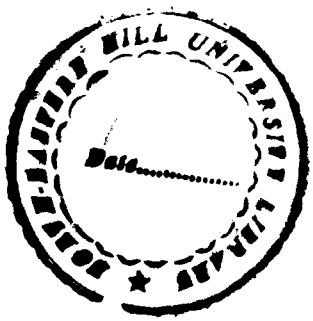
ESTABLISHMENT OF PHYLOGENETIC RELATIONSHIP IN *MYRICA* SPECIES OF MEGHALAYA AND DEVELOPMENT OF MOLECULAR MARKERS FOR EARLY SCREENING OF *MYRICA* TREES

By

MHATHUNG YANTHAN



THESIS SUBMITTED
IN FULFILMENT OF THE DEGREE OF
DOCTOR OF PHILOSOPHY IN BOTANY

NORTH-EASTERN HILL UNIVERSITY
SHILLONG-793022, INDIA
2009

Thesis

Dedicated to mum and dad

The North-Eastern Hill University

February, 2009

DECLARATION

I, Mhathung Yanthan, hereby declare that the subject matter of this thesis entitled "**Establishment of phylogenetic relationship in *Myrica* species of Meghalaya and development of molecular markers for early screening of *Myrica* trees**" is the record of work done by me, that the contents of this thesis did not form any basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the thesis has not been submitted by me for any research degree in any other University/ Institute.

This is being submitted to the North-Eastern Hill University for the award of the degree of Doctor of Philosophy in Botany.

(Mhathung Yanthan)

(Head of Department)

Head
Department of Botany
School of Life Sciences
N.E.H.U., Shillong-22

(Supervisor)

ACKNOWLEDGEMENTS

Foremost, I express my earnest gratitude and deep respect to none other than my supervisor, Prof. Arvind Kumar Misra, under whose erudite guidance and refined expertise this present study could finally take the shape of a thesis. All his valuable suggestions and insightful conversations were forthcoming during the development of this thesis. My gratefulness shall ever remain.

I thank Prof. N. K. Chrungoo (former Head, Botany) for allowing me access to his lab facilities unreserved and willing as and when approached.

I thank Prof. M. S. Dkhar, Head, Botany, for providing me the facilities to conduct experiments in the Department.

My earnest thanks are due to Prof. R. S. Tripathi (Retd. and former Head, Botany) and Prof. H. N. Pandey (Retd. and former Head, Botany) for their much needed moral encouragement and permission to use the Departmental facilities and also to Prof. S. K. Barik for the help and encouragement received during the course of this work.

Due thanks are to Dr. Anindita Khan for being a constant source of help and encouragement during the formative years of my research.

My labmates Mr. David Lalsama Biate, Mr. Syntiew Manik Syiemlieh, Ms. Meena Elangbam, Ms. Asem Mipeshwaree, Ms. Jubanri Nongkynrih and Ms. Insan Ara Rahman are thanked wholeheartedly for all the encouragement, selfless help and the much needed laughter received during the course of our working together. Special mention of Mr. David and Mr. Manik are made for bringing me closer to the world of computers and also for their enduring accompaniment during the collection trips even in the worst of weathers. My gratitude shall ever remain.

I also thank all the members of Molecular Physiology Lab, Botany, especially, Dr. Sangeeta Bharali, Dr. Anusuya Rout, Mr. Lalliansanga, Ms. Cressida Jamir, Mr. Tiatemsu Mollier, Ms. Shiny Sangma and Ms. Nabanita Devadasan for their incessant cooperation received at all times.

I wish to thank Ms. Lucy B. Nongbri and Mr. Arun Chettri (Forest Ecology lab, Botany) for helping me with the references and the ecological experiments.

I wish to thank Dr. P. B. Gurung, STA, Botany, for lending his expert reviews to my taxonomical queries and also Mr. Kumar Sohlang for rendering his photography expertise at the time of my need.

I also wish to thank all the non-teaching staffs of the Department for their cooperation and help received during the course of my work.

I thank CSIR for providing me financial assistance in the form of NET Fellowship during the course of my research.

It would be incomplete if I fail to mention Bah Mudin (Retd.) and Bah Desh, Departmental drivers, for their ever willingness to drive me to my study field whenever needed. I thank them both with gratitude.

Last but not the least I thank the almighty God for my parents and dear ones at home who have endured and filled me with their unceasing love all through my years of research.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

CHAPTER 1

INTRODUCTION

The availability of mineral nutrients, most importantly nitrogen, determines the growth of plants. Despite its presence in large quantities in the atmosphere (approximately 80%), plants cannot directly utilize free nitrogen. Nitrogen is required in large amounts as an important component of proteins, nucleic acids and other cellular constituents and is frequently found limiting to the growth of the green plants. This is because of the continual loss of nitrogen from soil reserves by processes such as microbial denitrification, soil erosion, leaching, chemical volatilization and removal of nitrogen-containing crop residues from the land. The nitrogen reserve of the agricultural soils must therefore be replenished periodically in order to maintain an adequate (non-growth limiting) level for crop production. This replacement of soil nitrogen is brought about by either adding nitrogen containing chemical fertilizers directly in the soil or by the activity of Biological Nitrogen Fixation (BNF) systems. The conversion of nitrogen gas or atmospheric nitrogen into nitrate, a form of nitrogen that can be metabolized into amino acids and proteins by terrestrial plants, is called nitrogen fixation. In the living world major conversion of N_2 into ammonia is brought about by certain prokaryotes by a process called as Biological Nitrogen Fixation. The biological nitrogen fixation involves two types of micro-organisms; non-symbiotic or free living micro-organisms like *Clostridium, Klebsiella, Azotobacter*, etc. and symbiotic micro-organisms like *Frankia*

and *Rhizobium*. The non-symbiotic organisms bring about nitrogen fixing process independent of a host requirement whereas the symbiotic organisms require a host plant to actually affect nitrogen fixation. In symbiotic fixation the nitrogen fixed by these micro-organisms is made available to the host plant whereas the microsymbionts, in turn, use organic compounds supplied by the plant as an energy source (Raven *et al.* 1986). Nitrogen fixing plants are key constituents in many natural ecosystems in the world. Many nitrogen fixing plants are woody perennials, or nitrogen fixing trees (NFTs), most of these being found in the tropics. In temperate areas, the nitrogen fixers tend to be herbaceous. *Rhizobium* inoculates trees in the families Leguminosae and Ulmaceae, while *Frankia* inoculates the actinorhizal trees.

