of diversity using unordered alleles and frequency of haplotypes) were calculated using the HAPLODIV software (Pons and Petit 1995). The HAPLONST software was used to calculate N_{ST} (the level of population subdivision for ordered alleles) and v_T and v_S (the analogues of h_T and h_S) (Pons and Petit 1996). The program PERMUT ver. 2.0 was used to test the difference between G_{ST} and N_{ST} through random permutations of haplotype identity (Burban et al. 1999). The estimators of these parameters proposed by Pons and Petit (1995, 1996) are based on a random model of population variation which takes into consideration the variation due to sampling of populations, unlike fixed model procedure as assumed by Nei and Chesser (1983). To elucidate the relationship between geographic distance and the genetic distance, Mantel correlation coefficient was calculated using the software Mantel Nonparametric Test Calculator ver. 2.00 (Liedloff 1999). Conventional F_{ST} (Wright 1931) for population subdivision based on chlorotypes and mitotypes was estimated using Arlequin ver. 3.11 (Excoffier et al. 2005). To describe the geographical distribution of diversity, the data was analysed in two stages. First, the diversity parameters were calculated for the whole data set and in second step, the populations were subdivided regionwise into two groups: NEI and SI.

The influence of spatial separation on the degree of differentiation among populations was investigated by calculating pairwise G_{ST} and N_{ST} values with the program DISTON (http://www.pierroton.inra.fr/genetics/labo/software/) and plotting means of these parameters against geographical distance classes. Analysis of

Molecular Variance (AMOVA) was used to partition the variance among regions, populations and within population using the software Arlequin ver. 3.11 (Excoffier et al. 2005). A minimum spanning tree (MST) for chlorotypes and minimum spanning network (MSN) for mitotypes were drawn using Arlequin while Treeview ver. 1.6.6 (Page 1996), was used to view the minimum spanning tree/network.

2.A.3 Results

2.A.3.1 Identification of chlorotypes and mitotypes

The screening of twelve cp and ten mt DNA primers resulted in identification of ten cp and eight mt DNA primers giving good amplicon. The amplicons were digested with various tetra and hexa cutting restriction enzymes as detailed in Materials and methods and only those combinations which gave polymorphic restriction patterns were used for the analysis. To assess cp and mt DNA polymorphism in *S. laurina*, a subset of 60 individuals was screened using PCR-RFLP. Five individuals per population were chosen randomly as representative of sample range. Overall 80 cp and 64 mt DNA primer-enzyme combinations (PEC) were tested. On the basis of initial screening, three cp and eight mt PEC that gave good PCR yield and non-ambiguous polymorphism were selected and used for all the populations. For the study, eleven (three from cpDNA and eight from mtDNA) polymorphic patterns representing six pairs of primers (Fig. 2.A.2) and twelve populations from the two biodiversity hotspots in India and EG comprising 218 individuals were considered (Table 2.A.1).

The size range of polymorphic fragments was between 88 bp and 1431 bp. The size range of amplicons using the cp and mt DNA specific primer pairs, namely: *trn*C- *trn*D, *psb*C-*trn*S, *trn*S-*trn*T, *nad*1B-*nad*1C, *nad*4/1-*nad*4/2 and *nad*4/3-*nad*4/4r used for further restriction enzyme digestion in the present study are shown in Fig. 2.A.2. Further, Figs. 2.A.3, 2.A.4 and 2.A.5 show the restriction patterns in *S. laurina* using the chloroplast specific primer pairs *trn*C-*trn*D / *Hinf*I, *psb*C-*trn*C / *Alu*I and *trn*T-*trn*F / *Taq*I, respectively. Figs 2.A.6 - 2.A.12 show the restriction patterns in *S. laurina* using the mitochondrial DNA specific primer pairs.

Table 2.A.2 List of chloroplast DNA and mitochondrial DNA PECs used for analysis of *S. laurina* populations

Primers	Size of amplicons (in kb)	Restriction enzymes	No. of fragments obtained	No. of polymorphic fragments	Percent polymorphism	Patterns contributed	References
Chloroplast DNA							
*trn*C-*trn*D	3.00	*Hinf*I	5	2	40.0	3	1
*psb*C-*trn*C	1.50	*Alu*I	11	3	27.0	3	1
*trn*T-*trn*F	1.80	*Taq*I	11	4	36.0	3	2
		Total	27	9			
Mitochondrial DNA							
*nad*4/1-*nad*4/2	2.00	*Msp*I	7	6	85.6	4	1
*nad*4/2-*nad*4/4	3.35	*Hha*I	18	18	100.0	4	1
		*Rsa*I	16	8	50.0	6	
		*Msp*I	13	3	23.0	4	
*nad*4/3-*nad*4/4r	3.00	*Hind*III	7	5	71.0	6	3
		*Taq*I	10	3	30.0	2	
		*Xba*I	7	6	85.7	7	
		*Hae*III	14	2	14.3	3	
		Total	92	51			

1. Demesure et al. (1995) 2. Taberlet et al. (1991) 3. Dumolin Lapègue et al. (1997b)

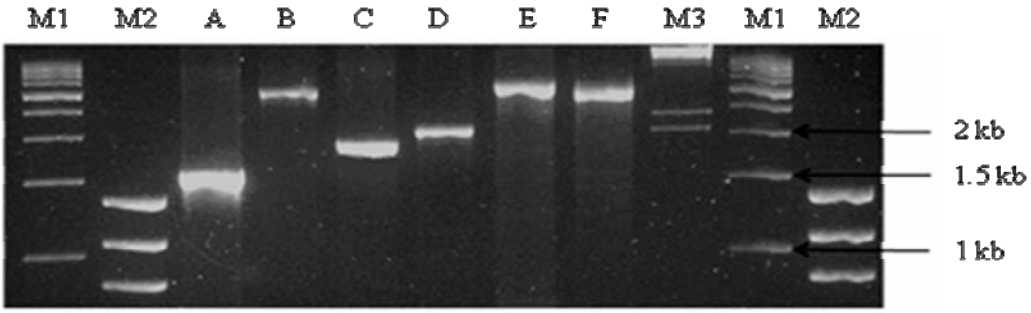


Fig. 2.A.2 Amplicons of cp and mt DNA region analyzed in the present study

Lane M1: Molecular weight marker 500 bp ladder, lane M2: molecular weight markers ØX174/ *Hae*III, lane A: *psb*C-*trn*C, lane B: *trn*C-*trn*D, lane C: *trn*T-*trn*F, lane D: *nad*4/1-*nad*4/2, lane E: *nad*4/3-*nad*4/4r lane F: *nad*4/2-*nad*4/4 and lane M3: λ DNA-*Hind*III digest molecular weight marker

The size range of polymorphic fragments was between 88 bp and 1431 bp. The size range of amplicons using the cp and mt DNA specific primer pairs, namely: *trn*C- *trn*D, *psb*C-*trn*S, *trn*S-*trn*T, *nad*1B-*nad*1C, *nad*4/1-*nad*4/2 and *nad*4/3-*nad*4/4r used for further restriction enzyme digestion in the present study are shown in Fig. 2.A.2. Further, Figs. 2.A.3, 2.A.4 and 2.A.5 show the restriction patterns in *S. laurina* using the chloroplast specific primer pairs *trn*C-*trn*D/*Hinf*I, *psb*C-*trn*C/*Alu*I and *trn*T-*trn*F/*Taq*I, respectively. Figs. 2.A.6 - 2.A.12 show the restriction patterns in *S. laurina* using the mitochondrial DNA specific primer pairs.

2.A.3.2 Distribution of chlorotypes and mitotypes of *S. laurina*

Three cp PEC revealed polymorphism in 12 populations of *S. laurina* and yielded 27 restriction fragments of which nine were polymorphic (percentage polymorphic bands = 33%) which defined nine chlorotypes (Tables 2.A.2 and 2.A.3). Similarly, initial screening of mtDNA resulted in identification of 92 fragments of which 51 were polymorphic (percentage polymorphic bands = 52.2%) and defined 24 mitotypes in 218 individuals from 12 populations of *S. laurina* (Tables 2.A.2 and 2.A.4). Most of these mutations are apparently insertions/deletions (indels), although it is difficult to ascertain in case of cpDNA, given that a single restriction endonuclease was used for each fragment. Further, some variation might have gone unseen if the restriction fragments containing the mutation were too small to be detected on the gel.

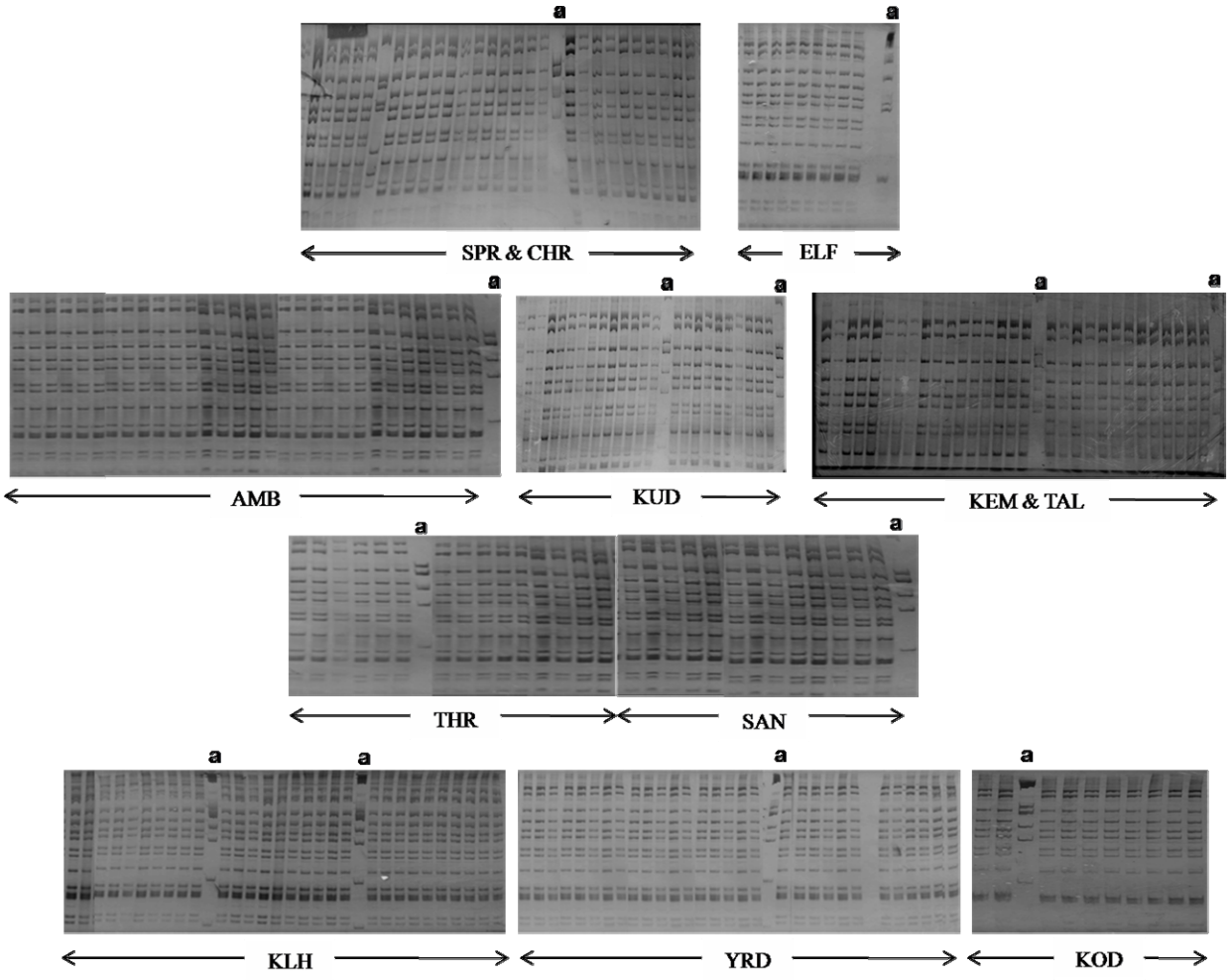


Fig. 2.A.3 Restriction pattern in *S. laurina* using primer pair *trn*C-*trn*D and *HinfI* restriction enzyme

Lanes marked 'a' above them indicate molecular weight marker ØX174/*Hae*III digest. Labels below the gel picture represent loading order of samples from the 12 populations.

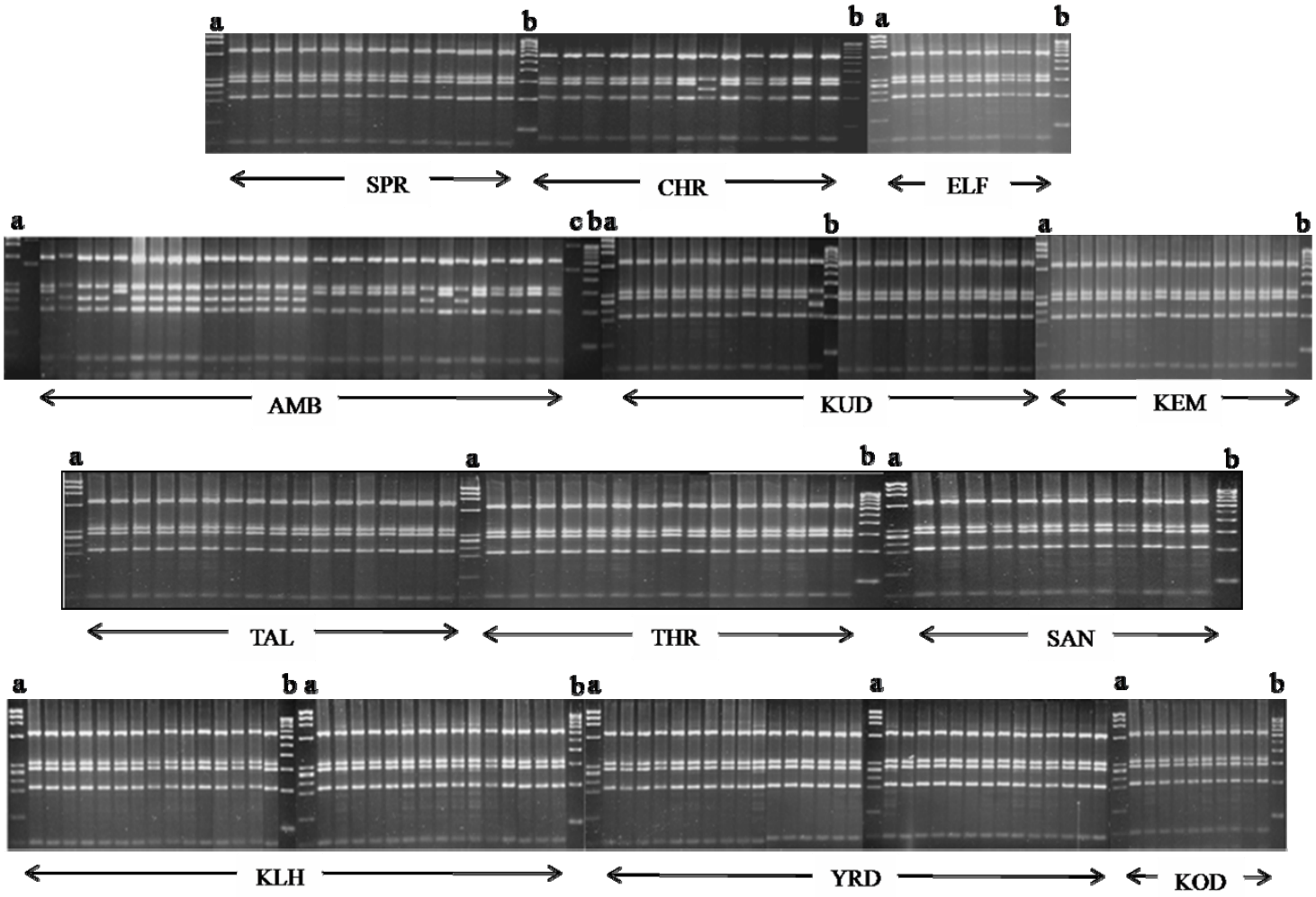


Fig. 2.A.4 Restriction pattern in *S. laurina* using primer pair *pbs*C-*trn*C and *Alu*I restriction enzyme

Lanes with molecular weight markers are marked as a, b and c above them, indicate ØX174/*Hae*III digest, 100 bp ladder and 500 bp ladder, respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

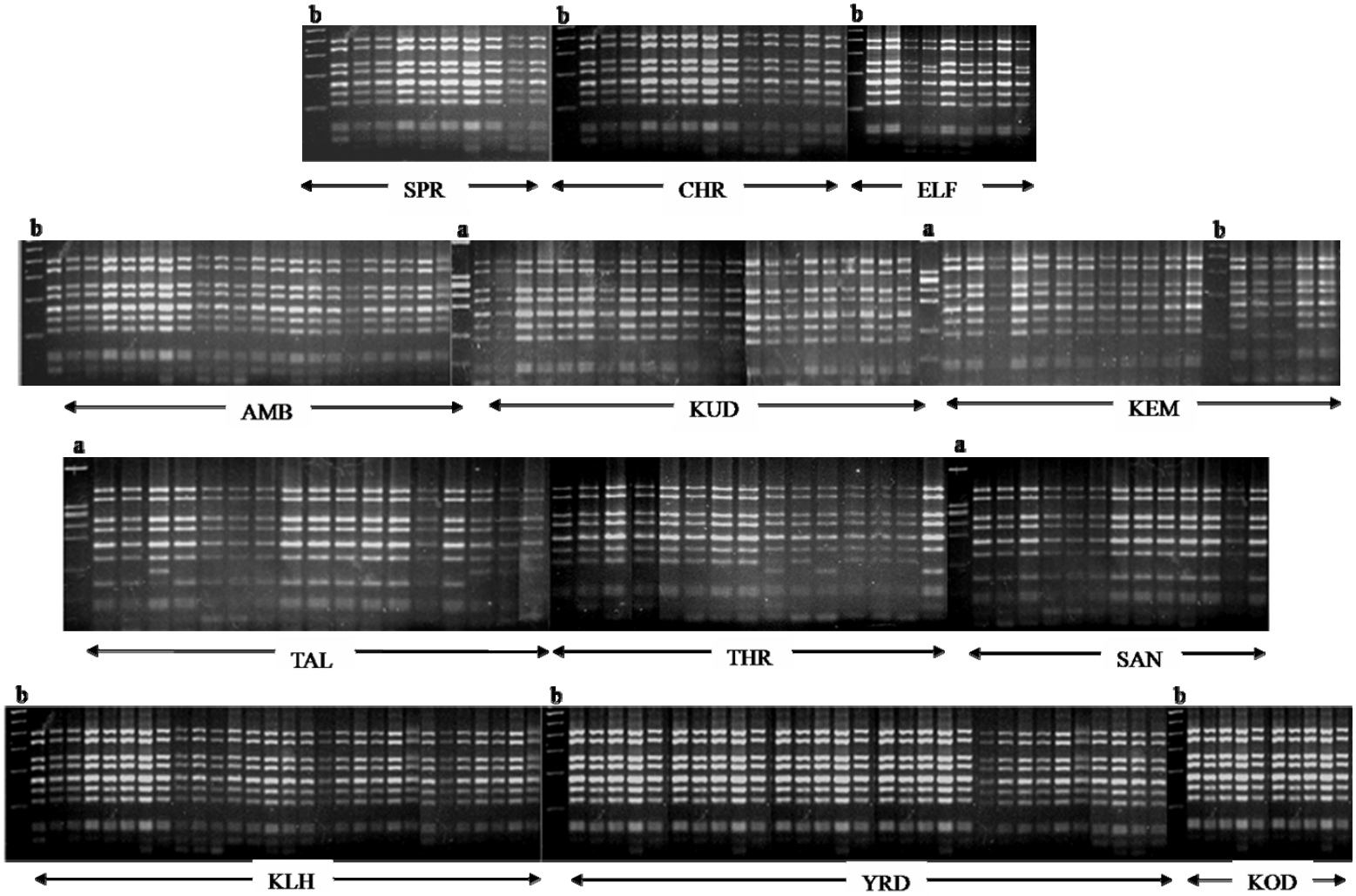


Fig. 2.A.5 Restriction pattern in *S. laurina* using primer pair *trn*T-*trn*F and *Taq*I restriction enzyme

Lanes with molecular weight markers are marked as a, b and c above them indicate, ØX174/*Hae*III digest, 100 bp ladder and 500 bp ladder, respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

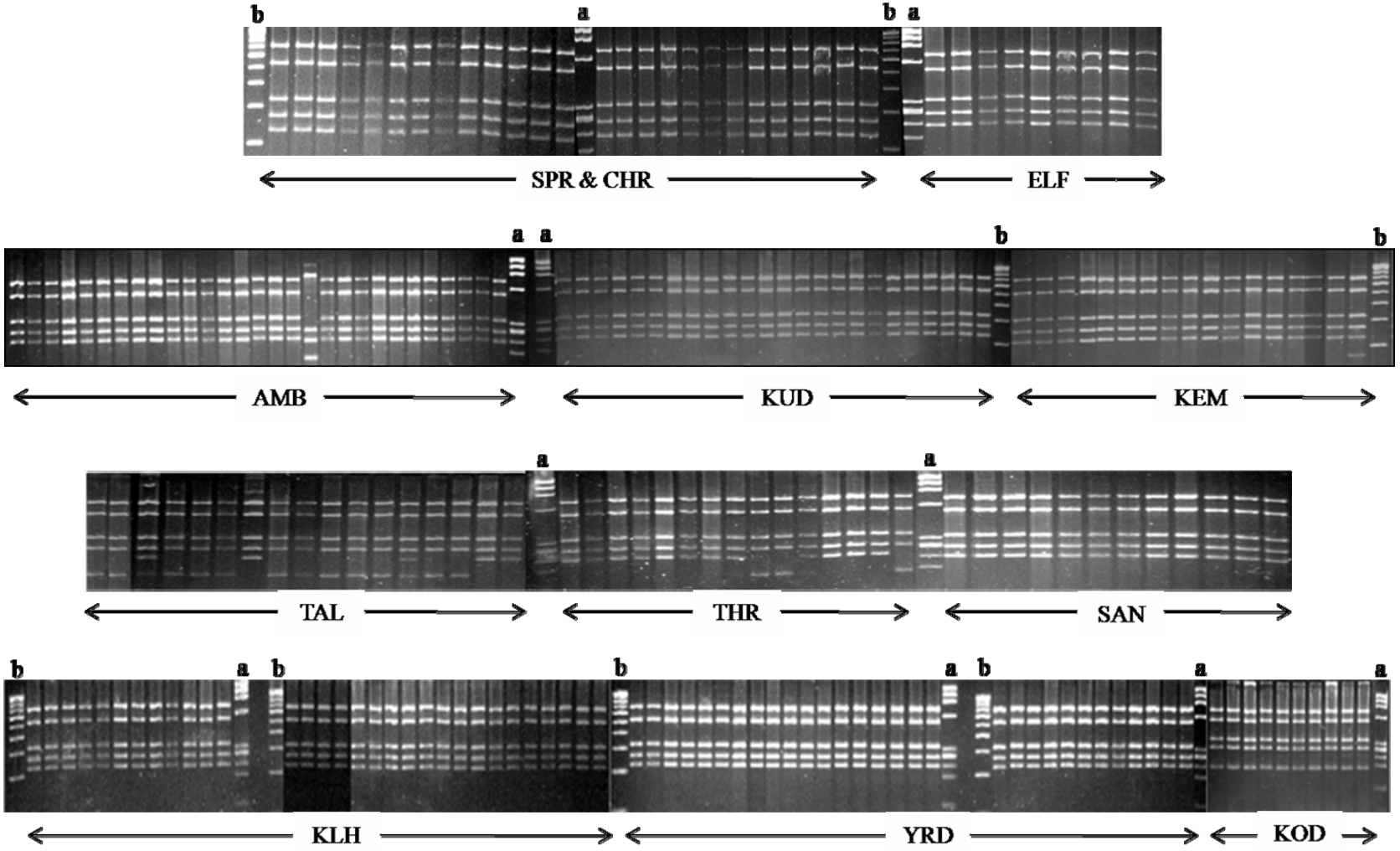


Fig. 2.A.6 Restriction pattern in *S. laurina* using primer pair *nad*4/1-*nad*4/2 and *Msp*I restriction enzyme

Lanes with molecular weight markers are marked as a and b above them indicate ØX174/*Hae*III digest and 100 bp ladder, respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

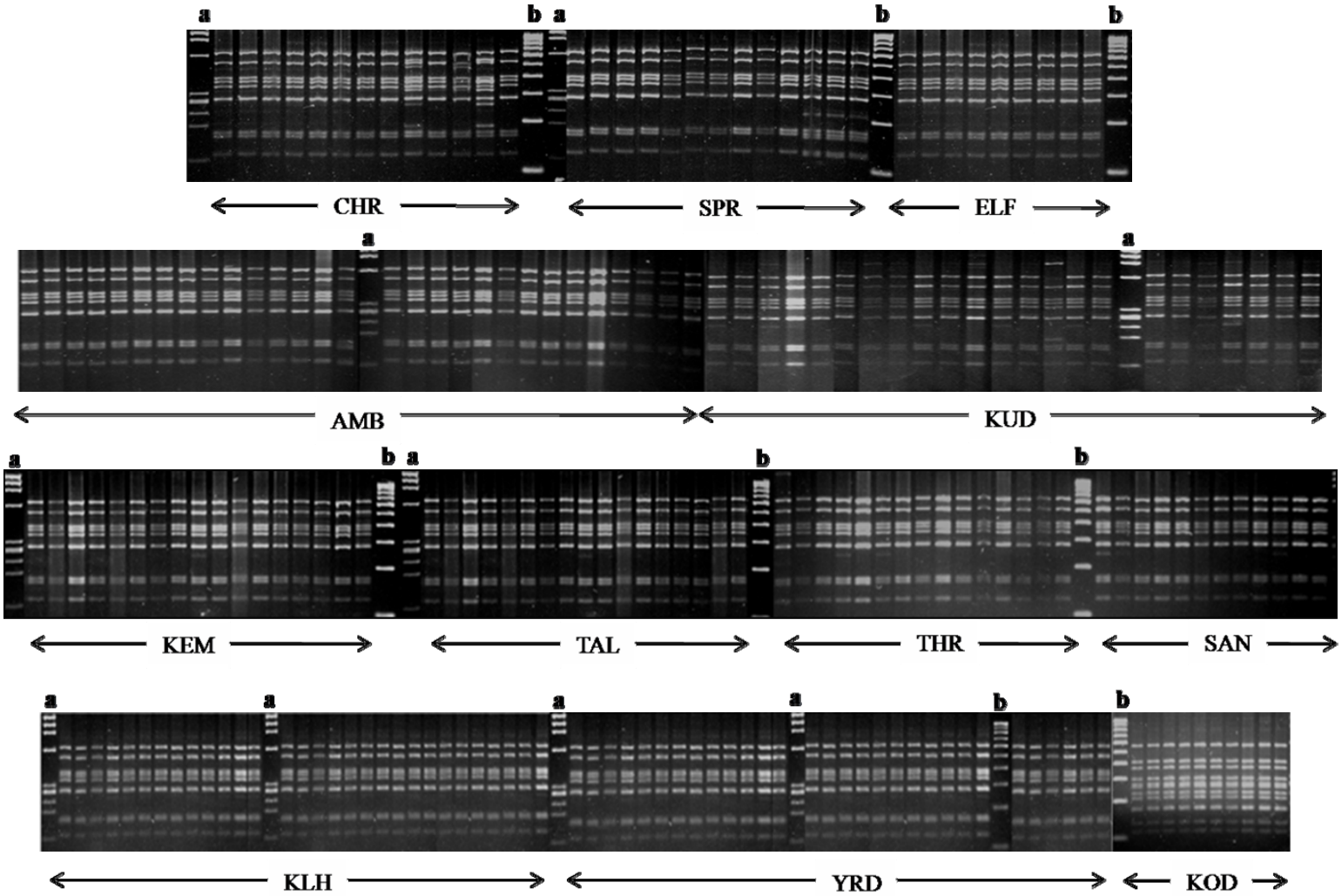


Fig. 2.A.7 Restriction pattern in *S. laurina* using primer pair *nad*4/2-*nad*4/4 and *Rsa*I restriction enzyme

Lanes with molecular weight markers are marked as a and b above them indicate ØX174/*Hae*III digest and 100 bp ladder, respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

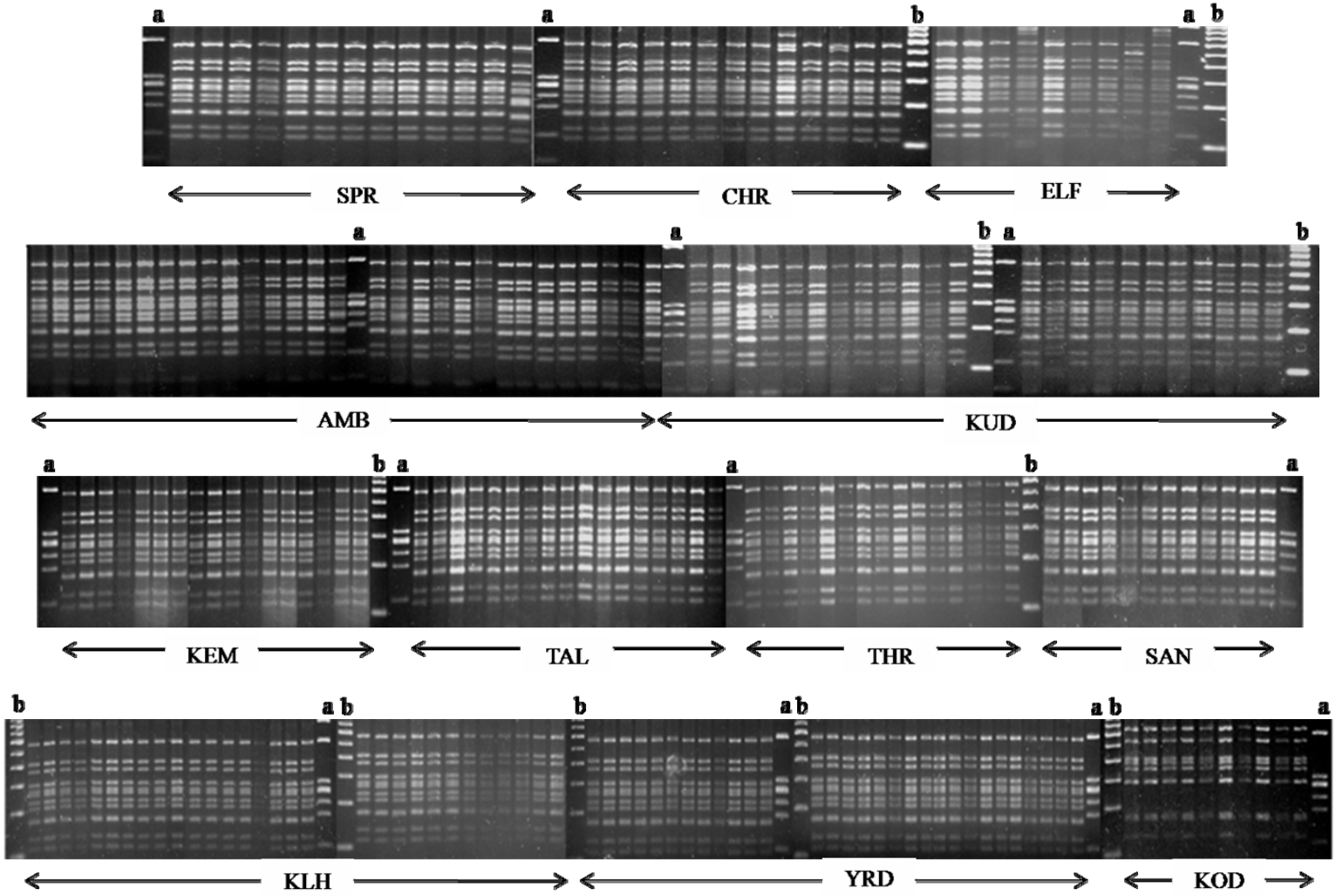


Fig. 2.A.8 Restriction pattern in *S. laurina* using primer pair *nad*4/2-*nad*4/4 and *Msp*I restriction enzyme

Lanes with molecular weight markers are marked as a and b above them indicate ØX174/*Hae*III digest and 100 bp ladder, respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

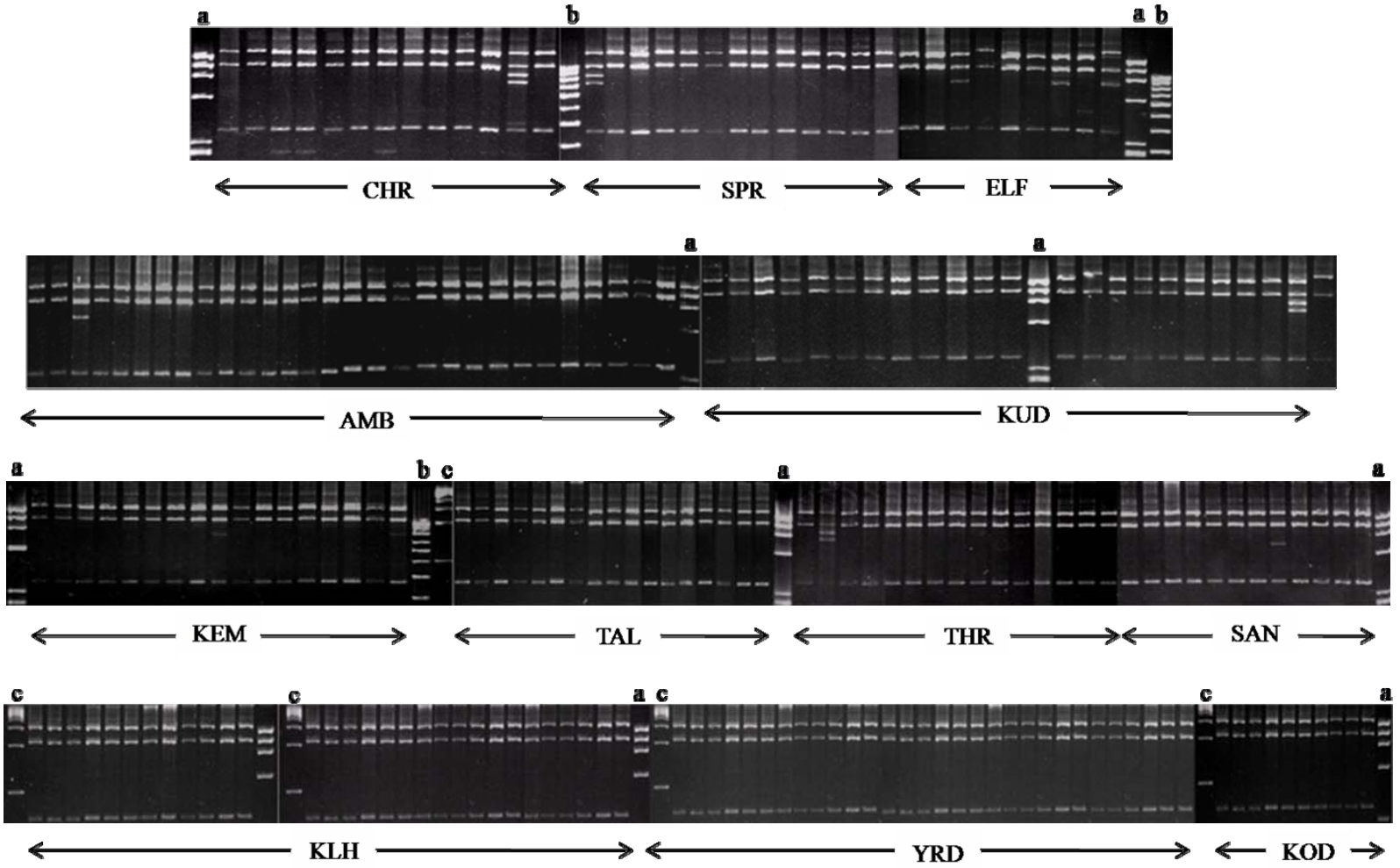


Fig. 2.A.9 Restriction pattern in *S. laurina* using primer pair *nad*4/3-*nad*4/4r and *Hind*III restriction enzyme

Lanes with molecular weight markers are marked as a, b and c above them indicate, ØX174/*Hae*III digest, 100 bp ladder and 500 bp ladder, respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

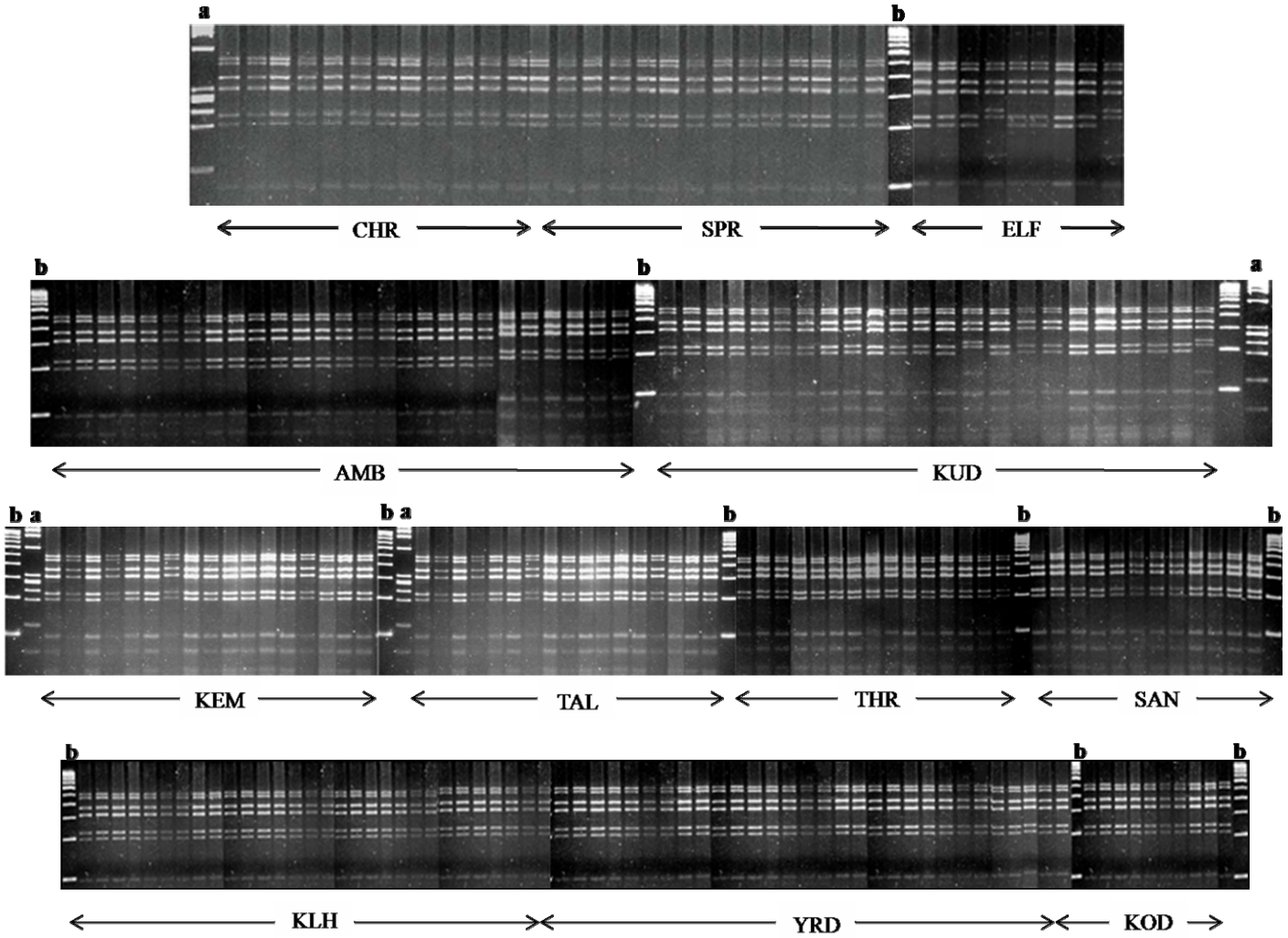


Fig. 2.A.10 Restriction pattern in *S. laurina* using primer pair *nad*4/3-*nad*4/4r and *Taq*I restriction enzyme

Lanes with molecular weight markers are marked as a and b above them indicate ØX174/*Hae*III digest and 100 bp ladder, respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