Symbiotic biological nitrogen fixation involving the actinomycete *Frankia* nodulates roots of dicotyledonous plants belonging to 8 plant families and 25 genera called as actinorhizal genera. These 8 plant families are Betulaceae, Casuarinaceae, Coriariaceae, Datiscaceae, Elaeagnaceae, Myricaceae, Rhamnaceae and Rosaceae. Although taxonomically diverse, the actinorhizal plants have some common features. All are perennial dicots and all, except *Datisca* which has herbaceous shoots, are woody shrubs or trees (Tjepkema *et al.*, 1986). Actinorhizal plants are widely distributed and are found in every continent except Antartica. They are mostly found in the temperate zone. Some species of Casuarinaceae and Myricaceae are considered tropical. Countries nearer the pole (Scandinavia, Canada and New Zealand) are home to many of these trees. In the subcontinent India, they are confined to higher altitudes in Himachal Pradesh, Jammu and Kashmir, Arunachal Pradesh, Sikkim, West Bengal hills,

Meghalaya and to some extent in the coastal regions and plains. Some of the genera commonly found in India are *Alnus*, *Casuarina*, *Coriaria*, *Elaeagnus*, *Hippophae* and *Myrica*. Actinorhizal plants are pioneers on nitrogen poor soils including sandy and gravelly sites, shores of streams and lakes, wetlands and exposed raw mineral soils.

Actinorhizal plants find economical use as timber, fuel wood, in land reclamation, biomass production and in forestry (Chauhan, 2000). In Western Europe actinorhizal plants are used primarily for reclamation of industrial wastelands and for land stabilization. Actinorhizal trees such as *Alnus* have been used in the reclamation of mine spoils in Britian (Schwintzer and Tjepkema, 1990) and *Alnus*, *Elaeagnus* and *Hippophae* have been widely used for land stabilization. In Eastern Europe and China, *Hippophae rhamnoides* is cultivated for its fruits for human consumption. Several actinorhizal trees like *Alnus* and *Casuarina* are used as timber and fuel wood. They are also planted extensively as windbreaks and to stabilize dunes against wind erosion. Actinorhizal trees are also valuable for reclamation of mine spoils and rehabilitation of wastelands.

About 97 species of *Myrica* Linn (Myricaceae) are reported globally and this genus is widely distributed in both the temperate and sub-tropical regions with maximum species diversity in Africa and Boreal America (Index Kewensis, 1977-1985) and only one representative has been reported in Australia (*Myrica australiasica* F. Muell). In Asia it is confined mostly to Indo-China and Java regions. Its geographical distribution in India can be seen in sub-tropical Indian Himalaya from Ravi eastwards to Assam and in Khasi, Jaintia, Naga and Lushai Hills at altitudes between 900-1200

meters above mean sea level (Osmaston, 1987). However, they have also been seen in higher reaches of Himachal Pradesh.

Trees of *Myrica* grow well in nitrogen depleted soils and are common associates of pine (*Pinus sp.*) and Oak (*Quercus leucotrichophora*). They are also found in mixed forests and in agricultural and marginal lands (Bhatt *et al.*, 2000).

Myrica has been variously used. Fruits of this tree are used in making jams, syrups, and juices and can also be consumed raw. Bark is used in making paper and rope. Mention of *Myrica esculenta* in the medicinal system of *ayurveda* is noteworthy. The extract from the root and bark is known to be astringent, carminative and antiseptic and especially a decoction of the bark is considered useful in asthma, toothache, diarrhoea, fever, lung infection, cough, chronic bronchitis and dysentery. Bark is chewed to relieve toothache and headache (Kirtikar and Basu, 1984) and also as fish poison (The Wealth of India, 1962). Tannin extracted from the bark is used as a tanning and dyeing agent (Dhyani and Dhar, 1994). The oil obtained from the flowers of *Myrica* is also reported to have medicinal effect especially in ear-ache, inflammation and paralysis (Kirtikar and Basu, 1984). Myriconol isolated from stem bark is reported to have lesser toxicity than related rotenone (Rastogi and Mehrotra, 1991).

Smith (1977) gave the following classification for genus *Myrica*: Division- Magnoliophyta; Class- Magnoliopsida; Subclass- Hamamelidae; Family- Myricaceae; Order- Myricales. The subclass Hamamelidae comprises of Myricales, Fagales, Juglansadales, Hamamelidales, Urticales, Leitneriales, and Casuarinales. This is a phylogenetic grouping of orders which are characterized by strongly reduced, often

unisexual flowers which either lack or produce a poorly developed perianth (Cronquist, 1978). Similar morphological features are observed in the order Myricales and those of the Casuarinales, Fagales, and Juglandales, which are thought to have been derived from the order Hamamelidales (Takhtajan, 1969). Phylogenetically, the Myricales are thought to be most closely related to the Juglandales and the Fagales (Takhtajan, 1969; Cronquist, 1978). The Myricaceae is considered to be an ancient family by taxonomists, dating to the Tertiary Epoch of the Cretaceous Period with the living members representing relics of once extensive tracts of subtropical forest that spread across the territory that is now central and southern Europe (Sporne, 1975; Takhtajan, 1969). Plants of the family Myricaceae are considered to be promiscuous hosts because several species are effectively nodulated by most isolated strains of *Frankia*. The base chromosome number throughout the family is eight, with various levels of ploidy present (Macdonald, 1989).

Myrica esculenta Buch.-Ham. ex D. Don (Syn. *M. farquhariana* Wall., *M. sapida* Wall., *M. nagi* Thunb., *M. integrifolia* Roxb.) belongs to the family Myricaceae and is commonly known as 'Soh-Phi' in Khasi, 'Nagatenga' in Assamese and 'Kaiphal' in Hindi. Common name in English is Box myrtle. This is the only species of the genus *Myrica* reported to be found in India (Haridasan and Rao, 1987; Kanjilal, 1940; Bor, 1953). This tree is distributed in the Chinese-Japanese region including the Sub-Himalayan tract, Khasi hills, Sylhet and Southwards up to Singapore and in the Malayan islands at an altitude of 1600-2000 m above mean sea level. They have been recognized

by many pioneer workers (Watt, 1891; Tanaka, 1976; Zeven and Wet, 1982; Arora, 1985; Arora and Pandey, 1996; Pareek *et al.*, 1998).

Myrica esculenta is a small to moderate sized, evergreen, dioecious tree with height ranging from 3-15 m. Leaves are crowded and aromatic and are lanceolate to oblanceolate or obovate, nearly entire or sharply serrate, obtuse or acute, almost glabrous with resinous dots beneath. New leaves usually start sprouting in February-March. Flowers are minute, unisexual, male and female flowers are borne on different trees. Male flowers occur in catkin and the female flowers in axillary erect spikes. Fruits are edible and ellipsoid or ovoid drupes of the size of cherry or bigger, tubercled, reddish or cheese coloured when ripe with rugose nut. They are covered with a crust of white waxy material, permeated with brown and black spots. They ripen during summer and possess a pleasant sourish sweet taste. Fruiting starts at the beginning of April and lasts till the month of June.

Interestingly, morphological diversity within *Myrica esculenta* exists with regard to fruit size, fruit colour, leaf serration, leaf size, etc. This has led to confusion among various workers. Some workers claim that the different morphotypes of the plant belong to different species. One of the morphotypes is referred to as *Myrica nagi* (Thunberg in Murray, Syst. Veg., ed. 14, 884. 1784) by them and the other as *Myrica esculenta* (Buchanan-Hamilton ex D. Don, Prodr. Fl. Nepal. 56. 1825). Others claim that this name is synonym of the same species, *M. esculenta* (Haridasan and Rao, 1987; Kanjilal, 1940; Dhyani and Dhar, 1994). These claims are based on morphological descriptors. Therefore, in an effort to resolve this dispute a need was felt to study molecular

phylogeny of this genus using nucleotide sequence data of different morphotypes. The first molecular phylogenetic study on the family Myricaceae was done by Huguet *et al.* (2005) where the molecular phylogeny of 13 species of the family Myricaceae was established based on *rbcL* gene and the 18S-26S ITS. Their results showed that some species of the genus *Myrica*, such as *Myrica gale* and *Myrica hartwegii*, and genus *Comptonia*, belong to a distinct phylogenetic cluster distinct from some other *Myrica* species. They transferred the latter *Myrica* species to a new genus, *Morella*. However, the taxonomy within this family is highly controversial because of morphological variation exhibited by many species.

Approach to both fundamental and applied biological problems have been transformed by the emergence of many new techniques. Amplification of specific regions of DNA to a million fold has been made possible by the Polymerase Chain Reaction (PCR) technique. This method was first invented by Kary Mullis in 1983 and was originally applied by a group in the Human Genetics Department at Cetus to amplify human β-globin DNA and prenatal diagnosis of sickle-cell anaemia (Elrich, 1989). PCR has revolutionized approaches to molecular study in many fields. Different modifications of the PCR technique are available for studying polymorphisms in DNA. In 'Alu-PCR', primers specific for the human *Alu* repeat allow the selective amplification of human sequences from hybrid cell lines containing both human and rodent genomic DNA (Gusella *et al.*, 1980). Another modification of PCR method known as 'inverse PCR' has been developed for the analysis of sequences that flank a known region (Triglia *et al.*, 1988; Ochman *et al.*, 1988). Other VNTR regions that have

been analyzed by PCR are single-locus markers with many alleles (Horn *et al.*, 1989). PCR based techniques like Asymmetric PCR, RT-PCR, Anchored-PCR, etc. are also available for studying polymorphisms in DNA at various levels. Restriction site analysis of amplified DNA (Mullis and Faloona, 1987; Kogan *et al.*, 1987) is a valuable method for detecting genetic variation in some cases. Length polymorphisms in the amplified product resulting from Variable Number of Tandem Repeats (VNTR) in the template have also been used as informative genetic markers (Jeffreys *et al.*, 1988; Horn *et al.*, 1989).

The ribosomal RNA genes family comprises of very conserved regions (i.e., the 18S and 26S gene) that can be used to infer phylogeny at higher taxonomic levels, as well as more rapid evolving segments (i.e., ITS, IGS) that may be useful at the generic, specific and even (in case of IGS) at the population level. Ribosomal DNA cistrons typically are located in the nucleolar organizing region (NOR) and may be present on several different chromosomes (Thompson and Flavell, 1988). The internal transcribed spacer (ITS) region of the 18S–5.8S–26S nuclear ribosomal cistron is one of the most popular sequences for phylogenetic inference at the generic and infrageneric levels in plants (Varghese *et al.*, 2003). The ITS-1 and ITS-2 regions are part of the nuclear rDNA transcript but are not incorporated into ribosomes but appear to play a role in the maturation of nuclear rRNAs, bringing the large and small subunits into close proximity within a processing domain (reviewed in Baldwin *et al.*, 1995). The need for sequence data from nuclear genome at lower taxonomic levels makes the ITS region a popular site for this study. Since the ITS region is G–C rich and prone to secondary structure,

sequencing can be difficult (Baldwin *et al.*, 1995). Different protocols have been used to amplify and sequence the ITS regions (Baldwin, 1992; Wen and Zimmer, 1996; Soltis and Kuzoff, 1995). Since ITS sequences lie in-between highly conserved regions of the ribosomal RNA genes, it is possible to design primers that are complementary to the conserved sequence for amplification purpose.