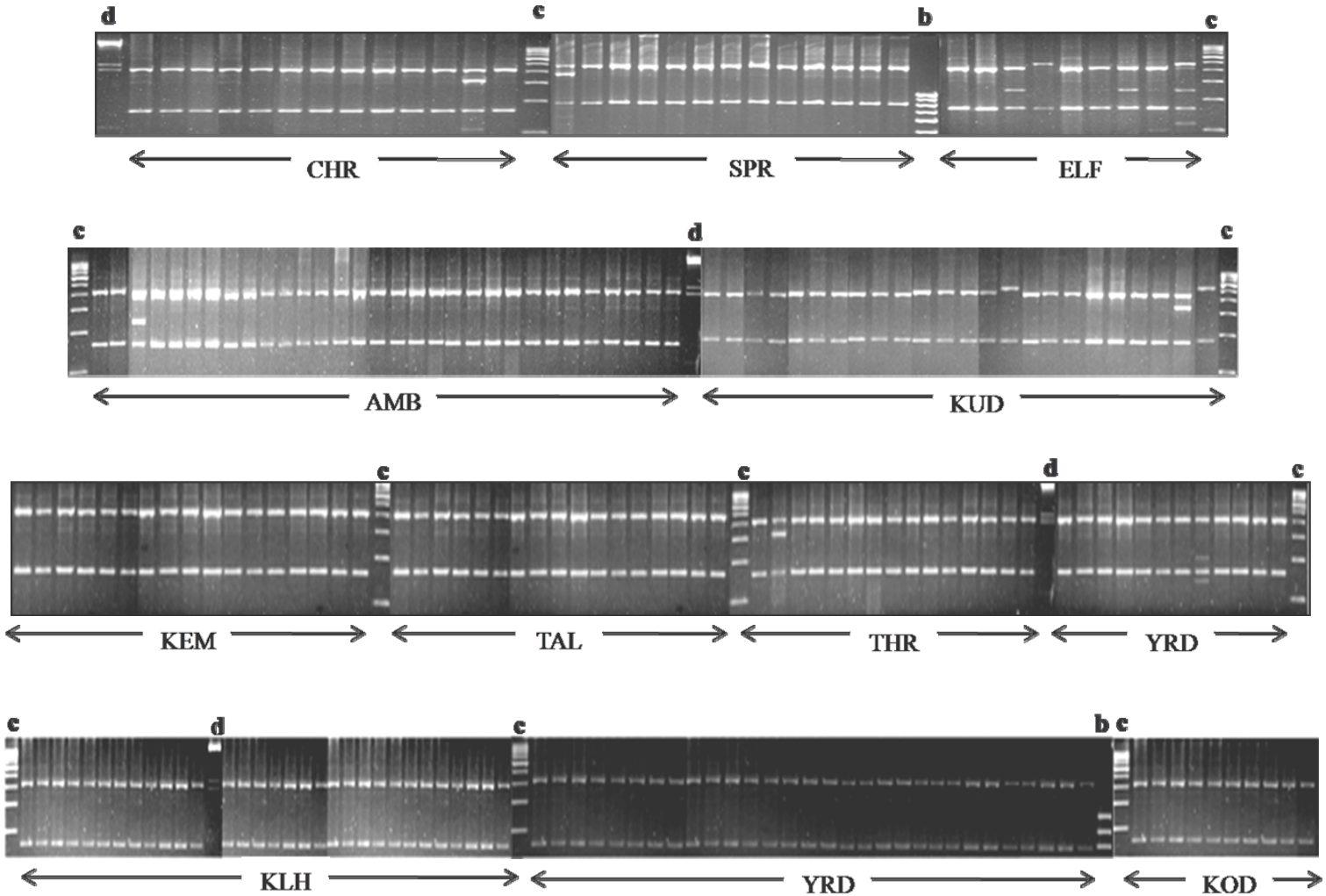


Fig. 2.A.11 Restriction pattern in *S. laurina* using primer pair *nad*4/3-*nad*4/4r and *Xba*I restriction enzyme

Lanes with molecular weight markers are marked as b, c and d above them indicate 100 bp ladder, 500 bp ladder and λ/*Hind*III respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

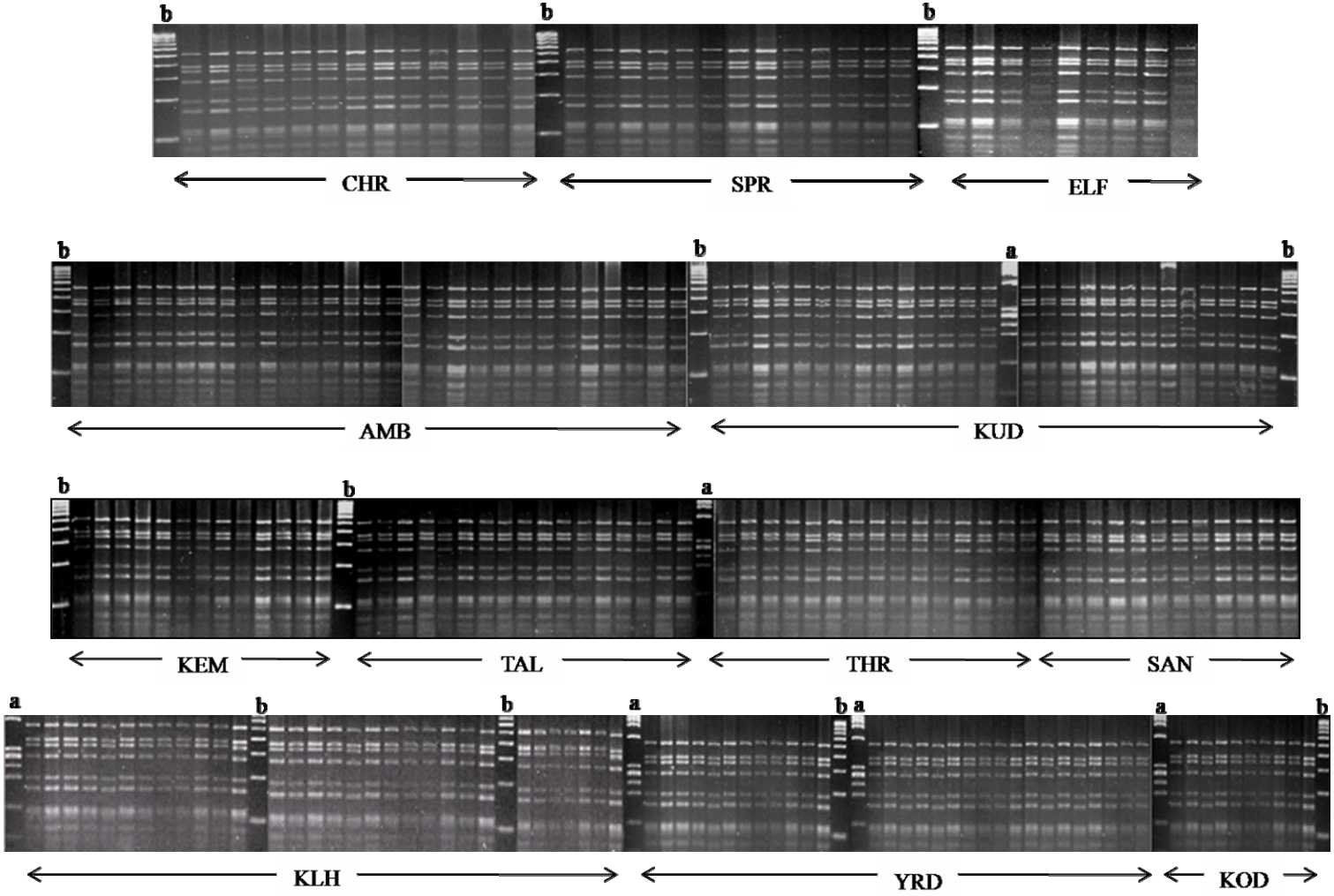


Fig. 2.A.12 Restriction pattern in *S. laurina* using primer pair *nad*4/3-*nad*4/4r and *Hae*III restriction enzyme

Lanes with molecular weight markers are marked as a and b above them indicate ØX174/*Hae*III digest and 100 bp ladder, respectively. Labels below the gel picture represent loading order of samples from the 12 populations.

Table 2.A.3 Distribution of PCR-RFLP chlorotypes in 12 populations of *S. laurina*

Chlorotypes	Populations												Total ($N^{\#}$)	Frequency
	Northeast India			Western Ghats						Eastern Ghats				
	SPR	CHR	ELF	AMB	KUD	KEM	TAL	THR	SAN	KOD	KLH	YRD		
CH1	13	13	7	14	19	17	3	11	12	-	-	-	109	0.587
CH2	-	-	-	3	2	-	-	1	-	-	-	-	5	0.016
CH3	-	-	-	-	**2**	-	-	-	-	-	-	-	2	0.007
CH4	-	-	-	-	-	-	**14**	-	-	-	-	-	14	0.069
CH5	-	-	-	-	-	-	-	**2**	-	-	-	-	2	0.012
CH6	-	-	-	-	-	-	-	**1**	-	-	-	-	1	0.006
CH7	-	-	2	-	-	-	-	-	-	-	30	30	62	0.186
CH8	-	-	-	-	-	-	-	-	-	**10**	-	-	10	0.083
CH9	-	-	-	**12**	-	-	-	-	-	-	-	-	12	0.035
Total	13	13	9	29	23	17	17	15	12	10	30	30	218	
n*	1	1	2	3	3	1	2	4	1	1	1	1		

n* - Number of chlorotype/s per population, $\mathbf{N}^{\#}$ - Total number of samples analysed per population. Private chlorotypes have been indicated in bold text.

Table 2.A.4 Distribution of PCR-RFLP mitotypes in 12 populations of *S. laurina*

	Northeast India			Western Ghats						Eastern Ghats				
Mitotypes	**SPR**	**CHR**	**ELF**	**AMB**	**KUD**	**KEM**	**TAL**	**THR**	**SAN**	**KOD**	**KLH**	**YRD**	**Total ($N^{\#}$)**	**Frequency**
MH1	-	-	-	26	20	16	3	13	10	-	-	-	88	0.382
MH2	-	-	-	-	**1**	-	-	-	-	-	-	-	1	0.004
MH3	-	-	-	-	**1**	-	-	-	-	-	-	-	1	0.004
MH4	-	-	-	-	**1**	-	-	-	-	-	-	-	1	0.004
MH5	-	1	1	-	-	1	2	-	-	-	-	-	5	0.030
MH6	-	-	-	-	-	-	**12**	-	-	-	-	-	12	0.059
MH7	-	-	-	-	-	-	-	**1**	-	-	-	-	1	0.006
MH8	-	-	-	-	-	-	-	-	**2**	-	-	-	2	0.014
MH9	-	-	-	-	-	-	-	-	-	-	30	30	60	0.167
MH10	-	-	-	-	-	-	-	-	-	**10**	-	-	10	0.083
MH11	8	-	3	-	-	-	-	-	-	-	-	-	11	0.079
MH12	**2**	-	-	-	-	-	-	-	-	-	-	-	2	0.013
MH13	1	-	-	-	-	-	-	1	-	-	-	-	2	0.012
MH14	**1**	-	-	-	-	-	-	-	-	-	-	-	1	0.007
MH15	**1**	-	-	-	-	-	-	-	-	-	-	-	1	0.007
MH16	-	**3**	-	-	-	-	-	-	-	-	-	-	3	0.019
MH17	-	9	1	-	-	-	-	-	-	-	-	-	10	0.067
MH18	-	-	**1**	-	-	-	-	-	-	-	-	-	1	0.010
MH19	-	-	**1**	-	-	-	-	-	-	-	-	-	1	0.010
MH20	-	-	**1**	-	-	-	-	-	-	-	-	-	1	0.010
MH21	-	-	**1**	-	-	-	-	-	-	-	-	-	1	0.010
MH22	-	-	-	**1**	-	-	-	-	-	-	-	-	1	0.003
MH23	-	-	-	**1**	-	-	-	-	-	-	-	-	1	0.003
MH24	-	-	-	**1**	-	-	-	-	-	-	-	-	1	0.003
Total	13	13	9	29	23	17	17	15	12	10	30	30	218	
n*	5	3	7	4	4	2	3	3	2	1	1	1		

n*- Number of chlorotype per population, **$N^{\#}$** - Total number of samples analysed per population. Private mitotypes have been indicated in bold text.

The geographical distribution of chlorotypes was found to be non-random and none of the chlorotypes was shared by all the populations (Table 2.A.3). The NEI was represented by two chlorotypes while WG by seven and EG by two. Only two chlorotypes (CH7 and CH8) were obtained from EG even though sample size was maximum in these populations, while six private chlorotypes (CH3, CH4, CH5, CH6, CH8 and CH9) restricted to SI (WG + EG) were detected. Five populations (ELF, AMB, KUD, TAL and THR) revealed more than one chlorotype, while others were fixed for one haplotype only. The CH1 chlorotype, most frequent over the whole range (frequency = 0.587), was found in most of the populations except KOD, KLH and YRD (populations from EG) while CH6 showed minimum frequency of 0.006. Surprisingly, in *S. laurina* the number of mitotypes were found to be very high. In case of distribution of mitotypes, the mitotype MH1 occurred with the highest frequency of 0.382 and was common to WG populations only (Table 2.A.4). The MH11 and MH17 mitotypes (frequency = 0.079 and 0.067, respectively) occurred frequently in NEI, whereas the EG populations (KOD, KLH and YRD) were fixed for two mitotypes (MH9 and MH10). NEI, WG and EG showed presence of twelve, nine and two mitotypes, respectively, while none of the mitotypes was shared by all the populations. Eighteen private mitotypes were detected of which eight were located in NEI, nine in WG and one in EG (Table 2.A.4).

2.A.3.3 Genetic structure and population differentiation in cp and mt DNA

Within population diversity for chlorotypes was observed to be low (h_S = 0.174, v_S = 0.147) whereas total diversity was high (h_T = 0.653, v_T = 0.655; Table 2.A.5). Similar to this trend mitotypes also showed low within population diversity (h_S = 0.305, v_S = 0.294) and a high total diversity (h_T = 0.850, v_T = 0.851) when calculated for all the populations taken together (Table 2.A.5).

Subdivision of the chlorotype dataset into two (NEI and SI) and three (NEI, WG and EG) regions, respectively showed low within population diversity with absence of diversity in EG populations (Table 2.A.6). However, the h_T was almost similar (0.148) to h_S in NEI populations but was high in SI population. Similar subdivision for mitotype dataset showed a different trend, where NEI exhibited high h_S and h_T, while SI population showed low h_S and high h_T, the EG populations did not exhibit any diversity at all (Table 2.A.6).

Table 2.A.5 Genetic diversity indices for 12 populations of *S. laurina* in India

	***h* type parameters**			***v* type parameters**		
	h_S	h_T	G_{ST}	v_S	v_T	N_{ST}
Chlorotypes	0.174 (±0.066)	0.653 (±0.112)	0.734 (±0.112)	0.147 (±0.055)	0.655 (±0.150)	0.776 (±0.085)
Mitotypes	0.305 (±0.082)	0.850 (±0.075)	0.642 (±0.089)	0.294 (±0.097)	0.851 (±0.124)	0.655 (±0.082)

Numbers in parenthesis indicate standard error, *h* type: ordered allele, *v* type: unordered allele.

The haplotype differentiation measured for all the populations of *S. laurina* for the chlorotype was relatively high (G_{ST} = 0.734) and was even slightly higher when the relationships between the chlorotypes were taken into account (Table 2.A.5). The regionwise separation showed the highest differentiation in EG region (Table 2.A.6). The value of N_{ST} showed similar trends as G_{ST} for regionwise divided dataset. Population differentiation values for the whole dataset for mitotypes showed relatively high G_{ST} and N_{ST} values (Table 2.A.5). The differentiation values for the divided dataset were low for NEI, however SI was found to possess high G_{ST} value (Table 2.A.6). It could be probably contributed by EG populations, as this region showed the highest differentiation for unordered alleles (G_{ST} = 1.000) as compared to WG (G_{ST} = 0.375). Overall the N_{ST} and G_{ST} values also indicated the same trend for the mitotypes. Although N_{ST} was slightly higher than G_{ST} for both the chlorotypes and mitotypes, permutation test indicated that the difference between N_{ST} and G_{ST} was insignificant at $P < 0.001$, implying the absence of phylogeographical structure at this scale (Pons and Petit 1996). Similar trends were detected for regionwise divided dataset for chlorotypes and mitotypes.

Table 2.A.6 Genetic diversity indices for *S. laurina* for different geographical regions

Genome	Regions	No. of haplotypes	No. of individuals	h_S	h_T	G_{ST}	N_{ST}
cpDNA	Northeast India (NEI)	2	35	0.130 (±0.130)	0.148 (±0.126)	0.125 (nc)	0.125 (nc)
	South India (SI:WG+EG)	9	183	0.189 (±0.080)	0.772 (±0.087)	0.756 (±0.112)	0.793 (±0.081)
	Western Ghats (WG)	7	113	0.283 (±0.999)	0.521 (±0.152)	0.457 (±0.136)	0.577 (±0.077)
	Eastern Ghats (EG)	2	70	0.000 (±0.000)	0.667 (±0.222)	1.000 (nc)	0.667 (±0.472)
mtDNA	Northeast India (NEI)	12	35	0.682 (±0.123)	0.903 (±0.033)	0.245 (±0.213)	0.172 (±0.124)
	South India (SI: WG+EG)	14	183	0.166 (±0.053)	0.727 (±0.104)	0.772 (±0.090)	0.481 (±0.179)
	Western Ghats (WG)	12	113	0.269 (±0.050)	0.430 (±0.153)	0.375 (nc)	0.258 (±0.092)
	Eastern Ghats (EG)	2	70	0.000 (±0.000)	0.667 (±0.222)	1.000 (nc)	1.000 (nc)

Numbers in parenthesis indicate standard error, nc - not calculated.

The pairwise genetic differentiation based on chlorotypes and mitotypes between two populations plotted against geographical distance showed that both G_{ST} and N_{ST} tend to vary irregularly with geographical distance (Fig. 2.A.13), which implies that the differentiation in *S. laurina* is independent of geographical distance. The highest differentiation for cpDNA for both G_{ST} and N_{ST} was obtained in the distance class of 200-300 km, while for mtDNA it was the distance class of 200-300 km for G_{ST} and 400-500 km for N_{ST}.

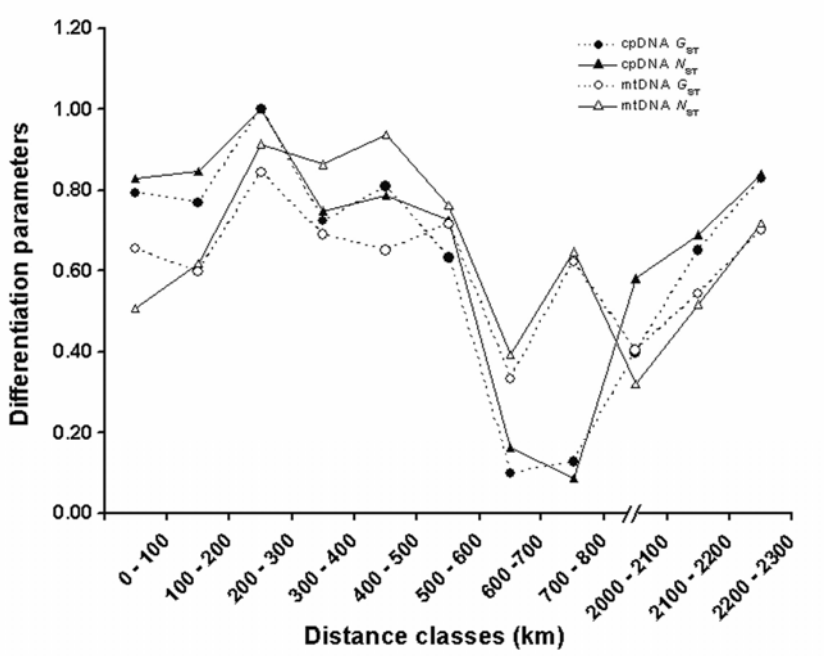


Fig. 2.A.13 Relationship of average pairwise G_{ST} and N_{ST} values with the geographical distance separating populations for cpDNA and mtDNA haplotypes

2.A.3.4 Pattern of partitioning of variation and phylogenetic relationship between the chlorotypes and mitotypes

A hierarchical AMOVA of the chlorotypes with populations grouped into three regions depicted that 45.83% of the variation partitioned among regions, 39.3% among populations within regions and the remaining 14.87% variation within populations, all partitions were significant (Table 2.A.7). The fixation index (F_{ST} = 0.851) for the whole data set was very high. The hierarchical AMOVA for mitotypes with all the populations divided into three regions showed that 89.94% of mtDNA

Table 2.A.7 Analysis of molecular variance (AMOVA) for the whole range of *S. laurina* populations from three geographical regions using cpDNA and mtDNA PCR-RFLP data

	Region analysed and source of variation	d.f.	SS	Variance components	Percentage of variation	
cpDNA	Among regions	2	98.811	0.599	45.83*	F_{ST} = 0.851
	Among populations within regions	9	83.144	0.520	39.30*	
	Within populations	206	40.059	0.195	14.87*	
	Total	217	222.014	1.308		
mtDNA	Among regions	2	830.187	6.246	89.94*	F_{ST} = 0.930
	Among populations within regions	9	41.465	0.236	3.39*	
	Within populations	206	95.467	0.463	6.67*	
	Total	217	967.119	6.945		

d.f., degrees of freedom; SS, sum of squares; *$P < 0.001$.

variation partitioned among regions, 3.39% among populations within regions and the rest 6.67% within populations (Table 2.A.7). The partitioning of the variance was significant. The fixation index (F_{ST}) for the whole data set was also very high for the mitotypes (F_{ST} = 0.93).

The relationships between the chlorotypes are shown in Fig. 2.A.14. The chlorotype CH1 with wide distribution formed the internal node, while the private haplotypes CH3, CH4, CH8 and CH9 occupied the edges of the MST. The chlorotype CH8, separated by three mutational events from CH7 which in turn was separated by two mutational events from CH1, formed a separate group within the cpDNA MST having distinct geographical distribution (EG restricted) with its link with ELF population. The MSN for the mitotypes further supported the phylogenetic distinctiveness of EG populations as shown in Fig. 2.A.15. The mitotype MH1 forming the internal node in the MSN was connected to other mitotypes. The mitotype MH9 from EG separated from MH1 by 17 mutational steps along with MH10 forming a distinct group as also observed with cpDNA MST. The private mitotypes (MH2, MH4, MH10, MH18, MH21, and MH24) formed the external nodes. Three alternative links were found, between the haplotype MH17 (from ELF and CHR) and MH16 (from CHR) separated by one mutational event, MH19 (ELF) and MH22 (AMB) separated by three mutational events and between MH14 (from SPR) and MH20 (from ELF) separated by one mutational event.

2.A.3.5 Association between chloroplast and mitochondrial genomes in *S. laurina*

In the absence of any specific record on inheritance of chloroplast and mitochondrial DNA in *S. laurina*, they were assumed to be maternally inherited as described for most angiosperms. Maternally inherited genomes are known to be inherited together and are expected to behave as one unit, however in *S. laurina* no distinct association was observed between the chlorotypes and mitotypes. Moreover, they showed regionwise association (Table 2.A.8). The chlorotype CH1 was associated with 17 mitotypes. The CH1-MH1 association was the most frequent association; however, it was represented only in the WG. CH1-MH5 and CH2-MH1 were the second most predominant associations present in four and three populations, respectively. Two

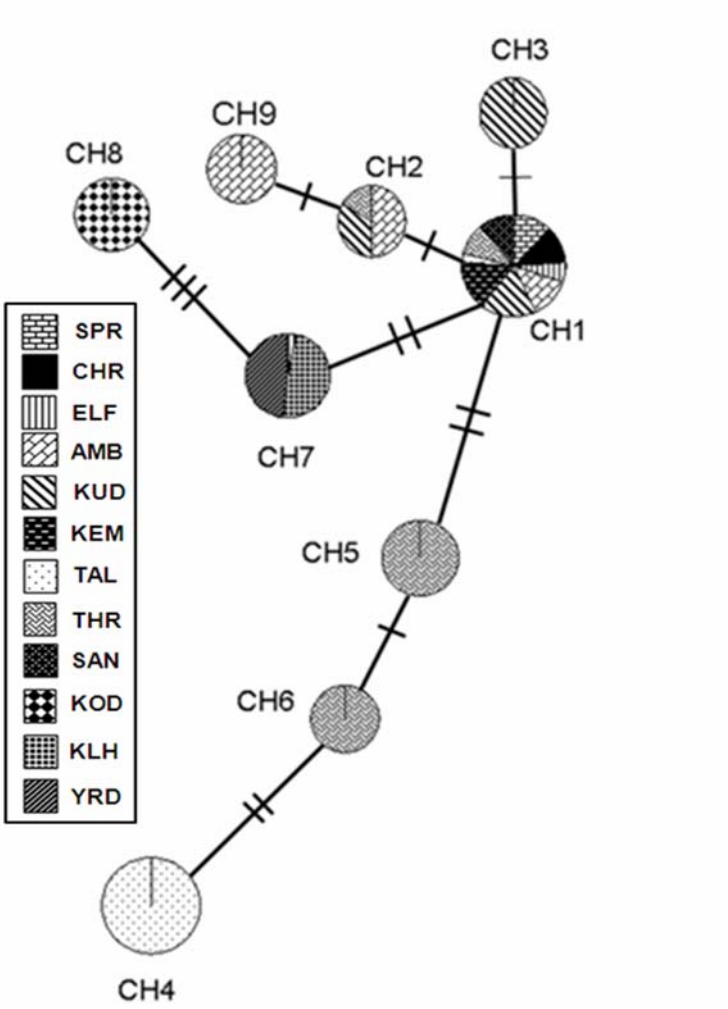


Fig. 2.A.14 Minimum Spanning Tree of *S. laurina* showing relationship between the chlorotypes

The circles in the figure indicate the chlorotype. The shades within the circle indicate the contribution of each population to the chlorotype and size of the circles are proportional to the frequency of the chlorotype. The hatching lines within the lines connecting the circles indicate the mutations involved.

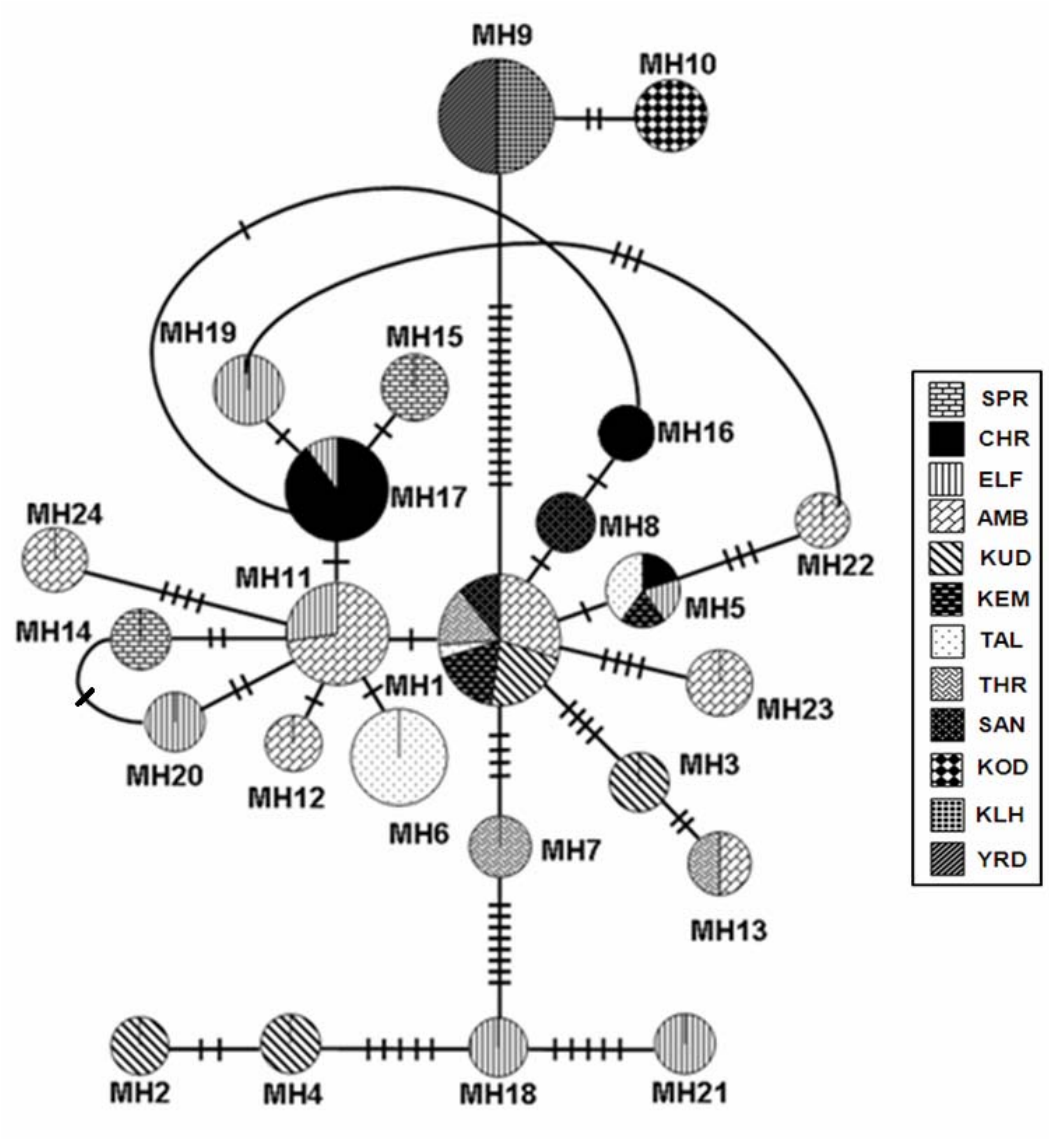


Fig. 2.A.15 Minimum Spanning Network of *S. laurina* showing relationship between the mitotypes

The circles in the figure indicate the type of mitotype. The shades within the circle indicate the contribution of each population to the mitotype and the size of circles are proportional to the frequency of the mitotype. The hatching lines within the lines connecting the circles indicate the mutations involved. The curved lines indicate alternative routes.

Table 2.A.8 Association between chlorotypes and mitotypes in *S. laurina* from India

	Populations	Chlorotypes identified	Mitotypes identified	Chlorotype-mitotype associations (No. of individuals)
1	SPR	1	11, 12, 13, 14, 15	CH1[MH11(8), MH12(2), MH13(1), MH14(1), MH15(1)]
2	CHR	1	5, 16, 17	CH1[MH5(1), MH16(3), MH17(19)]
3	ELF	1, 7	5, 11, 17, 18, 19, 20, 21	CH1[MH5(1), MH11(2), MH17(1), MH19(1), MH20(1), MH21(2)]; CH7[MH18(1)]
4	AMB	1, 2, 9	1, 22, 23, 24	CH1[MH1(13), MH24 (1)]; CH2[MH1(2), MH23(1)]; CH9[MH1(11), MH22(1)]
5	KUD	1, 2, 3	1, 2, 3, 4, 23	CH1[MH1(17), MH2(1), MH4(1)]; CH2[MH1(1),MH3(1)];CH3[MH1(2)]
6	KEM	1	16, 5	CH1[MH1(16), MH5(1)]
7	TAL	1, 4	1, 5, 6	CH1[MH1(2), MH5(1)]; CH4[MH(1), MH5(1), MH6(12)]
8	THR	1, 2, 5, 6	1, 7, 13	CH1[MH1(10), MH7(1)] ; CH2[MH1(1)]; CH5[MH1(2)]; CH6[MH13(1)]
9	SAN	1	1, 8	CH1[MH1(10), MH8(2)]
10	KOD	8	10	CH8[MH10(10)]
11	KLH	7	9	CH7[MH9(30)]
12	YRD	7	9	CH7[MH9(30)]

unique cases were observed of which CH8-MH10 was restricted to only KOD population, while CH7-MH9 was restricted to KLH and YRD populations. Many associations obtained were restricted to a particular population like CH6-MH13, CH1-MH7, CH5-MH1 to THR, CH1-MH12, MH13, MH14 and MH15 to SPR, CH1-MH19, MH20, MH21 and CH7-MH18 to ELF and many others (Table 2.A.8). It was noteworthy that in NEI no cp-mt haplotype association was common. SPR and ELF showed one common association (CH1-MH11) while CHR and ELF showed two common associations (CH1-MH5 and CH1-MH17). In NEI, many mitotypes were available to associate with two chlorotypes, however it was opposite in case of WG. Out of six cases of uncoupling (CH2-MH3, CH4-MH5, CH6- MH13, CH9-MH22, CH2-MH23 and CH7-MH18), five were restricted to WG population. The uncoupling might have resulted from mutations either in the chlorotypes or in the mitotypes. For example, CH2-MH3 uncoupling could have arisen by mutation of MH3 from MH1, which was found to be true as they were separated from each other by four mutation steps as shown in Fig. 2.A.5.

2.A.4 Discussion

2.A.4.1 Genetic diversity and differentiation within *S. laurina* in India using cytoplasmic markers

The chloroplast genome of *S. laurina* was found to possess a moderate level of haplotype diversity (nine chlorotypes in 218 plants) similar to levels observed in other tree species, *e.g*. *Alnus glutinosa* (13 haplotypes, King and Ferris 1998). The haplotype diversity level in mtDNA in *S. laurina* (24 mitotypes in 218 plants) was similar to three other flowering species *viz*., *Hevea brasiliensis* (212 mtDNA RFLP variants in 395 accessions screened; Luo et al. 1995), *Thymus vulgaris* (50 mitotypes from 400 plants; Manicacci et al. 1996) and *Beta vulgaris* ssp. maritime (20 mtDNA haplotypes from 414 plants; Desplanque et al. 2000). Genetic variation in *S. laurina* in both the cp and mt DNA was found to exist among populations rather than within populations as depicted by low diversity value for the whole dataset as well as the subdivided dataset. The high total diversity was a result of among population difference which is shown by the partitioning of variance among the regions (45.83% for cpDNA and 89.74% for mtDNA for the whole data set). The EG populations were

fixed for the chlorotype (CH7, CH8) and the mitotype (MH9, MH10) respectively, depicting lack of diversity in *S. laurina* from the EG region.

2.A.4.2 Population structure of *S. laurina* in India

For many models of population structure, it has been shown theoretically that the level of genetic differentiation among populations is higher for maternally inherited markers than the nuclear genes (Birky et al. 1989; Petit et al. 1993; Hu and Ennos 1997). The level of differentiation for *S. laurina* populations in present study for all the regions taken together was high for cpDNA (G_{ST} = 0.734 and N_{ST} = 0.776) and mtDNA markers (G_{ST} = 0.642 and N_{ST} = 0.655). Such high values of G_{ST} have been reported for cpDNA in *Argania spinosa* (G_{ST} = 0.60) (El Mousadik and Petit 1996); in *Hedera* spp. (G_{ST} = 0.642 and N_{ST} = 0.827) (Grivet and Petit 2002) and in *Frangula alnus* (G_{ST} = 0.81 and N_{ST} = 0.91) (Hampe et al. 2003) and for mtDNA in *Pinus flexilis* (G_{ST} = 0.68) (Latta and Mitton 1997) and in *Picea mariana* (G_{ST} = 0.671 and N_{ST} = 0.726) (Jaramillo-Correa et al. 2004). The higher differentiation in EG as compared to the rest of the populations probably contributes to the trend seen for G_{ST} and N_{ST} in *S. laurina* (Table 2.A.5 and 2.A.6). In spite of high differentiation among the regions, insignificant difference between G_{ST} and N_{ST} showed that the populations are phylogenetically equivalent and phylogeographically unstructured (Pons and Petit 1996).

The analysis of cpDNA and mtDNA haplotypes and their relationships revealed that the three regions are occupied by two different groups: one Northeast-Western Ghat group (NEI-WG group present in the populations belonging to NEI and WG) and the other Eastern Ghat group (EG group, found only in EG) (Figs. 2.A.14 and 2.A.15). The phylogenetic affinities of the groups are irrespective of the geographical distances that exist between them. However, the absence of dominant haplotype of WG (MH1, assumed to be ancient haplotype from its central position in MSN) from NEI poses an important question on the rate of mutation accumulated by mitochondrial genome in *S. laurina*.

In most of the angiosperm species, chloroplast and mitochondrial DNA are maternally inherited (Reboud and Zeyl 1994). As a consequence, the two organellar genomes should remain associated; as if completely linked (Schnabel and Asmussen 1989) and are expected to give similar information on dispersal and gene flow by

seeds. However, such association between the two genomes has not been explicitly studied, except for a few reports (Dumolin-Lapègue et al. 1998; Mohanty et al. 2003). In the present study, *S. laurina* showed low chloroplast diversity in NEI, whereas it was high in WG while, KLH, YRD and KOD were fixed for one type. This was reverse in case of mitotypes, where NEI showed many different mitotypes which were unique to it, resulting in associations confined to NEI only. Availability of a number of chlorotypes and mitotypes in WG produced many unique associations restricted to WG (Table 2.A.8). The geographical distribution of haplotypes and especially the localization of unique cpDNA and mtDNA combination in two populations of AMB and KUD of WG region and a few associations represented by only one or two plants explain the evolution of these mutations at local scale in the chloroplast genome of *S. laurina*. The absence of association, CH1-MH1 from NEI and EG populations indicates that these populations are evolving independently in their respective location and have lost the ancestral association type. It appears that recurrent mutation events, rather than intragenic recombination and/or gene conversion, generated the dissociations in our data. Moreover, each genome is subject to its own unique history of stochastic events; hence are expected to have independent genealogical histories and independent haplotype trees.

2.A.4.3 *S. laurina* population structure: postulations on postglacial migration/ dispersal hypothesis

To understand the present day distribution of *S. laurina*, it is imperative to take into consideration the glacial cycles during the Quaternary that played an important role in the assembly of communities and the distribution of populations within species (Hewitt 2000). According to De Terra and Paterson (1939) the glacial phases of Pleistocene could have descended to an altitude of 4000-5000 feet. Thus glaciation being restricted to Himalayan region was not felt in SI and SriLanka, but this could have lowered the mean annual temperature by 5°C to 7°C (Rajguru 1969). The lowered temperature and evaporation might have increased the pressure for migration on species existing close to Western Himalayas such as NEI as well as species existing in SI and SriLanka. Although there are no direct evidences for the presence of *S. laurina* fossils from WG, paleo-palynological studies in Nilgiris (WG) reveal colonization of the area by *Rhododendron* spp. (which is observed to be an associate of *S. laurina* in WG) about 35,000 years ago in the late Pleistocene (Vishnu-

Mittre and Gupta 1970). *S. laurina* is often a pioneer species that invades shrubby thickets to form the shola forest (Meher-Homji 1967), thus is subjected to extinction/recolonization dynamics. The present day population structure of *S. laurina* could be an amalgamation of many historic processes, shaped variously by environmental, topographial and climatic factors.

Based on cp and mt DNA data and the theories put forth by different authors, two conditions-dispersal (postglacial migration) or vicariance can be hypothesized to elucidate the underlining events. The existence of chlorotype CH1 and CH7 in NEI and the mitotypes MH5 and MH13 in SI acting as linking haplotypes, indicate the migration of *S. laurina* from NEI to SI. Such overlap in genotype distribution is predicted under the condition of dispersal (Hewitt 1996). The absence of geographical structuring at a broad scale but the local differentiation of EG with private haplotypes and low genetic diversity (Figs. 2.A.14 and 2.A.15), indicate that the populations of *S. laurina* of EG might have undergone many bottlenecks having been highly fragmented for several years.

The long distance involved in the migration from NEI to SI seems difficult to understand as an explanation for direct migration of flora and fauna from NEI to SI. Moreover, the evidences like the existence of pollen of *Symplocos* spp. in central India obtained in Dongar-Sabar swamp in central India (Chauhan 2002) along with the palynological studies from the Chhota Nagpur plains of central India by Bande and Chandra (1990) suggest the existence of evergreen forest in the central Indian plains in past. Additionally, the existence of wet temperate climate in central India at that period as suggested by the Brij hypothesis might have played a significant role (but till now undefined) and favoured the migration of *Symplocos* spp. from NEI to central India. The existence of a transitional refugium in central India from which, *Symplocos* spp. may have invaded the WG and EG separately by the routes postulated by Hora (1949), seems a tenable explanation for the large differences between WG and EG. However, in absence of *S. laurina* in other parts of EG especially in the Orissa and Andhra Pradesh region and in central India, the hypothesis does not give a definite elaboration and thus needs to be scrutinized stringently. There are many observations in the data which do not support the above interpretations. The high genetic differentiation observed in the present data remains partially answered if the above hypotheses are taken into consideration. The complete differentiation between

NEI and WG populations at mtDNA, contrary to cpDNA data makes the interpretation more complex. The postglacial creation of high diversity especially in WG and NEI and spatial patterns in a rather limited number of generations (species being a long living with long generation cycles: even though it has pioneering characteristics) is also difficult to decipher. Secondly, inspite of having an endozoochorous seed dispersal system, surprisingly high G_{ST} and N_{ST} points towards isolation of regions from a long time rather than contact. Moreover, the absence of assumed ancient mitotype, MH1 from NEI doesn't seem to support the prediction that plants have migrated from NEI to SI and the absence of populations from central India does not allow us to make a definitive comment on its role on the postglacial migration/expansion of *S. laurina*. However, the presence of connecting haplotypes such as CH7, MH5 and MH13 in some populations of NEI and SI might be explained on the basis of homoplasy. This is possible as the haplotypes involved are present in very low frequency.

2.A.4.4 Evidences for vicariance as an explanation for the present day population structure of *S. laurina*

The high genetic diversity and among population divergence obtained in the present study reflect the existence of *S. laurina* populations in NEI, WG and EG regions long before the formation of central arid zone, complying with the second hypothesis namely vicariance model. The high diversity in NEI and SI could have accumulated either from newly evolved alleles and/or from other sources that join the population. An area showing higher genetic diversity could imply that it could have been a refugium; an area with a stable ecological habitat during the fluctuation of environment that led to the accumulation of genetic diversity (Tzedakis et al. 2002). In our case the three regions probably represent the putative glacial refugia for *S. laurina* which arose due to climatic oscillations during Quaternary and probably enhanced fragmentation contributing further in isolation of these putative refugia. The diversity present in the NEI and WG could be a result of faster recolonization events from a source population (refugium) through a stepping stone mode. As the NEI and WG populations are from the regions identified as two biodiversity hotspots, they might be the center of diversity for the plant under study. The high genetic variation coupled with genetic differentiation and partitioning of variation among regions suggests that NEI and WG regions harboured a sufficient number of populations of *S.*

laurina to maintain genetic variation throughout the Quaternary climatic changes. The contemporary populations of *S. laurina* in India thus might have descended from the remnant populations that survived in these glacial refugia during cold periods of the Pleistocene. The existence of linking chlorotype (CH1) between NEI and WG can be accounted by possible contacts between the populations in past. It is likely that ancestral haplotype CH1 became fixed by genetic drift in NEI and WG refugial populations.

The existence of two lineages in SI within a distance of 300 km suggests the presence of two major source populations for *S. laurina* during glaciation and reveals that the natural barriers were not overcome to allow seed exchange, even though WG and EG are predicted to be more or less linked by Biligirirangan hills; the link is supposed to be near Nilgiri hills, to which the populations SAN and THR belong (Legris and Meher-Homji 1984). As suggested by Hewitt (1996) refugial expansions encounter mountains, rivers and other barriers with various permeabilities. The absence of WG haplotypes from EG, although they are geographically close, indicate that topographical components are more influensive than geographical barriers.

The altitudinal range shift represents another important factor for the maintenance of tree species against climatic changes. The cooler climate during glaciation might have led to colonization of open areas (low lying areas like river valleys and ravines close to the mountains with more humid climates) congenial for survival by the plant. As the three putative refugial areas for *S. laurina* are associated with different orographical systems, range expansion into high altitude areas (mostly hilly areas) would have occurred during interglacial period as a response to temperature changes and increase in aridity. As montane vegetations are known to be maintained by harsh factors (such as strong wind, fire and frost) that inhibit shola forest formation (Agarwal et al. 1961) the persistence of these populations in the refugium indicate that the populations of *S. laurina* must have tolerated these and other stresses with various efficacies.

The EG populations might have also served as refugium; although no evidence of refugia was found because of absence of diversity in these populations. The extant refugial population may have undergone a strong bottleneck caused by altitudinal shift (moving up and down the mountains) *via* range contraction and range expansion during glacial and interglacial period. This might have eliminated many populations

so that only recently established populations survived in locally suitable places (Hewitt 1996) and might have become homozygous at many loci, thus erasing the genetic diversity in EG populations. According to Hewitt (1999), population surviving on a small mountain is labile to undergo a series of bottlenecks. Whether montane taxa descended during cooling due to low temperature, high precipitation or some other factors such as lower CO_2 concentration of the glacial atmosphere are all plausible explanations for altitudinal movements in *S. laurina*. Similar altitudinal shifts have been reported in other tree species such as *Pinus longaeva* and *Picea mexicana* (Ledig et al. 2002) in northern hemisphere. Thus, various geographical, climatic and other phenomena such as disturbance by fire, slope rejuvenations would have played their role in historical distribution of *S. laurina*, establishing the current genetic structure.

Section B: Analysis of populations with chloroplast microsatellite markers

OUTLINE

2.B.1 Introduction

Rapid advances in molecular techniques and algorithms used to model population genetic processes have opened up new opportunities for understanding the evolutionary history of organisms. Detection of relatively rapidly mutating sites in the chloroplast genome has resulted in a surge of interest in tracing molecular lineages that help to uncover historical relationships among plant populations. In particular, chloroplast simple sequence repeat (cpSSR) loci have proven invaluable for their universal application that overcomes the high investment needed for development of nuclear microsatellites.

Chloroplast microsatellites typically consist of mononucleotide motifs that are repeated eight to 15 times. Level of polymorphism in cpSSRs is quite variable across loci and across species, and some loci have been found to be monomorphic in all the species studied (Li et al. 2002). There are two important features that differentiate chloroplast from nuclear microsatellites. First, chloroplasts are uniparentally inherited; some species have maternal inheritance of the chloroplast, others have paternal. This means that cpSSRs provide information for the lineages of only one of the sexes. Also, the chloroplast chromosome is a nonrecombinant molecule and therefore, all cpSSR loci are linked. The genotyping of cpSSRs will result in haplotypes that will be composed of the combination of alleles found at each cpSSR locus. Mutation rates for length variation in microsatellites have been found to be higher (10^{-2} - 10^{-6}) than point mutations rates (Li et al. 2002). In order to explain this difference, two kinds of mutational mechanism have been proposed; replication slippage (Tachida and Iizuka 1992) and recombination with out-of-phase aligning (Harding et al. 1992). Both the processes result in changes in the number of repeat units which is compatible with the observed size polymorphism of microsatellites. One consequence of these mutational mechanisms is that the same genetic state (*i.e.* number of repeats) may evolve in two different microsatellite lineages through independent mutational events, a phenomenon known as homoplasy.

Chloroplast SSRs (cpSSRs) have been used in population and systematic studies in a variety of species (Powell et al. 1996b; Provan et al. 1999a). Knowledge of mutation rates at SSR loci is important because they determine levels of variability within populations, hence greatly influence estimates of population structure. The

lower substitution rate of cpDNA compared to the nuclear genome has been documented (Wolfe et al. 1987) and mutational processes at simple repeat loci in the chloroplast genome also occur less frequently than those in the nuclear genome.

Conserved primers for the amplification of cpSSRs have been reported for conifers (Vendramin et al. 1996), gramineae (Provan et al. 2004) and dicotyledons (Weising and Gardner 1999), but it is among conifers, for studies of population genetics, that cpSSR markers have mainly been used (*e.g*. Cuenca et al. 2003; Fady et al. 2003; Gómez et al. 2003). Microsatellite loci are particularly useful for intraspecific genetic comparisons, but their high mutation rate makes them less useful for interspecific analyses. Chloroplast microsatellite markers (cpSSRs) have higher polymorphism and are now frequently used in phylogeographic analyses of forest tree (Palmé and Verdramin 2002; Collevatti et al. 2003; Grivet and Petit 2003) and other species (Grivet and Petit, 2002; Mengoni et al. 2001, 2003).

In the present study cpSSR markers have been utilized in order to study population structure of *S. laurina*. For ease of presentation and discussion, the terms 'locus' to refer to a cpSSR site (defined by the termini of a PCR primer pair) and 'alleles' to refer to length variants at a cpSSR site have been used.

2.B.2 Materials and methods

2.B.2.1 Plant material

A total of 195 plants were sampled from 14 locations from the two biodiversity hotspots, NEI and WG along with EG. The details of sampling are given in Table 2.B.1. The collection details are given in Section 2.A.2.1.

2.B.2.2 DNA extraction and identification of cpSSRs

Total genomic DNA was extracted from leaf material using modified CTAB method (Richards et al. 1994) as detailed in Section 2.A.3.2. A set of 10 cpSSRs representing various microsatellite regions in the chloroplast genome described by Weising and Gardner (1999), were selected for the present study. Initial screening was carried out using these primers with representative samples from each population. PCR amplifications were carried out using a PTC 200 thermal cycler (MJ Research, Waltham, Massachusetts, USA). Each reaction contained 0.2 mM of each dNTPs, 2.5 mM of $MgCl_2$, 0.2 μM of each primer, 10X reaction buffer, 40 ng of template DNA

and 1 unit of *Taq* polymerase (Bangalore Genei, Bangalore, India,) in a total volume of 25 µl. PCR profile was as follows: 5 min denaturation at 95°C followed by 25 cycles of 1 min denaturation at 94°C, 30 s - 1 min annealing at 50°C - 65°C (Table 2.B.2) and 1 min extension at 72°C, with a final extension step of 72°C for 5 min. After amplification, PCR products were mixed to a loading buffer. These samples were electrophoretically separated first on 3.5% MetaPhor agarose stained with ethidium bromide and simultaneously on a 6% native polyacrylamide gel and stained using silver nitrate (Creste et al. 2001).

Table 2.B.1 Details of the samples used in the cpSSR analysis

	Population	Regions	Sample sizes	Latitude	Longitude	Altitude (m)
1	Shillong Peak Road (SPR)	NEI	11	E 91.85528	N 25.53548	1903
2	Cherrapunji Road (CHR)	NEI	13	E 91.73666	N 25.33120	1661
3	Elephanta Falls (ELF)	NEI	6	E 91.82544	N 25.53368	1754
4	Amboli (AMB)	WG	27	E 74.06654	N 15.99816	825*
5	Kudremukh (KUD)	WG	17	E 75.18670	N 13.20770	836
6	Kemmangundi (KEM)	WG	14	E 75.74579	N 13.47867	1369
7	Talakaveri (TAL)	WG	10	E 75.48790	N 12.38866	1243
8	Thaishola Road (THR)	WG	16	E 76.62558	N 11.24087	1877
9	Sandynallah (SAN)	WG	12	E 76.06321	N 11.43310	2227#
10	Munnar (MUN)	WG	10	E 77.15291	N 11.81318	1606
11	Ponmudi (PON)	WG	3	E 77.74813	N 8.74813	1066
12	Kodaikanal (KOD)	EG	9	E 77.47947	N 10.23795	2131
13	Kolli hills (KLH)	EG	20	E 78.36096	N 11.30846	1303
14	Yercaud (YRD)	EG	27	E 78.22367	N 11.81318	1293

* Population showing the lowest altitude from where the plants were sampled,
# Population showing the highest altitude from where the plants were sampled.

Each PCR product was genotyped twice/thrice on gels to check the reproducibility. The results obtained in the initial screening are depicted in Table 2.B.2. The amplification patterns of the samples were variable; some primers showed many minor bands other than the intense product hence they were excluded from the study. The screening resulted in identification of four primers ccmp 3, ccmp 5, ccmp 6 and ccmp 10 that produced unambiguous products. After intial screening, 195 samples belonging to 14 populations were analyzed with these four primers.

2.B.2.3 Data analysis

Diversity and differentiation parameters used in this study included the average intrapopulation diversity (h_S), total diversity (h_T) and the differentiation for unordered alleles (G_{ST}) and for ordered alleles (N_{ST}). They were quantified according to Pons and Petit (1995 and 1996) using PERMUT ver. 2.0 (Burban et al. 1999). The following population genetic parameters were computed for each population using Arlequin ver. 3.11 (Excoffier et al. 2005): number of haplotypes (n_h), number of polymorphic loci (n_p) and Nei's unbiased haplotypic diversity (H_e; Nei 1987). A hierarchical analysis of variance was used to partition the total variance into covariance components consisting of within population variation, variation among populations within regions, and variation among regions. The significance of covariance components was tested using permutation tests (1000 permutations) at different levels (haplotypes among populations among regions, haplotypes among populations within regions and populations among regions). Only P-values lower than 0.05 were considered significant. Observed number of alleles (na), effective number of alleles (ne), Nei's (1973) gene diversity and Nei's unbiased genetic distance were calculated for all population pairs (Nei 1978) using the PopGene (ver. 1.32; Yeh et al. 2000; http://www.ualberta.ca/fyeh/index.htm). The neighbor joining (Saitou and Nei 1987) dendrogram was computed using Nei's (1978) genetic distance based on cpSSRs data with the MEGA 3.1 software (Kumar et al. 2004). Evolutionary relationships between haplotypes were estimated with the Network 4.1.0.9 (available at www.fluxus-engineering.com) using the median-joining method ($\epsilon = 0$, Bandelt et al. 1999*).* Mantel's test (Mantel 1967) was applied to examine whether genetic and geographic distance matrices correlate significantly (Bohonak 2002).

2.B.3 Results

2.B.3.1 Identification of cpSSR haplotypes

As summarized in Table 2.B.2, out of 10 cpSSR markers analyzed four primers

Table 2.B.2 Primers used in the study from Weising and Gardner (1999) and the PCR results obtained in the study using 14 populations of *S. laurina* from India

S. no	Code	Primers (Forward and reverse)	Location	Annealing temperature	Polymorphism status
1	ccmp 1	5'-CAGGTAAACTTCTCAACGGA -3' 5'-CCGAAGTCAAAAGAGCGATT-3'	*trn*K intron	-	No amplification
2	ccmp 2	5'-GATCCCGGACGTAATCCTG-3' 5'-ATCGTACCGAGGGTTCGAAT-3'	5' to *trn*S	61 °C for 30 s	Multiple bands
3	ccmp 3	5'CAGACCAAAAGCTGACATAG-3' 5'-GTTTCATTCGGCTCCTTTAT-3'	*trn*G intron	60 °C for 30 s	**Polymorphic**
4	ccmp 4	5'-AATGCTGAATCGAYGACCTA-3' 5'-CCAAAATATTBGGAGGACTCT-3'	*atp*F intron	65 °C for 30 s	Multiple bands
5	ccmp 5	5'-TGTTCCAATATCTTCTTGTCATTT-3' 5'-AGGTTCCATCGGAACAATTAT-3'	3' to *rps*2	57 °C for 1 min	Monomorphic
6	ccmp 6	5'-CGATGCATATGTAGAAAGCC-3' 5'-CATTACGTGCGACTATCTCC-3'	ORF 77-ORF 82 intergenic	61 °C for 30 s	**Polymorphic**
7	ccmp 7	5'-CAACATATACCACTGTCAAG-3' 5'-ACATCATTATTGTATACTCTTTC-3'	*atp*B-*rbc*L intergenic	50 °C for 30 s	Multiple bands
8	ccmp 8	5'-TTGGCTACTCTAACCTTCCC-3' 5'-TTCTTTCTTATTTCGCAGDGAA-3'	*rp*l20-*rps* 12 intergenic	-	No amplification
9	ccmp 9	5'-GGATTTGTACATATAGGACA-3' 5'-CTCAACTCTAAGAAATACTTG-3'	ORF 74b-*psb*B intergenic	-	No amplification
10	ccmp 10	5'-TTTTTTTTTAGTGAACGTGTCA-3' 5'-TTCGTCGDCGTAGTAAATAG-3'	*rp*l2-*rps* 19 intergenic	52 °C for 30 s	**Polymorphic**

showed amplification. The representative picture showing the amplicons is depicted in Fig. 2.B.1. Out of these four, three primers *viz*., ccmp 3, ccmp 6 and ccmp 10 were found to be polymorphic. In all three alleles per polymorphic cpSSR locus were detected.

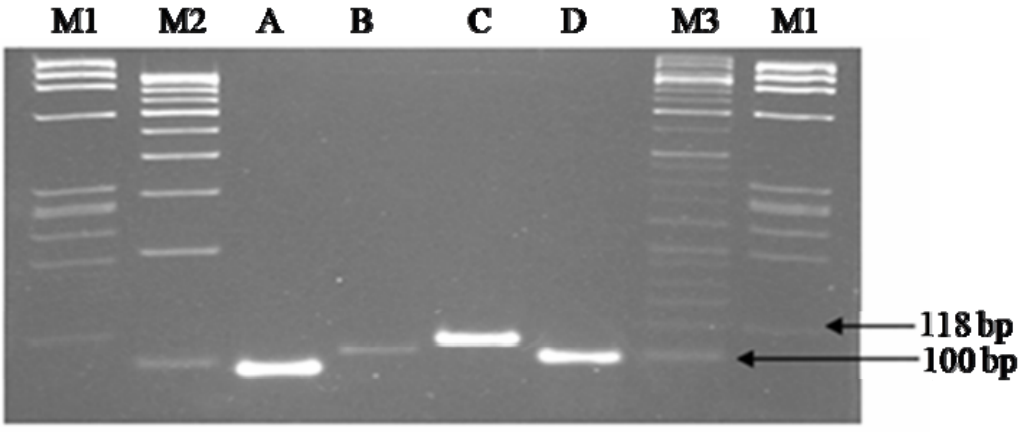


Fig. 2.B.1 Sizes of the amplicons used in the study

Lane M1: Molecular weight marker ØX 174/*Hae*III, lane M2: molecular weight marker 100 bp ladder, lane A-D: products of primers used ccmp 5, ccmp 3, ccmp 6 and ccmp 10, respectively lane M3: 20 bp ladder; lane M1: molecular weight marker ØX 174/*Hae*III marker

The gene frequencies of the alleles are given in Table 2.B.3 and they ranged from 0.0051 to 0.9846. The second allele of ccmp 3 and ccmp 6 was found to be abundant while, first allele of the ccmp 10 occurred with the highest frequency among the 195 samples from 14 populations. The summary of genic variation statistics for the polymorphic loci is depicted in Table 2.B.4. The mean observed number of alleles was 3.0 while, effective number of alleles in the study was 1.0496. The primer ccmp 10 showed the highest genetic diversity, followed by ccmp 6 and ccmp 3. In the present study a very low Nei's genetic diversity was observed (0.0467).

Table 2.B.3 Gene frequencies of cpSSR alleles calculated using PopGene

Allele/Locus	ccmp 3	ccmp 6	ccmp 10
Allele 1	0.0051	0.0103	0.9590
Allele 2	0.9846	0.9846	0.0308
Allele 3	0.0103	0.0051	0.0103

Table 2.B.4 Summary of genic variation statistics for all loci (Nei 1987)

Locus	Observed no. of alleles (na)	Effective no. of alleles (ne)	Nei's genetic diversity (h)
ccmp 3	3.0000	1.0314	0.0304
ccmp 6	3.0000	1.0314	0.0304
ccmp 10	3.0000	1.0861	0.0793
Mean	3.0000	1.0496	0.0467
Std. dev	0.0000	0.0316	0.0282

The polymorphic primers produced three alleles that ranged from 98 bp to 112 bp in size as shown in Table 2.B.5. The three primers revealing a total of nine alleles were used to define seven different haplotypes (C1-C7) from all the 195 samples (Table 2.B.5).

Table 2.B.5 Haplotype definition resulting from the combination of nine microsatellite alleles recorded from three polymorphic loci

Code	ccmp 3			ccmp 6			ccmp 10		
	112 bp	110 bp	108 bp	111 bp	110 bp	98 bp	105 bp	104 bp	103 bp
C1	0	1	0	0	1	0	1	0	0
C2	0	1	0	0	0	1	1	0	0
C3	0	1	0	0	1	0	0	1	0
C4	0	1	0	1	0	0	1	0	0
C5	0	0	1	0	1	0	1	0	0
C6	1	0	0	0	1	0	1	0	0
C7	0	1	0	0	1	0	0	0	1

'1' denotes the presence and '0' the absence of a fragment from its expected position on gel.

2.B.3.2 Distribution of cpSSR haplotypes in *S. laurina* populations

The seven haplotypes identified in the study were distributed randomly. The haplotype C1 was particularly common in every population (92% of the samples), C3 was the second most frequent haplotype (3%) that occurred in four populations throughout the distribution range of *S. laurina*. ELF, AMB, KOD, MUN and PON were found to be monomorphic. The frequency of haplotypes in the 14 populations ranged from 0.005 to 0.872 as presented Table 2.B.6. The distribution of the haplotypes is depicted in Fig. 2.B.2.

Table 2.B.6 Distribution of *S. laurina* cpSSR haplotypes in 14 populations

	Haplotypes identified							
Populations	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**Total**
CHR	11	0	1	0	1	0	0	13
SPR	10	0	1	0	0	0	0	11
ELF	6	0	0	0	0	0	0	6
AMB	27	0	0	0	0	0	0	27
KUD	13	1	2	0	1	0	0	17
KEM	12	0	2	0	0	0	0	14
TAL	9	0	0	1	0	0	0	10
THR	15	0	0	0	0	1	0	16
SAN	11	0	0	1	0	0	0	12
MUN	10	0	0	0	0	0	0	10
PON	3	0	0	0	0	0	0	3
KOD	9	0	0	0	0	0	0	9
KLH	19	0	0	0	0	0	1	20
YRD	26	0	0	0	0	0	1	27
Total	181	1	6	2	2	1	2	195
Frequency	0.872	0.005	0.0256	0.010	0.005	0.005	0.010	

Table 2.B.7 Genetic diversity within different populations of *S. laurina* using CpSSR

Populations	**n_h**	**n_p**	**H_e**
CHR	3	4	0.061538
SPR	2	2	0.036364
ELF	1	1	0.000000
AMB	1	1	0.000000
KUD	4	6	0.091176
KEM	2	2	0.052747
TAL	2	2	0.040000
THR	2	2	0.025000
SAN	2	2	0.033333
MUN	1	0	0.000000
PON	1	0	0.000000
KOD	1	0	0.000000
KLH	2	2	0.020000
YRD	2	2	0.014815

n_h: Number of haplotypes; n_p: Number of polymorphic loci; H_e: Nei's unbiased haplotypic diversity

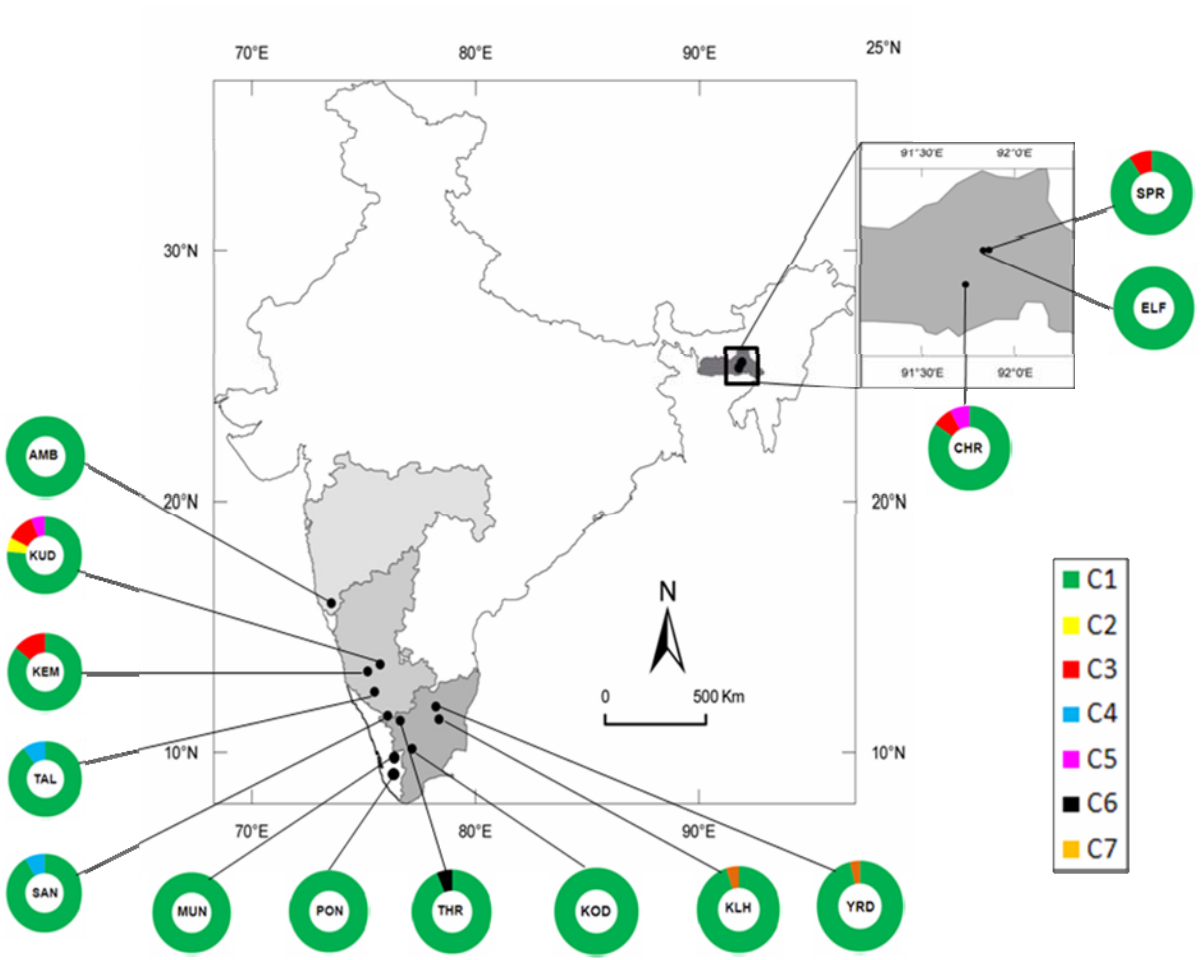


Fig. 2.B.2 Map of India and the cpSSR haplotype composition of each population

The circles projecting out of respective approximate geographical loactions are irrespective of the sample size of the populations and colours represent the different haplotypes as shown in the key. The NEI regions has been projected out to provide more clarity.

2.B.3.3 Analysis of diversity parameters in the study

Population genetic diversity (Nei's unbiased haplotypic diversity) in the *S. laurina* using cpSSR was found to be extremely low, probably because of the low number of polymorphic loci detected in the study. The haplotype diversity ranged from zero in single haplotype population to 0.091176. The highest level of haplotypic diversity was concentrated in populations KUD (0.091176) and CHR (0.061538) (Table 2.B.7).

The total chloroplast DNA diversity (h_T) and mean within population diversity (h_S) were 0.1313 and 0.130, respectively, and the differentiation among populations (G_{ST}) was 0.0074 (Table 2.B.8). N_{ST} was calculated to investigate whether related haplotypes clustered in their geographical distribution. The level of differentiation among populations taking into account the distances between haplotypes (N_{ST}) was not significantly different from G_{ST}. This result suggested that groups of related *S. laurina* haplotypes were not restricted to particular geographical regions.

Table 2.B.8 Diversity indices calculated using PERMUT

	h_S	h_T	G_{ST}	N_{ST}
Total	0.1304	0.1313	0.0074	0.008
NEI	0.1589	0.1562	-0.0174	-
WG	0.1468	0.1499	0.0204	-
EG	0.0580	0.0568	-0.0217	-

Table 2.B.9 of AMOVA highlights the high levels of genetic differentiation within populations (100%) when the dataset was divided into three regions NEI, WG and EG and very low genetic differentiation among regions.

Table 2.B.9 Analysis of molecular variance (AMOVA)

Source of variation	d.f.	Variance	%	*F* statistics
Among region	2	0.276	0.06	F_{CT} = 0.00064
Among populations within region	11	1.479	-0.36	F_{ST} = -0.00363
Within populations	181	25.567	100.30	F_{SC} = -0.00299

Table 2.B.10 Nei's unbiased measures of genetic distance

Populations	CHR	SPR	ELF	AMB	KUD	KEM	TAL	SAN	THR	MUN	PON	KOD	KLH	YRD
CHR	****													
SPR	-0.0011	****												
ELF	0.0009	0.0009	****											
AMB	0.0009	0.0009	0.0000	****										
KUD	-0.0021	-0.0014	0.0021	0.0021	****									
KEM	0.0000	-0.0016	0.0036	0.0036	-0.0013	****								
TAL	0.0029	0.0025	0.0011	0.0011	0.0031	**0.0056**	****							
SAN	0.0024	0.0021	0.0007	0.0007	0.0028	0.0051	-0.002	****						
THR	0.0006	0.0016	0.0004	0.0004	0.0025	0.0046	0.0019	0.0014	****					
MUN	0.0009	0.0009	0.0000	0.0000	0.0021	0.0036	0.0011	0.0007	0.0004	****				
PON	0.0009	0.0009	0.0000	0.0000	0.0021	0.0036	0.0011	0.0007	0.0004	0.0000	****			
KOD	0.0009	0.0009	0.0000	0.0000	0.0021	0.0036	0.0011	0.0007	0.0004	0.0000	0.0000	****		
KLH	0.0006	0.0002	0.0002	0.0002	0.0015	0.0024	0.0016	0.0012	0.0008	0.0002	0.0002	0.0002	****	
YRD	0.0006	0.0003	0.0001	0.0001	0.0016	0.0027	0.0014	0.0011	0.0007	0.0001	0.0001	0.0001	-0.0004	****

**** represent no genetic distance, the largest genetic distance has been indicated in bold.

Table 2.B.11 Geographical distance (in km) between the 14 populations

Populations	CHR	SPR	ELF	AMB	KUD	KEM	TAL	SAN	THR	MUN	PON	KOD	KLH	YRD
CHR	0													
SPR	25.3	0												
ELF	24.6	3.02	0											
AMB	2116	2137	2134	0										
KUD	2193	2216	2213	325	0									
KEM	2189	2151	2149	327	67.1	0								
TAL	2228	2252	2249	419	94.9	123	0							
SAN	2213	2237	2234	569	225	244	163	0						
THR	2228	2252	2250	587	266	264	177	10	0					
MUN	2282	2307	2305	732	535	544	310	156	137	0				
PON	2401	**2426**	2424	868	872	872	441	302	281	152	0			
KOD	2248	2272	2270	732	414	406	325	164	148	40.5	169	0		
KLH	2100	2124	2122	691	402	371	334	187	187	187	314	148	0	
YRD	2064	2088	2086	640	363	325	305	179	185	221	360	192	59.7	0

The largest geographical distance between populations has been indicated in bold.

The calculation of Nei's unbiased genetic distances between populations showed that the genetic distance between populations ranged from 0.0001 to 0.0056 (Table 2.B.10). The negative genetic distance values obtained in the analysis could be because of sampling error (Nei 1978). From the Nei's matrix it is interesting to observe that the highest genetic distance was between TAL and KEM (0.0056) although geographical distance between them was not the largest (Table 2.B.11). An effect of geographical isolation on population structure was studied through the correlation of genetic and geographical distances. The Mantel's test for independence between geographic and genetic distance was insignificant.

2.B.3.4 Relationship among the populations and cpSSR haplotypes

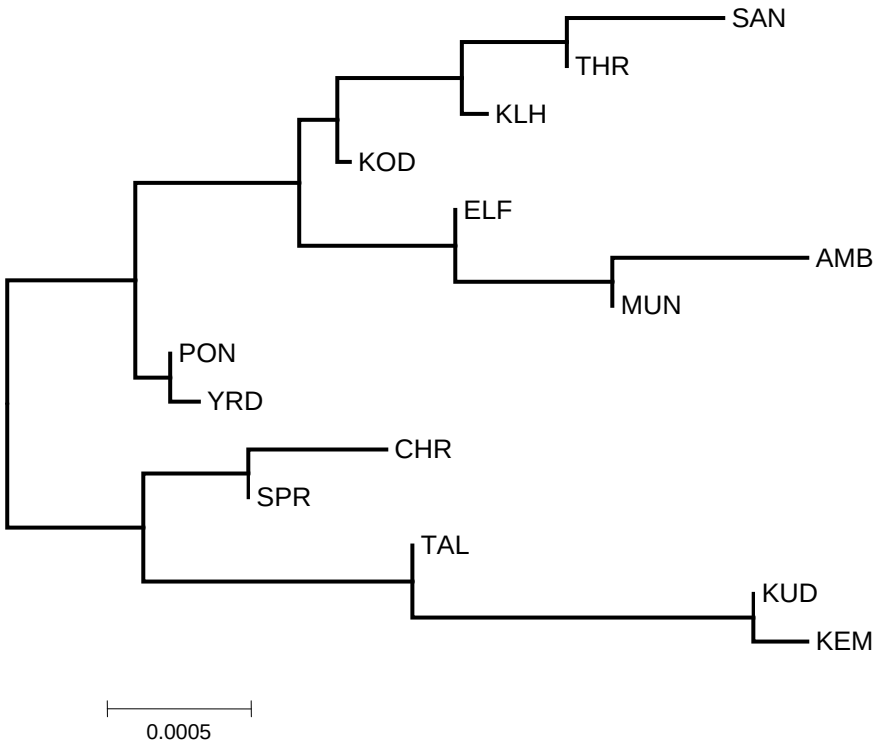


Fig. 2.B.3 Unrooted Neighbour Joining tree depicting the relationship between the populations based on cpSSR markers using Nei's genetic distance

The neighbor joining dendrogram constructed using Nei's (1978) genetic distance based on cpSSR data with the MEGA 3.1 software (Kumar et al. 2004) showed grouping of populations irrespective of the geographical distances between them (Fig 2.B.3). The populations formed two major clusters, the first cluster

comprised populations from NEI (ELF), WG (SAN, THR, AMB, MUN and PON) and EG (KOD, KLH and YRD), while the second cluster comprised three populations of WG (TAL, KUD and KEM) and two populations of NEI (SPR and CHR).

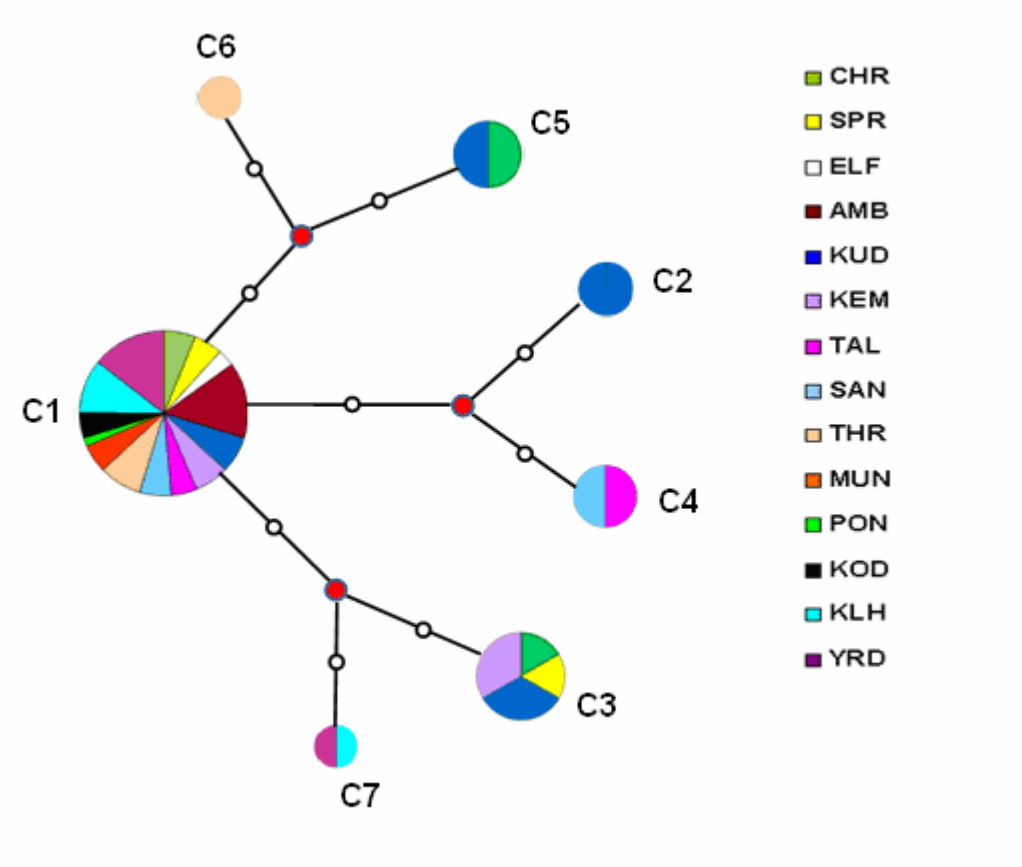


Fig. 2.B.4 Median-joining network based on the cpSSR markers

Each circle represents one haplotype, and its size is near about proportional to its frequency. The colours within the circles represent the contribution of different populations to that haplotype. The small white circles in the connecting line indicate the mutations involved. The key to colour codes of each population is provided separately.

Median-joining network was used to understand the relationship between the haplotypes (Fig. 2.B.4). The C1 haplotype occupied the central position from which all other haplotypes arose. The network showed three branches which were not linked to each other, the branches arose from the dominant haplotype C1. The other haplotypes were connected to C1 through three different medians, which are the connecting haplotypes that had not been detected in the study.

2.B.4 Discussion

In spite of an increased interest in plant phylogeography, most of the studies to date have been carried out with temperate species using such cpSSR markers (*e.g.* Dumolin-Lapègue et al. 1997a; Latta and Mitton 1999; Clark et al. 2000; Belahbib et al. 2001; Gugerli et al. 2001). Surprisingly, only a few reports were published for neotropical species (*e.g.* Hamilton 1999; Caron et al. 2000). Till date not a single report has been published on montane species/subtropical species from India using cpSSR markers, though high species diversity has been reported and also the study area has been declared as biodiversity hotspots emphasizing its conservation importance.

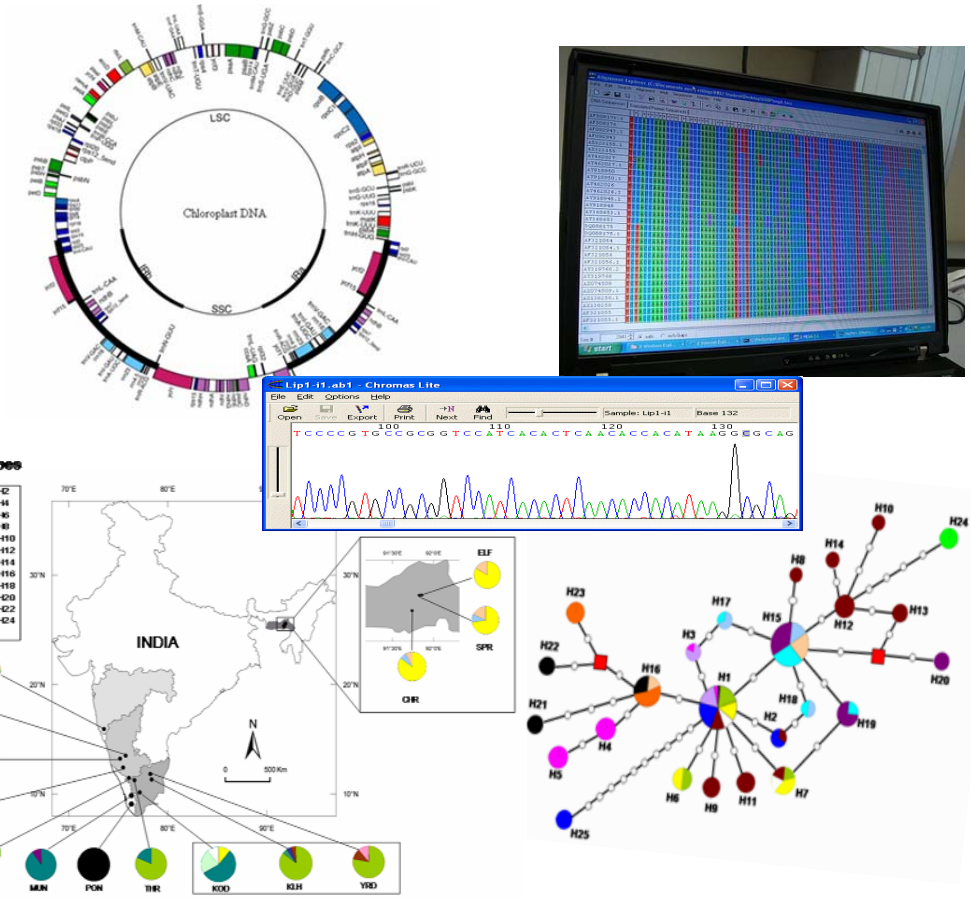
Of the ten primers tested in the present study, only four produced amplification. A relatively constant rate of mutation among cpSSR loci was evident, as the three polymorphic cpSSR loci produced three alleles each. The combination of these three polymorphic markers allowed us to detect seven haplotypes among the 195 individual plants which is consistent with the previous studies on angiosperms *e.g.* six haplotypes in *Pinus resinosa* (Walter and Epperson 2001), eight haplotypes in *Caesalpinina echinata* (Lira et al. 2003), seven haplotypes in *Vitellaria paradoxa* (Fontaine et al. 2004) etc. Preponderance of haplotype C1 throughout the 14 populations is one of the most striking features of the pattern observed in *S. laurina*. The distribution of C1 haplotype in the present study is more or less comparable to the CH1 detected in the PCR-RFLP analysis of chloroplast DNA as described in the previous Section, although in the present analysis C1 was also observed in EG populations unlike PCR-RFLP analysis. The presence of single and most frequent haplotype in all the populations suggests a common lineage for *S. laurina*. A few published studies concerning the diversity of chloroplast microsatellite markers in angiosperm tree species with respect to the allelic diversity have shown a wide range of variability. The average number of alleles exhibited per locus in the present study (1.0372) was less than those reported in natural populations of other angiosperms such as *Silene paradoxa* (2.8; Mengoni et al. 2001), soybean (3.84; Xu et al. 2002), oak (2.33; Deguilloux et al. 2003), *Caesalpinina echinata* (3.28; Lira et al. 2003), *Salix reinii* (3.2; Lian et al. 2003) and *Alyssum bertolonii* (4.8; Mengoni et al. 2003).