Three different morphotypes of *Myrica* sp. are found in Meghalaya. Morphotype 1 trees are considered as *Myrica nagi* by some workers. Morphotype 2 trees are considered as *Myrica esculenta*. Morphotype 3 trees have not been described separately and may have been considered as variants of morphotype 2 trees. As stated above, some workers consider all these three morphotypes as members of the species *Myrica esculenta.* With a view to resolve the dispute related to classification of this genus and to develop molecular tools for early screening of better nitrogen fixers, following objectives were set for this study:

I. Phylogenetic analysis of the DNA sequence data of the variable 18S-28S ITS region and the conserved 18S ribosomal RNA gene of the nuclear genome of different morphotypes of *Myrica* sp.

II. Development of molecular markers based on Amplicon Restriction Pattern (ARP) for early screening of superior nitrogen fixer seedlings of *Myrica* sp.

REVIEW OF LITERATURE

CHAPTER 2

REVIEW OF LITERATURE

Recent advances in the field of phylogenetic analysis have been made possible by the advent of new molecular techniques that have revolutionized the traditional methods previously used to study plant systematic and genetics, which were earlier based on morphological and geographical variations between organisms (Joshi *et al.*, 1999). One of the most significant techniques is the polymerase chain reaction (PCR) developed by Kary Mullis in 1983. The rationale behind the principle is the denaturing and annealing of DNA using specific primers combined with other components like *Taq* polymerase, dNTPs, enzyme buffer, etc. and subsequently amplifying the target DNA to a million fold (Bagasra and Hansen, 1997). Initially, PCR relied on the thermolabile Klenow fragment of DNA Polymerase I which degraded at higher temperatures required for DNA denaturation, thereby requiring a fresh addition of the enzyme every cycle. With the introduction of thermo-stable *Taq* polymerase, extracted from the bacterium *Thermus aquaticus,* the cyclic process has been made easier and less time consuming (Lakobashvili and Lapidot, 1999).

One of the most commonly used genetic marker techniques is RFLP (restriction fragment length polymorphism) that has its origin in the DNA rearrangements, occurring due to mutations, insertions or deletions (Burr *et al.*, 1983) and unequal crossing over (Schlotterer and Tautz, 1992). In the RFLP technique, genomic DNA is first digested

with restriction enzymes and the resultant fragments separated by electrophoresis on an agarose gel which is then blotted onto a filter and finally hybridized with specific probes. RFLP are co-dominant markers yielding highly reproducible patterns. But the technique is time consuming and requires large quantity of good quality DNA and also good supply of probes as well (Karp *et al.*, 1996).

PCR-based studies of molecular markers have gained popularity over the years. Randomly-amplified Polymorphic DNA (RAPD) is a PCR-based genetic assay developed by Welsh and McClelland in 1991 that uses a single primer of arbitrary nucleotide sequence to detect nucleotide sequence polymorphisms. The single species of primer anneals to the genomic DNA at two different sites on complementary strands of the template DNA allowing amplification of several discrete fragments in the genome. The amplification products are run on agarose gels stained with ethidium bromide and resolved under ultraviolet light (Williams *et al.*, 1990). RAPDs have been used to screen polymorphism among individuals (Tingey *et al.*, 1993), in molecular ecology (Hardys *et al.*, 1992) and for taxonomy (Karp *et. al.*, 1996). DAF (DNA amplification fingerprinting) and AP-PCR (arbitrary primed PCR) are PCR-based modified RAPD techniques. DAF employs the use of short arbitrary primers, which may be as short as 5 bases, to amplify DNA using PCR. The band patterns are reproducible and are usually analyzed by polyacrylamide gel electrophoresis and also by silver staining (Caetano-Anolles *et al.*, 1991). AP-PCR is a special type of RAPD in which single primers of 10-50 bases in length are used to generate amplification products from genomic DNA which are then analyzed by autoradiography as also by ethidium bromide stained

agarose gels (Welsh and McClelland, 1991). Another recent technique called AFLP (amplified fragment length polymorphism) designed by Zabeau *et al.* (1993) detects DNA restriction fragments by means of PCR amplification. There is no requirement of any prior information of the sequence to generate AFLP fingerprints. AFLP technique involves three basic steps: Restriction digestion of genomic DNA using appropriate enzymes, selective amplification of the restriction fragments by PCR and gel analysis of the amplified fragments (Joshi *et al.*, 1999). Regardless of the origin and complexity of the genome, AFLP generates unique fingerprints that can be exploited as landmarks in genetic and physical mapping (Vos *et al.*, 1995). CAPS (cleaved amplified polymorphic DNA) or PCR-RFLP, is a fast approaching technique where the amplified product is digested with specific restriction enzymes and the profiles directly visualized on an ethidium bromide stained agarose gel (Akopyanz *et al.*, 1992). The requirement of only a minute amount of DNA is one of the advantages of this technique. However, prior sequence information is needed to tag the desired DNA fragment. PCR-RFLP technique has been employed by many workers to study a number of crop and forest species for which extensive sequence information is available (Chen *et al.*, 1994; Ghareyazie *et al.*, 1995; Tsumura *et al.*, 1997; Perry *et al.*, 1999; Chauhan and Misra, 2002). Many other techniques like VNTR (Variable Number of Tandem Repeats), SSR (Single Sequence Repeats), RLGC (Restriction Landmark Genomic Scanning), STS (Sequence-tagged Sites), EST (Expressed Sequence Tag Markers), STMS (Sequence-tagged Microsatellite site markers), TGGE (Thermal Gradient Gel Electrophoresis), etc. have their own

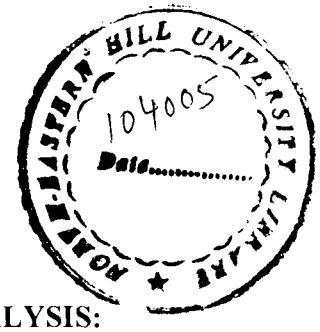
applicability at different resolving levels in carrying out molecular studies of the species concerned.