Hamrick et al. (1991) suggested that species that are long lived woody perennials, out-crossing and widespread are likely to have high within-population variability (h_s) which, however, was not found in *S. laurina* when analyzed using cpSSR markers (Table 2.B.8). A strong differentiation between populations or between groups is expected from the classical result for forest trees especially angiosperms (Raspé et al. 2000) although in the present study a very low level of differentiation (G_{ST}) was detected in *S. laurina* (Table 2.B.8).

The low levels of intra- and inter- population genetic diversity (five of the seven populations had a single haplotype) suggested very low levels of seed-mediated gene flow within and between the populations over an extended period. It is unlikely that sampling of maternally related individuals was responsible for low levels of within population diversity as the majority of trees sampled were well over the minimum 50 m apart. Thus, the present study was found to provide very little information from the phylogeographic point of view. Analysis of maternally inherited genomes with more sensitive sequencing data can throw light on the genetic structure of montane plants in India.

Chapter 3

Understanding the distribution of genetic diversity in *S. laurina* based on chloroplast genome sequence to elucidate the phylogeographical structure



OUTLINE

3.1 Introduction

Genetic and geographical structures of natural plant populations are the consequences of life history, which is largely shaped by both, intrinsic factors (such as migratory capabilities and mating system) and extrinsic forces including ecological nature of habitats and historical events (*e.g.* vicariance). As a result, the level of genetic variation varies among species with different histories and life strategies. Each species undergoes various processes such as vicariance, dispersal, population bottlenecks, etc., due to historical changes in the environment, including orogeny, soil genesis and climatic changes such as glaciations. Pleistocene glaciation has commonly been considered a powerful force in biotic history and the existence of glacial refugia might have impacted both the species and the ecosystems (Ferris et al. 1998; Hewitt 2000; Weider and Hobæk 2003). By testing the hypotheses of genetic structure expected in populations that may have survived in refugia or may have undergone recent dispersal, it is possible to build an understanding of the history of the ecosystem and provide a foundation for the conservation of cryptic biodiversity and the management of forest resources.

In the Indian subcontinent, the last glacial age might have been an important factor in shaping the genetic structures and phylogeographical patterns of indigenous plant species. Although the land in India may have never been covered by ice sheets, the tremendous climatic changes might have influenced species' distributions and evolution. The montane forests of India form an important ecosystem, which have survived the LGM with variable permeabilities. These ecosystems currently span geographic regions that greatly differ from the last ice age, resulting in great spatial differences in genetic structure and diversity. The montane forests of SI along with NEI are of great interest to evolutionary biologists. Particularly, the shola forests are known to harbor many plant species that have disjunct distribution across its range and are found in patches in the regions identified as biodiversity hotspots. However, the effects of glaciations on the genetic structure of these organisms having disjunct distribution in India are largely unexplored. Phylogeographic analyses examining the temporal and spatial distribution of gene lineages (Avise 2000) can throw light on the effect of glaciations on the shola forest ecosystem of India using a montane plant system like *Symplocos laurina*. As stated earlier in Chapter 1, vicariance and the

dispersal are the two most influential hypotheses put forth to understand the effects of glaciations on the distribution of flora and fauna in the Indian subcontinent.

Molecular techniques have provided the ideal tools for studying the phylogeography or migratory footprints for many species (Avise 2000). In plants, chloroplast DNA (cpDNA) is thought to evolve slowly, with low mutation (Wolfe et al. 1987; Li 1997) and recombination rates (Clegg and Zurawaski 1992). It is known to be maternally inherited in most angiosperms and the gene flow or haplotype exchange occurs exclusively by seeds (Corriveau and Coleman 1988). Maternally inherited markers generally reveal much greater genetic structure than the nuclear markers and have a smaller effective population size (Petit et al. 1993; Hare 2001). In view of this, cpDNA markers are considered to be useful tools in the study of phylogeography for discussing spatio-temporal dynamics of plant species and for identifying glacial refugia and postglacial colonization routes (Dumolin-Lapègue et al. 1997a; Ferris et al. 1998; Matyas and Sperisen 2001; Csaikl et al. 2002). Moreover, for assessing and discerning the relative contribution of these factors, noncoding intergenic spacers (Harris and Ingram 1991), which are nearly neutral, provide possible resolutions (Schaal et al. 1998).

In the present study, sequencing of noncoding spacer regions in cpDNA of 14 populations of *S. laurina* was carried out to address the following questions: (a) is the present day plant population a result of vicariant event or dispersal event? (b) did the species survive the last glaciation in a unique refuge or did it persist in more locations (several glacial refugia)? (c) is the genetic structure of the Northeast Indian populations and the South Indian populations the same?

3.2 Materials and methods

3.2.1 Plant material

The plant material was obtained from 14 populations from three different regions- NEI, WG and EG during field collections performed between October 2003 and May 2006 in collaboration with Botanical Survey of India (BSI), Western Circle, Pune, India as described in Section 2.A.2.1 and details of sampling are given in Table 2.B.1. The sampling sites are shown in Fig. 3.1. The additional samples from Kerala have been included in this study.

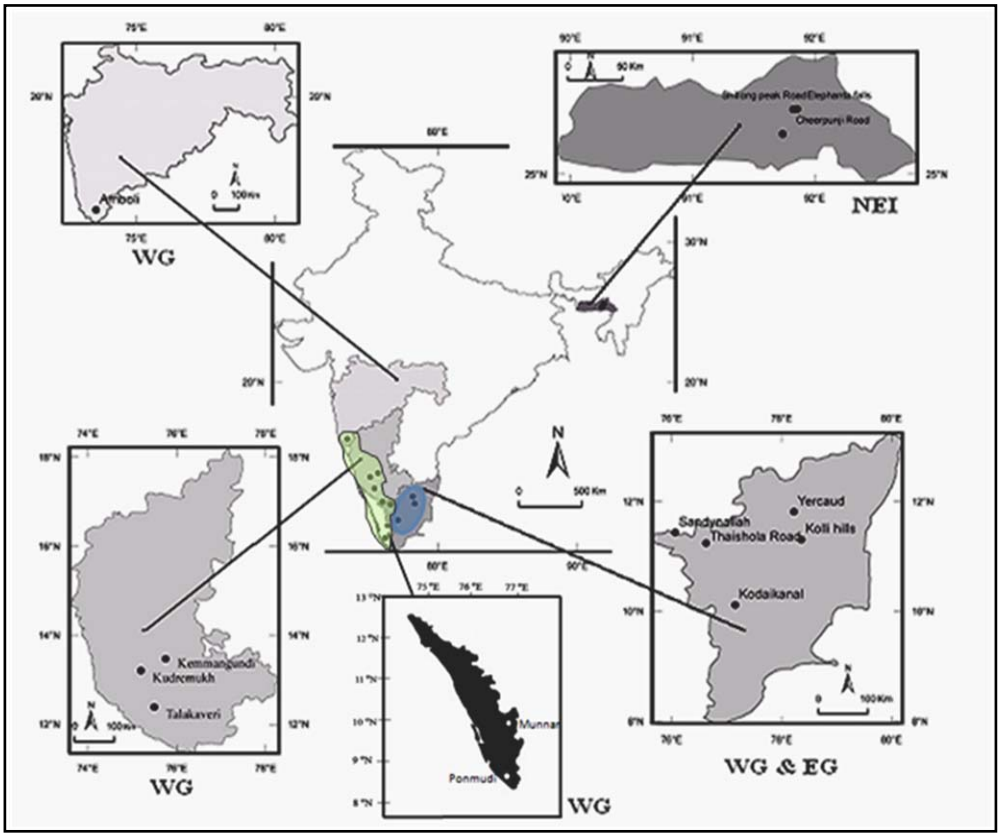


Fig. 3.1 Map of collection sites (populations) in India

The area covering the WG populations and EG populations have been indicated in light green and light blue shades within the Indian map. The states from which populations have been collected are projected out.

3.2.2 DNA extraction, amplification and sequencing

Genomic DNA was extracted using a modified cetyltrimethyl ammonium bromide (CTAB) method as described in Section 2.A.3.2. For the chloroplast intergenic region analysis, eleven cpDNA fragments were investigated with a subset of samples comprising representatives from the populations under study. The intergenic spacers *trn*H-*trn*K, *trn*S-*trn*fm, *trn*S-*trn*T, *atp*H-*atp*I, *psb*B-*psb*B, *trn*S-*trn*G, *pet*B-*pet*D and *trn*L-*trn*F revealed no or extremely low levels of variability. The plastid intergenic spacers *trn*C-*pet*N, *trn*D-*trn*T and *trn*H-*psb*A showed significant intraspecific variability and hence were chosen for further analysis of the populations. For amplification and sequencing of the spacers, following sets of primers were used: (i) CCAGTTCAAATCTGGGTGTC (forward) CCATTAAAGCAGCCCAAGCAAGAC (reverse), for *trn*C {tRNA-Cys(GCA)}-*pet*N spacer, as described by Demesure et al.

(1995) and Lee and Wen (2004) (ii) ACCAATTGAACTACAATCCC (forward) and CTACCACTGAGTTAAAAGGG (reverse), for the intergenic region between *trn*D {tRNA-Asp (GUG)} and *trn*T {tRNA-Thr(GGU)}, as described by Demesure et al. (1995) and (iii) ACTGCCTTGATCCACTTGGC (forward) and CGAAGCTCCATCTACAAATGG (reverse), for the *trn*H-*psb*A spacer, as described by Hamilton (1999).

The polymerase chain reaction (PCR) amplifications were performed in 25 µL reaction volume consisting of 1U *Taq* DNA polymerase with 1X buffer (Bangalore Genei, India), 2 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of each primer and 50 ng of template genomic DNA. The amplification conditions for *trn*C-*pet*N spacer were as follows: initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 55^{o}C for 30 s and 72°C for 2 min, and final extension at 72°C for 5 min. For the *trn*D-*trn*T and *trn*H-*psb*A spacers, the amplification conditions were the same, except for annealing, which was 60°C for 1 min 45 s. The PCR amplifications were carried out in PTC 200 thermal cycler (MJ Research, Waltham, Massachusetts, USA) and the PCR products were electrophoresed in 1% agarose gel, followed by staining with ethidium bromide. DNA sequencing was performed using the MegaBACE DYEnamic ET dye terminator kit and the MegaBACE 1000 DNA analysis system (Amersham Biosciences, Illinois, USA), in accordance with the manufacturer's specifications.

3.2.3 Data analysis

The sequences were aligned using CLUSTALX ver. 1.81 (Thompson et al. 1997) and were manually edited using of MEGA ver. 3.1 (Kumar et al. 2004) and CHROMAS LITE ver. 2.01 (available at http://www.technelysium.com.au/chromas_lite.html). Irrespective of their sizes, all the contiguous insertion/deletion events (indels) were treated as one mutational event and the singletons (polymorphic sites found in only one individual) were removed from the data for final analysis, as they were not considered phylogenetically informative within the dataset and may represent sequencing artifacts. The sequences were deposited into GenBank prior to removal of singletons to allow an unbiased comparison with other datasets.

3.2.4. Analysis of population structure

The measures of genetic diversity and population differentiation were analyzed with HAPSTEP ver. 2001 (available at http://www.pierroton.inra.fr/genetics/labo/Software) as described by Pons and Petit (1995, 1996). The parameters included the mean within-population gene diversity (h_S), the total gene diversity (h_T) and the coefficient of genetic differentiation over all populations (G_{ST}) as well as other equivalent parameters (v_S, v_T and N_{ST}) obtained by taking into account the similarities among the haplotypes. G_{ST} depends only on the frequencies of the haplotypes; however, N_{ST} is influenced by both the haplotype frequencies and the genetic distance between the haplotypes. The haplotype diversity (h) and nucleotide diversity (π) parameters were calculated using DnaSP ver. 4.0 (Rozas et al. 2003). The neutrality tests, Tajima's '*D*' (Tajima 1989), Fu's *F*s (Fu 1997) and a pairwise mismatch distribution were used to test for population expansion (Rogers and Harpending 1992). Conventional F_{ST} (Wright 1931) based on cpDNA sequences for population subdivision was estimated using the analysis of molecular variance (AMOVA) as implemented in Arlequin ver. 3.11 (Excoffier et al. 2005). The distribution of inter- and intra-population genetic variation was also calculated using Arlequin.

In addition, the level of divergence of each population from the remaining populations was also calculated as mean values of pairwise F_{ST} of the population against the rest of the populations. The contributions to total diversity (CT) of each population were calculated using CONTRIB ver. 1999 (available at http://www.pierroton.inra.fr/genetics/labo/Software) according to Petit et al. (1998). The two components of CT, within population diversity (*CS*) and its differentiation (*CD*), were also estimated for the total diversity. Evolutionary relationships among the haplotypes were estimated using Network 4.1.0.9 (available at http://www.fluxus-engineering.com) using the median-joining method ($\varepsilon = 0$, Bandelt et al. 1999). Nucleotide weighting (ω) was adjusted to reflect the difference in mutational frequency among indels ($\omega = 30$), transversions ($\omega = 20$), and transitions ($\omega = 10$), where the least-common event received the highest value. The influence of spatial separation on the degree of differentiation among populations was investigated by calculating pairwise G_{ST} and N_{ST} values using DISTON (available at

http://www.pierroton.inra.fr/genetics/labo/Software), and plotting the means of these parameters against geographical distance classes.

3.3 Results

3.3.1 Chloroplast DNA polymorphism

Eleven noncoding *S. laurina* chloroplast DNA regions were initially sequenced from the representative samples from each collection site. However, only three regions exhibited an optimum level of sequence divergence. The amplicons obtained are shown in the Fig. 3.2, the sizes of the amplicons varied from 450 bp to 1.2 kb.

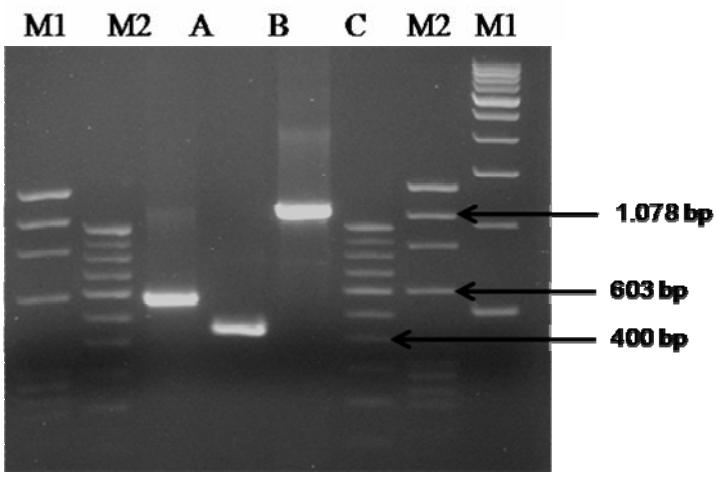


Fig. 3.2 PCR amplicons of markers used in the study

Lane M1: Molecular weight marker ØX174/*Hae*III digest, lane M2: molecular weight marker 100 bp ladder, lane M3: molecular weight marker 500 bp ladder, lane A: *trn*C-*pet*N (600 bp), lane B: *trn*H-*psb*A (450 bp) and lane C: *trn*D-*trn*T (1.2 kb) amplicons from chloroplast DNA.

Hence, these three cpDNA regions were chosen for further analysis and were sequenced from all the 195 samples. These three examined sequences consisted of *trn*C-*pet*N region (542 bp, GenBank accession nos. EU769953 to EU770148), *trn*H-*psb*A intergenic spacer (404 bp, GenBank accession nos. EU769757 to EU769952) and *trn*D-*trn*T intergenic spacer (968 bp, GenBank accession nos. EU769561 to EU769756). Thus, a consensus length of 1914 bp was analyzed in the present study.

A total of 40 polymorphic sites (4.37%), with 29 parsimony informative sites, were detected within these three intergenic spacers. Among them, four parsimony informative sites were present in the *trn*C-*pet*N region (0.34%; G+C), sixteen in the *trn*H- *psb*A region (0.26%; G+C) and nine in the *trn*D-*trn*T spacer (0.36%; G+C). Among the polymorphic sites, 27 were caused by point mutations, while two were the result of indels that included one 8-bp (GAAAGGGG) and one 7-bp (ATCCCTC) indels in *trn*C-*pet*N region. The point mutations detected in the present analysis included one insertion, one deletion, nine transitions and sixteen transversions (Table 3.1).

3.3.2 Haplotype distribution, polymorphism, diversity and test of neutrality

Twenty-five cpDNA haplotypes (H1-H25) were identified from the 29 polymorphic sites of 195 sequences using DnaSP (Table 3.1). Among these, ten haplotypes occurred in more than one populations; whereas the other 15 haplotypes (H4, H5, H8, H9, H10, H11, H12, H13, H14, H20, H21, H22, H23, H24 and H25) were singletons, occurring only in one population (Fig. 3.1). The H1 haplotype had widespread distribution and was detected in eight populations, followed by H7 and H15 (4 populations each). Similarly, three dominant haplotypes with high haplotype frequencies were detected; the haplotypic frequency of H15 was the highest (60), followed by H1 (57) and H16 (17). The region wise subdivision of the data showed presence of the highest number of haplotypes in WG (20 haplotypes) followed by EG (9) and NEI (3) (Table 3.2).

From the distribution of haplotypes (Fig. 3.3), it can be observed that some of the haplotypes were restricted to particular locations and showed geographical patterns in their distribution *e.g.* the haplotype H6 was restricted to SPR and CHR, while the haplotype H3 was observed only in KEM and TAL. Moreover, the dominant haplotypes (H1, H15 and H16) also followed the similar pattern with H1 being present only in NEI and northern populations of WG (AMB, KUD, KEM and TAL), H15 in SAN, THR and EG populations (KLH and YRD), while MUN and KOD were dominated by haplotype H16. The boundary line for the distribution of haplotypes could not be demarked according to geographic regions as the SAN and THR populations of WG shared the dominant haplotype H15 found in the EG populations of KLH and YRD.

Table 3.1 Chloroplast DNA haplotypes detected in *S. laurina* samples

	***trn*C-*pet*N**				***trn*H (GUG)-*psb*A**															***trn*D-*trn*T**									
Haplotype	**101-108**	**139**	**340**	**428-434**	**551**	**556**	**563**	**564**	**581**	**595**	**630**	**637**	**645**	**683**	**700**	**707**	**724**	**748**	**866**	**889**	**947**	**983**	**989**	**1024**	**1182**	**1255**	**1325**	**1597**	**1914**	**No. found**
H1	GAAAGGGG	A	A	ATCCCTC	T	C	-	T	C	C	A	A	A	T	T	C	G	A	T	T	G	A	T	T	C	G	T	A	G	57
H2	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.	.	.	.	G	.	.	3
H3	. . . . .	.	.	-	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	5
H4	. . . . .	.	.	-	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5
H5	-	.	.	-	.	T	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
H6	-	.	.	. . . . .	.	T	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
H7	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	5
H8	. . . . .	.	.	. . . . .	.	.	-	.	.	.	C	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
H9	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	2
H10	. . . . .	.	.	. . . . .	.	.	-	.	.	.	C	.	G	.	.	.	.	.	G	.	.	.	.	.	.	A	.	.	.	2
H11	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	5
H12	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	5
H13	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	2
H14	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	.	1
H15	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	60
H16	-	.	.	. . . . .	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	17
H17	-	.	.	. . . . .	.	.	-	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	3
H18	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	2
H19	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	4
H20	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	G	.	.	.	.	.	3
H21	. . . . .	.	.	. . . . .	.	.	-	.	.	.	.	.	.	.	.	A	.	.	.	.	C	.	.	.	.	.	G	.	.	2
H22	-	.	.	. . . . .	.	.	-	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.	.	1
H23	-	.	.	. . . . .	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	.	.	.	1
H24	. . . . .	.	.	. . . . .	C	T	-	-	.	.	.	.	G	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	3
H25	. . . . .	G	C	. . . . .	.	.	A	.	A	T	T	.	G	.	C	.	A	C	.	.	.	.	.	.	.	.	.	C	.	2
Type of mutations	Indel	t	T	Indel	t	t	I	D	T	t	T	T	t	T	t	T	t	T	T	T	T	T	T	T	T	t	T	T	t	
	Indel - 2 (8 bp and 7 bp) Transition – 1(A–G) Transversion – 1 (A-C)				Insertion – 1 (A), Deletion – 1 (T) Transition – 6 {2(T-C), 2(C-T), 1(A-G), 1(G-A)} Transversion – 8 {3(C-A), 1(A-T), 2(A-C), 2(T-G)}															Transition – 2 {2(G-A) Transversion - 7 {1(G-C), 3(T-G), 1(C-G), 2(A-C)}									

Numbers in the second row refer to the positions in the aligned *trn*C-*pet*N, *trn*H(GUG)-*psb*A and *trn*D-*trn*T region. The positions are numbered from the 5' end to the 3' end. Dot (.) indicates that the character state is same as in haplotype H1, '-' indicates absence of the base in the respective site (deletion). The single letter codes for "Type of mutations" are - I: Insertion, D: Deletion, t: Transition and T: Transversion.

Table 3.2 Distribution of cpDNA haplotypes in the populations and their haplotype diversity (*h*) and nucleotide diversity (π)

	cpDNA haplotypes																											
Location	**H1**	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	**H8**	**H9**	**H10**	**H11**	**H12**	**H13**	**H14**	**H15**	**H16**	**H17**	**H18**	**H19**	**H20**	**H21**	**H22**	**H23**	**H24**	**H25**	**Total**	***h***	***π***
CHR	11	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13	0.4727	0.00027
SPR	8	-	-	-	-	1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11	0.2949	0.00016
ELF	5	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	0.4727	0.00027
AMB	7	1	-	-	-	-	1	1	2	2	5	5	2	1	-	-	-	-	-	-	-	-	-	-	-	27	0.8746	0.00111
KUD	13	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	17	0.4118	0.00127
KEM	10	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14	0.4396	0.00023
TAL	2	-	1	5	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	0.7333	0.00029
SAN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	-	2	1	-	-	-	-	-	-	-	12	0.325	0.00017
THR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13	3	-	-	-	-	-	-	-	-	-	16	0.4394	0.00025
MUN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-	1	-	-	10	0.2000	0.00021
PON	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	3	0.0000	0.00000
KOD	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	2	1	-	-	-	9	0.6944	0.00079
KLH	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	-	1	1	1	-	-	-	-	-	-	20	0.2842	0.00016
YRD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	21	-	-	-	3	3	-	-	-	-	-	27	0.3846	0.00032
Total	**57**	3	5	5	2	2	5	1	2	2	5	5	2	1	**60**	17	3	2	4	3	2	1	1	3	2	195	0.8108	0.00078

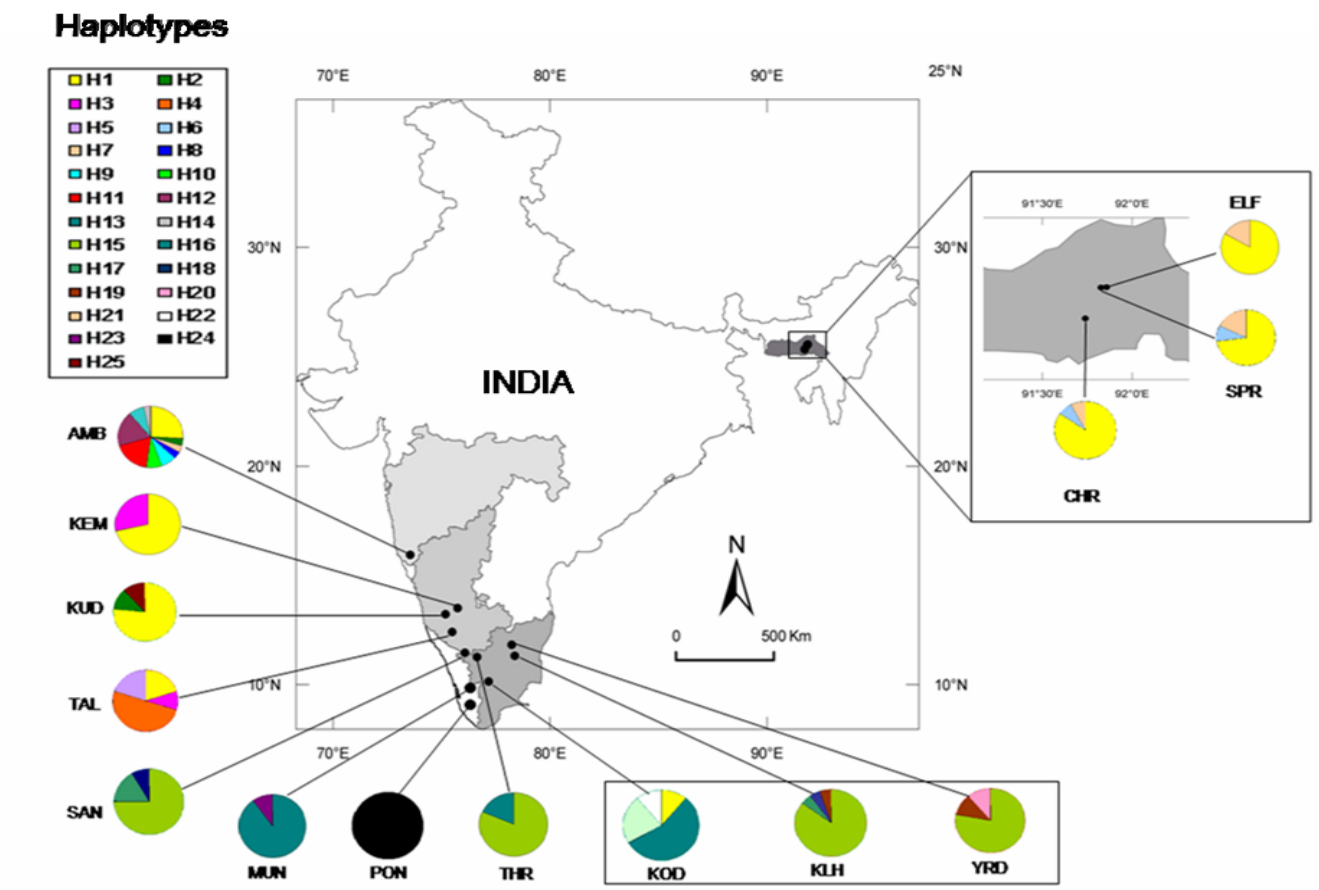


Fig. 3.3 The approximate geographical locations of the collection sites and the haplotypic composition of each population is depicted in the form of pie charts

Circle sizes are not correlated with sample sizes of the populations, however, the coloured slices within the pie charts correspond to the proportion of constituent haplotypes (as depicted in the legend). The NEI region has been projected out to provide more clarity, refer Table 2.B.1 for abbreviations of the location names.

Haplotype diversity (*h*) in the total population ranged from zero to 0.8746 with an average of 0.8108. The WG region, comprising 19 haplotypes, showed prominently high haplotype diversity (0.8564), followed by EG (0.5305) and NEI (0.3494) (Table 3.3). When considered location wise, SPR in NEI possessed the highest haplotype diversity (0.4727), while in WG, AMB with ten haplotypes, was the most variable population (*h* = 0.8746). Similarly, in EG region, KOD showed the highest haplotype diversity (0.6944).

Table 3.3 Estimates of diversity parameters of the *S. laurina* populations sampled

	No. of polymorphic sites	No. of haplotypes	*h*	π	*D*	F_S
NEI	2	3	0.3494	0.00019	- 0.55597	- 0.524
WG	23	19	0.8564	0.00092	- 1.75025	- 8.827
EG	8	9	0.5305	0.00050	- 1.19836	- 2.995
Total	29	25	0.8108	0.00078	- 1.81072*	- 13.812

h: haplotype diversity, π: nucleotide diversity, *D*: Tajima's test of neutrality, *F*s: Fu's test of neutrality,* Statistically significant at $P < 0.05$.

The nucleotide diversity (π), when calculated separately for each sequenced spacer, was the highest (0.00233) for the *trn*H-*psb*A intergenic region, followed by *trn*D-*trn*T (0.00052) and *trn*C-*pet*N (0.00008) spacer. When all the three sequenced spacers were considered together, the nucleotide diversity among the 14 populations was 0.00078, while considered region wise, it was 0.00019 in NEI, 0.00092 in WG and 0.00050 in EG. The range of variation was zero (in PON) to 0.00127 (in KUD), with an average of 0.00078 (Table 3.2 and 3.3). In the region wise subdivided data, SPR (0.00027), KUD (0.00127) and KOD (0.00079) revealed the highest nucleotide diversity in NEI, WG and EG, respectively.

Tajima's *D* (Tajima 1989) and Fu's *F*s (Fu 1997) neutrality tests were used to determine whether a locus was evolving neutrally and thus making it appropriate for a phylogeographical study (Caicedo and Schaal 2004). Tajima's *D* (test of neutrality of sequence evolution) revealed a significant negative value for the total dataset (*D* = -1.81072, $P < 0.05$) as well for the subdivided dataset. However, the values were

insignificant for the region wise divided dataset (Table 3.A.3). This result implied that there were rare nucleotide site variants than would be expected under a neutral model of evolution (Rand 1996). Similarly, the *F*s was negative and differed from zero for the total dataset (*F*s = -13.812) as well as for the region wise divided dataset (Table 3.3). The negative coefficients of Tajima's *D* and Fu's *F*s might be suggestive of population growth, although significant deviation from the neutral model was observed.

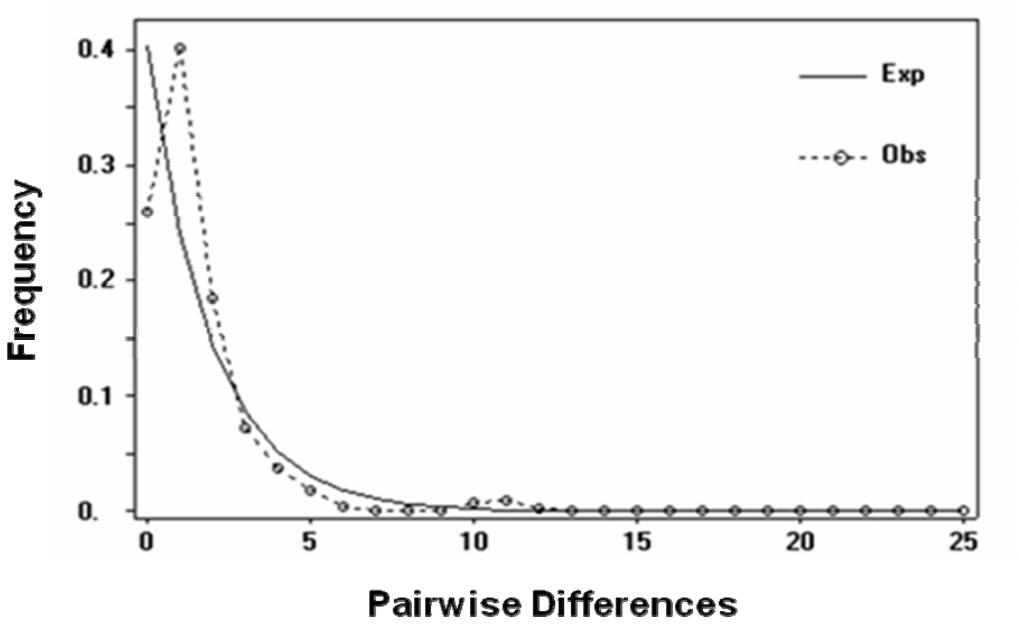


Fig. 3.4 Mismatch distribution established for *S. laurina* populations

The solid line represents the expected mismatch distribution of a stationary population, while the dotted line represents the observed mismatch from segregating sites of the aligned sequences of *trn*C-*pet*N, *trn*H (GUG)-*psb*A and *trn*D-*trn*T intergenic spacers of the cpDNA.

To test the hypothesis of population expansion of *S. laurina*, the distribution of pairwise differences from segregation sites of cpDNA haplotypes was calculated using DnaSP. Under a constant population model, the general distribution follows a gradually declining curve (Rogers and Harpending 1992). The observed mismatch distribution (Fig. 3.4) departed from the constant population model by showing a single main peak while the raggedness statistics *r*, a measure of the smoothness of the

distribution (Harpending 1994), was 0.0815, as expected for a population having experienced recent growth (Fig. 3.4).

3.3.3 Genetic diversity and differentiation

In the present study, the average gene diversity (h_S) for the whole dataset was 0.421, while considering region wise, the EG populations possessed the highest average gene diversity followed by WG and NEI. The total gene diversity was also very high (h_T = 0.856) for the whole dataset. While, in the region wise divided dataset, WG (0.925) was highly diverse followed by EG and NEI. High proportion of the variation resulted from the differences among populations within groups, G_{ST} = 0.509 (Table 3.4). The region wise divided dataset also showed the highest differentiation in WG populations followed by EG and NEI. G_{ST} was negative in NEI. The ordered counterparts of h_S, v_S was low (0.311) for the total data set; however, in the region wise divided dataset, NEI showed the highest v_S followed by EG and WG unlike h_S. v_T, the ordered counterpart of h_T, showed trends similar to h_T for region wise divided dataset with total of 0.864 for the whole dataset. N_{ST} can be used to investigate whether the geographical distribution is related to haplotype distribution. The N_{ST} value for whole dataset was 0.131 while for the region wise divided dataset it was high in EG (0.176) followed by WG (0.134). The permutation test showed that N_{ST} was significantly higher than G_{ST} ($P < 0.05$). The populations in WG (G_{ST} = 0.53, h_T = 0.925) exhibited greater differentiation and total diversity than the populations of EG (G_{ST} = 0.416, h_T = 0.778) and NEI (G_{ST} = - 0.091, h_T = 0.336). The values for N_{ST} - G_{ST} for WG and EG were significantly different from zero indicating the presence of geographical structure in the distribution of haplotypes. Additionally, genetic distance correlated with geographical distance when evaluated by Mantel's test. The results obtained from the CONTRIB indicated that the populations AMB, TAL and KOD contributed the most to the diversity component of the total diversity (Fig. 3.5) due to their high haplotype diversity, while the populations MUN, PON and KLH contributed the most to the differentiation component of the total diversity. The later was due to presence of high frequency of specific rare alleles in those populations.

Table 3.4 Estimation of average gene diversity within *S. laurina* population

Diversity parameters	NEI	WG	EG	Total
h_S	0.367 (±0.054)	0.428 (±0.098)	0.454 (±0.124)	0.421 (±0.060)
h_T	0.336 (±0.053)	0.925 (±0.032)	0.778 (±0.215)	0.856 (±0.034)
G_{ST}	- 0.091(NC)	0.538 (±0.113)	0.416 (NC)	0.509 (±0.073)
v_S	0.366 (±0.060)	0.309 (±0.093)	0.340 (±0.167)	0.311 (±0.074)
v_T	0.337 (±0.057)	0.941 (±0.234)	0.832 (±0.312)	0.864 (±0.208)
N_{ST}	- 0.087 (±0.055)	0.672 (±0.152)	0.592 (NC)	0.640 (±0.123)
N_{ST} - G_{ST}	NC	0.134*	0.176*	0.131*

Within population gene diversity (h_S), total gene diversity (h_T), interpopulation differentiation (G_{ST}) and their ordered counterparts: v_S, v_T and N_{ST} (mean ± SE in parentheses), with analysis done in two stages: (i) populations divided into three regions in India and all populations taken together. * Indicates significantly different from zero at the P > 0.05 level.

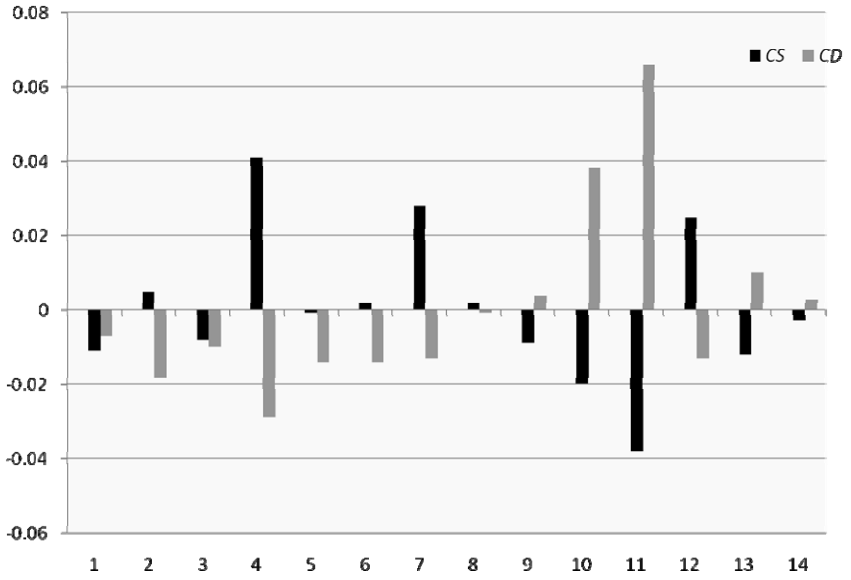


Fig. 3.5 Contribution of each population of *S. laurina* to the total diversity (CT) for chloroplast haplotypes

Closed bars (*CS*) and Grey bars (*CD*) represent contribution of diversity and differentiation, respectively. Numbers corresponding to the populations are as follows 1-CHR; 2-SPR; 3-ELF; 4-AMB; 5-KUD; 6-KEM; 7-TAL; 8-SAN; 9-THR; 10-MUN; 11-PON; 12-KOD; 13-KLH and 14-YRD.

3.3.4 Correlation between differentiation and geographical distance along with partitioning of genetic variation

A hierarchical AMOVA of the chlorotypes analyzed using Arlequin with populations grouped into three regions (NEI, WG and EG) depicted partitioning of 1.45% of the variation among the regions, 51.10% among populations within the regions and the remaining 47.45% variation within populations and all the partitions were significant (Table 3.5). The molecular variance was attributable to the differences among populations within groups. The fixation index (F_{ST} = 0.5255) for the whole data set was very high. The pattern of partitioning also remained the same when WG and EG populations were considered together as SI populations as against the NEI population. The genetic population differentiation was also analyzed through calculation of average F_{ST} values for individual populations in comparison to every other population and most of the population pairs differed significantly for pairwise F_{ST} values. In this analysis ELF and PON were genetically the most distinct populations, followed by CHR with MUN or SAN. This led us to further correlate genetic differentiation (G_{ST} and N_{ST}) among populations with the geographical distance (Fig. 3.6). The populations lying within 100 km distance showed minimum differentiation, while the populations with maximum geographic distance showed maximum differentiation. However, there was no linear correlation between the geographic distance and the genetic differentiation.