2.1. CLASSIFICATION OF ACTINORHIZAL PLANTS:

Basic molecular phylogenetic tools of the present day were not available in the early days and therefore most of the early theories of evolution were based on morphological and geographical characters that existed among organisms (Joshi *et al.*, 1999). Great taxonomic diversity exists among actinorhizal plants based primarily on comparative morphology and anatomy implying distant relationships. None of these classifications is actually based on phylogenetic analysis of morphological characters (Swensen and Mullin, 1997). The existing classification system of the actinorhizal plants has been based on the well-known schemes of Cronquist (1981), Dahlgreen (1980), Takhtajan (1980) and Thorne (1992). About 194 different plant species have been identified as symbiotic hosts for the actinomycete *Frankia.* Most of them are woody plants classified among eight different families and four subclasses of the flowering plants according to Cronquist (1981) (Table 2.1).

TABLE 2.1. CLASSIFICATION OF ACTINORHIZAL PLANT GENERA ACCORDING TO CRONQUIST (1981)

Subclass	Family	Genus	Number of species
Hamamelidae	Betulaceae	*Alnus*	47
Hamamelidae	Casuarinaceae	*Allocasuarina*	54
		Casuarina	16
		Ceuthostoma	2
		Gymnostoma	18
Hamamelidae	Myricaceae	*Comptonia*	1
		Myrica	28
Rosidae	Elaeagnaceae	*Elaeagnus*	38
		Hippophae	2
		Shepherdia	2
Rosidae	Rhamnaceae	*Caenothus*	31
		Colletia	4
		Discaria	5
		Kentrothamnus	1
		Retanilla	2
		Talguenea	1
		Trevoa	2
Rosidae	Rosaceae	*Cercocarpus*	4
		Chamaebatia	1
		Cowania	1
		Dryas	1
		Purshia	2
Magnoliidae	Coriariaceae	*Coriaria*	5
Dilleniidae	Datiscaceae	*Datisca*	2



2.2. CHOICE OF GENES FOR PHYLOGENETIC ANALYSIS:

The vast array of genes available today to study the phylogeny of any organism is a matter of choice depending on what level of evolutionary study it is proposed for. Given the choice of molecular approaches available to carry out such study it is imperative to choose the appropriate gene for drawing correct inferences. Choices range from plastid, mitochondrial to genomic DNA.

In the early years much emphasis was given on the chloroplast genome with the nuclear gene sequences slowly gaining its popularity in the recent years (Soltis and Soltis, 1998). Chloroplast genome, which is circular in structure, has two inverted repeats separating the genome into a large and a small single copy region. This structure is found in many land plants with some exceptions (Soltis and Soltis, 1998). Chloroplast genome being small in size (approximately 120 to 200 kb), it is relatively easy to study the entire genome through restriction site analysis (Soltis *et al.*, 1992) since most of the genes are single copy (Palmer, 1985a, 1985b, 1986). The *rbcL* gene, which is located in the large single-copy region of the chloroplast genome, is one of the most popular genes studied. It encodes the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase (RUBISCO). Extensive work on *rbcL* gene for phylogenetic analysis has been emphasized by many workers (Palmer *et al.*, 1988a; Clegg, 1993; Chase *et al.*, 1993). The *rbcL* gene has been most widely used to infer phylogeny at the family level and above because of its conservative evolution (Soltis and Soltis, 1998). However, at the lower taxonomic levels the use of chloroplast DNA for inferring phylogeny has its disadvantage because of the potential occurrence of chloroplast genome transfer from

one species to another by introgression (Harris and Ingram, 1991; Rieseberg and Soltis, 1991; Risieberg and Brunsfeld, 1992). Another highly conserved region, the 16S rRNA gene is located within the inverted repeat of the chloroplast genome and encodes for the small subunit of the ribosomal RNA. Because of its conserved nature they have been used to reconstruct phylogeny (Wolters and Erdmann, 1986; Pace *et al.*, 1986). The 16S ribosomal RNA gene is approximately 1,600 bp in length (Soltis and Soltis, 1998) and has been used to conduct broad phylogenetic analysis of cyanobacteria (Woese, 1987; Giovannoni *et al.*, 1988; Urbach *et al.*, 1992) and also in fern (Manhart, 1995) and *Frankia* (Normand *et al.*, 1996). A complete set of amplification and sequencing primers of the 16S ribosomal RNA has been provided by Manhart (1995). Other chloroplast genes which have been used to study phylogenetic relationships in plants include the *atpB* (Ritland and Clegg, 1987), *matK* (Johnson and Soltis, 1995), *ndhF* (Kim and Jansen, 1995), *trnL* (Gielly and Taberlet, 1996), *trnF* (Gielly and Taberlet, 1996), *rps2* (dePamphilis *et al.*, 1997) and *rps4* (Nadot *et al.*,1994).

Mitochondrial genome has been a major focus of phylogenetic studies more in animals (Harrison, 1989; Hillis *et al.*, 1996; Mitton, 1994) than in plants. The reason being its fairly large size, high variability, presence of many foreign DNA, particularly chloroplast DNA sequences, appearance of large duplications, high recombination frequency and slow rate of nucleotide substitution (Palmer, 1992). However, the frequent rearrangements in the mitochondrial DNA have an advantage since it can be easily distinguished by restriction patterns (Palmer, 1992). This proves valuable for studying genome types at the lower taxonomic levels (Palmer, 1992; Palmer and

Hebron, 1988). The occasional loss of mitochondrial genes and introns could also be exploited for systematic purposes (Doyle *et al.*, 1995).

Despite its large size the nuclear genome has been studied extensively by many workers with special attention to the ribosomal DNA cistron. The rDNA (Fig. 1) of the nuclear genome consists of variable regions (i.e., ITS, IGS) interspersed with highly conserved sequence (i.e., the 18S, 26S/28S genes) that can be utilized to infer phylogeny at different taxonomic levels (Soltis and Soltis, 1998). The 18S gene is approximately 1,800 bp (Nickrent and Soltis, 1995), the 26S/28S gene about 3,300 bp (Bult *et. al.*, 1995) and the 5.8S gene equals 160 bp (Takaiwa *et al.*, 1985; Tanaka *et al.*, 1980). The highly conserved coding regions have useful applicability at the family level and above, whereas rapidly evolving regions find importance for comparing species and closely related genera (Soltis and Soltis, 1998).

2.3. USE OF THE 18S RIBOSOMAL DNA AS A TAXONOMIC TOOL:

In eukaryotic nuclear genomes, the ribosomal RNA gene (rDNA) occurs as multiple copies of tandemly repeated units. Each transcription unit consists of an 18S, 5.8S and 28S rRNA gene separated by non-coding intergenic spacers (IGS) (Schmidt-Puchta *et al.*, 1989). These spacers are called as internally transcribed spacers (ITS) because they are part of primary transcript. In nature rRNA coding sequences have been highly conserved during evolution, indicating a strong positive selection pressure (Nairn *et al.*, 1988). Due to its conserved nature, 18S rDNA sequences have been used to infer phylogenetic relationships at the higher taxonomic level (Soltis *et al.*, 1997). The 18S

rDNA with an approximate size of 1,800 bp is more amenable to PCR amplification and sequencing and has been much more extensively used in contrast to the 26S/28S rDNA, which is more than 3,000 bp, making it difficult to obtain its complete sequence (Soltis and Soltis, 1998). Typically, only one 18S rDNA sequence type is found in an organism, excepting a few cases, where two types of 18S rDNA are found, as in *Dugesia mediterranea* (flatworm), of which only one is transcribed although both types appear functional (Carranza *et al.*, 1996). Nickrent and Soltis (1995) described angiosperm 18S rDNA as a region containing highly conserved segments interspersed with highly variable regions that undergo multiple substitutions per site. The highly conserved regions appear to occur near the stem region of the secondary structure model which is somewhere about 50 nucleotides in length. The highly variable regions are located in the terminal helices which corresponds to about 5-10 nucleotides in length (Soltis and Soltis, 1998). This mosaic pattern of ribosomal DNA, highly conserved regions interspersed with variable regions, is also observed in other organisms as reported by Van de Peer and co-workers (1996). Takaiwa *et al.* (1984) developed the first sequence of 18S ribosomal RNA of an angiosperm in rice followed by Messing *et al.* (1984) in maize and later in soybean by Eckenrode *et al.* (1985). Following the work by Nairn and Ferl (1988), Nickrent and Franchina (1990) examined the phylogenetic relationships in the parasitic plant order Santalales using the complete 18S ribosomal DNA sequences to support the monophyly of the order Santalales and also confirming the basal position of Olacaceae within the order, indicating that complete 18S rRNA sequence contains sufficient information to allow phylogenetic comparisons to be made at the family level

and above. The first detailed analyses of molecular evolution of 18S rRNA genes in angiosperm was made possible by the availability of complete 18S rDNA sequences obtained by Soltis *et al.* (1997) and Nickrent *et al.* (1995). The 18S rDNA yielded sufficient number of characters for broad scale phylogenetic taxonomy in angiosperm, sidelining the earlier view that 18S rDNA were prone to deletion and insertion making sequence alignment difficult. 18S rDNA region has been widely employed to study angiosperm systematics by Zimmer *et al.* (1989), Hamby and Zimmer (1992), Nickrent and Soltis (1995) and Soltis *et al.* (1997). In order to resolve the confusion of names of five species of micro-algae *Dunaliella,* specific primers were designed by Olmos *et al.* (2000) to study the 18S rDNA sequence for characterization and its proper identification. Their study revealed three species with introns in the 18S rDNA forming one group of strains, whereas the other two species without intron were grouped differently. Savard and Lalonde (1991) obtained the full sequence of 18S rDNA gene of *Alnus glutinosa* with nucleotide size of 1698 bp and compared with that of four angiospermic plant species, two dicots, tomato (96%) and soyabean (96%) and two monocots, maize (94%) and rice (94%). Their comparative studies revealed that the18S rDNA gene of *Alnus glutinosa* showed sequence similarity with that of other angiospermic plants, in this case more affinity to the dicots than that of the monocots.