Table 3.5 Analysis of molecular variance (AMOVA) of *S. laurina* samples based on noncoding chloroplast DNA sequences

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among geographical regions	2	19.471	0.01490	1.45
Among populations within the regions	11	82.007	0.52564	51.10
Within populations	181	88.337	0.48805	47.45
Total	194	189.815	1.02858	
Fixation indices				
F_{SC}: 0.51854; F_{ST}: 0.52551; F_{CT}: 0.01448				

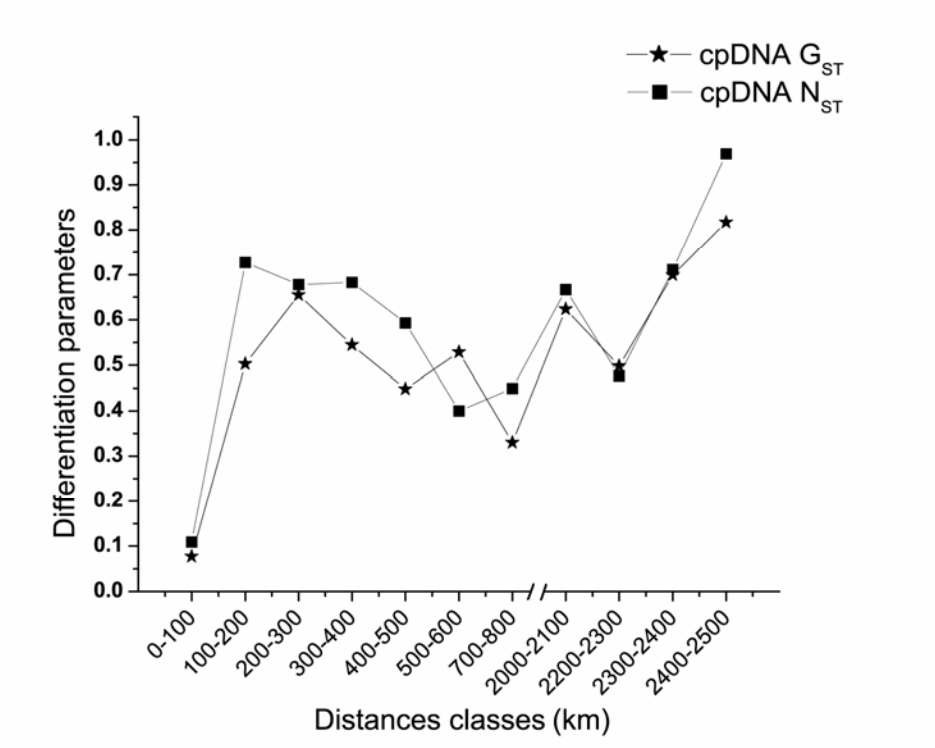


Fig. 3.6 Relationship of average pairwise G_{ST} and N_{ST} values with the geographical distance separating the populations for cpDNA haplotypes

3.3.5 Relationships among the haplotypes

From the median-joining network drawn using Network software (Fig. 3.7), the haplotypes H1 (57 individuals) and H15 (60 individuals) were identified as the dominant haplotypes occupying the interior positions in the network. These two haplotypes (differentiated by a T-G transversion in the *trn*H-*psb*A spacer at 683 site) showed region specific distribution, with H1 being distributed in NEI and northern populations of WG, while H15 was spread out in SAN, THR and EG populations (KLH and YRD). The third dominant haplotype, H16, was a major haplotype in KOD and MUN populations and was derived from H1 by 8 bp indel in the *trn*C-*pet*N region (101-108 bp). Five haplotypes (H4, H5, H21, H22 and H23) generated from this haplotype were uniquely located at TAL (H4 and H5), KOD (H21 and H22) and MUN (H23), thus representing diverse locations. The haplotype H25, which is a unique haplotype found in KUD population, was also derived from H1 *via* 10 mutational changes (with two mutational changes in the *trn*C-*pet*N region, seven in the *trn*H-*psb*A region and one in the *trn*D-*trn*T region). Haplotype H6 (of CHR and SPR), H9 and H11 (of AMB population) were also found to arise from H1 and were

restricted to NEI region and AMB populations, respectively. Interestingly, the haplotype H24 of PON population was derived from H12 by four mutational changes. The alternative routes/networks detected in the study were: H3-H17, H2-H18 and H17-H19 linked to H1 and H15. In the haplotype network, all exterior haplotypes were unique to particular populations. However, the linking of the haplotypes H8, H14, H12, H10 and H13 (of AMB) to the major haplotype H15, which was altogether absent in the AMB population.

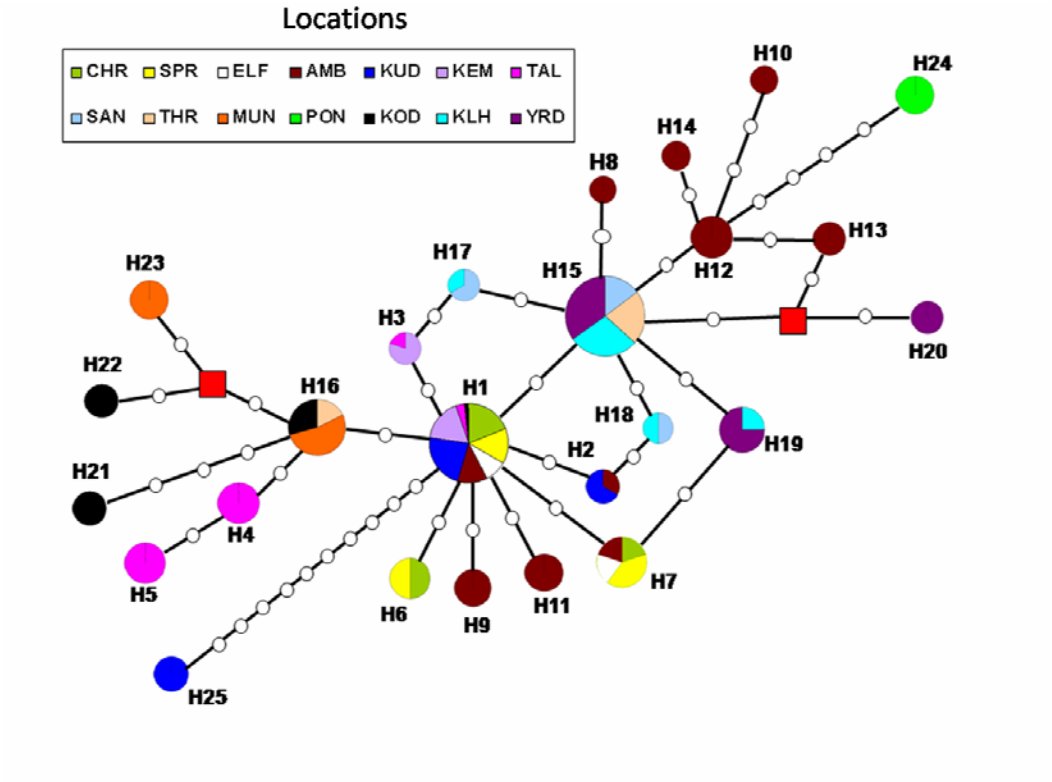


Fig. 3.7 Median-joining network based on the joint analysis of the three cpDNA sequences

Each circle represents one haplotype, and their sizes are proportional to their frequencies. The coloured slices within the circles represent contribution of different populations to the particular haplotype. Each of the small white circles on the lines connecting the haplotypes indicate one mutational event.

3.4 Discussion

A chronicle of the intraspecific evolutionary history of a species can be inferred from the degree and distribution of genetic variation among populations in the absence of detailed fossil record. Such studies are the focus of the burgeoning field of

phylogeography (Avise 2000). In the present study an attempt has been made to explain the phylogeography of *S. laurina* based on the nucleotide sequence diversity in noncoding regions of chloroplast DNA.

3.4.1 Comparison of diversity parameters in *S. laurina* and other species

The nucleotide diversity (π = 0.00078, Table 3.3) of the *S. laurina* populations for all the three regions of cpDNA taken together, was greater than that reported for other tree species such as *Cyclobalanopsis glauca* (π = 0.00065, studied using *trn*T-*trn*L, *trn*V-*trn*M and *pet*G-*trn*P; Huang et al. 2002) and *Trochodendron aralioides* (π = 0.00052, from a combination of *pet*G-*trn*P and *pet*A-*psb*J; Huang et al. 2004), while it was less than that reported for *Cunninghamia konishii* (π = 0.0019, obtained from combination of *trn*V intron, *trn*D-*trn*T, *trn*L-*trn*F and *pet*G-*trn*P; Hwang et al. 2003) and *Castanopsis carlesii* (π = 0.00095, studied using *trn*L intron and *trn*V-*trn*M IGS; Cheng et al. 2005a). The level of chloroplast differentiation among the populations was high (G_{ST} = 0.509) and was near to the average cytoplasmic differentiation (G_{ST} = 0.64) reported for angiosperm species (Petit et al. 2005) as well as close to many tree species like *Eucalyptus nitens* (G_{ST} = 0.68, Byrne and Moran 1994) and *Eucalyptus kochii/ horistes* (G_{ST} = 0.62, Byrne and Macdonald 2000). However, the analysis of *S. laurina* samples collected from the same study area with random nuclear markers (ISSR) by Deshpande et al. (2001) displayed a lower nuclear differentiation (G_{ST} = 0.463) than that observed with the cytoplasmic markers in the present study. The estimation of population differentiation, which takes into account the similarities among the haplotypes (N_{ST} = 0.640) was significantly greater than G_{ST}. Significant N_{ST} - G_{ST} difference is more common in species with higher G_{ST} values (G_{ST} > 0.5; Petit et al. 2005) and indicates presence of relationship between the phylogeny and the geographical distribution of haplotypes (Pons and Petit 1996) as observed in the present study.

The low genetic differentiation among NEI, WG and EG (1.45%) obtained using AMOVA, suggested sharing of haplotypes among the regions. For example, the haplotype H1 was a major haplotype in NEI and was also present in northern populations of WG (AMB, KUD, KEM and TAL), while H7 was shared between NEI populations and the AMB population (WG). Similarly, another major haplotype, H15, and the haplotypes H17 and H18 were shared by SAN (WG) and KLH populations

(EG). The presence of H1 in many populations and its central position in the network suggested its ancestral nature, supporting the coalescence theory (Posada and Crandall 2001). Further, H15 could also be considered as an ancient haplotype, since it was found in high frequency in EG and THR and SAN populations. The southern populations of MUN and KOD were dominated by H16, which was also present in THR population. Whereas, the occurrence of H1 in KOD population might be due to accidental dispersal of seeds by birds or animals.

3.4.2 Identification of glacial refugia

The sampling of *S. laurina* in the present study was performed from three geographically distinct regions *viz*., NEI, WG and EG within India. However, the plastid DNA sequences showed a subdivision of the sampled populations into three genetically differentiated groups: NEI-northern WG group, EG group and southern group showing overlaps in the geographic regions. Among the three predominant haplotypes (H1, H15 and H16) obtained in the present study, H1 was observed in SPR, CHR and ELF (of NEI region), AMB, KUD, KEM and TAL (of WG region) forming the NEI-northern WG group. Similarly, H15 was present in SAN and THR (of WG region) along with KLH and YRD (from EG region) forming the EG group, while H16 was present in MUN (of WG) and KOD (of EG) forming the southern group. With the exception of H16 found in THR and H1 found in one KOD individual, all other haplotypes were specific to either of these groups only.

Various factors could be associated with the subdivision of *S. laurina* populations into three genetically differentiated areas/groups. Considering that the temperature in the glacial maximum in India was 5-7°C lower than that observed today (Rajguru 1969; Behling 1998), the temperature in the higher altitude zones of NEI and SI locations would have been similar to the present day temperatures in Himalayan region. Field visits to Himalayan regions did not support the presence of *S. laurina*, suggesting that the occurrence of this species might have been restricted to the regions with mild climate at lower altitude during last glaciation. Furthermore, the species is a part of shola community, which is known to be maintained by a delicate balance of frost (in winter) and forest fire (in summer). The frost seems to be one of the selective factors influencing the floristic composition of shola forest in WG and EG. Due to colder climatic conditions during the glacial time, the frost sensitive shola

vegetation might have remained in the pockets (refugia) where frosts were less or rare. After savanna vegetation and climatic conditions were re-established the species would have probably expanded to suitable locations. The restricted distribution of *S. laurina* in small isolated populations in NEI and SI during the glacial period was probably maintained through limited seed dispersal. It is reported that the flora of these regions are indeed of ancient origin and survived in different pockets within the unglaciated area, thus supporting the hypothesis of glacial refugium for *S. laurina*.

3.4.3 Could there be many refugia for *S. laurina* in India?

Gene genealogies could be used to trace the phylogenetic relationships of the geographical distribution of the haplotypes. As seen in Fig. 3.7, interior positions coupled with high frequency indicated that H1 and H15 could be the ancestral haplotypes. According to the coalescent theory (Crandall and Templeton 1993), these haplotypes may have a great probability of producing mutational derivatives and may represent relict ancestral genotypes. Following the glaciation, the ancestral haplotypes may have probably become widespread, making it difficult to trace the sites of glacial refugia. However, the derived or unique haplotypes along with the ancestral haplotypes may help to identify the refuge. Secondly, the population divergence or genetic differentiation can also be a useful criterion for locating the region of glacial refugium. Petit et al. (2003) tested the hypothesis that glacial refuge areas harbour a large fraction of intraspecific diversity and the plant populations in glacial refuge areas have high genetic divergence and uniqueness rather than a high number of haplotypes. This view has been further supported by the studies in common ash by Heuertz et al. (2004).

Among the populations of the northern WG (AMB, KUD, KEM and TAL), AMB showed many (seven out of fifteen) divergent and unique haplotypes and also showed the highest diversity in this region. Considering the haplotype uniqueness and nucleotide diversity, we propose AMB as a potential refuge for the region covering the WG populations. Moreover, the AMB and TAL populations contributed highest to the total diversity component (Fig. 3.5) observed in *S. laurina*. This also supports the importance of northern part of WG in pooling of many rare private alleles. Why many unique chlorotypes observed in the AMB refuge did not extend to higher elevation (with the exception of the ancestral haplotype H1) is difficult to understand. It is

usually assumed that refugia harbor the highest levels of diversity and part of this diversity is lost during migration due to several founding events (Hewitt 1996; Comps et al. 2001).

The existence of H1 haplotype only in the northeastern Indian (NEI) population and northern population of WG is very intriguing. It poses an important question on the origin of haplotype H1 to ascertain the location from which it might have arisen or it was present in both the regions before glaciations having a glacial refuge also in NEI. The low number of populations in NEI under the present study, even though I attempted collections from many more sites in Himalayan region (Arunachal Pradesh, Sikkim and Darjeeling), makes the unbiased conclusion difficult. The absence of *S. laurina* in the central arid zone of India has further made the interpretation difficult. However, from the present data, it seems that the geneflow between these regions date back to pre LGM since these populations retain the signatures of that period in the form of the H1 haplotype. The other populations of WG (SAN and THR) showed ancestral haplotype H15, similar to EG populations (KLH and YRD). Furthermore, KLH and YRD shared the haplotype H19 while, SAN and KLH shared the haplotypes H17 and H18 among themselves. On the other hand, MUN and KOD shared the major haplotype H16 with THR, which seems to arise from H1 by one mutation. Many haplotypes H21, H22 and H23 derived from H16 were predominant in MUN and KOD populations. This strongly supports the presence of another glacial refugium near AMB and THR within WG. Moreover, the existence of haplotypes in the NEI and in the northern part of WG different from those in southern part of WG (along with EG) and the high level of differentiation in WG (Table 3.4) revealed using chloroplast markers strongly suggest that southern WG and EG populations may not have originated in northern WG refugia. The possibility of occurrence of several refugia during the last ice age thus seems more plausible. The orological barriers such as rivers, valleys and the temperature conditions of the regions might have acted as barriers limiting the refugial pockets to these populations only, thus conforming to vicariant event.

3.4.4 Range expansion from the glacial refuge

According to Avise (2000), a starburst phylogeographic pattern in the network, particularly considering the separate geographic groups is an expected signature for a

species that has expanded its population and geographic range from a small number of refugia. In the present study, the distribution of the haplotype in the genealogical tree with haplotypes arising from the ancestral haplotypes H1 and H15 also supports range expansion hypothesis. The present geographic distribution of haplotypes reflects range expansion events from the refugia after the glacial period. More warm and wet interglacial period (compared to the glacial) due to the intensification of summer monsoons, might have led to the expansion of *S. laurina* from their glacial refugia. Furthermore, the distribution of number of pairwise nucleotide differences is known to be affected by the demographic history of the population. The significant Tajima's *D* neutrality test (Table 3.3), indicates a recent population expansion from the refugial populations. The relatively milder conditions in postglacial period might have helped for range expansion of *S. laurina* from AMB to other northern regions in WG such as KUD, KEM and TAL and from THR to the remaining populations in SI.

It is interesting to note that no ancestral type was found in the southern populations of MUN and KOD. A similar condition was observed in the study of *Cyclobalanopsis glauca* in Taiwan by Huang et al. (2002). Genetic drift might possibly explain the elimination of the ancestral haplotype during the course of colonization. However, this event requires a strong drift effect, which is unlikely to occur in a dominant tree species like *S. laurina* producing many seeds. Instead, the present distribution pattern seems to be the result of a founder effect. A new colonizer (haplotype H16)from the nearby THR population, probably got established at MUN and KOD and then inhibited the further establishment of later arrivals through intraspecific competition (Silvertown 2004). The H16 haplotype then might have evolved into unique haplotypes H21, H22 and H23 at these locations (Table 3.2).

3.4.5 Identification of suture zone in *S. laurina* population

A region exhibiting high genetic diversity might be the result of either of the two phenomena: first, a refuge area with a stable ecological habitat during environmental fluctuations fostering the accumulation of genetic diversity (Tzedakis et al. 2002); or second, an intermediate zone (suture zone) that receives an admixture of organisms from different sources, resulting in higher diversity than the original sources (Petit et al. 2003). The THR populations in the study seem to comply with both. In addition, the EG region is known to be connected to Nilgiri hill population by the

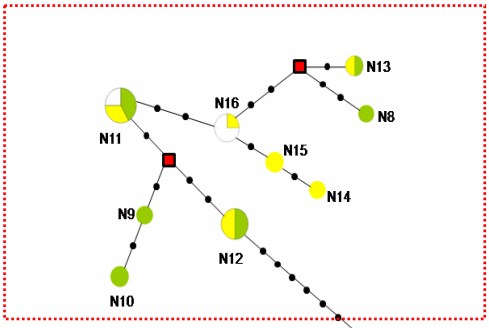
Billigirirangan hills, which run east from the WG to the Kaveri river and forms a forested ecological corridor that connects the EG and WG and probably allows gene flow between the two regions as evident from the studies on elephants from this region by Vidya et al. (2005). In the present data, the THR population probably represents an ancient hybrid zone between the WG and EG, where the divergent plastid lineages of *S. laurina* co-exist.

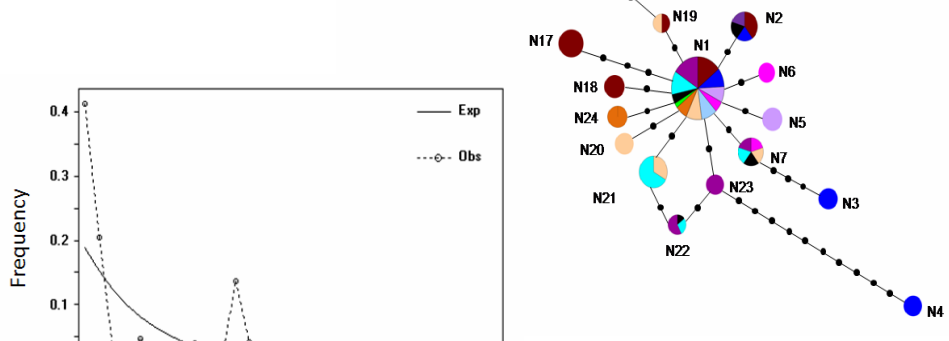
3.4.6 Implications of the study

The total genetic diversity of a species is a key factor in its persistence and in conservation considerations (Rauch and Bar-Yam 2005). The maintenance of genetic diversity is critical for a long term survival of a species (Frankel and Soulé 1981), because loss of variation may largely limit the adaptability of populations to changing environments (Ge et al. 2005). The maintenance of effective population sizes and reduction of human disturbance are thus the priority requisites for conservation. Based on this study, the evolutionary history of *S. laurina* seems to include ancient fragmentation and because of low vagility, localized range expansion from these refugia. These findings have important implications for *in situ* conservation of genetic diversity of *S. laurina*, as the protection of single area would not capture a significant amount of total diversity. To preserve the majority of genetic diversity, several areas possessing different haplotypes need to be targeted.

Chapter 4

Assessing the phylogeographic pattern of genetic structure in *S. laurina* using nuclear markers





OUTLINE

4.1 Introduction

The distribution of genetic variation within a species reflects both historical relationships among populations and ongoing processes of gene flow and isolation. As a consequence, genetic similarity observed in a given pair of populations may be due to recent common ancestry, or it may be a result of gene flow subsequent to population divergence or due to both. Current population structure is typically assumed to reflect an equilibrium level of gene flow per generation (*e.g*. Wright 1951). However, this equilibrium perspective can be misleading, particularly for species that have undergone Pleistocene range expansion, where recent history is a major determinant of population structure.

Severe climatic shifts during Pleistocene ice ages produced large changes in geographical distribution of species and their abundance resulting in range changes (Dynesius and Jansson 2000; Hewitt 2000). The present population of such species thus reflects not only the patterns of genetic exchange, but also the history of gene flow, range fragmentation and isolation among population lineages. Historical events such as population bottlenecks, expansion on wide geographical areas and recurrent geneflow between lineages have been important evolutionary forces (Ellstrand et al. 1999). Studying the geographical distribution of plant lineages (phylogeography) can help synthesize both history and current genetic exchanges and provide insights into different factors that shape the genetic diversity of a species (Avise 2000; Morrell et al. 2003).

Till date, cytoplasmic DNA remains the popular tool for phylogeographic analysis. Chloroplast DNA data have been successfully used to identify glacial refugia and postglacial migration routes and to reconstruct past historical events influencing the genetic structure. However, in few instances the variation was found to be too low for recovering precisely the pattern of past population processes (Jordan et al. 1997; Holderegger et al. 2002; Griffin and Barrett 2004). In recent time, single-copy (or low-copy number) genes of plant nuclear DNA (nDNA) encompassing intron sequences are emerging as a promising tool with potentially high levels of possibly neutral polymorphism for phylogeographic analysis (*e.g.* Strand et al. 1997; Caicedo and Schaal 2004). In contrast to other sources of variation from the nuclear genome, like fragment data, allozymes and microsatellites, such nuclear intron sequences

provide an explicitly historical tool for their construction of allelic (haplotypic) phylogenies at the intraspecific level (Hare 2001; Zhang and Hewitt 2003). Thus, in conjunction with their alleged high levels of polymorphism, nuclear intron sequences may offer a long-sought venue for the study of plant phylogeography at Quaternary timescales (Riddle 1996; Schaal et al. 1998).

Nuclear DNA is considered as the fastest evolving among the three genomes (nuclear, chloroplast and mitochondrial) in the plant cell (Wolfe et al. 1987). Nuclear introns and genes represent a largely untapped source of genetic variation for population genetics and phylogeography. As nuclear genes are often present as gene families, the potential exists to amplify paralogs related by gene duplications and to generate recombinant amplification products mediated by PCR. While the challenge presented by paralogs has been widely recognized the problem of PCR-mediated recombination can not be neglected. The analysis of among-population nDNA intron variation from single/low-copy genes has hitherto received limited attention, partly because of the difficulty in haplotype determination due to interallelic recombination, heterozygosity, and/or homology (Schaal et al. 1998; Avise 2000; Zhang and Hewitt 2003).

In most phylogeographical studies in plants, though cytoplasmic DNA has been used to characterize population structure and evolutionary history, limited use has been made of nuclear genes as a marker, *e.g. Glyceraldehyde-3-phosphate dehydrogenase* (*G3pdh*: Olsen and Schaal 1999; Tani et al. 2003) and vacuolar invertase (Caicedo and Schaal 2004). Studies with these traditional phylogeographical goals have often used nuclear data to corroborate initial results based on cytoplasmic loci (Hare 2001). There are three distinct isoforms of G3PDH (GAPDH) enzyme in higher plants and one of these is cytosolic G3PDH which catalyzes the oxidation of glyceraldehyde 3-phosphate to 3-phosphoglycerate in the cytosol (Cerff 1982). The *Glyceraldehyde-3-phosphate dehydrogenase gene* (*GapC*), which encodes a NADP-dependent nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase enzyme in the cytosol, is a classical low-copy nuclear gene marker and is commonly used in molecular evolutionary studies. The *Glyceraldehyde-3-phosphate dehydrogenase gene* (*GapC*) has been successfully used in molecular evolutionary studies (Olsen and Schaal 1999; Olsen 2002; Pérusse and Schoen 2004).

The purpose of this study is to analyze the population structure and evolutionary history of *Symplocos laurina*, along its entire geographical range in India using DNA sequence variation in the *Glyceraldehyde-3-phosphate dehydrogenase* nuclear gene region. I am particularly interested in developing a polymorphic nuclear DNA marker system for inferring nuclear phylogeography of *S. laurina.*

In the present study, the first set of polymorphic nuclear markers for *Glyceraldehyde-3-phosphate dehydrogenase* gene for *S. laurina* has been reported and its utility in analysis of genetic variation has been demonstrated. I have addressed specific issues *viz*., what is the population structure of *S. laurina*? Is the variation in accordance to geographical region and are identifiable historical processes reflected in the current population structure of *S. laurina*? Till date no population-genetic studies using low-copy nuclear markers have been conducted for any member of the *Symplocos*, yet such information may be critical in phylogeographical context.

4.2 Materials and methods

4.2.1 Sampling of plant populations

Leaves of *S. laurina* were collected from a total of 195 plants sampled from 14 locations (as detailed in Chapter 2, Section 2.B.2.1) from the two biodiversity hotspots, NEI and WG along with EG. The details of sampling are given in Table 2.B.1 and Fig. 3.1 of Section 3.2.1. Isolation of DNA was carried out as detailed in Section 2.A.2.2.

4.2.2 PCR and sequencing

The primers reported for nuclear genome by various authors (Table 4.1) were tested for amplification with *S. laurina* DNA. Of the twenty three pairs of primers, only three pairs (primers GPDX7F and GPDX9R for G3PDH, primers for *PHY A* gene and primers AATX5 F and AATX7 R for *Aat* gene) gave amplification when tested with arbitrarily chosen *S. laurina* samples. The polymerase chain reaction (PCR) was performed in 25 µL reaction volumes consisting of 1U *Taq* polymerase and 1X Buffer (Bangalore Genei, Bangalore, India), 2 mM $MgCl_2$, 0.2 mM dNTP, 0.2 µM of each primer and 50 ng of template genomic DNA, in a PTC 200 thermal cycler (MJ Research, Waltham, Massachusetts, USA). The PCR amplification conditions included initial denaturation of 5 min at 95°C, then 35 cycles of 1 min at 95°C, 1 min

to 45 s at 60°C and 2 min at 72°C, followed by 5 min extension at 72°C. The PCR products were checked by electrophoresis on 1% agarose gel, stained with ethidium bromide. DNA sequencing was performed using the MegaBACE DYEnamic ET dye terminator kit and the MegaBACE 1000 DNA Analysis System (Amersham Biosciences, Illinois, USA), in accordance with the manufacturer's specifications.

The amplicons using *Aat* and *PHYA* gene primers failed to produce readable sequences and hence were not considered for the study; however, amplicons with *G3pdh* gene primers produced partially readable sequence. An aliquot of the product amplified using *G3pdh* gene primers in PCRs was checked and visually quantified on 1.5% agarose gel and was ligated with pGEM-T Easy vector system I (Promega Madison, Wisconsin, USA) as per manufacturer's instructions. The ligated product was stored at 4°C overnight and then used to carry out the transformation of competent *E. coli* Top10 cells (Invitrogen, USA) using heat shock method (Sambrook et al. 1989). Transformed cells were selected under ampicillin (20 µg/mL) by blue-white screening on Luria Bertonii (LB) medium and confirmed by carrying out colony PCR with the same gene specific primers. Colony PCR positive clones were grown in 3 mL LB broth with ampicillin overnight and plasmids were isolated by alkali lysis method, followed by RNAase treatment and quantification (Sambrook et al. 1989). Presence of the inserts in the plasmids was confirmed by carrying out restriction digestions of the recombinant plasmids with *EcoR*1 enzyme. Clones (eleven in number) were sequenced using M13 forward and reverse primers in the MegaBACE sequencer (Amersham Biosciences, Illinois, USA) which yielded a set of two sequences *viz., Symplocos laurina* clone NCL1_11 and *Symplocos laurina* clone NCL2_12 which have been deposited in GenBank with accession numbers EU770149 and EU770150. The two sequences were blasted for sequence homology and were found to correspond to gene *Gap*C encoding the cytosolic G3PDH enzyme. Primers were designed from the clone NCL2_12 which comprised one terminal forward, one internal forward and one terminal reverse primer. The sequences of the designed primers are as follows:

NCL2_12 F: GATGCTTGACCTGCATAAAACA

NCL2_12INF: GTCCACAACTGAAACATCCACA and

NCL2_12R: GATTTGGAATTGTTGAGGGTCT

The *Symplocos* specific primers NCL2_12F and NCL2_12R were used to amplify the *GapC* region of all the 195 samples using PCR conditions described above. A single amplicon of 1020 bp size was obtained and was sequenced using the same primers; however, in some cases NCL2_12 INF was used to obtain a sequence where the reverse primer failed.

Table 4.1 Nuclear genes used in the study and their respective primer pairs

	Gene name	Reference	Primers	Primer sequence (5'– 3')
1	Alcohol dehydrogenase (*ADH* gene)	Innan et al. (1996)	F	GAGCTTGTGAGAAAGATTAAT
			R	GGATCTTGATCGGTATAAGG
	ADH gene	Strand et al. (1997)	ADHX2F	TACTTTTGGGAACGIAAGGTA
			ADHX4R	TCICCIACACTCTCICAAT
2	Chalcone synthase (*chs* gene)	Strand et al. (1997)	CHSX1F	AGGAAAAATTCAAGCGCATG
			CHSX2R	TTCAGTCAAGTGCATGTAACG
3	Chalcone isomerase (*chi* gene)	Strand et al. (1997)	CHIX1 F	TNNTTCCTCGGCGGGCGC
			CHIX4 R	TCCCCGATNATGGNCTCCA
4	Glyceraldehyde 3 phosphate dehydrogenase (*G3PDH* gene)*****	Strand et al. (1997)	GPDX7 F	GATAGATTTGGAATTGTTGAGG
			GPDX9 R	AAGCAATTCCAGCCTTGG
5	Phytochrome A gene (*PHY A*) *****	Mathews and Donoghue (1999)	F	CCYTAYGARGRNCCYATGACWGC
			R	GDATDGCRTCCATYTCRTAGTC
6	Triose phosphate isomerase (*Tpi*)	Strand et al. (1997)	TPIX4FN F	AAGGTCATTGCATGTGTTGG
			TPIX6FN R	CTTTACCAGTTCCAATAGCCC
7	Phosphoglucose isomerase (*Pgi)*	Strand et al. (1997)	PGIX12F	TCTCTICAGTAIGGCTT
			PGIX14R	AATGATACATTCCATCACCT
8	*Pep*C intron region	Gaskin and Schaal (2002)	PPCL1 F	GTCCCTAAGTTTCTGCGTCG
			PPCL1 R	TTCAGGTGTTACTCTTGGG
9	Nitrate reductase	Howarth and Baum (2002)	F	TCBGTGATTACGACGCCGTGTCATGA
			R	GACCARAARCACCARCACCARTAYT
			F	AARTAYTGGTGYTGGTGYTTYTGGTC
			R	GAACCARCARTTGTTCATCATDCC
10	Leafy gene	Shu et al. (2000)	001F	ATGGATCCTGAAGGTTTCACG
			1198R	ACAGCTAATACCGCCAACTAA
11	Floricaula/ leafy *(FLO/FLY*) gene	Grob et al. (2004)	Flint 2 F	CTTCCACCTCTACGACCAGTG
			Flint 2 R	TCTTGGGCTTGTTGATGTAGC

	Gene name	Reference	Primers	Primer sequence
12	Aspartate aminotransaminase (*Aat* gene) *****	Strand et al. (1997)	AATX5F AATX7R	GCTATTCAAGAGAACAG TCAACACCAGTAGGGTTA
13	Calmodulin	Strand et al. (1997)	CAMI F CAMI R	AGCCTNTTCGACAAGGATGG AGTGANCGCATCACAGTT
14	Acidic chitinase	Kawabe et al. (1997)	F R F R	ACATTTTCCATACTACTCAAG ATACTAAAAGCGAATCCATAA AAAGTCAACACCACAAGAGA ATACTAAAAGCGAATCCATAA
15	Malate synthase	Lewis and Doyle (2001)	F R F R	GGAAGATGRTCATCAAYGCNCTYAAYT TTCRTAYYTNAKCCAYTGCCAGTTYTG GGAAGATGRTCATCAAYGCNCTYAAYT GTCTTNACRTAGCTGAADATRTARTCC
16	FAH1	Aguade (2001)	FAH1 F FAH1 R	GAACTTTGCCTCCTGACAAC TTCCACCCCTAATTGACACA
17	Thioglucoside glucohydrolase (myrosinease enzyme) *TGGI* gene	Stranger and Mitchell-Olds (2005)	TGGI F TGGI R	CTTTTCAACAGTGGCAATTTC CATAAACGTTACCGACGAAGACTC
18	Flavanone- 3-hydrolase gene (*F3H*)	Aguade 2001	F3H F F3H R	ACACCGCGCCTAGCATAATT ACACCGCGCCTAGCATAATT
19	Starch synthase gene (*GBBSI* gene)	Mason-Gamer et al. (1998)	F R	TGCGAGCTCGACAACATCATGCG GGCGAGCGGCGCGATCCCTCGCC

***** : produced amplification in *S. laurina*, F: Forward, R: Reverse.

4.2.3 Genealogical analyses

The DNA sequences were edited manually by inspecting electrophorograms using CHROMAS LITE ver. 2.01 (available at http://www.technelysium.com.au/chromas lite.html). Sequence alignment was performed using CLUSTALX 1.83 (Thompson et al. 1997) and MEGA ver. 3.1 (Kumar et al. 2004). Nucleotide diversity (π), haplotype diversity (h), tests of neutrality and determination of associated significance were carried out using DnaSP ver. 4.10 (Rozas et al. 2003). Haplotype (gene) diversity and its sampling variance were based on Nei (1987, eqs. 8.4 and 8.12). Nucleotide diversity, π, the average number of nucleotide differences per site between the two sequences, was based on Nei (1987, eqs. 10.5 or 10.6), and its sampling variance was based on equation 10.7. Tests for departure from neutrality on the total number of segregating sites were based on the '*D*' test of Tajima (1989), and the *D** and *F** tests of Fu and Li (1993). The mismatch distribution of substitutional differences between

pairs of haplotypes was calculated using DnaSP. To survey the possible recombination events, the minimum number of recombination events (Rm) based on the four-gamete test (Hudson and Kaplan 1985) was estimated using DnaSP. Haplotypes and their relationships were determined using the computer program Network 4.1.0.9 (available at http://www.fluxus-engineering.com using the median-joining method (ε = 0, Bandelt et al. 1999).

4.2.4. Analyses of the genetic structure

The measures of diversity and population differentiation were analyzed with the HAPSTEP (ver. 2001, http://www.pierroton.inra.fr/genetics/labo/Software) as described in Pons and Petit (1996). The parameters included the mean within-population gene diversity (h_S), the total gene diversity (h_T) and the coefficient of genetic differentiation over all the populations (G_{ST}) as well as other equivalent parameters (v_S, v_T and N_{ST}) obtained by taking into account similarities between the haplotypes. G_{ST} depends only on the frequencies of the haplotypes. However, N_{ST} is influenced by both the haplotype frequencies and the distance between haplotypes. All the parameters of diversity were calculated for three datasets, firstly for all the populations together, then division of a dataset into NEI and SI and division of a dataset into NEI, WG and EG. A test of the matrix of pairwise genetic distance against the matrix of pairwise geographic distance was performed using a software tool for simple and partial Mantel's test (Bonnet and Van dePeer 2002). To test if genetic differentiation might be geographically related, the populations were separated into three groups and were tested with analysis of molecular variance (AMOVA) as implemented in Arlequin ver. 3.11 (Excoffier et al. 2005). AMOVA with statistical significance was determined by permutation analyses to partition the genetic variation at different levels and the significance was determined by 1000 permutations.

The genetic distance, D_{XY}, incorporating the allelic frequencies for compared populations, and the mean number of nucleotide substitutions per site between the populations using one-parameter substitution model (Jukes and Cantor 1969) were estimated using DnaSP. As D_{XY} provides more information than F_{ST} regarding divergence comparisons, the amount of divergence for each population from the

remaining populations was calculated as a mean of pairwise values of D_{XY} for each population against the remaining populations.

4.3 Results

4.3.1 Identification of primers showing polymorphism

Twenty three pairs of universal primers reported in various studies were tested as detailed in Table 4.1. Of these, primers for the *Phytochrome A* gene (*PHYA* gene), *Aspartate aminotransaminase* gene (*Aat* gene) and *Glyceraldehyde-3-phosphate dehydrogenase* (*G3pdh* gene**)** produced amplification as shown in Fig. 4.1. The size range of the amplicons varied from 610 bp in *Aat* gene to 1.3 kb in *PHYA* gene as shown in Fig 4.1.

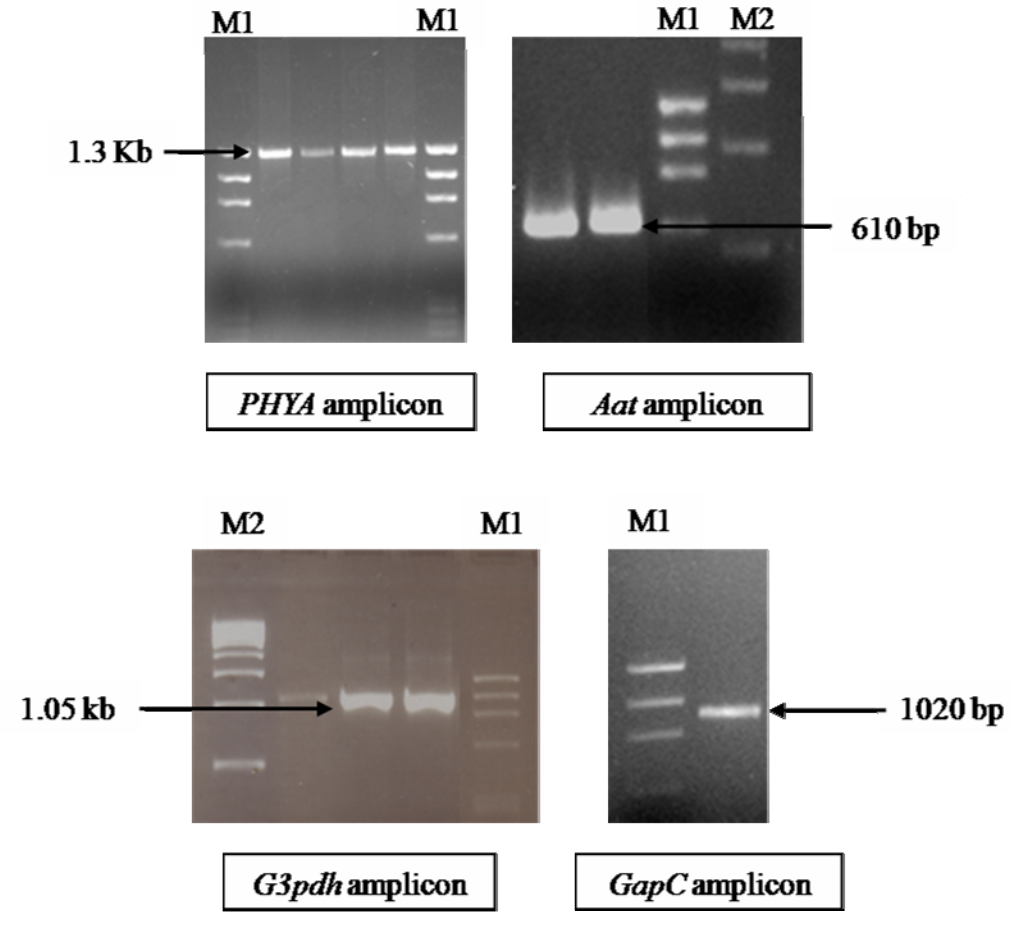


Fig. 4.1 Amplicons of the *PHYA*, *Aat, G3pdh and GapC* gene

Lane M1: Molecular weight marker ØX174/ *Hae*III digest and lane M2: molecular weight marker 500 bp ladder.

Of these three nuclear genic regions only *G3pdh* was considered for final analysis due to its uniform and reproducible amplification in all the samples of the subset used for screening and its utlity as reported in literature. The primers designed from Clone NCL2_12 produced amplicons of the size 1020 bp as shown in Fig. 4.1. An analysis of arbitrarily chosen samples from each population detected variation in sequences. Hence the analysis was extended to all the 195 samples from the 14 populations. The amplicons obtained from the study were subjected to sequence analysis using the primers described in Section 4.2.2.

4.3.2 Sequence analysis of *GapC* gene

A 881 bp sequence was obtained that comprised both intronic and exonic region of *GapC* (cytosolic *Glyceraldehyde 3 phosphate dehydrogenase*) gene. The sequences obtained from the 195 samples have been submitted to GenBank with accession nos. EU784448 to EU784642. The sequence analysis detected 35 parsimonious polymorphic sites in the 195 samples. In the present dataset, 18 transitions and 17 transversions were detected; indels were not detected in the *GapC* sequence analyzed. The haplotypes were defined by 35 parsimonious polymorphic sites and the composition of haplotypes for these sites is shown in Table 4.2. In all, 24 haplotypes were obtained and were assigned as N1 to N24 (Table 4.2 and 4.3). The distribution of these haplotypes within each population was variable, SPR and CHR were represented by the highest number of haplotypes (six) while, SAN and PON by only one haplotype as shown in Table 4.3 and Fig. 4.2.

4.3.3 Spatial distribution of haplotypes and diversity patterns

The number of haplotypes obtained in the present study ranged from one in PON and SAN populations to maximum of six in CHR and SPR populations. Fourteen unique haplotypes were identified among the 14 populations, of which five were distributed in NEI and nine were distributed in the SI region (one in EG and eight in WG). The haplotypes produced a clear cut distribution pattern. The haplotype N1 was represented by as many as 124 individuals with wide spread distribution among all the populations of *S. laurina* in WG and EG (*i.e*. SI) and accounted for 63.5% of the total haplotypes identified in the study. It was altogether absent from the NEI populations.

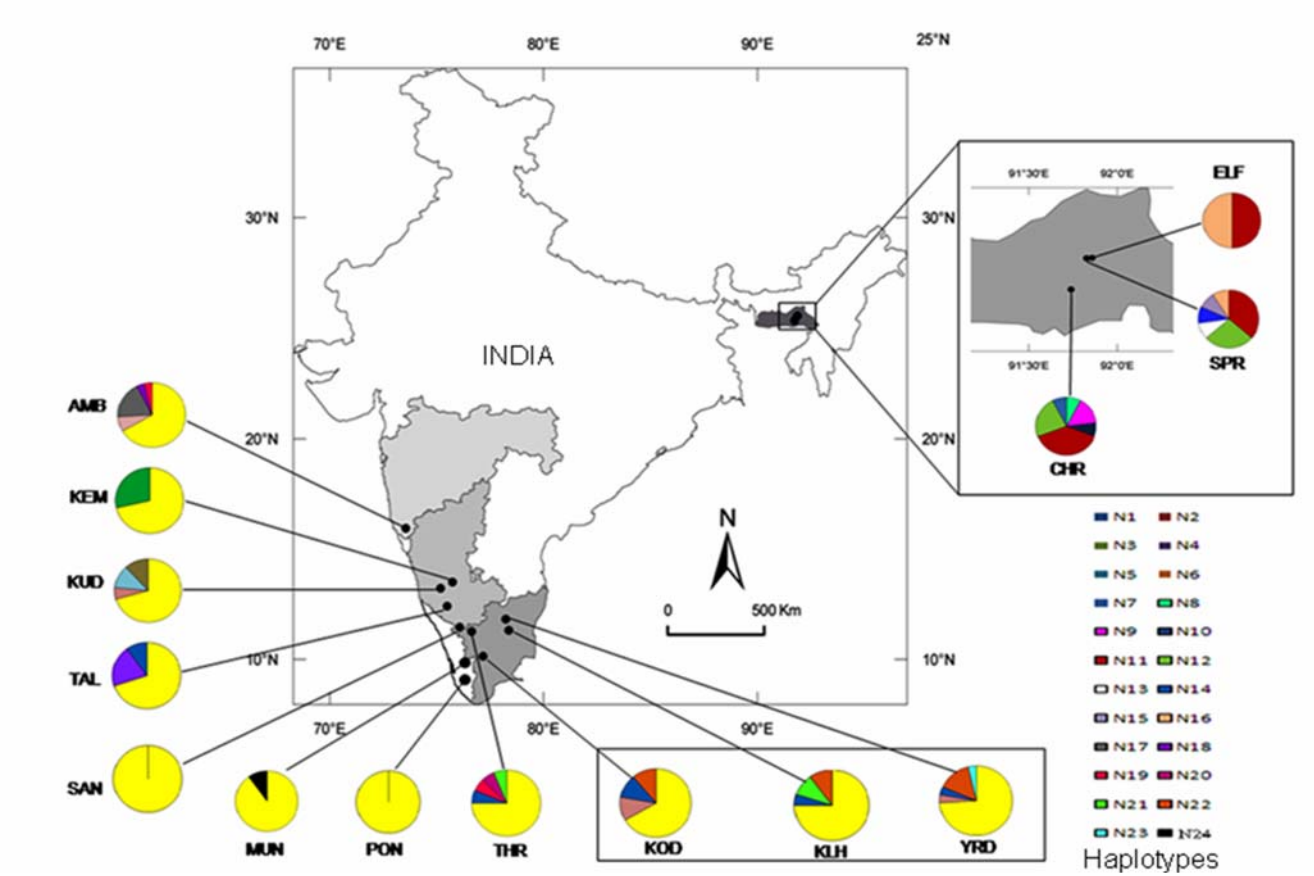


Fig. 4.2 Distribution of *S. laurina* haplotypes obtained with the analysis of *GapC* nuclear gene

The colour codes of the populations are given in the legend bar. The NEI populations have been projected out to provide a better resolution. The pie chart corresponds to each population and show the composition of haplotype and frequency of the haplotypes.

Table 4.2 Definition of nuclear DNA haplotypes in *S. laurina*

Haplotypes	51	63	92	102	105	131	137	148	152	158	168	170	201	259	278	371	398	429	516	636	654	667	677	743	623	836	839	858	862	868	872	874	877	878	881	Haplotypes frequencies
N1	C	G	T	A	C	T	C	T	C	C	A	G	A	G	A	C	C	A	T	C	T	C	A	A	A	A	A	A	G	A	A	A	C	T	124
N2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6
N3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	G	G	.	C	.	.	.	.	2
N4	T	C	.	.	A	G	G	G	.	.	.	.	T	.	G	T	.	.	.	.	C	.	G	.	.	.	.	.	.	.	.	.	.	.	2
N5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	4
N6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	2
N7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	5
N8	.	.	A	G	.	.	.	.	G	.	G	A	.	.	G	.	T	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	1
N9	.	.	A	.	.	.	.	.	G	T	G	A	.	.	G	.	T	.	.	T	C	.	.	.	G	.	G	.	.	.	.	.	.	.	2
N10	.	.	A	.	.	.	.	.	G	T	G	A	.	.	G	.	T	.	.	.	C	.	.	.	G	.	G	.	.	.	.	.	.	.	1
N11	.	.	A	.	.	.	.	.	G	T	G	A	.	.	G	.	T	.	.	T	C	.	.	.	G	.	.	.	A	.	.	.	.	.	12
N12	.	.	A	.	.	.	.	.	G	.	G	A	.	.	G	.	.	.	.	T	C	.	.	.	G	.	.	.	.	.	.	.	.	.	6
N13	.	.	A	G	.	.	.	.	G	T	G	A	.	.	G	.	T	.	.	.	C	.	G	.	.	.	.	.	.	.	.	.	.	.	2
N14	.	.	A	G	.	.	.	.	G	T	G	A	.	.	G	.	T	C	G	.	C	.	.	.	.	G	.	.	A	.	.	.	.	.	1
N15	.	.	A	G	.	.	.	.	G	T	G	A	.	.	G	.	T	C	.	.	C	.	.	.	.	G	.	.	A	.	.	.	.	.	1
N16	.	.	A	G	.	.	.	.	G	T	G	A	.	.	G	.	T	.	.	.	C	.	.	.	.	G	.	.	A	.	.	.	.	.	4
N17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	A	C	5
N18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
N19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	1
N20	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3
N21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	7
N22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
N23	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
N24	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
Wt given in network analysis	10	20	20	10	20	20	20	20	20	10	10	10	20	20	10	10	10	20	20	10	10	10	10	20	20	10	20	10	10	10	20	20	10	20	10	
Types of mutation	t	**T**	**T**	t	**T**	**T**	**T**	**T**	**T**	t	t	t	**T**	**T**	t	t	t	**T**	**T**	t	t	t	t	**T**	**T**	t	**T**	t	t	t	**T**	**T**	t	**T**	t	

Numbers in the first line refer to the position in the aligned *Gap*C region. Sequences are numbered from the 5' end to the 3' end. Dot (.) indicates that the character states are same as haplotype N1 in the respective site. T: transversion and t: transition.

Table 4.3 Haplotypes in *S. laurina* populations identified using *GapC* nuclear sequence

	Haplotypes identified																							
Location	N1*	N2	**N3**	**N4**	**N5**	**N6**	N7	**N8**	**N9**	**N10**	N11*	N12	N13	**N14**	**N15**	N16	**N17**	**N18**	N19	N20	N21	N22	**N23**	**N24**	Total
CHR	-	-	-	-	-	-	-	1	2	1	5	3	1	-	-	-	-	-	-	-	-	-	-	-	13
SPR	-	-	-	-	-	-	-	-	-	-	4	3	1	1	1	1	-	-	-	-	-	-	-	-	11
ELF	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	3	-	-	-	-	-	-	-	-	6
AMB	18	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	1	1	-	-	-	-	-	27
KUD	12	1	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17
KEM	10	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14
TAL	7	-	-	-	-	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10
SAN	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12
THR	12	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	1	1	-	-	-	16
MUN	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	10
PON	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
KLH	15	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	2	2	-	-	20
YRD	20	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	1	-	27
KOD	6	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	9
Total	124	5	2	2	4	2	5	1	2	1	12	6	2	1	1	4	5	1	2	1	3	7	1	1	195

The dominant haplotypes are indicated with asterisk mark and the unique haplotypes are indicated in bold.

Table 4.4 Diversity estimates and test statistics for *S. laurina* populations in India based on mutations identified in *GapC* gene

	No. of samples	No. of haplotypes	*h*	*π*	*D*	*D**	*F**	*Fs*
Total	195	24	0.588	0.00407	-1.10280	2.08412*	0.89678	-4.232
NEI	30	11	0.828	0.00447	-0.28162	1.22531	0.87949	-1.171
SI	165	16	0.441	0.00141	-2.31837**	1.38200	-0.18256	-8.356
WG	109	14	0.431	0.00180	-2.27514 **	1.05305	-0.32150	-5.132
EG	56	6	0.451	0.00060	-1.21787	0.10616	-0.36365	-2.948
CHR	13	6	0.821	0.00341	0.13553	0.50976	0.46900	-0.087
SPR	11	6	0.836	0.00371	0.27350	0.15974	0.21303	-0.321
ELF	6	2	0.600	0.00136	1.75324	1.27971	1.43414	1.938
AMB	27	4	0.527	0.00174	-0.04498	0.46634	0.36714	1.432
KUD	17	4	0.500	0.00389	-1.07003	1.22122	0.66354	3.297
KEM	14	2	0.440	0.00050	0.84228	0.71557	0.84693	0.944
TAL	10	3	0.511	0.00063	-0.69098	-0.28020	-0.42293	-0.594
SAN	12	1	-	-	-	-	-	-
THR	16	5	0.450	0.00057	-1.83088 ***	-2.40923 ***	-2.58115 ***	-3.314
MUN	10	2	0.200	0.00023	-1.11173	-1.24341	-1.34668	-0.339
PON	3	1	-	-	-	-	-	-
KLH	20	4	0.437	0.00054	-1.15810	-0.12425	-0.46829	-1.663
YRD	27	5	0.442	0.00061	-1.27896	-1.90894	-2.00162	-2.422
KOD	9	4	0.583	0.00076	-1.51297	-1.68268	-1.82046	-1.892

*Significant at P < 0.02, **Significant at P < 0.01, ***Significant at P < 0.005, Haplotype diversity (*h*), nucleotide diversity (π), Tajima's test of neutrality (*D*), Fu and Li's test statistic (*D**), Fu and Li's test statistic (*F**), Fu's test of neutrality (*F*s).

However, N11, the dominant haplotype had restricted distribution and was found only in NEI populations.

The haplotype diversity (*h*) across all the populations was 0.588 (Table 4.4). While SPR possessed the maximum haplotype diversity (h = 0.836), SAN and PON revealed no diversity. The NEI region was rich in haplotype diversity followed by EG and WG in the regionwise divided dataset. The nucleotide diversity for all the populations taken together was 0.00407. The KUD population had the highest nucleotide diversity (0.00389) while the populations SAN and PON did not possess any diversity as they were represented by a single haplotype only. The dataset divided into three regions showed presence of high nucleotide diversity in NEI region (0.00447) followed by WG and EG (Table 4.4).

Table 4.5 Gene diversity estimates in *S. laurina* populations

Diversity parameters	Total	NEI	SI	WG	EG
h_S	0.030 (±0.0204)	0.497 (±0.0946)	0.150 (±0.0407)	0.117 (±0.0527)	0.865 (±0.0682)
h_T	0.378 (±0.1039)	0.541 (±0.0579)	0.151 (±0.0572)	0.118 (±0.0772)	0.883 (±0.0913)
G_{ST}	0.921 (±0.0510)	0.082 (NC)	0.012 (NC)	0.012 (NC)	0.021 (NC)
v_S	0.039 (±0.0268)	0.240 (±0.0468)	0.178 (±0.0664)	0.105 (±0.0568)	1.045 (±0.2149)
v_T	0.511 (±0.1670)	0.279 (±0.0339)	0.185 (±0.0662)	0.105 (±0.0545)	1.147 (±0.2774)
N_{ST}	0.923 (±0.0578)	0.139 (NC)	0.036 (NC)	0.008 (NC)	0.089 (NC)

Within-population gene diversity (h_S), total gene diversity (h_T), interpopulation differentiation (G_{ST}) and their ordered counter parts v_S, v_T and N_{ST} (mean ± SE in parentheses) with analysis done in two stages: all populations taken together and populations divided into three regions in India.

For the *GapC* sequences, h_S (intrapopulational diversity) and h_T (total diversity) was low for the total populations being 0.030 (±0.0204) and 0.378 (±0.1039), respectively (Table 4.5). The genetic differentiation among populations (G_{ST}) was very high (0.921 ±0.0510). The values for v_S were also low (0.039) for the

total dataset while, v_T was found to be high (0.511). For both the subdivided datasets the values of v_S and v_T were low. The degree of substructure was also very high (N_{ST} = 0.923 ± 0.0578) when calculated for the total population. The permuted value of G_{ST} was lower than the N_{ST} at P > 0.05 indicating presence of geographic structure in the population and that, genetic distance are correlated with geographic distance. The subdivision of data into three regions showed a variable trend in distribution of diversity. Average gene diversity and total gene diversity were high in EG population, where as differentiation parameters were low in this dataset. NEI was found to be more diverse, with both the average gene diversity and total diversity being high when the dataset was divided into two groups NEI and SI (Table 4.5). Mantel's test indicated presence of correlation between genetic distance and geographical distance.

To determine if the *GapC* locus evolves neutrally, the neutrality test was performed for the entire dataset and the population datasets separately. These tests show different degrees of sensitivity to deviation from neutrality caused by demography or selection. Tajima's '*D*'- test of neutrality of sequence evolution revealed a significant negative value for the total dataset (*D* = -1.10280) as well for the subdivided dataset (Table 4.4). However, the values were significant for the regionwise divided dataset of SI and WG, while for others they were insignificant (Table 4.4). The *F*s was observed to be negative and differed from zero for the total dataset (*F*s = -4.232) as well as for the regionwise divided dataset (Table 4.4). The negative coefficients of Tajima's '*D*' and Fu's F_S may be suggestive of population growth, although we observed significant deviation from the neutral model. This result implies that there were more rare nucleotide site variants than would be expected under a neutral model of evolution (Rand 1996).

To analyze the partitioning of genetic variance in *S. laurina*, AMOVA was carried out in two steps. Division of data into three regions (NEI, WG and EG) showed that 75.65% of the genetic variation was among the regions. While 22.45% was within the populations. Division of data into two regions NEI and SI showed that 87.39% variation was among the regions and 11.89% was within the population, thus, pointing towards the differentiation of NEI populations from SI populations. Variation among the populations within region for both the datasets was very low while the F_{ST} values were high (Table 4.6).

Table 4.6 Analysis of molecular variance of *S. laurina* samples based on *GapC* sequences

The populations were divided into three regions NEI, WG and EG in section A. In section B populations are divided into two regions NEI and SI (includes WG and EG).

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Section A				
Among regions	2	226.974	1.98117	75.65
Among populations within regions	11	11.892	0.03617	1.38
Within populations	181	108.888	0.60159	22.45
Total	194	347.754	2.61894	
Fixation Indices F_{SC}: 0.05672, F_{ST}: 0.77029, F_{CT}: 0.75648				
Section B				
Among regions	1	225.578	4.42268	87.39
Among populations within regions	12	13.288	0.03661	0.72
Within populations	181	108.888	0.60159	11.89
Total	194	347.754	5.06089	
Fixation Indices F_{SC}: 0.05737, F_{ST}: 0.88113, F_{CT}: 0.97389				

A mismatch distribution of substitutional differences between the pairs of haplotypes was calculated within the populations and compared with a fit to the Poisson model using DnaSP (Rozas and Rozas 1997). A multimodal mismatch distribution characterizing old population of a constant size (Fig. 4.3) was obtained. The raggedness index (*r*) could be used to quantify the smoothness of mismatch distribution (Harpending 1994). Under the population growth model, the *r* values are expected to be typically low and can be tested for deviation from the constant population size model by simulation implemented in DnaSP using mismatch distribution. In present case the raggedness index was found to be 0.1012. Moreover, possible recombinations were investigated by the minimum number of recombination events (Rm) based on the four-gamete test (Hudson and Kaplan 1995) which showed presence of a minimum seven recombination events in *S. laurina* populations.

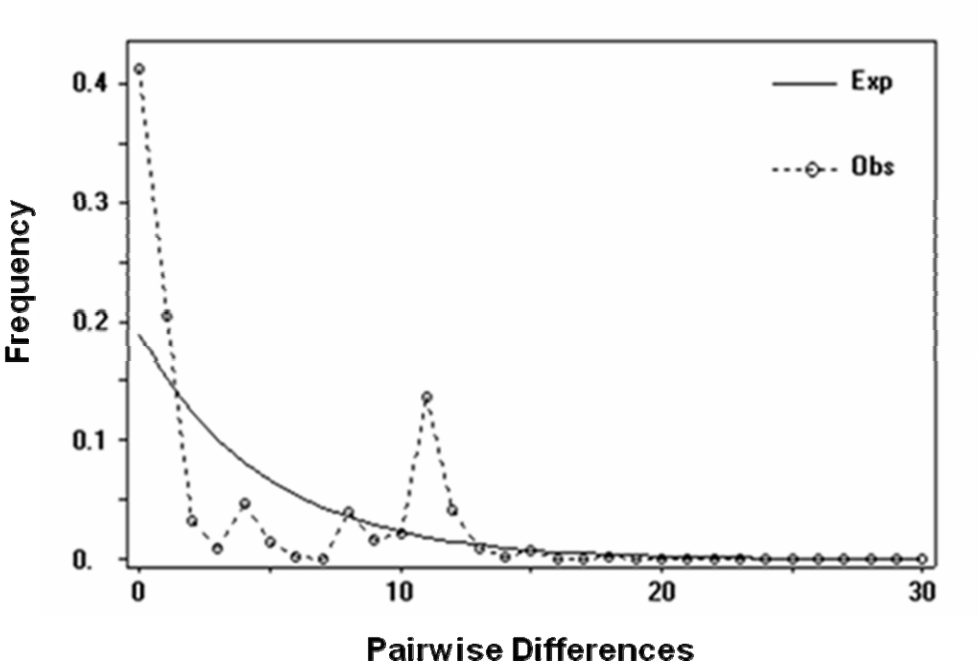


Fig. 4.3 Mismatch distribution established for *S. laurina*, the thin line represents the expected mismatch distribution of a stationary population

The dotted line represents the observed mismatch from segregating sites of the aligned sequences of *GapC* gene.

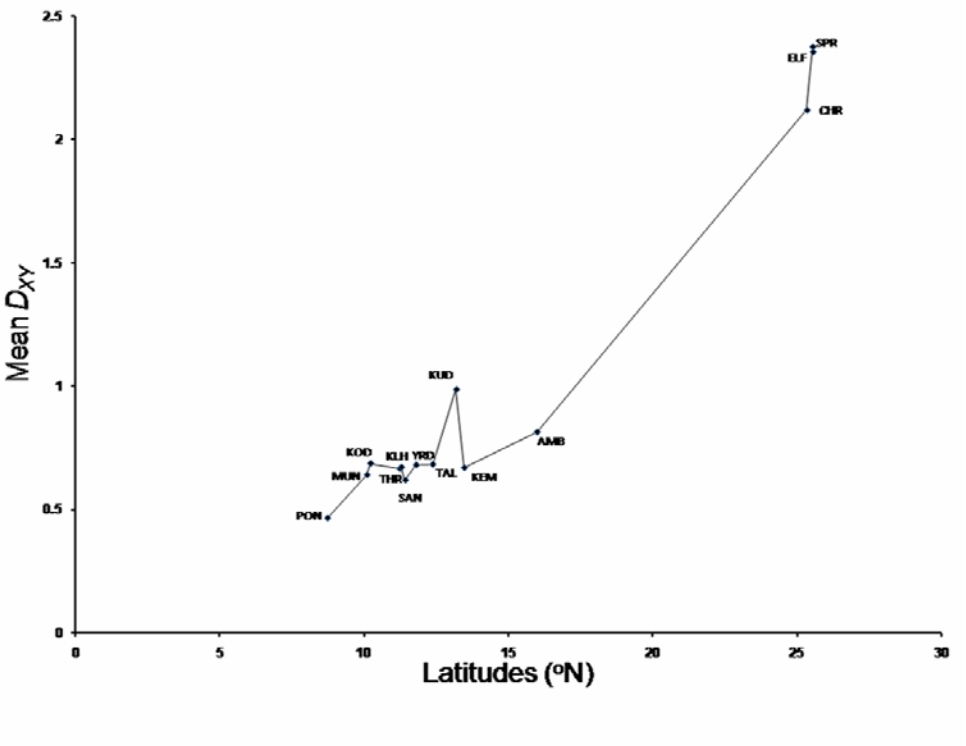


Fig. 4.4 Plot of the mean D_{XY} values of each population compared with all other vs the population latitude in *S. laurina* using *Gap*C sequences

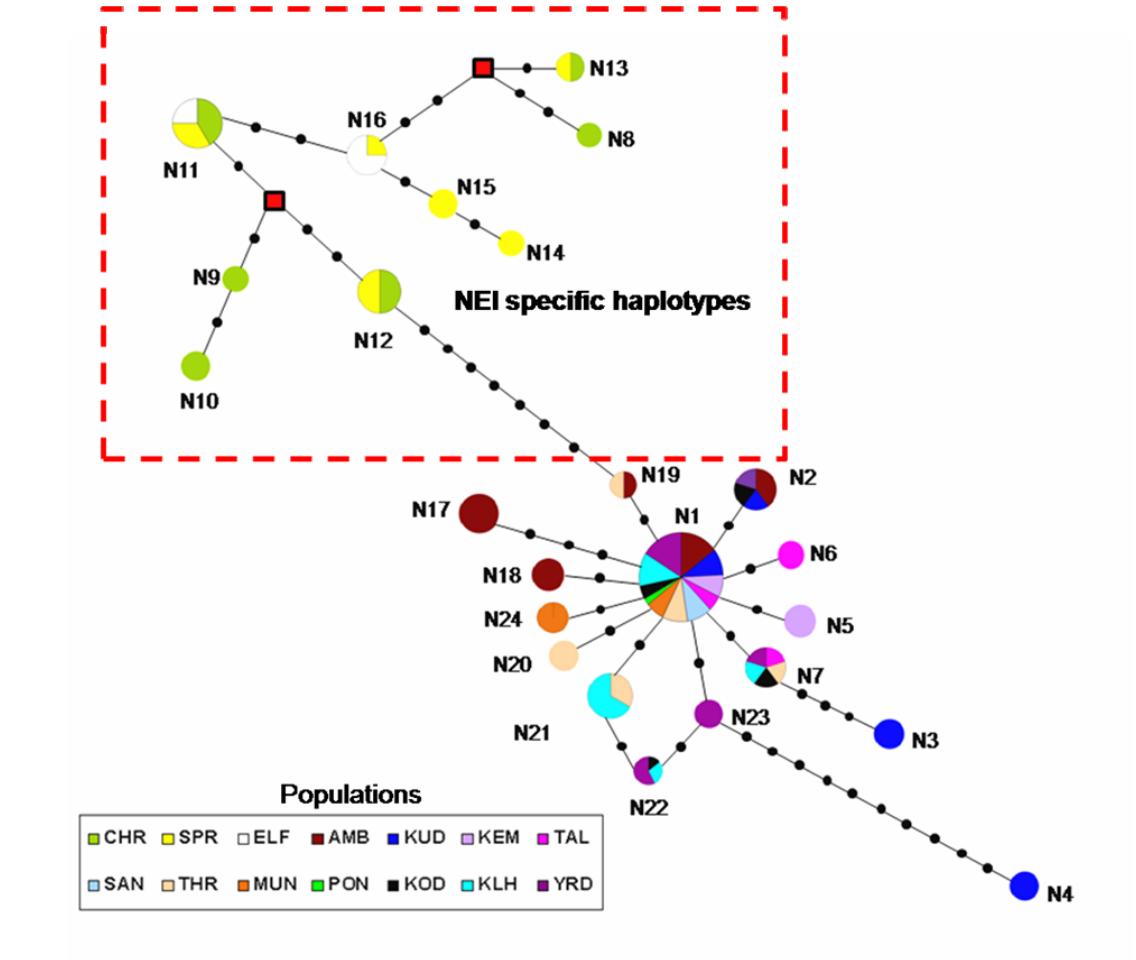


Fig. 4.5 Median-joining network of *S. laurina* haplotypes based on *GapC* sequences

Each piechart represents one haplotype, and their sizes are near about proportional to their frequencies. Colours within the circle represent contribution of different populations to that haplotype. Small black circles in the connecting line indicate the mutations involved. A key to the colour codes of each population is provided separately. Square red boxes represent the medians.

4.3.4 Pairwise D_{XY} statistics according to nucleotide differences and population divergence

The mean D_{XY} values according to the *GapC* region of each population compared with all other populations were small and varied from 0.463562 (PON) to 2.376504 (SPR). Only one peak of average D_{XY} values was found at KUD (0.98502) which decreased to low values for the KEM and AMB populations followed by a sharp upsurge till SPR population (Fig. 4.3). The genetic divergence among SI populations was small, and was absent between the WG and EG populations while the NEI population showed differentiation from the SI populations as depicted in Fig. 4.4**.**

4.3.5 Relationships among the haplotypes

The Median-joining network drawn using the Network software showed grouping of haplotypes according to two broad groups NEI and SI (Fig. 4.5). N1, the dominant haplotype of SI occupied an interior position in the network while the dominant haplotype of NEI, N11 was connected to N1 via N19 and N12 with 10 mutational steps. The haplotypes N12 and N19 were the connecting links between the two dominant haplotypes. The haplotypes obtained in SI formed a star shaped pattern with the exception of N3, N4 and N23. In case of NEI, the distribution of haplotypes was not the same as SI. The haplotypes N8, N13, N14 and N15 were linked to N16. The EG specific haplotypes N22 and N23 were linked to N4 haplotype of KUD population. The network showed presence of two medians. The median mv1 was linked to N9 and N10 which were unique haplotypes of CHR population. The haplotypes N13 (of CHR and SPR) and N8 unique haplotypes from CHR were linked to the network via the median mv2.

4.4 Discussion

Low-copy nuclear genes with rapidly evolving introns provide a good source of DNA sequence data for phylogenetic reconstruction at the intraspecific level (Schaal and Olsen 2000). They have already shed light on the evolution of a few crop plants. Olsen and Schaal (1999) studied the domestication of cassava (*Manihot esculenta* subsp. *esculenta*) using sequences of a single-copy nuclear gene *G3pdh*. They observed that the level of phylogenetically informative variation in the *G3pdh* region far exceeded the levels typically observed in the organellar genomes of plants at the

intraspecific level. Comparison of *G3pdh* haplotypes of the crop and wild species indicated that *Manihot esculenta* subsp. *flabellifolia* alone could account for the genetic variation observed in cassava.

4.4.1 Level and pattern of nucleotide variation at *GapC* region in *S. laurina*

Patterns of DNA polymorphism at the *GapC* locus were analyzed, in samples of *S. laurina* collected over its entire range in India. The nucleotide diversity was fairly low (π = 0.00407) for the populations taken together. Similar level of nucleotide diversity was observed in *Solanum pimpinellifolium* (π = 0.0064; Caicedo and Schaal, 2004). *Abies kawakamii* (π = 0.00648; Shih et al 2007) and *Citrullus colocynthis* (π = 0.0049; Dane et al. 2007) when they were analyzed using the *GapC* gene. However, nucleotide diversity was high for *GapC* region in *Manihot esculenta* (π = 0.49; Olsen 2002) and *Passiflora alata* (π = 0.01051; Koehler-Santos et al. 2006b).

4.4.2 Genetic diversity and genetic differentiation in *GapC* region in *S. laurina* populations

The haplotype diversity was found to be moderate when calculated for the populations taken together. The regionwise division of diversity was not uniform, NEI being more diverse possessing high number of unique haplotypes/alleles than SI. The populations were represented by variable number of haplotypes, some being very rich in haplotype diversity *e.g.* SPR (h = 0.836) while the populations like PON and SAN were represented by only one haplotype with no nucleotide diversity. The low haplotype diversity in MUN and the presence of only one haplotype in SAN and PON resulted in lowering of the average haplotype diversity in *S. laurina* when all the populations were taken together. The haplotype diversity was high in NEI population though they were represented by only 30 individuals indicating accumulation of high number of mutations in NEI region. The variable number of haplotypes across the populations and presence of population specific haplotypes resulted in high differentiation among the populations in *S. laurina.*

The average genetic diversity within population (h_S) was low for the total dataset; probably because of presence of two monomorphic populations along with the low number of haplotypes detected in populations like ELF, KEM and MUN. Additionally, in the regionwise divided dataset, the overall number of haplotypes detected were low, maximum being six. The total diversity (h_T) followed the same

trend as average genetic diversity (h_S). In the EG population, three common haplotypes were present in all the three populations though very less number of unique haplotypes were detected thus contributing to high within population diversity as well as total diversity in EG when compared with the other regions.

The populations studied were significantly differentiated at the *GapC* locus. Highly significant values of F_{ST} and G_{ST} were obtained for all the population pairwise comparisons, independent of whether the compared populations are from different regions or not (Table 4.5). The differentiation was low for the regionwise dataset; however, for the total population it was high probably because of presence of two divergent haplotypes in NEI and SI.

4.4.3 Divergence among populations

In this study, selectively neutral genetic variation at nuclear genomes was used in order to draw inferences about population divergence and dispersal. The nearly neutral model (Ohta 2002) predicts that lineages with small population size have higher substitution rates because mutations are more easily fixed due to genetic drift.

The degree of average genetic divergence of each population in comparison with the remaining populations can be used to examine the consequence of historical and contemporary geographical population subdivisions on evolutionary processes (Johnson et al. 2000). It is important for reconstructing phylogeographical histories that have evolved during pre- and postcolonization events (Grant and Grant 1997).

In the study, the genetic divergence in terms of D_{XY} of each population from the remaining populations was obtained for the *GapC* marker. The NEI populations diverged significantly from the SI populations. A closer examination of the SI population revealed that KUD was divergent from the other SI populations. There was no evidence of divergence among the populations of EG and WG (Fig. 4.4). The deeply divergent lineages restricted to portions of the species range are indicative of imprint of drift or selections. The species is likely to have undergone habitat shifts and climatic oscillations resulting in the population fragmentation. Similar observations have been made in *Hordeum* and *Manihot* studied using *G3pdh* sequence information (Olson and Schaal 1999; Morrell et al. 2003).

4.4.4 Test of neutrality: interpretation of demographic processes

In order to test whether nucleotide variation has been shaped by natural selection the neutrality tests were applied to the dataset. The *GapC* site-frequency spectrum exhibited a significant excess of rare variants relative to that expected from panmictic and neutral population. In the sample of 195 accessions, Tajima's *D*, Fu and Li's *D** and *F** revealed a significant excess of rare variants relative to neutral expectation. Consistent with the results of the Tajima's *D*, Fu and Li's *D** and *F** were negative in sign, though only Tajima's *D* was significant. Several aspects of the data demonstrate that the pattern of nucleotide variation at *GapC* is incompatible with a neutral equilibrium model. The overall pattern of nucleotide variation was compatible with the expectation of a selective sweep (Charlesworth et al. 1993; Braveman et al. 1995), wherein a new, favourable mutation increases in frequency, eventually becoming fixed in population or species. Population growth (or genetic hitchhiking, which can produce a similar signal) can be detected from the patterns of significance of these tests: given a population expansion, F_S is expected to deviate significantly from the null expectation, while *D** and *F** are less sensitive to population growth and usually do not show significant deviation. Non-significant values of *D** and *F** justify exclusion of background selection (Ramos-Onsins and Rozas 2002). During a selective sweep, linked neutral variants are also swept to fixation, resulting in reduced polymorphism near the site under selection (Smith and Haigh 1974; Kreitman 2000). New mutations that occur following a selective sweep create polymorphisms initially present at low frequency, leaving a characteristic footprint in the site frequency spectrum that can be revealed by Tajima's *D* and Fu and Li's *D** statistics.

A star-like distribution of haplotypes indicating population expansion in the SI region is also supported by Tajima's *D* test. This type of pattern in Tajima's *D* test has been observed for several local subtropical and montane species *e.g. Quercus glauca* (Hunag et al. 2002), *Cunninghamia konishii* (Chung et al. 2004), *Castanopsis carlesii* (Cheng et al. 2005a) and *Machilus thunbergii* (Wu et al. 2006). The Tajima's *D* test of neutrality evaluates the significance of the amount of recent mutations. Significant excess of such mutations (negative values of Tajima's *D*) may suggest selective sweep or population expansion events, while significant lack of recent mutations (positive values) could be attributed to balancing selection, population structure, or decline.

The star-like phylogeny of haplotypes indicates positive selection or a population bottleneck in the historical demography of *S. laurina*.

Secondly, historical demographic expansions were investigated by examination of frequency distributions of pairwise differences between sequences (mismatch distribution). The mismatch distribution is a widely accepted molecular approach to examine genealogical diversity by calculating the number of mutational differences between individual haplotypes within a population (Rogers and Harpending 1992; Johannesen et al. 2005). The *GapC* marker presented a large number of pairwise mismatch differences in this study. The multimodal mismatch distribution obtained in the study suggested that the populations might either be old and are at demographic equilibrium or likely to have high mutation rate.

4.4.5 Historical processes

In a geographical context, the structuring of genetic variation is a product of current genetic exchange within a species, as well as historical relationships between the populations. By taking into account the genealogical relationships between haplotypes as well as their geographical distribution, phylogeographical methods can potentially determine the historical and recurrent population-level processes that shape current patterns of variation.

The presence of genealogical structure in the geographical distribution of the *GapC* haplotypes was the most striking observation in the present data. Geographical structuring which is evident by visual inspection of the haplotype network (Fig. 4.5) indicates that the populations that are geographically proximal, share closely related or derived haplotypes. This pattern of variation is consistent with a possible fragmentation of an ancestral population, followed by gradual accumulation of new mutations by different populations. The presence of two dominant haplotypes in two specific regions implies that the populations in these regions might have originated from two different source populations or refugia.

The strong differentiation between the NEI and SI populations indicate long term persistence of these plants in the respective regions. This differentiation can happen only if the populations spread from two different refugia. This is indicated by the high differentiation value of G_{ST}. As *S. laurina* is long living perennial plant with long generation time, the mutations accumulated by the NEI and SI populations

(which are sharply different as depicted by presence of different haplotypes in NEI and SI) might have taken a long time. This differentiation probably predates last Pleistocene ice age. The data obtained suggests that the populations are in demographic equilibrium. Thus, the mutations accumulated in the populations might have occurred over many years reaching the present genetic structure. Since the age of *S. laurina* in Indian biogeography is unknown, as the fossils evidence is lacking, it is difficult to comment on the time since it existed in the Indian peninsula and NEI.

In *S. laurina* populations, a relatively large number of local populations were fixed for one haplotype. Founder events, in particular, can lead to the local spread of a single haplotype, which may then be followed by migration and natural selection increasing the genetic variation. It is expected that due to founder effects, populations towards the periphery of the species distribution are more likely to be fixed for a certain haplotype than more central populations. This has been suggested for *Arabidopsis* populations from Great Britain, Japan and Scandinavia (Abbott and Gomes 1989; Todokoro et al. 1995; Kuittinen et al. 1997), which is true in the case of PON populations in the present study. Also, the monomorphic populations were found both in the peripheral regions (PON) and within the centre of the species range (SAN).

S. laurina is a perennial cross pollinating plant with dissemination occurring primarily through seeds. It is generally assumed that dispersal occurs through self-fertilized seeds. However, it may also be mediated by wind which could be important for long-distance gene flow. For example, the existence of the haplotype N2 in AMB, KUD, KOD and YRD as well as the N21 in KLH and THR might be a result of long distance dispersal as the haplotypes are present in specific populations but are absent from populations proximal to these populations. However, it is not clear whether this is an important mechanism for gene flow between local populations.

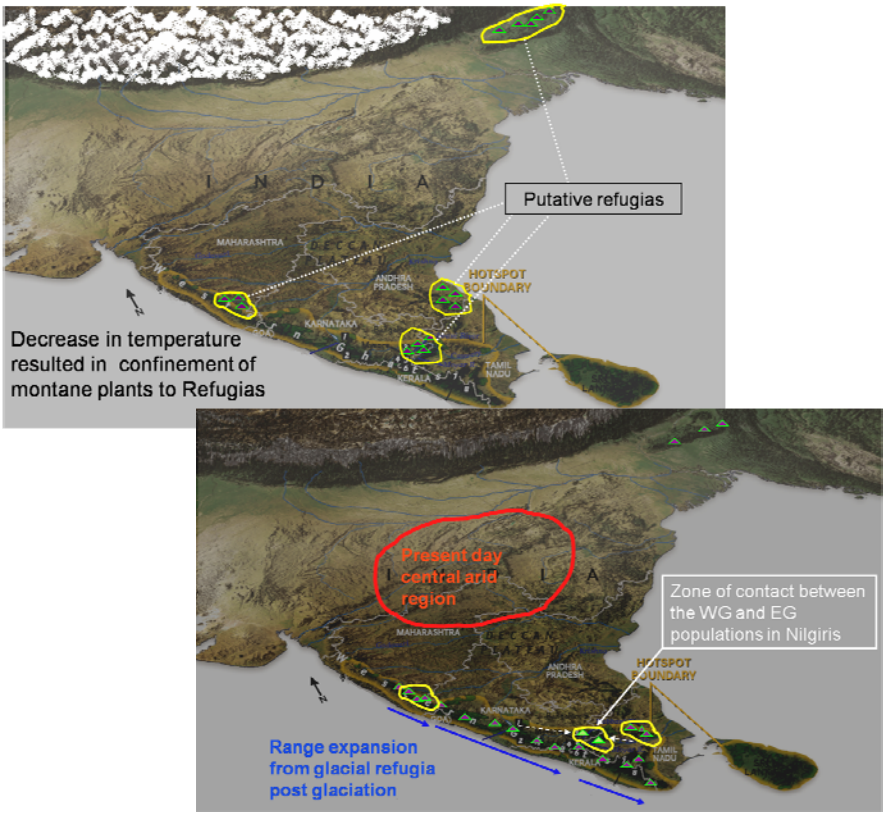
The effect of glaciation and postglacial climatic changes on the montane ecosystem is still largely unknown. The climatic fluctuations of the Pleistocene are considered to play a major role in the distribution of flora and fauna (Meher-Homji 1972, 1975). The increase in the aridity post glaciation in the central Indian region probably did not favour the geneflow or migration between populations of SI and NEI. However, the less xeric condition across WG and EG probably allowed the

unhindered spread of the haplotypes either *via* local dispersal or long distance dispersal resulting in the dominance of a single type of haplotype across SI region.