2.4 USE OF ITS SEQUENCE OF RIBOSOMAL DNA AS A TAXONOMIC TOOL:

The Internally Transcribed Spacers (ITS) are non-coding regions flanked by highly conserved sites. They are highly variable and therefore accumulate lots of nucleotide changes even within closely related organisms. Owing to this characteristic feature, ITS regions have been used to infer phylogenetic relationships at the lower taxonomic level especially at the sub-generic level (White *et al.*, 1990). In order to resolve doubts over generic relationships in Winteraceae, Suh *et al.* (1993) did sequence analysis of the ITS region of *rrn* operon. Nucleotide sequence of ITS1 in the family ranging from 235 to 252 bases in size and from 213 to 226 bases of ITS2 was reported. They observed that the range of sequence divergence of ITS1 and ITS2 between pairs of genera within winteraceae is relatively low as compared with other plant families. Lashermes *et al.* (1997) studied the phylogenetic relationships of *Coffea* species inferred from ITS sequences of nuclear rDNA. They reported high sequence variation in the ITS2 region within the accessions of different coffee taxa which could be used for phylogenetic investigation among closely related species. Gomes *et al.* (2002), analysed the ITS sequence of ribosomal DNA of 26 isolates ectomycorrhizal fungi belonging to 8 genera and 19 species using PCR-RFLP. They reported the potential of ITS region PCR-RFLP for molecular characterization of ectomycorrhizal fungi and their identification, with the interspecific variation relatively higher than intraspecific variations within this region. In order to assess the relationships of Araliaceae in India, the Internal Transcribed Spacer (ITS) region of ribosomal DNA for forty two accessions belonging

to nine genera was obtained by Pandey *et al.* (2004). Two major clades of Indian Araliaceae sharing close relationship was suggested which was supported by a strong bootstrap value of 89%. ITS sequences have also been used for phylogenetic reconstruction in angiosperms (Baldwin *et al.*, 1995), algae (Bakker *et al.*, 1995; Coleman *et al.*, 1994), and ferns (Stein *et al.*, unpubl.). Use of ITS sequences to infer phylogeny at the lower taxonomic levels in other organisms, include fungi (Vilgalys and Sun, 1994) and insects (Campbell *et al.*, 1993; Fritz *et al.*, 1994; Schlötterer *et al.*, 1994; Vogler and DeSalle, 1994).

Navarro *et al.* (2003) studied the molecular phylogeny of 19 species of *Alnus* inferred from the ITS region of the nuclear ribosomal DNA. Their study reported that all species of *Alnus* formed a monophyletic group close to *Betula,* a non-actinorhizal plant, and supported the view of Murai (1964) who proposed the division of the genus into two subgenera, *Alnacaster* (*Alnobetula*) and *Gymnothyrsus* (*Alnus*). A similar study by Savard *et al.* (1993), based on the results obtained from ITS sequences, proposed a monophyletic origin of *Alnus* and *Betula.* Their observation also supported the hypothesis that both *Alnus* and *Betula* diversified during the late Senonian. PCR/RFLP study of the ITS region of the nuclear 18S-28S ribosomal DNA of some *Alnus nepalensis* population was done by Chauhan and Misra (2002) using primers AnpITS1 (designed by them) and ITSC26A (Wen and Zimmer, 1996). Their study revealed that restriction digestion of the amplicon with *Scr*F1 separated the genus into nine different groups. They also did a comparative study of the PCR/RFLP of the 18S-28S ITS region with that of the nitrogenase activitiy of *Alnus nepalensis* and developed molecular

markers that could be used to eliminate genotypes with low nitrogenase activity. Distribution of *Frankia* genotypes with respect to varying altitudinal zones was studied for 90 *Alnus nepalensis* trees from three altitudes at nine different locations by utilizing the 16S-23S ITS region in a study made by Khan *et al.* (2007). The forward primer FGPS 989ac, designed by Bosco *et. al.* (1992), and the reverse primer FGPL 2054', designed by Simonet *et al.* (1991), were used to amplify the ITS region located between the distal part of 16S rRNA and the initial part of 23S rRNA genes of *Frankia*. Amplified recombinant DNA restriction analysis (ARDRA) patterns were generated utilizing the restriction enzyme *Rsa*1 that produced 17 different patterns corresponding to different altitudinal heights. Their study indicated that microclimatic factors such as temperature and moisture acted as primary determinants of distribution of *Frankia* genotypes. Huguet *et al.* (2005) used primers F502 ITSC26A (Wen and Zimmer, 1996) and F682 ITS1 plant (Navarro *et al.*, 2003) in conjunction with internal primers F1617 ITSpl and F1618 ITSp2 designed by them to amplify the nuclear 18S-26S ITS gene of 13 species of Myricaceae. They obtained an average GC content of 58% in ITS sequence. Based on ITS sequence their study confirmed the division of Myricaceae into two main groups with a bootstrap value of 75% for the group containing *C. peregrina, M. gale* and *M. hartwegii* and an 82% bootstrap value for the other group containing other *Myrica* species. Based on the sequences of the internal transcribed spacers (ITS) and the 5.8S coding region of the nuclear ribosomal DNA of fifteen species representing major subfamilies and tribes of Hamamelidaceae, Shi *et al.* (1998) constructed the phylogeny of Hamamelidaceae. They reported that the total length of the ITS1. 5.8S and

the ITS2 regions of Hamamelidaceae ranged from 638 to 676 bp with the ITS1 longer than ITS2. They performed phylogenetic analyses with PAUP* 4.0 (Swofford, 1997) using the maximum parsimony (Swofford *et al.*, 1996), the maximum likelihood (Felsenstein, 1981) and the neighbor-joining (Saitou and Nei, 1987). Results from their study of ITS data strongly supported the monophyly of both the Hamamelidoideae and the Altingioideae and also suggested a close relationship between the tribe Hamamelideae and the tribe Fothergilleae. Tree generated from ITS data in their study also showed that *Distylium* formed a monophyletic group with *Shaniodendron* of the Fothergilleae. In a study made by Varghese *et al.* (2003), the hypervariable ITS region of *rrn* operon of *Frankia* was used for developing PCR-RFL profiles to discriminate between closely related Frankiae. Restriction patterns generated as a result of enzymatic digestion of the amplicons allowed them to differentiate between very similar endosymbionts of alder. Taking advantage of the sequence variability of the nuclear *rrn* operon ITS regions Chauhan and Misra (2002) also designed an alder-specific PCR primer which was able to amplify the *Alnus nepalensis* DNA.