The populations of SI following their expansion or spread into the present distribution areas might be facing different kinds of ecological pressures. As the plant is known to grow in forest fringes, it is highly susceptible to anthropogenic activities. The recent increase in urbanization and deforestation may have exacerbated the local extinction of haplotypes in this region. Thus, the distribution of *GapC* haplotypes in *S. laurina* reflects the interaction between historical processes and current population level processes. The low diversity detected in the present study hindered the identification of refugial pockets for *S. laurina* populations using *GapC* markers in both SI and NEI.

Chapter 5

Thesis summary and future directions



OUTLINE

5.1 Phylogeography in India

5.2 Vicariance: a possible model explaining phylogeography of *S. laurina* in India as deciphered from PCR-RFLP analysis using cytoplasmic markers

5.3 Possible refugia of *S. laurina*: evidences from sequence analysis of the chloroplast genome

5.4 Establishment of new populations through range expansion in *S. laurina*: evidences from sequence analysis of nuclear genome

5.5 Synthesis of historical events for the distribution of *S. laurina* in India

5.6 Factors affecting present day distribution of *S. laurina* in India

5.7 *S. laurina*: a key to analyse and conserve the diversity present in disjunct populations in India

5.8 Future directions

5.1 Phylogeography in India

Since the inception of the term ‘‘phylogeography’’, there has been a continued interest in knowing the distribution of organismal diversity. This is especially true for plants that experienced range fluctuations forced by major climate changes, which were caused by glaciation and deglaciation cycles (Hewitt 2000). The present day distribution of plants in India has been implicated to be the result of Pleistocene glaciation. However, very few efforts have been made to characterize the historical events contributing to the present day disjunct distribution of wetzone species in Indian biogeography. Further, no molecular evidences have been provided till date in favour or against the models for population structure of such plants. The research carried out in the purview of the present thesis could assess the models by capturing molecular evidences in favour/against the postulates to identify the best fit model explaining disjunct distribution and the effect of glaciation on montane plants. Genetic diversity of the montane tree *Symplocos laurina* Wall. collected from the two biodiversity hotspots Indo-Burma and Western Ghats, along with Eastern Ghats was analyzed using cytoplasmic as well as nuclear markers. A comprehensive view of evolutionary relationships and biogeography for this ecologically important plant species has been provided below.

5.2 Vicariance: a possible model explaining phylogeography of *S. laurina* in India as deciphered from PCR-RFLP analysis using cytoplasmic markers

The genetic diversity analysis of twelve populations (218 plants) of *S. laurina* with three cpDNA and eight mtDNA specific primer-enzyme combinations analyzed using PCR-RFLP technique detected nine cpDNA (chlorotypes) and 24 mtDNA (mitotypes) haplotypes. The evidences obtained from the statistical parameters such as high total diversity (cpDNA h_T = 0.653; mtDNA h_T = 0.850), high differentiation (cpDNA G_{ST} = 0.734; mtDNA G_{ST} = 0.642) and partitioning of variation among the regions strongly suggest the persistence of the present species in these locations prior to last glaciation. The haplotype network for both the genomes revealed presence of two main groups: Northeastern-Western Ghat (NEI-WG) group and Eastern Ghat (EG), group largely without overlapping distributions. The study refutes the hypothesis of post glacial migration of plants from NEI to SI and provides evidence in support of the presence of *S. laurina* in SI preglaciations supporting the relict flora or vicariance

model. The study also identifies climatic fluctuations of Pleistocene glaciation and postglaciation as the crucial decisive factors influencing the distribution of *S. laurina* in montane forests (more specifically in Shola communities) of India.

5.3 Possible refugia of *S. laurina*: evidences from sequence analysis of the chloroplast genome

Sequence analysis of the three noncoding regions of chloroplast genome *trn*C-*pet*N (542 bp), *trn*H-*psb*A (404 bp) and *trn*D-*trn*T (968 bp) of chloroplast genome captured more mutations than PCR-RFLP. The sequence analysis of 1914 bp detected 29 informative sites and 25 haplotypes from fourteen populations. Phylogeographic analysis did not show a clear-cut association between the haplotype network and geographical distribution of the haplotypes. The study revealed existence of three genetically differentiated groups, NEI-northern WG group, southern group and EG group with overlapping geographical ranges. The high differentiation obtained for the dataset confirms the long time separation of the populations. The studies suggest that during the glacial times, the *S. laurina* population may have retreated to lower altitudes and the population might have survived in putative refugial pockets in Amboli, Thaishola of WG and in Northeast India from which new populations could have established postglaciation by range expansion upon arrival of congenial climatic conditions.

5.4 Establishment of new populations through range expansion in *S. laurina*: evidences from sequence analysis of nuclear genome

To characterize the present refuge populations of *S. laurina* intraspecific genetic diversity was assessed using the *GapC* nuclear gene. Using the 881 bp sequence of *GapC*, 24 haplotypes was detected among 195 individuals. Low within population genetic diversity, high differentiation, a high number of population specific haplotypes and neutral evolution characterized the populations of *S. laurina*. The analysis of molecular variance (AMOVA) indicated presence of geographic structure within the haplotype distribution. The statistical analysis (mismatch distribution, median-joining network) showed the signature of range expansions. Thus, the nuclear genome analysis of *S. laurina* populations using *GapC* region identified expansion of populations from previously existing populations in India.

5.5 Synthesis of historical events for the distribution of *S. laurina* in India

The analysis of genetic variation and the high differentiation obtained in chloroplast, mitochondrial and nuclear genome analyses led us to identify vicariance as the most probable model explaining the distribution of *S. laurina* in India. The study infers existence of *S. laurina* in the NEI and SI region long before the formation of central arid zone. The populations of *S. laurina* are thus relict populations as aptly described by Meher-Homji (1975). While PCR-RFLP analysis of cp and mt DNA identified the WG and EG regions as genetically distinct, a more sensitive sequence analysis of cpDNA identified the suture zone (contact zone) between WG and EG in THR population. The analyis of nuclear genome, which is considered to evolve faster than cytoplasmic genome, detected recent geneflow among the populations of WG and EG and total separation of NEI from SI region. Interestingly, the analysis of chloroplast genome detected presence of linking haplotype between the two biodiversity hotspot regions (NEI and WG), which hints towards existence of preglacial geneflow between the regions. Thus, the study suggests existence of *S. laurina* in NEI and SI long before LGM with occasional contact between them. The climatic changes during LGM severely affected the distributional pattern of the plant species restricting them in congenial habitat pockets (refugia). During the events like altitudinal shifts, phenomenon like bottlenecks might have wiped out few populations in the process. The arrival of the congenial temperature postglacial provided the plant species the opportunity to expand via various processes from the refugial pockets. The study identifies Amboli, Thaishola and Northeast India (Khasi Hills) as the probable refugia from which *S. laurina* might have expanded to occupy new habitats. The range expansion might have taken place through bird/animal mediated long as well as short dispersals. The genetic structure of the newly established populations depended upon various factors ranging from biotic to abiotic; the frost, fire and wind being the most prominent ones; as highlighted in Chapter 1. Postglacial changes in the climate as well as geography might have resulted in the genetic isolation of NEI populations from the southern populations as detected by the nuclear gene analysis. However, the barriers in geneflow within SI were overcome postglaciation as indicated by the presence of a common haplotype in EG and WG by nuclear gene analysis. Further, the anthropogenic activities like deforestation, collection of materials for medicinal purpose etc. might have further enhanced the genetic isolation of populations. Thus,

this study captures the genetic processes that have shaped the present day distribution of haplotypes in populations of *S. laurina*, providing us a scope to extrapolate and base to suggest similar kind of effects on the genetic structure of other plants belonging to shola fringes in the two biodiversity hotspots.

5.6 Factors affecting present day distribution of *S. laurina* in India

S. laurina is a part of montane ecosystem complex also known as sholas, and is a principal tree species or even the climax species in subtropical evergreen broad-leaved forests (Guo et al. 1997). The species is important both in the tree and shrub layers (Ma et al. 2002) and therefore, very important for the regeneration of evergreen broad-leaved forests. The plant is known to reproduce by either clonal or sexual reproduction and the fitness of the two reproductive methods varies with different habitats. Clonal reproduction predominates in habitats with enough water and fertility and high canopy cover, while sexual reproduction predominates in environments with less water, fertility and stronger sunlight. The bottlenecks of the two reproductive methods are different; the bottleneck is from seed to seedling for sexual reproduction, while it is seedling to mature plant for clonal reproduction. *S. laurina* occupies a new habitat first by invasion of grown-up plantlets and then occupies the space quickly through ramets because of the ease of clonal reproduction. The plant being among the pioneering stock is subjected to a wide range of factors namely frost, fire, wind, precipitation, differences in day and night temperatures etc. Moreover, the species is known to grow in the fringes (border) of the shola forests, making it susceptible to anthropogenic destruction (by exploitation for firewood as well as for medicinal properties). All these factors severly affect establishment of the plant in its natural habitat.

5.7 *S. laurina*: a key to analyse and conserve the diversity present in disjunct populations in India

The importance of conservation could be ascribed to a region on the basis of past evolutionary history or future evolutionary potential (Bowen 1999). The elevated genetic variation in the NEI and WG populations of *S. laurina* highlights the importance of these populations as reservoir of genetic diversity. As refugial areas, these are valuable populations under a global climate-change scenario. The approaches for setting conservation priorities are becoming a matter of concern as the

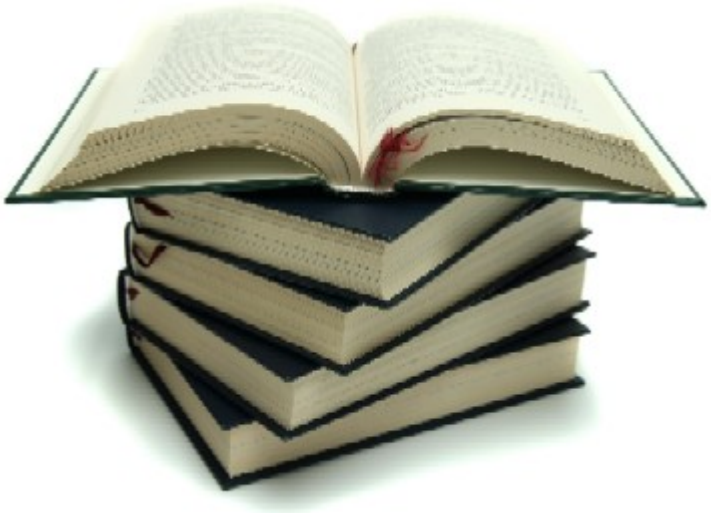
accelerating and potentially catastrophic loss of biotic diversity, unlike other environmental threats, is irreversible (Mittermeier et al. 1998). Primary forests of Asia, particularly those of the Western Ghats and the Eastern Ghats of peninsular India, are disappearing at an alarming rate (*e.g.* 25.6% of the total forest area, 19.5% of the dense forest cover and 33.2% of open forests of WG have undergone deforestation during the last 22 years) due to anthropogenic activities and are replaced by forests comprising inferior species or due to changed land use pattern of these forests (Jha et al. 2000; Parthasarathy 1999). The disappearance of tropical forests comes at a time when our knowledge on their structure and dynamics is woefully inadequate (Hubbell and Foster 1992). In particular, the detection of centers of genetic diversity could be useful for designing conservation strategies of genetically diverse long-lived species such as *S. laurina*. The knowledge of population structure is important for *ex-situ* and *in-situ* conservation of natural populations (Williams and Hamrick 1996) by maintaining the total evolutionary potential and minimizing consanguinity. Also, as *S. laurina* is a good colonizer: local genetic stock should be used for the restoration of degraded habitats. Thus, planning of conservation strategies on the basis of population structure can be initiated for the present system under study.

5.8 Future directions

Comparative phylogeographic approaches are based on the assumption that species with similar habitats and a comparable distribution might share a similar biogeographic history (Gugerli and Holderegger 2001). This approach has been successfully used to determine the Pleistocene refugia on a local scale for high alpine taxa (Schönswetter et al. 2002, 2003; Tribsch et al. 2002; Tribsch and Schönswetter 2003). These patterns indicated several Pleistocene glacial refugia in Europe (Konnert and Bergmann 1995; Hewitt 1996, 1999; Comes and Kadereit 1998; Taberlet et al. 1998) as well as several glacial colonization routes (Koch et al. 1998; Taberlet et al. 1998; Sharbel et al. 2000). However, phylogeographic studies of species occurring in the shola forest are still scarce. It is apparent that more phylogeographic studies with other species from the shola community are needed to obtain better understanding of the influence of the Quaternary climatic changes on the evolutionary history of the flora of this biome. Because *S. laurina* has a strong affinity and association with other shola community plants (like *Neolitsea* spp., *Olea* spp. etc.), we predict that a similar

pattern of divergence will be found in these species. Selection of species with similar distributional range corresponding to *S. laurina* will allow better insight into the relationships of possible refugia and the colonization patterns among these species. Since *S. laurina* has much more restricted distribution than most of its forest associates, analysis of other species might suggest different refugial compartments. Comparative phylogeography will not only help defining conservation-worthy genetic variation in these species, but will also circumscribe regions likely to possess abundant cryptic biodiversity in both common and non-charistmatic species.

Bibliography



List of references cited in the thesis

Abbott RJ and Brochmann C (2003) History and evolution of the arctic flora: in the footsteps of Eric Hultén. Molecular Ecology 11: 299-313.

Abbott RJ and Comes HP (2004) Evolution in the Arctic: a phylogeographic analysis of the circumarctic plant, *Saxifraga oppositifolia* (purple saxifrage). New Phytologist 161: 211-224.

Abbott RJ and Gomes MF (1989) Population structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. Heredity 62: 411-418.

Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K and Balfour J (2000) Molecular analysis of plant migration and refugia in the arctic. Science 289: 1343-1346.

Abdulali H (1949) Some peculiarities of avi-faunal distribution in Peninsular India. Proceedings of the National Institute of Sciences of India 15: 387-393.

Afzal-Rafii Z and Dodd RS (2007) Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in Western Europe. Molecular Ecology 16: 723-736.

Agarwal SC, Madan US, Chinamani S and Rege ND (1961) Ecological studies in the Nilgiris. Indian Forester 87: 376-389.

Aguade M (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H g*enes, in *Arabidopsis thaliana.* Molecular Biology and Evolution 18: 1-9.

Aizawa M, Yoshimaru H, Saito H, Katsuki T, Kawahara T, Kitamura K, Shi F and Kaji M (2007) Phylogeography of a northeast Asian spruce, *Picea jezoensis*, inferred from genetic variation observed in organelle DNA markers. Molecular Ecology 16: 3393-3405.

Albaladejo RG and Aparicio A (2007) Population genetic structure and hybridization patterns in the Mediterranean endemics *Phlomis lychnitis* and *P. crinita* (Lamiaceae). Annals of Botany 100: 735-746.

Ali S (1935) The ornithology of Travancore and Cochin. Journal of Bombay Natural History Society 37: 814-843.

Almeida SM (1990) The Flora of Savantwadi, Maharashtra, India. Scientific Publishers, Jodhpur, India.

Anderson LL, Hu FS, Nelson DM, Petit RJ and Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. Proceedings of the National Academy of Sciences USA 103: 12447-12450.

Apte GS, Bahulikar RA, Kulkarni RS, Lagu MD, Kulkarni BG, Suresh HS, Rao PSN and Gupta VS (2006) Genetic diversity analysis in *Gaultheria fragrantissima* Wall. (Ericaceae) from the two biodiversity hotspots in India using ISSR markers. Current Science 91: 1634-1640.

Ashton PS and Gunatilleke CVS (1987) New light on the plant geography of Ceylon. Historical plant geography. Journal of Biogeography 14: 249-285.

Auden JB (1949) A geological discussion on the Satpura hypothesis and Garo-Rajmahal gap. Proceedings of the National Academy of Sciences USA 8: 315-340.

Avise JC (1998) The history and purview of phylogeography: a personal reflection. Molecular Ecology 7: 371-379.

Avise JC (2000) Phylogeography: The history and formation of species. Harvard University Press, Cambridge, Massachusetts, pp. 447.

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

Bakker EG, Van Dam BC, Van Eck HJ and Jacobsen E (2001) The description of clones of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. with microsatellites and AFLP in an ancient woodland. Plant Biology 3: 616-621.

Balloux F and Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Molecular Ecology 11: 155-166.

Bande MB and Chandra S (1990) Early tertiary vegetational reconstructions around Nagpur-Chhindwara and Mandla, central India. Palaeobotanist 38: 196-208.

Bandelt HJ, Forster P and Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37-48.

Bartish IV, Jeppsson N and Nybom H (1999) Population genetic structure in the dioecious pioneer plant species *Hippophae rhamnoides* investigated by RAPD markers. Molecular Ecology 8: 791-802.

Behling H (1998) Late quaternary vegetational and climatic changes in Brazil. Review of Paleobotany and Palynology 99: 143-156.

Belahbib N, Pemonge MH, Ouassou A, Sbay H, Kremer A and Petit RJ (2001) Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. Molecular Ecology 10: 2003-2012.

Bermingham E and Moritz C (1998) Comparative phylogeography: concepts and applications. Molecular Ecology 7: 367-369.

Birky Jr. CW, Fuerst P and Maruyama T (1989) Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells and comparison to nuclear genes. Genetics 121: 613-627.

Blasco F (1970) Aspects of flora and ecology of savannas of the South Indian hills. Journal of Bombay Natural History Society 67: 522-534.

Blasco F (1971) Orophytes of South India and Himalayas. Journal of Indian Botanical Society 50: 377-381.

Bleeker W and Hurka H (2001) Introgressive hybridisation in *Rorippa* (Brassicaceae): gene flow and its consequences in natural and anthropogenic habitats. Molecular Ecology 10: 2013-2022.

Blouin MS, Parsons M, Lacaille V and Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology 3: 393-401.

Bohonak AJ (2002) IBD (Isolation By Distance): A program for analyses of isolation by distance. Journal of Heredity 93: 153-154.

Bonnet E and Van De Peer Y (2002) zt: a software tool for simple and partial Mantel tests. Journal of Statistical Software 7: 1-12.

Bowen BW (1999) What is wrong with ESUs? The gap between evolutionary theory and conservation principles. Journal of Shellfish Research 17: 1355-1358.

Braverman J, Hudson R, Kaplan N, Langley CH and Stephan W (1995) The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140: 645-683.

Breton C, Tersac M and Berville A (2006) Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. Journal of Biogeography 33: 1916-1928.

Brochmann C, Gabrielsen TM, Nordal I, Landvik JY and Elven R (2003) Glacial survival or tabula rasa? The history of North Atlantic biota revisited. Taxon 52: 417-450.

Brown AHD and Allard RW (1970) Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. Genetics 66: 133-145.

Bryan GJ, McNicoll J, Ramsay G, Meyer RC and De Jong WS (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theoretical and Applied Genetics 99: 859-867.

Bucci G, González-Martínez SC, Le Provost G, Plomion C, Ribeiro MM, Sebastiani F, Alía R and Vendramin GG (2007) Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. Molecular Ecology 10: 2137-2153.

Burban C and Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. Molecular Ecology 12: 1487-1495.

Burban C, Petit RJ, Caracreff E and Jactel H (1999) Rangewide variation of the maritime pine bast scale *Matsococcus feytauti* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. Molecular Ecology 8: 1593-1602.

Burkill IH (1924) The botany of the abor expedition. Records of the Botanical Survey of India 10: 155-420.

Bush GL (1975) Modes of animal speciation. Annual Review of Ecological Systems 6: 339-364.

Bussell JD (1999) The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotoma petraea* (Lobiliaceae). Molecular Ecology 8: 775-789.

Byrne M and Macdonald B (2000) Phylogeography and conservation of three oil mallee taxa, *Eucalyptus kochii* ssp. *kochii*, ssp. *plenissima* and *E. horistes*. Australian Journal of Botany 48: 305-312.

Byrne M and Moran GF (1994) Population divergence in the chloroplast genome of *Eucalyptus nitens*. Heredity 73: 18-28.

Caicedo AL and Schaal BA (2004) Population structure and phylogeography of *Solanum pimpinellifolium* inferred from a nuclear gene. Molecular Ecology 13: 1871-1882.

Cannon CH and Manos PS (2003) Phylogeography of the Southeast Asian stone oaks (*Lithocarpus*). Journal of Biogeography 30: 211-226.

Caris P, Decraene LPR, Smets E and Clinckemaillie D (2002) The uncertain systematic position of *Symplocos* (Symplocaceae): evidence from a floral ontogenetic study. International Journal of Plant Science 163: 67-74.

Caron H, Dumas S, Marque G, Messier C, Bandou E, Petit RJ and Kremer A (2000) Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. Molecular Ecology 9: 1089-1098.

Cerff R (1982) Separation and purification of NAD- and NADP-linked glyceraldehyde-3-phosphate dehydrogenases from higher plants. In: Edelman M, Hallick RB and Chua NH (eds.) Methods in chloroplast molecular biology. Elsevier/North Holland, Amsterdam, pp. 683-694.

Chapman JL and Reiss MJ (1992) Ecology: Principles and Applications, Cambridge University Press, Cambridge, pp. 294.

Charlesworth B, Morgan MT and Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289-1303.

Chase MR, Moller C, Kesseli R and Bawa KS (1996) Distant gene flow in tropical trees. Nature 383: 398-399.

Chauhan MS (2002) Holocene vegetation and climatic changes in southeastern Madhya Pradesh, India. Current Science 83: 1444-1445.

Chen S, Wu G, Zhangi D, Gao Q, Duan Y, Zhangi F and Chen S (2008) Potential refugium on the Qinghai–Tibet Plateau revealed by the chloroplast DNA phylogeography of the alpine species *Metagentiana striata* (Gentianaceae). Botanical Journal of the Linnean Society 157: 125-140.

Cheng YP, Hwang SY and Lin TP (2005a) Two potential refugia in Taiwan revealed by the phylogeographical study of *Castanopsis carlesii* Hayata (Fagaceae). Molecular Ecology 14: 2075-2085.

Cheng YP, Hwang SY, Chiou WL and Lin TP (2005b) Allozyme Variation of Populations of *Castanopsis carlesii* (Fagaceae) Revealing the Diversity Centres and Areas of the Greatest Divergence in Taiwan. Annals of Botany 98: 601-608.

Chiang TY, Chiang YC, Chen YJ, Chou CH, Havaanond S, Hong TN and Huang S (2001) Phylogeography of *Kandelia candel* in East Asiatic mangroves based on nucleotide variation of chloroplast and mitochondrial DNAs. Molecular Ecology 10: 2697-2710.

Chung JD, Lin TP, Tan YC, Lin MY and Hwang SY (2004) Genetic diversity and biogeography of *Cunninghamia konishii* (Cupressaceae), an island species in Taiwan: a comparison with *Cunninghamia lanceolata*, a mainland species in China. Molecular Phylogenetics and Evolution 33: 791-801.

Clark CM, Wentworth TR and O'Malley DM (2000) Genetic discontinuity revealed by chloroplast microsatellites is eastern North American *Abies* (Pinaceae). American Journal of Botany 87: 774-782.

Clegg MT and Zurawski G (1992) Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In: Soltis DE, Soltis PS and Doyle JJ (eds.) Molecular Systematics of Plants, Chapman and Hall, pp. 1-13.

Clegg MT, Ritland K and Zurawski G (1986) Processes of chloroplast DNA evolution. In: Karlin S and Nevo E (eds.) Evolutionary processes and theory, Academic Press, New York, pp. 275-294.

Collevatti RG, Grattapaglia D and Hay JD (2003) Evidence for multiple lineages of *Caryocar brasiliense* populations in the Brazilian Cerrado based on the analysis of chloroplast DNA sequence and microsatellite haplotype variation. Molecular Ecology 12: 105-115.

Comes HP and Kadereit JW (1998) The effects of Quaternary climatic changes on plant distribution and evolution. Trends in Plant Science 3: 432-438.

Comps B, Gomory D, Letouzey J, Thiebaut B and Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. Genetics 157: 389-397.

Conservation International (2005) Biodiversity Hotspots Conservation International, Washington, DC. http://www.biodiversityhotspots.org/xp/Hotspots/.

Corriveau JL and Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. American Journal of Botany 75: 1443-1445.

Crandall KA and Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics 134: 959-969.

Creste S, Tulmann Neto A and Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Molecular Biology Reporter 19: 299-306.

Cruse Sanders JM and Hamrick JL (2004) Genetic diversity in harvested and protected populations of wild American ginseng, *Panax quinquefolius* L. (Araliaceae). American Journal of Botany 91: 540-548.

Cruzan MB and Templeton AR (2000) Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. Trends in Ecology and Evolution 15: 491-496.

Csaikl UM, Burg K, Fineschi S, Konig AO, Matyas G and Petit RJ (2002) Chloroplast DNA variation of white oaks in the Alpineregion. Forest Ecology and Management 156: 131-145.

Cuenca A, Escalante AE and Piñero D (2003) Long-distance colonization, isolation by distance, and historical demography in a relictual Mexican pinyon pine (*Pinus nelsonii* Shaw) as revealed by paternally inherited genetic markers (cpSSRs). Molecular Ecology 12: 2087-2097.

Dahanukar N, Raut R and Bhat A (2004) Distribution, endemism and threat status of freshwater fishes in the Western Ghats of India. Journal of Biogeography 31: 129-136.

Dane F, Liu J and Zhang C (2007) Phylogeography of the bitter apple, *Citrullus colocynthis*. Genetic Resources and Crop Evolution 54: 327-336.

Daniel JC (2002) The book of Indian reptiles and amphibians, Oxford University Press, Mumbai, pp. 248.

Daniels RJR (1992) Geographical distribution patterns of amphibians in the Western Ghats, India. Journal of Biogeography 19: 521-529.

Daniels RJR (1997) A field guide to the birds of Southwestern India, Oxford University Press, Delhi, pp. 217.

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

Das I (1996) Biogeography of the Reptiles of South Asia, Krieger Publishing Company, Malabar, Florida, pp. 87.

Das I (2002) An introduction of the amphibians and reptiles of tropical Asia, Natural History publications (Borneo), Kota Kinabalu, pp. 207.

De Queiroz A (2005) The resurrection of oceanic dispersal in historical biogeography. Trends in Ecology and Evolution 20: 68-73.

De Terra H and Paterson TT (1939) Studies on the Ice age in India and associated human cultures, Publ. No. 493. Carnegie Institute, Washington.

Deguilloux MF, Dumolin-Lapègue S, Gielly L, Grivet D and Petit RJ (2003) A set of primers for the amplification of chloroplast microsatellites in *Quercus*. Molecular Ecology Notes 3: 24-27.

Demesure B, Comps B and Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. Evolution 50: 2515-2520.

Demesure B, Sodzi N and Petit RJ (1995) A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. Molecular Ecology 4: 129-131.

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

Desplanque B, Viard F, Bernard J, Forciolo D, Smitou-Laprade P, Cugen J and Van Dijk H (2000) The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in *Beta vulgaris* ssp. Maritime (L.): the usefulness of both genomes for population genetics studies. Molecular Ecology 9: 141-154.

Despres L, Loriot S and Gaudeul M (2002) Geographic pattern of genetic variation in the European globeflower *Trollius europaeus* L. (Ranunculaceae) inferred from amplified fragment length polymorphism. Molecular Ecology 11: 2337-2347.

Diamond J (1997) Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 Years. Chatto and Windus, UK.

Dijk PPV, Aston P and Ma J (1999) Indo-Burma In: Mittermeier RA, Myers N, Robles-Gil P and Mittermeimer CG (eds.) Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions. CEMEX/Agrupación Sierra Madre, 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: 125-127.

Dixon CJ, Schönswetter P and Schneeweiss GM (2007) Traces of ancient range shifts in a mountain plant group (*Androsace halleri* complex, Primulaceae). Molecular Ecology 16: 3890-3901.

Dow BD and Ashley MV (1998) High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. Journal of Heredity 89: 62-70.

Dumolin-Lapègue S, Demesure B, Fineshi S, Le Corre V and Petit RJ (1997a) Phylogenetic structure of white oaks throughout the European continent. Genetics 146: 1475-1487.

Dumolin-Lapègue S, Pemonge M-H and Petit RJ (1997b) An enlarged set of consensus primers for the study of organelle DNA in plants. Molecular Ecology 6: 393-397.

Dumolin-Lapègue S, Pemonge M-H and Petit RJ (1998) Association between chloroplast and mitochondrial lineages in oaks. Molecular Biology and Evolution 15: 1321-1331.

Dynesius M and Jansson R (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. Proceedings of the National Academy of Sciences USA 97: 9115-9120.

Edh K, Widen B and Ceplitis A (2007) Nuclear and chloroplast microsatellites reveal extreme population differentiation and limited gene flow in the Aegean endemic *Brassica cretica* (Brassicaceae). Molecular Ecology 16: 4972-4983.

El Mousadik A and Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. Molecular Ecology 5: 547-555.

Ellstrand NC, Prentice H and Hancock JF (1999) Geneflow and introgression from domesticated plants into their wild relatives. Annual Review of Ecological Systems 30: 539-563.

Ellsworth DL, Rittenhouse KD and Honeycutt RL (1993) Artifactual variation in randomly amplified polymorphic DNA banding patterns. Biotechniques 14: 214-217.

Endler JA (1977) Geographic variation, speciation and clines. Monographs in Population Biology 10: 1-246.

Excoffier L, Laval G and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.

Fady B, Lefèvre F, Vendramin GG, Ambert A, Régnier C and Bariteau M (2008) Genetic consequences of past climate and human impact on eastern Mediterranean *Cedrus libani* forests implications for their conservation. Conservation Genetics 9: 85-95.

Fady B, Levèvre F, Reynaud M, Vendramin GG, Bou Dagher-Kharrat M, Anzidei M, Pastorelli R, Savouré A and Bariteau M (2003) Gene flow among different taxonomic units: evidence from nuclear and cytoplasmic markers in Cedrus plantation forests. Theoretical and Applied Genetics 107: 1132-1138.

Ferris C, King RA, Vainola R and Hewitt GM (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. Heredity 80: 584-593.

Ferris C, Oliver RP, Davy AJ and Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. Molecular Ecology 2: 337-344.

Fontaine C, Lovett PN, Sanou H, Maley J and Bouvet J-M (2004) Genetic diversity of the shea tree (*Vitellaria paradoxa* C.F. Gaertn.), detected by RAPD and chloroplast microsatellite markers. Heredity 93: 639-648.

Frankel OH and Soulé ME (1981) Conservation and Evolution. Cambridge University Press, Cambridge.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915-925.

Fu YX and Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133: 693-709.

Fujii N (2007) Chloroplast DNA phylogeography of Pedicularis *ser.* Gloriosae *(Orobanchaceae) in Japan. Journal of Plant Research 120: 491-500.*

Fujii N, Tomaru N, Okuyama K, Koike T, Mikami T and Ueda K (2002) Chloroplast DNA phylogeography of *Fagus crenata* (Fagaceae) in Japan. Plant Systematics and Evolution 232: 21-33.

Fujii N, Ueda K, Watano Y and Shimizu T (1997) lntraspecific sequence variation of chloroplast DNA in *Pedicularis chamissonis* Steven (Scrophulariaceae) and geographic structuring of the Japanese "alpine" plants. Journal of Plant Research 110: 195-207.

Gao LM, Moller M, Zhang XM, Hollingsworth ML, Liu J, Mill RR, Gibby M and Li DZ (2007) High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. Molecular Ecology 16: 4684-4698.

Gaskin JF and Schaal BA (2002) Hybrid Tamarix widespread in US invasion and undetected in native Asian range. Proceedings of the National Academy of Sciences USA 99: 11256-11259.

Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. Evolutionary Biology 30: 93-120.

Ge XJ, Zhou XL, Li ZC, Hsu TW, Schaal BA and Ching TY (2005) Low genetic diversity and significant population structuring in the relict *Amentotaxus argotaenia* complex (Taxaceae) base on ISSR fingerprinting. Journal of Plant Research 118: 415-422.

Gerber S, Mariette S, Streiff R, Bodénès C and Kremer A (2000) Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. Molecular Ecology 9: 1037-1048.

Gielly L, Debussche M and Thompson JD (2001) Geographic isolation and evolution of endemic *Cyclamen* in the Mediterranean basin: insights from chloroplast *trn*L (UAA) intron sequence variation. Plant Systematics and Evolution 230: 75-88.

Gómez A, González-Martínez SC, Collada C, Gil L and Climent J (2003) Complex population genetic structure in an endemic Canary Island pine using chloroplast microsatellite markers. Theoretical and Applied Genetics 107: 1123-1131.

Grant PR and Grant BR (1997) Genetics and origin of bird species. Proceedings of the National Academy of Sciences USA 94: 7768-7775.

Grassi F, Labra M, Imazio S, Ocete Rubio R, Failla O, Scienza A and Sala F (2006) Phylogeographical structure and conservation genetics of wild grapevine. Conservation Genetics 7: 837-845.

Griffin SR and Barrett SCH (2004) Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in eastern North America: inferences from phylogeography. American Journal of Botany 91: 465-473.

Grivet D and Petit RJ (2002) Phylogeography of the common ivy (*Hedera* sp.) in Europe: genetic differentiation through space and time. Molecular Ecology 11: 1351-1362.

Grivet D and Petit RJ (2003) Chloroplast DNA phylogeography of the hornbeam in Europe: evidence for a bottleneck at the outset of postglacial colonization. Conservation Genetics 4: 47-56.

Grob GBJ, Gravendeel B and Eurlings MCM (2004) Potential phylogenetic utility of the nuclear *FLORICAULA/LEAFY* second intron: comparison with three chloroplast DNA regions in Amorphophallus (Araceae). Molecular Phylogenetics and Evolution 30: 13-23.

Gugerli F and Holderegger R (2001) Nunatak survival, tabula rasa and the influence of the Pleistocene ice-ages on plant evolution in mountain areas. Trends in Plant Science 6: 397-398.

Gugerli F, Sperisen C, Büchler U, Magni F, Geburek T, Jeandroz S and Senn J (2001) Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. Molecular Ecology 10: 1255-1263.

Guo Q-B, Liu Y-C and Li X-G (1997) The niche of dominant populations in forest secondary sere in Mt. Jinyun. Journal of Southwest Normal University 22: 73-78.

Gupta AK, Anderson DM and Overpeck JT (2003) Abrupt changes in the Asian southwest monsoon during the Holocene and their links to the North Atlantic ocean. Nature 421: 354-357.

Gupta HP (1990) Sholas in south Indian montane: Past, present and future. Paleobotanist 38: 394-403.

Gupta RK (1994) Arcto-Alpine and Boreal Elements in the High Altitude Flora of the North West Himalaya. In: Pangtey YPS and Rawal RS (eds.) High Altitudes of the Himalaya. (Biogeography, Ecology and Conservation). Gyanodaya Prakashan, Nainital, Delhi.

Hajra PK and Mudgal V (1997) Plant diversity hotspots in India, an overview. Botanical Survey of India, Calcutta, India.

Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Molecular Ecology 8: 521-523.

Hampe A, Arroya J, Jordano P and Petit RJ (2003) Rangewide phylogeography of a bird-dispersed Eurasian shrub: contrasting Mediterranean and temperate glacial refugia. Molecular Ecology 12: 3415-3426.

Hamrick JL and Godt MJW (1989) Allozymes diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL and Weir BS (eds.) Plant Population Genetics, Breeding and Genetic Resources, Sinauer, Sunderland, Massachusetts, USA, pp. 43-63.

Hamrick JL and Godt MJW (1996) Effects of life history traits on genetic diversity in plants. Proceedings of the Royal Society of London B 351: 1291-1298.

Hamrick JL and Holden LR (1979) Influence of microhabitat heterogeneity on gene frequency distribution and gametic phase disequilibrium in *Avena barbata* (wild oats, California). Evolution 33: 521-533.

Hamrick JL, Godt MJW, Murawski DA and Loveless MD (1991) Correlations between species traits and allozyme diversity: implications for conservation biology. In: Falk DA and Holsinger KE (eds.) Genetics and Conservation of Rare Plants, Oxford University Press, New York, pp. 75-86.

Harding RM, Boyce AJ and Clegg JB (1992) The evolution of tandemly repetitive DNA: recombination rules. Genetics 132: 847-859.

Hare MP (2001) Prospects for nuclear gene phylogeography. Trends in Ecology and Evolution 16: 700-706.

Harpending H (1994) Signature of ancient population growth in a low resolution mitochondrial DNA mismatch distribution. Human Biology 66: 591-600.

Harris SA and Ingram R (1991) Chloroplast DNA and biosystematics: the effect of intraspecific diversity and plastid transmission. Taxon 40: 393-412.

Heuertz M, Hausman JF, Hardy OJ, Vendramin GG, Frascaria-Lacosite N and Vekemans X (2004) Nuclear microsatellites reveal contrasting patterns of genetic structure between western and southern European populations of the common ash (*Fraxinus excelsior* L.). Evolution 58: 976-988.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58: 247-276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68: 87-112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405: 907-913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences 359: 183-195.

Holderegger R, Stehlik I and Abbott RJ (2002) Molecular analysis of the Pleistocene history of *Saxifraga oppositifolia* in the Alps. Molecular Ecology 11: 1409-1418.

Honjo M, Ueno S, Tsumura Y, Washitani I and Ohsawa R (2004) Phylogeographic study based on intraspecific sequence variation of chloroplast DNA for the conservation of genetic diversity in the Japanese endangered species *Primula sieboldii*. Biological Conservation 120: 215-224.

Hooker JD and Thomson T (1855) Introductory essay to the Flora Indica, W Pamplin, London.

Hooker JD (1904) A Sketch of the Flora of British India. Eyre and Srottiswoode, London.

Hora SL (1949) Satpura hypothesis of the distribution of the Malayan fauna and flora to peninsular India. Proceedings of the National Institute of Sciences India 15: 309-314.

Hore DK (1990) Symplocaceae. In: Nayar MP, Tholhathri K and Sanjappa M (eds.) Fascicles of flora of India 20: pp. 155.

Howarth DG and Baum DA (2002) Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. Molecular Phylogenetics and Evolution 23: 525-528.

Hu XS and Ennos RA (1997) On estimation of the ratio of pollen to seed flow among plant populations. Heredity 79: 541-552.

Huang S, Chiang YC, Schaal BA, Chou CH and Chiang TY (2001) Organelle DNAs phylogeography of *Cycas taitungensis*, a relict species in Taiwan. Molecular Ecology 10: 2669-2681.

Huang S-F, Hwang S-Y and Lin T-P (2002) Spatial pattern of chloroplast DNA variation of *Cyclobalanopsis glauca* in Taiwan and East Asia. Molecular Ecology 11: 2349-2358.

Huang S-F, Hwang S-Y, Wang J-C and Lin T-P (2004) Phylogeography of *Trochodendron aralioides* (Trochodendraceae) in Taiwan and its adjacent areas. Journal of Biogeography 31: 1251-1259.

Hubbell SP and Foster RB (1992) Short-term dynamics of a neotropical forest: Why ecological research matters to tropical conservation and management. Oikos 63: 48-61.

Hudson RR and Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111: 147-164.

Hugall A, Moritz C, Moussalli A and Stanisic J (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis*. Proceedings of the National Academy of Sciences USA 99: 6112-6117.

Hwang SH, Lin TP, Ma CS, Lin CL, Chung JD and Yang JC (2003) Postglacial population growth of *Cunninghamia konishii* (Cupressaceae) inferred from phylogeographical and mismatch analysis of chloroplast DNA variation. Molecular Ecology 12: 2689-2695.

Ikeda H and Setoguchi H (2007) Phylogeography and refugia of the Japanese endemic alpine plant, *Phyllodoce nipponica* Makino (Ericaceae). Journal of Biogeography 34: 169-176.

Ikeda H, Senni KEI, Fujii N and Setoguchi H (2006) Refugia of *Potentilla matsumurae* (Rosaceae) located at high mountains in the Japanese archipelago. Molecular Ecology 15: 3731-3740.

Innan H, Tajima F, Treauchi R and Miyashita NT (1996) Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana.* Genetics 143:1761-1770.

Ishii T and McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Graminae species. Theoretical and Applied Genetics 100: 1257-1266.

Jain A, Pandit MK, Elahi S, Jain A, Bhaskar A and Kumar V (2000) Reproductive behaviour and genetic variability in geographically isolated populations of *Rhododendron arboreum* (Ericaceae). Current Science 79: 1377-1381.

Jaramillo-Correa JP and Bousquet J (2003) New evidence from mitochondrial DNA of a progenitor-derivative relationship between black spruce and red spruce (Pinaceae). American Journal of Botany 90: 1801-1806.

Jaramillo-Correa JP, Beaulieu J and Bousquet J (2004) Variation in mitochondrial DNA reveals multiple glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. Molecular Ecology 13: 2735-2747.

Jarne P and Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution 11: 424-429.

Jha CS, Dutt CBS and Bawa KS (2000) Deforestation and land use change in Western Ghats, India. Current Science 79: 231-238.

Johannesen J, Lubin Y, Laufs T and Seitz A (2005) Dispersal history of a spider (*Stegodyphus lineatus*) across contiguous deserts: vicariance and range expansion. Biological Journal of the Linnean Society 84: 739-754.

Johansen AD and Latta RG (2003) Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. Molecular Ecology 12: 293-298.

Jøhnk N and Siegismund HR (1997) Population structure and postglacial migration routes of *Quercus robur* and *Quercus petraea* in Denmark, based on chloroplast DNA analysis. Scandinavian Journal of Forest Research 12: 130-137.

Johnson KP, Adler FR and Cherry JL (2000) Genetic and phylogenetic consequences of island biogeography. Evolution 54: 387-396.

Jønsson KA and Fjeldså J (2006) Determining biogeographical patterns of dispersal and diversification in oscine passerine birds in Australia, Southeast Asia and Africa. Journal of Biogeography 33: 1155-1165.

Jordan WC, Courtney MW and Neigel JE (1997) Low levels of infraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). American Journal of Botany 83: 430-439.

Jukes TH and Cantor CR (1969) Evolution of protein molecules. In: Munro HM (ed.) Mammalian protein metabolism. Academic Press, New York, pp. 21-120.

Karanth PK (2003) Evolution of disjunct distribution among wet-zone species of the Indian subcontinent: Testing various hypotheses using a phylogenetic approach. Current Science 85: 101-108.

Karunakaran PV, Rawat GS and Uniyal VK (1998) Ecology and Conservation of the Grasslands of Eravikulam National Park, Western Ghats. Wildlife Institute of India, Dehradun, India.

Kawabe A, Innan H, Terauchi R and Miyashita NT (1997) Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana.* Molecular Biology and Evolution 14: 1303-1315.

Khoshoo TN (1996) India needs a national conservation board. Current Science 71: 506-513.

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

Knowles LL and Maddison WP (2002) Statistical phylogeography. Molecular Ecology 11: 2623-2635.

Koch M and Bernhardt K-G (2004) Comparative biogeography of the cytotypes of annual *Microthlaspi perfoliatum* (Brassicaceae) in Europe using isozymes and cpDNA data: refugia, diversity centers and postglacial colonization. American Journal of Botany 91: 115-124.

Koch M, Mummenhoff K and Hurka H (1998) Molecular biogeography and evolution of the *Microthlaspi perfoliatum* s.l. polyploid complex (Brassicaceae): chloroplast DNA and nuclear ribosomal DNA restriction site variation. Canadian Journal of Botany 76: 382-396

Koehler-Santos P, Lorenz-Lemke AP, Muschner VC, Bonatto SL, Salzano FM and Freitas BL (2006a) Molecular genetic variation in *Passiflora alata* (Passifloraceae), an invasive species in southern Brazil. Biological Journal of the Linnean Society 88: 611-630.

Koehler-Santos P, Lorenz-Lemke AP, Salzano FM and Freitas LB (2006b) Ecological-evolutionary relationships in *Passiflora alata* from Rio Grande do Sul, Brazil. Brazilian Journal of Biology 66: 809-816.

Konnert M and Bergmann F (1995) The geographical distribution of genetic variation of silver fir (*Abies alba*, Pinaceae) in relation to its migration history. Plant Systematics and Evolution 196: 19-30.

Kreitman M (2000) Methods to detect selection in populations with applications to the human. Annual Review of Genomics and Human Genetics 1: 539-559.

Kropf M, Kandereit JW and Comes HP (2003) Differential cycles of range contraction and expansion in European high mountain plants during the late Quaternary: insight from *Pritzelago alpine* (L.) O. Kuntze (Brassicaceae). Molecular Ecology 12: 931-949.

Kuittinen H, Mattila A and Savolainen O (1997) Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. Heredity 79: 144-152.

Kumar S, Tamura K and Nei M (2004) MEGA 3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5: 150-163.

Kumar V (1983) Pleistocene glaciation and evolutionary divergences in Himalayan rhododendrons. In: Swaminathan MS, Chopra VL, Joshi BC, Sharma RP and Bansal HC (eds.) Proceedings XV International Congress of Genetics, New Delhi. Oxford and IBH Publishing Company, New Delhi, pp. 444.

Latta RG and Mitton JB (1997) A comparison of population differentiation across four classes of gene marker in Limber pine (*Pinus flexilis* James). Genetics 146: 1153-1163.

Latta RG and Mitton JB (1999) Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. Evolution 53: 769-776.

Ledig FT, Hodgskiss PD and Jacob-Cervantes V (2002) Genetic diversity, mating system and conservation of a Mexican subalpine relict, *Picea mexicana* Martínez. Conservation Genetics 3: 113-122.

Lee C and Wen J (2004) Phylogeny of Panax using chloroplast *trn*C-*trn*D intergenic region and the utility of *trn*C-*trn*D in interspecific studies of plants. Molecular Phylogenetics and Evolution 31: 894-903.

Legris P and Blasco F (1969) Variability des facteurs du climat: Cas des montagnes du sud de l'Inde et de Ceylan. Institut Francais de Pondichery Torrey Section Science and Technology 8:1-195.

Legris P and Meher-Homji VM (1984) The Eastern Ghats: phytogeographic aspects. Indian Review of Life Sciences 4: 115-136.

Lewis CE and Doyle JJ (2001) Phylogenetic utility of the nuclear gene malate synthase in the palm family (Arecaceae). Molecular Phylogenetics and Evolution 19: 409-420.

Lewontin R (1974) The Genetic Basis of Evolutionary Change, Columbia University Press, New York.

Li J, Ge XJ, Cao HL and Ye WH (2007) Chloroplast DNA diversity in *Castanopsis hystrix* populations in south China. Forest Ecology and Management 243: 94-101.

Li WH (1997) Molecular Evolution. Sinauer Associates, Sunderland, Massachusetts.

Li Y-C, Korol AB, Fahima T, Beiles A and Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology 11: 2453-2465.

Lian CL, Oishi R, Miyashita N, Nara K, Nakaya H, Wu B, Zhou Z and Hogetsu T (2003) Genetic structure and reproduction dynamics of *Salix reinii* during primary succession on Mount Fuji, as revealed by nuclear and chloroplast microsatellite analysis. Molecular Ecology 12: 609-618.

Liedloff A (1999) A Mantel nonparametric test calculator for windows version 2.00. Tropical Ecosystems Research Centre, Winnellie, Northern Territory, Australia.

Lihová J, Kucera J, Perny M and Marhold K (2007) Hybridization between two polyploid *Cardamine* (Brassicaceae) species in north-western Spain: discordance between morphological and genetic variation patterns. Annals of Botany 99: 1083-1096.

Lira CF, Cardoso SRS, Ferreira PCG, Cardoso MA and Provan J (2003) Long-term population isolation in the endangered tropical tree species *Caesalpinia echinata* Lam. revealed by chloroplast microsatellites. Molecular Ecology 12: 3219-3225.

López de Heredia U. Carrión JS, Jiménez P, Collada C and Gil L (2007) Molecular and palaeoecological evidence for multiple glacial refugia for evergreen oaks on the Iberian Peninsula. Journal of Biogeography 34: 1505-1517.

López-Pujol J, Font J, Simon J and Blanché C (2007) Can the preservation of historical relicts permit the conservation of endangered plant species? The case of *Silene sennenii* (Caryophyllaceae). Conservation Genetics 8: 903-912.

Lorenz-Lemke AP, Mäder G, Muschner VC, Stehmann JR, Bonatto SL, Salzano FM and Freitas BL (2006) Diversity and natural hybridization in a highly endemic species of *Petunia* (Solanaceae): a molecular and ecological analysis. Molecular Ecology 15: 4487-4497.

Luo H, van Coppenolle B, Seguin M and Boutry M (1995) Mitochondrial DNA polymorphism and phylogenetic relationships in *Hevea brasiliensis*. Molecular Breeding 1: 51-63.

Lynch JD (1989) The gauge of speciation: on the frequencies of modes of speciation. In: Otte D and Endler JA (eds.) Speciation and its consequences, Sunderland, Massachusetts, Sinauer, pp. 527-553.

Ma D-W, Zhang G, Wang Y-H and Wang D-M (2002) Studies on species diversity of forest vegetation on Qingcheng Mountain. Journal of Sichuan University (Nat. Sci.) 39: 115-123.

Mani MS (1974) Ecology and biogeography in India. Dr. W. Junk b.v. Publishers, The Hague, The Netherlands.

Manicacci D, Couvet D, Belhassen E, Gouyon PH and Atlan A (1996) Founder effects and sex ratio in the gynodioecious *Thymus vulgaris* L. Molecular Ecology 5: 63-72.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209-220.

Mardulyn P (2001) Phylogeography of the Vosges mountains populations of *Gonioctena pallida* (Coleoptera: Chrysomelidae): a nested clade analysis of mitochondrial DNA haplotypes. Molecular Ecology 10: 1751-1763.

Markwith SH and Parker KC (2007) Conservation of *Hymenocallis coronaria* genetic diversity in the presence of disturbance and a disjunct distribution. Conservation Genetics 8: 949-963.

Marshall HD, Newton C and Ritland K (2002) Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). Theoretical and Applied Genetics 104: 367-378.

Mason-Gamer RJ, Well CF and Kellogg EA (1998) Granulebound starch synthase: Structure, function, and phylogenetic utility. Molecular Biology and Evolution 15: 1658-1673.

Mathews S and Donoghue MJ (1999) The root of angiosperm phylogeny inferred from duplicate phytochrome genes. Science 286: 947-950.

Matyas G and Sperisen C (2001) Chloroplast DNA polymorphism provide evidence for postglacial re-colonisation of oaks (*Quercus* spp.) across the Swiss Alps. Theoretical and Applied Genetics 102: 12-20.

McCauley DE (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends in Ecology and Evolution 10: 198-202.

Medlicott HB and Blanford WT (1879) A manual of the geology of India. Government Printer, Calcutta.

Meher-Homji VM (1967) Phytogeography of the South Indian hill stations. Bulletin of Torrey Botanical Club 94: 230-242.

Meher-Homji VM (1972) Himalayan plants on South Indian hills: role of Pleistocene glaciation vs long distance dispersal. Science and Culture 38: 8-12.

Meher-Homji VM (1973) A phytosociological study of the *Albizia amara* Boiv. community of India. Phytocoenologia 1: 114-129.

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

Meher-Homji VM (1977) Vegetation - climate parallelism along Pondicherry - Mysore - Murkal transect, South India. Phytocoenologia 4: 206-217.

Meng L, Yang R, Abbott RJ, Miehe G, Hu T and Liu J (2007) Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan Plateau and adjacent highlands. Molecular Ecology 16: 4128-4137.

Mengoni A, Barabesi C, Gonnelli C, Galardi F, Gabbrielli R and Bazzicalupo M (2001) Genetic diversity of heavy metal-tolerant populations in *Silene paradoxa* L. (Caryophyllaceae): a chloroplast microsatellite analysis. Molecular Ecology 10: 1909-1916.

Mengoni A, Gonnelli C, Brocchini E, Galardi F, Pucci S, Gabrielli R and Bazzicalupo M (2003) Chloroplast genetic diversity and biogeography in the serpentine endemic *Ni*-hyperaccumulator *Alyssum bertolonii*. New Phytologist 157: 349-356.

Meudt HM and Bayly MJ (2008) Phylogeographic patterns in the Australasian genus *Chionohebe* (*Veronica* s.l., Plantaginaceae) based on AFLP and chloroplast DNA sequences. Molecular Phylogenetics and Evolution 47: 319-338.

Mittermeier RA, Myers N, Thomsen JB, da Fonesca GAB and Oliveri S (1998) Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. Conservation Biology 12: 516-520.

MoEF (1999) National policy and macrolevel action strategy on biodiversity. New Delhi: Ministry of Environment and Forests, Government of India.

Mohanty A, Martín JP and Aguinagalde I (2002) Population genetic analysis of European *Prunus spinosa* (Rosaceae) using chloroplast DNA markers. American Journal of Botany 89: 1223-1228.

Mohanty A, Martin JP, Gonzalez LM and Aguinagalde I (2003) Association between chloroplast and mitochondrial DNA haplotypes in *Prunus spinosa* L. (Rosaceae) populations across Europe. Annals of Botany 92: 749-755.

Molnar P (1986) The geologic history and structure of the himalaya. American Scientist 74: 144-154.

Morales JC and Melnick DJ (1998) Phylogenetic relationships of the macaques (Cercopithecidae: *Macaca*) as revealed by high resolution restriction site mapping of mitochondrial ribosomal genes. Journal of Human Evolution 34:1-23.

Morgante M and Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant Journal 3: 175-182.

Morrell PL, Lundy KE and Clegg MT (2003) Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. Proceedings of the National Academy of Sciences USA 100: 10812-10817.

Myers N (1988) Threatened biotas: hotspots in tropical forests. The Enviromentalist 8: 178-208.

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

Nagamitsu T, Ogawa M, Ishida K and Tanouchi H (2004) Clonal diversity, genetic structure, and mode of recruitment in a *Prunus ssiori* population established after volcanic eruptions. Plant Ecology 174: 1-10.

Nair NC and Daniel P (1986) The floristic diversity of the Western Ghats and its conservation: a review. Proceedings of Indian Academy of Science (Animal/Plant Science) Supplement 3: 127-163.

Nayar MP (1996) Hot-spot of endemic plants of India, Nepal and Bhutan. Tropical Botanic Garden and Research Institute, Thiruvananthapuram, Kerala, India

Nayar MP and Sastry ARK (1987) Red data book of Indian plants. vol. I. Botanical Survey of India.

Nayar MP and Sastry ARK (1988) Red data book of Indian plants. vol. II. Botanical Survey of India.

Nayar MP and Sastry ARK (1990) Red data book of Indian plants. vol. III. Botanical Survey of India.

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

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Nei M and Chesser RK (1983) Estimation of fixation indices and gene diversities. Annals of Human Genetics 47: 253-259.

Nelson G and Platnick N (1981) Systematics and biogeography: cladistics and vicariance. Columbia University Press, New York.

Nettel A and Dodd RS (2007) Drifting propagules and receding swamps: genetic footprints of mangrove recolonization and dispersal along tropical coasts. Evolution 61: 958-971.

Ohi T and Murata J (2000) Geographical variations in chloroplast DNA of *Stachyurus macrocarpus* Koidz. and *Stachyurus praecox* Sieb. et Zucc. (Stachyuraceae). Ogasawara Research 25: 1-24.

Ohta T (2002) Near-neutrality in evolution of genes and gene regulation. Proceedings of the National Academy of Sciences USA 99: 16134-16137.

Okaura T and Harada K (2002) Phylogeographical structure revealed by chloroplast DNA variation in Japanese Beech (*Fagus crenata* Blume). Heredity 88: 322-329.

Okaura T, Quang ND, Ubukata M and Harada K (2007) Phylogeographic structure and late Quaternary population history of the Japanese oak *Quercus mangolica* var. *cripsula*

and related species revealed by chloroplast DNA variation. Genes, Genetics and Systematics 82: 465-477.

Olsen KM (2002) Population history of *Manihot esculenta* (Euphorbiaceae) inferred from nuclear DNA sequences. Molecular Ecology 11: 901-911.

Olsen KM and Schaal BA (1999) Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Sciences USA 96: 5586-5591.

Ouborg NJ, Piquot Y and van Groenendael JM (1999) Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology 87: 551-568.

Page RD (1996) Treeview: An application to display phylogenetic trees on personal computers. Computer Applied Biosciences 12: 357-358.

Palmé AE and Verdramin GG (2002) Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. Molecular Ecology 11: 1769-1779.

Palmé AE, Su Q, Rautenberg A, Manni F and Lascoux M (2003) Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. Molecular Ecology 12: 201-212.

Pandit MK, Bhakar A and Kumar V (2000) Floral diversity of Goriganga Valley in the Central Himalayan highlands. Journal of Bombay Natural History Society 97: 184-192

Parathasarathy N (1999) Tree diversity and distribution in undisturbed and human-impacted sites of tropical wet evergreen forest in southern Western Ghats, India. Biodiversity and Conservation 8: 1365-1381.

Paula M and Leonardo G (2006) Multiple ice-age refugia in a southern beech of South America as evidenced by chloroplast DNA markers. Conservation Genetics 7: 591-603.

Pérusse JR and Schoen DJ (2004) Molecular evolution of the *GapC* gene family in *Amsinckia spectabilis* populations that differ in outcrossing rate. Journal of Molecular Evolution 59: 427-436.

Petit RJ and Grivet D (2002) Optimal randomization strategies when testing the existence of a phylogeographic structure. Genetics 161: 469-471.

Petit RJ, Aguinagalde I, Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Muller-Starck G, Demesure-Musch B, Palmé A, Pedro Marti J, Rendell S and Vendramin GG (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300: 1563-1565.

Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu J-L and Kremier A (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecology and Management 156: 49-74.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D and Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Molecular Ecology 14: 689-701.

Petit RJ, El Mousadik A and Pons O (1998) Identifying populations for conservation on the basis of genetic marker. Conservation Biology 12: 844-855.

Petit RJ, Kremer A and Wagner DB (1993) Finite island model for organelle and nuclear genes. Heredity 71: 630-641.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A and Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. Proceedings of the National Academy of Sciences USA 94: 9996-10001.

Pongratz N, Storhas M, Carranza S and Michiels NK (2003) Phylogeography of competing sexual and parthenogenetic forms of a freshwater flatworm: patterns and explanations. BMC Evolutionary Biology 3:23.

Pons O and Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity I. Haploid locus. Theoretical and Applied Genetics 90: 462-470.

Pons O and Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144: 1237-1245.

Posada D and Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology and Evolution 16: 37-45.

Powell W, Machray G and Provan J (1996a) Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1: 215-222.

Powell W, Morgante M, Doyle JJ, McNicol JW, Tingey SV and Rafalski AJ (1996b) Genepool variation in the genus *Glycine* subgenus *Soja* revealed by polymorphic nuclear and chloroplast microsatellites. Genetics 144: 792-803.

Powell W, Morgante M, McDevitt R, Vendramin GG and Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. Proceedings of the National Academy of Sciences USA 92: 7759-7763.

Provan J, Biss PM, McMeel D and Mathews S (2004) Universal primers for the amplification of chloroplast microsatellites in grasses (Poaceae). Molecular Ecology Notes 4: 262-264.

Provan J, Powell W and Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends in Ecology and Evolution 16: 142-147.

Provan J, Russell JR, Booth A and Powell W (1999a) Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. Molecular Ecology 8: 505-511.

Provan J, Soranzo N, Wilson NJ, Goldstein DB and Powell W (1999b) A low mutation rate for chloroplast microsatellites. Genetics 153: 943-947.

Puri GS, Gupta RK and Meher-Homji VM (1989) Forest Ecology Volume 2, New Delhi, Oxford and IBH Publishing Company, India.

Quiroga MP and Premoli AC (2007) Genetic patterns in *Podocarpus parlatorei* reveal the long-term persistence of cold tolerant elements in the southern Yungas. Journal of Biogeography 34: 447-455.

Rajguru SN (1969) On the late Pleistocene of the Deccan. Quaternaria 11: 241-253.

Ram J and Singh SP (1994) Ecology and conservation of Alpine meadows in central Himalaya, India. In: Pangtey YPS and Rawal RS (eds.) High Altitudes of the Himalaya. (Biogeography, Ecology and Conservation). Gyanodaya Prakashan, Nainital, Delhi.

Ramos-Onsins SE and Rozas J (2002) Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution 19: 2092-2100.

Rand DM (1996) Neutrality tests of molecular markers and the connection between DNA polymorphism, demography and conservation biology. Conservation Biology 10: 665-671.

Randhwa MS (1945) Progressive dessiccation of Northern India in historical times. Journal of Bombay Natural History Society 45: 558-565.

Ranganathan CR (1938) Studies on the ecology of the grassland vegetation of the Niligiri plateau. Indian Forester 64: 523-541.

Rao NM, Ganeshaiah KN and Uma Shaanker R (2007) Assessing threats and mapping sandal resources to identify genetic 'hot-spot' for *in-situ* conservation in peninsular India. Conservation Genetics 8: 925-935.

Rao RR (1993) Floristic diversity of Eastern Himalayas - A national heritage for conservation. In: Dhar U (ed.) Himalayan Biodiversity, Nainital, pp. 139.

Rao RR (1994) Biodiversity in India, floristic aspects. Bishen Singh Mahendra Pal Singh, Dehradun, India.

Raspé O, Saumitou-Laprade P, Cuguen J and Jacquemart A-L (2000) Chloroplast DNA haplotype variation and population differentiation in *Sorbus aucuparia* L. (Rosaceae: Maloideae). Molecular Ecology 9: 1113-1122.

Rauch EM and Bar-Yam Y (2005) Estimating the total genetic diversity of a spatial field population from a sample and implications of its dependence on habitat area. Proceedings of the National Academy of Sciences USA 102: 9826-9829.

Razi BA (1954) Some observations on the plants of the south Indian hill tops and their distribution. Proceedings National Institute of Sciences India 21: 79-89.

Reboud X and Zeyl C (1994) Organelle inheritance in plants. Heredity 72: 137-140.

Rhoads DD and Roufa DJ (1991) SEQAID II, version 3.81. Molecular Genetics Laboratory, Kansas State University, Kansas.

Richards E, Reichardt M and Rogers S (1994) Preparation of genomic DNA from plant tissues. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K (eds.) Current Protocols in Molecular Biology, John Wiley and Sons, New York, pp. 2.3.1-2.3.7.

Riddle BR (1996) The molecular phylogeographic bridge between deep and shallow history in continental biotas. Trends in Ecology and Evolution 11: 207-211.

Riedy MF, Hamilton WJ and Aquadro CF (1992) Excess non-parental bands in offspring from known pedigrees assayed using RAPD PCR. Nucleic Acids Research 20: 918.

Ripley D and Beehler B (1990) Patterns of speciation in Indian birds. Journal of Biogeography 17: 639-648.

Ripley D, Beehler B and Raju K (1986) Birds of the Visakhapatnam Ghats, Andhra Pradesh. Journal of Bombay Natural History Society 84: 540-558.

Rodgers WA, Panwar HS and Mathur VB (2000) Wildlife protected area network in India: a review (executive summary) Dehradun: Wildlife Institute of India, pp. 44.

Rogers AR and Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9: 552-569.

Ronikier M, Cieślak E and Korbecka G (2008) High genetic differentiation in the alpine plant *Campanula alpina* Jacq. (Campanulaceae): evidence for glacial survival in several Carpathian regions and long-term isolation between the Carpathians and the Alps. Molecular Ecology 17: 1763-1775.

Ronquist F (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Systematic Biology 46: 195-203.

Rosen DE (1978) Vicariant patterns and historical explanation in biogeography. Systematic Zoology 27: 159-188.

Rozas J and Rozas R (1997) DNASP, version 2.0: a novel software package for extensive molecular population genetic analysis. Computer Applied Biosciences 13: 307-311.

Rozas J, Sanchez-Delbarrio JC, Messeguer X and Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB and Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning: A Laboratory Manual, second edition, Cold Spring Habor Laboratory Press, Plainview, New York.

Sampson J (1998) Multiple paternity in *Eucalyptus rameliana* (Myrtaceae). Heredity 81: 349-355.

Schaal BA and Olsen KM (2000) Gene genealogies and population variation in plants. Proceedings of the National Academy of Sciences USA 97: 7024-7029.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT and Smith WA (1998) Phylogeographic studies in plants: problems and prospects. Molecular Ecology 7: 465-474.

Schierup MH and Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. Genetics 156: 879-891.

Schnabel A and Asmussen MA (1989) Definition and properties of disequilibria within nuclear-mitochondrial-chloroplast and other nuclear-dicytoplasmic systems. Genetics 123:199-215.

Schönswetter P, Stehlik I, Holderegger R and Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. Molecular Ecology 14: 3547-3555.

Schönswetter P, Tribsch A, Barfuss M and Niklfeld H (2002) Several Pleistocene refugia detected in the high alpine plant *Phyteuma globulariifolium* Sternb. & Hoppe (Campanulaceae) in the European Alps. Molecular Ecology 11: 2637-2647.

Schönswetter P, Tribsch A, Schneeweiss GM and Niklfeld H (2003) Disjunctions in relict alpine plants: phylogeography of *Androsace brevis* and *A. wulfeniana* (Primulaceae). Botanical Journal of the Linnean Society 141: 437-446.

Selkoe KA and Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9: 615-629.

Sharbel TF, Haubold B and Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and post-glacial colonization of Europe. Molecular Ecology 9: 2109-2118.

Shih F-L, Hwang S-Y, Cheng Y-P, Lee P-F and Lin T-P (2007) Uniform genetic diversity, low differentiation, and neutral evolution characterize contemporary refuge populations of Taiwan fir (*Abies Kawakamii*, Pinaceae). American Journal of Botany 94: 194-202.

Shrestha TB and Joshi RM (1997) Biodiversity gap analysis: terrestrial ecoregions of the Himalaya (Nepal). Draft Report submitted to WWF-Nepal Program, Lal Durbar, Kathmandu, Nepal.

Shu G, Weber A, Lena CH and Baum DA (2000) *LEAFY* and the evolution of rosette flowering in Violet cress (*Jonopsidium acaule*, Brassicaceae). American Journal of Botany 87: 634-641.

Silvertown J (2004) The ghost of competition past in the phylogeny of island endemic plants. Journal of Ecology 92: 168-173.

Singh G and Gill BS (1984) IOPB chromosome number reports LXXXII. Taxon 33: 126-134.

Smith JM and Haigh J (1974) The hitch-hiking effect of a favourable gene. Genetical Research 23: 23-35.

Smouse PE (1998) To tree or not to tree. Molecular Ecology 7: 399-412.

Soejima A and Nagamasu H (2004) Phylogenetic analysis of Asian *Symplocos* (Symplocaceae) based on nuclear and chloroplast DNA sequences. Journal of Plant Research 117: 199-207.

Solheim II WG (1972) An earlier agricultural revolution. Scientific American 266: 34-41.

Soranzo N, Provan J and Powell W (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. Genome 42: 158-161.

Speranza PR, Seijo JG, Grela IA and Solis Neffa VG (2007) Chloroplast DNA variation in the *Turnera sidoides* L. complex (Turneraceae): biogeographical implications. Journal of Biogeography 34: 427-436.

Strand AE, Leebens-Mack J and Milligan BG (1997) Nuclear DNA based markers for plant evolutionary biology. Molecular Ecology 6: 113-118.

Stranger BE and Mitchell-Olds T (2005) Nucleotide variation at the myrosinase-encoding locus, *TGG1*, and quantitative myrosinase enzyme activity variation in *Arabidopsis thaliana*. Molecular Ecology 14: 295-309.

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J and Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L and *Q. petraea* (Matt) Liebl. Molecular Ecology 8: 831-841.

Su Y-J, Wang T, Zheng B, Jiang Y, Chen G-P, Outang P-Y and Sun Y-F (2005) Genetic differentiation of relictual populations of *Alsophila spinulosa* in southern China inferred from cpDNA *trnL-F* noncoding sequences. Molecular Phylogenetics and Evolution 34: 323-333.

Sukumar R, Ramesh R, Pant RK and Rajagopalan G (1993) A $\delta^{13}C$ record of late Quaternary climate change from tropical peats in southern India. Nature 364: 703-706.

Taberlet P, Gielly L, Pautou G and Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17: 1105-1109.

Taberlet PL, Fumagalli L, Wustsaucy AG and Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7: 453-464.

Tachida H and Iizuka M (1992) Persistence of repeated sequences that evolve by replication slippage. Genetics 131: 471-478.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595

Takayama K, Kajita T, Murata J and Tateishi Y (2006) Phylogeography and genetic structure of *Hibiscus tiliaceus*- speciation of a pantropical plant with sea-drifted seeds. Molecular Ecology 15: 2871-2881.

Tani N, Tsumura NY and Sato H (2003) Nuclear gene sequences and DNA variation of *Cryptomeria japonica* samples from postglacial period. Molecular Ecology 12: 859-868.

Templeton AR, Routman E and Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA sequence data III cladogram estimation. Genetics 140: 767-782.

Terry RG, Nowak RS and Tausch RJ (2000) Genetic variation in chloroplast and nuclear ribosomal DNA in Utah juniper (*Juniperus osteosperma*, Cupressaceae): evidence for interspecific gene flow. American Journal of Botany 87: 250-258.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25**:** 4876-4882.

Tikader BK (1983) Threatened animals of India. Zoological Survey of India, Calcutta.

Todokoro S, Terauchi R and Kawano S (1995) Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. The Japanese Journal of Genetics 70: 543-554.

Torres-Diaz C, Ruiz E, González F, Fuentes G and Cavieres LA (2007) Genetic diversity in *Nothofagus alessandrii* (Fagaceae), an endangered endemic tree species of the coastal maulino forest of central Chile. Annals of Botany 100: 75-82.

Trewick SA (2000) Molecular evidence for dispersal rather than vicariance as the origin of flightless insect species on the Chatham Islands, New Zealand. Journal of Biogeography 27: 1189-1200.

Trewick SA, Morgan-Richards M, Russell SJ, Henderson S, Rumsey FJ, Pinter I, Barrett A, Gibby M and Vogel JC (2002) Polyploidy, phylogeography and Pleistocene refugia of the rockfern *Asplenium ceterach*: evidence from chloroplast DNA. Molecular Ecology 11: 2003-2012.

Tribsch A and Schönswetter P (2003) Patterns of endemism and comparative phylogeography confirm palaeo-environmental evidence in the Eastern Alps. Taxon 52: 477-498.

Tribsch A, Schönswetter and Stuessy T (2002) *Saponaria pumila* (Caryophyllaceae) and the ice age in the European Alps. American Journal of Botany 89: 2024-2033.

Truong C, Palme AE and Felber F (2007) Recent invasion of the mountain birch *Betula pubescens* ssp. *tortuosa* above the treeline due to climate change: genetic and ecological study in northern Sweden. Journal for Evolutionary Biology 20: 369-380.

Tsumura Y, Taguchi H, Suyama Y and Ohba K (1994) Geographical cline of chloroplast DNA variation in *Abies mariesii*. Theoretical and Applied Genetics 89: 92-96.

Tzedakis PC, Lawson IT, Forgley MR and Hewitt GM (2002) Buffered tree population changes in a Quaternary refugium: evolutionary implications. Science 297: 2044-2047.

Udvardy MDF (1969) Dynamic zoogeography. Van Nostrand Reinhold, New York.

Upadhyay RR (1999) Ecological problems due to shifting cultivation. Current Science 77: 1246-1250.

Valdiya KS (2002) Emergence and evolution of Himalaya: reconstructing history in the light of recent studies. Progress in Physical Geography 26: 360-399.

Van Rossum F, Bonnin I, Fénart S, Pauwels M, Petit D and Saumitou-Laprade P (2004) Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy metal-tolerant species. Molecular Ecology 13: 2959-2967.

Vendramin GG and Ziegenhagen B (1997) Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*. Genome 40: 857-864.

Vendramin GG, Anzidei M, Madaghiele A, Sperisen C and Bucci G (2000) Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). Genome 43: 68-78.

Vendramin GG, Lelli L, Rossi P and Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. Molecular Ecology 5: 595-598.

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

Vishnu-Mittre and Gupta HP (1968) A living fossil plant community in South Indian hills. Current Science 37: 671-672.

Vishnu-Mittre and Gupta HP (1970) The origin of Shola forest in the Nilgiris, South India. Paleobotanist 19: 110-114.

Vos P, Hogers R, Bleejer M, Reijans M, Lee T, van de Hornes M, Frijters A, Pot J, Peleman J and Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 11: 4407-4414.

Walter R and Epperson BK (2001) Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. Molecular Ecology 10: 103-111.

Wang Y, Fritsch PW, Shi S, Almeda F, Cruz BC and Kelly LM (2004) Phylogeny and infrageneric classification of *Symplocos* (Symplocaceae) inferred from DNA sequence data. American Journal of Botany 91: 1901-1914.

Warrier PK, Nambiar VPK and Ramankutty C (1996) Indian medicinal plants vol 5: a compendium of 500 species. Orient Longman, India, pp. 215-217.

Webb T and Bartlein PJ (1992) Global changes during the last 3 million years: climatic controls and biotic response. Annual Review of Ecology and Systematics 23: 141-173.

Weider LJ and Hobæk A (2003) Glacial refugia, haplotype distributions, and clonal richness of the *Daphnia pulex* complex in arctic Canada. Molecular Ecology 12: 463-473.

Weising K and Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42: 9-19.

Weising K, Nybom H, Wolff K and Meyer W (1995) DNA fingerprinting in plants and fungi. CRC Press, Boca Raton.

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

Williams CG and Hamrick JL (1996) Elite populations for conifer breeding and gene conservation. Canadian Journal of Forest Research 26: 453-461.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JK and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful genetic markers. Nucleic Acids Research 18: 6531-6535.

Wolfe KH, Li WH and Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondria, chloroplast and nuclear DNAs. Proceedings of the National Academy of Sciences USA 84: 9054-9058.

Wright S (1931) Evolution in Mendelian populations. Genetics 16: 97-159.

Wright S (1951) The genetical structure of populations. Ann Eugen 15: 323-354.

Wu M-J, Huang S-F, Huang T-C, Lee P-F and Lin T-P (2005) Evolution of the *Euphrasia transmorrisonensis* complex (Orobanchaceae) in alpine areas of Taiwan. Journal of Biogeography 32: 1921-1929.

Wu SH, Hwang CY, Lin TP, Chung JD, Cheng YP and Hwang SY (2006) Contrasting phylogeographic patterns of two closely related species, *Machilus thunbergii* and *Machilus kusanoi* (Lauraceae), in Taiwan. Journal of Biogeography 33: 936-947.

Xu DH, Abe J, Gai JY and Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soyabeans: evidence for multiple origins of cultivated soyabean. Theoretical and Applied Genetics 105: 645-653.

Yeh F, Yang RC and Boyle T (2000) PopGene (v.1.32) Microsoft windows-based freeware for population genetic analysis. http://www.ualberta.ca/~fyeh/Pop32.exe.

Zhang D-X and Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular Ecology 12: 563-584.

Zietkiewicz E, Rafalski A and Labuda D (1994) Genome fingerprinting by simple sequence repeats SSR - anchored PCR amplification. Genomics 20: 176-183.

Zink RM (1996) Comparative phylogeography in North American birds. Evolution 50: 308-317.

Zink RM (2002) Methods in comparative phylogeography and their application to studying evolution in the North American arid lands. Integrative and Comparative Biology 42: 953-959.

Sofia Banu

Date of birth: 20th Feb 1979
Email: sofiabanu2@gmail.com or sofia_jorhat@yahoo.com
Permanent Address:
C/o Mrs. Nurjahan Begum
Near Mission Leprosy Hospital
Jail Road, Jorhat
Assam- 785004
Correspondence Address:
C/o Dr. Vidya S. Gupta
Plant Molecular Biology Group
Division of Biochemical Sciences,
National Chemical Laboratory (NCL),
Dr. Homi Bhabha Road,
Pune-411 008, INDIA
+91-20-25902247 (Tel)
+91-20-25902658 (Fax)

Academic Qualifications

M.Sc. Biotechnology, 2002, Gauhati University, Assam, India.

B.Sc. Zoology (Hons.), 1999, Dibrugarh University, Assam, India.

Research Experience

Project Assistant (2002-2003), worked on project entitled "Mass propagation of certain rare and highly valued medicinal plant species of North East India using Tissue Culture technology" sponsored by NATP, New Delhi, India

As Masters Student (2000-2001), completed one year project entitled "*Micropropagation of Phyllanthus fraternus*".

Awards

1. Awarded National merit scholarship during the three year Graduate program for the duration 1996 -1999.
2. Awarded **Junior Research Fellowship (NET)** by CSIR, New Delhi, India in the year 2001(December).

Publications

1. "*In vitro* propagation of *Phyllanthus fraternus*"

S. Banu and P.J. Handique
Journal of Tropical Medicinal Plants, Vol. 4, No.1, (2004)

2. Book chapter entitled "Molecular elucidation of the history and evolution of montane plants of India" to be published in the book "**Diversity analysis**" by The **Ranipet Herbarium**, St. Joseph's College, Tiruchirapalli. (**In press**)
   S. Banu, M.D. Lagu and V.S. Gupta

3. "Molecular phylogeographical studies in populations of *Symplocos laurina* Wall. from India"
   S. Banu, M.D. Lagu and V.S. Gupta
   (Communicated to **Tree Genetics and Genomes**)

4. "Understanding phylogeography of *Symplocos laurina* Wall. Using sequences of non coding of chloroplast DNA"
   S. Banu, N. Kadoo, M.D. Lagu and V.S. Gupta
   (Communicated to **Journal of Plant Research**)

5. "Low genetic diversity and significant population structuring in the relict plant *Symplocos laurina* based on nuclear DNA analysis"
   S. Banu, N. Kadoo and V.S. Gupta
   (To be communicated)

6. "Genetic diversity analysis of *Symplocos racemosa* Roxb. (Symplocaceae) from the biodiversity hotspot - Western Ghats"
   S. Banu, R. Bhagwat, M.D. Lagu and V.S. Gupta
   (To be communicated)

7. "Chloroplast and mitochondrial DNA diversity in *Gaultheria fragrantissima* from the two biodiversity hotspots in India"
   G.S. Apte, R.S. Kulkarni, S.Banu, M.D. Lagu, B.G. Kulkarni, S.P. Joshi and V.S. Gupta
   (To be communicated)

Conferences

1. Awarded **best oral presentation** for the paper entitled "Genetic diversity in natural populations of *Gaultheria fragrantissima* in two biodiversity hotspots in India"
   S. Banu, G.S. Apte, M.D. Lagu, R.S. Kulkarni, S.G. Borse, B.G. Kulkarni, S.P. Joshi and V.S. Gupta presented at National Workshop on Biodiversity Resources Management and Sustainable use; 11-15 October 2004, Center for Biodiversity and Forest Studies, Madurai Kamraj University, Madurai, India.

2. Presented poster entitled "Genetic and Biomolecular prospecting of two medicinally potent species of *Symplocos* from the two biodiversity hotspots in India"
   S. Banu, R.S. Kulkarni, S.G. Borse, B.G. Kulkarni, M.D. Lagu, S.P. Joshi and V.S. Gupta at International conference on 'Plant Genomics and Biotechnology: Challenges and Opportunities"; 26-28 October 2005, IGAU, Raipur, India.

3. Presented Poster entitled "Genetic and Biomolecular prospecting of *Symplocos laurina* from the two biodiversity hotspots in India".
   S. Banu, R.S. Kulkarni, M.D. Lagu, S.P. Joshi and V.S. Gupta at Symposium on "National Biodiversity and Ecosystem Information Infrastructure: Challenges and Potentials"; January 30-Febuary 2 2006, National Chemical Laboratory, Pune, India.

4. Awarded **best poster presentation** for the paper entitled "*Symplocos laurina*: A glimpse into its present and past analyzed using molecular markers"
   S. Banu, R. Bhagwat, M.D. Lagu, S.P. Joshi and V.S. Gupta at Research Scholar Meet; 21 February 2007, National Chemical Laboratory, Pune, India.