2.5. PHYLOGENY OF ACTINORHIZAL PLANTS:

Work on recent developments in actinorhizal symbiosis in the context of molecular approaches has been reviewed by Misra *et al.*, (2004), with special emphasis on the molecular approaches to identification of the actinomycete *Frankia* using the *nif* genes and 16S rRNA. The use of 16S rRNA genes and nitrogenase *nif* genes as taxonomic tool is reported. Swensen and Mullin (1997) reviewed the recent molecular

systematics and phylogenetic relationships among actinorhizal plants and revealed that actinorhizal plants were more closely related than those viewed by current taxonomists. They had reviewed studies on molecular systematics of nodulating symbiotic higher plants, molecular systematics and evolution of nodulation in actinorhizal plants with comparative discussion of the evolution of nodulation in the legume-rhizobium system and comparison of the host and bacterial phylogenies. Further study of the geographical distribution of the host plants with in-depth phylogenetic analysis of the actinomycete *Frankia* and also analysis of estimated times of divergence among host plants and *Frankia* has been suggested. The chloroplast *rbc*L gene, which encodes the large subunit of the photosynthetic enzyme ribulose 1,5-bisphosphate carboxylase oxygenase, has been used for phylogenetic relationship studies in the subclass Hamamelidae by Maggia *et al.* (1994). Their work suggested that the three families namely Betulaceae, Myricaceae and Casuarinaceae, under the subclass Hamamelidae, appeared to be monophyletic in their origin and that Myricaceae appeared to derive first before Betulaceae and Casuarinaceae. This possible divergence of the Myricaceae before the Betulaceae and Casuarinaceae, which appeared in the Cenomanian stage of the Upper Cretaceous (Cronquist, 1988; Crane, 1989) and at the point between the Cretaceous and the Tertiary (Johnson and Wilson, 1989), is in concurrence with the fossil records where the Myricaceae appeared in the Cenomanian stage of the Upper Cretaceous (Gladkova, 1962; MacDonald, 1977). Studies by Maggia *et al.* (1994) suggest that evolution might have proceeded in the direction of narrower promiscuity since both molecular and traditional approaches indicate earlier diversion of the Myricaceae relative to the

Casuarinaceae and Betulaceae, which could be interpreted as a specialized feature towards possible evolution with resistance to *Frankia* infection, as was the case observed with *Allocasuarina* (Johnson & Wilson, 1989).

2.6. PHYLOGENY OF MYRICACEAE:

Scanty reports on the molecular phylogeny of Myricaceae are available. Till date no family-wide molecular phylogeny has been generated for the Myricaceae except for some fossil history documented by MacDonald (1989). Misidentification (Para-O, 2001) owing to its wide native distribution and polymorphic nature (Chevalier, 1901) has raised doubts over its authenticity (Adams, 1972). Raven and Axelrod (1974) proposed the hypothesis that Myricaceae might be of Laurasian origin but there was a doubt regarding its centre and time of origin. Studies by Chourey (1974) based on megafossil and pollen evidence suggested a late Cretaceous origin of the Myricaceae. This view was supported by Abbe (1963), who described the origin of the Myricaceae in Southeastern Asia based on floral evidence and that the Southeastern Asia species diverged first while the African species were the most derived. However, Chourey (1974) rejected Abbe's proposal on the basis that available fossil data to support such an ancient origin was invalid because of their lack of conservation. However, controversies regarding the division of the genus *Myrica* into two separate genera is still debated (Bornstein, 1997; Weakley, 2000; Wilbur, 1994).

Huguet *et al.* (2005) conducted molecular phylogenetic study on 13 species of Myricaceae using the *rbc*L gene and 18S-26S ITS region of the nuclear DNA aimed at

clarifying the controversial taxonomy of Myricaceae and to relate its specificity of association with the acinomycete *Frankia* with its evolutionary pattern. They used the primers F1615 NrbcLf and F1616 NrbcLr (designed by them) and also F1725 rbcLp1 and F1726 rbcLp2 (designed by them) to amplify the plastid *rbcL* gene. In order to amplify the nuclear 18S-26S ITS gene they used the primer F682 ITS1plant (Navarro *et al.*, 2003) in conjunction with F502 ITSC26A (Wen and Zimmer, 1996). The resultant amplicon obtained by them for *rbc*L gene was 1,450 bp and that of the ITS fragment was 785 bp. Analysis of their study showed that *Myrica gale*, *Myrica hartwegii* and *Comptonia peregrina* belong to a phylogenetic cluster as compared to other *Myrica* species, now replaced with a new genus name, *Morella* (Baird, 1969; Wilbur, 1994 and Brummitt, 1999). Their result provides molecular support to the work done by Chevalier (1901), Baird (1969) and Wilbur (1994), that the two species *M. gale* and *M. hartwegii* should be transferred to a new genus which is distinct from the rest. They have also reported that within the samples of *Morella cerifera,* separated on their geographic origin, a sequence divergence of 0.4% was recorded suggesting that two different species or subspecies within *M. cerifera* could be possible.

Molecular phylogeny of *Myrica esculenta* has not been done. However, a general survey of its economics and medicinal values was done by Dhyani and Dhar (1994). A brief report on the morphological variability that existed within the population of *Myrica esculenta* was also reported by Barua and Srivastava (2003). The family Myricaceae has been revised in the recent years by some authors by distributing the species under three genera, *Myrica, Morella* and *Comptonia*. According to them only two species, *hartwegii*

and *gale,* fall under the genus *Myrica* and the rest of the species fall under the genus *Morella.* The genus *Comptonia peregrina* retains the same name. This is still debatable (Bornstein, 1997; Weakley 2000; Wilbur 1994). However, for our convenience we shall be using the genus name *Myrica* in the present investigation.

MATERIALS AND METHODS

CHAPTER 3

MATERIALS AND METHODS

3.1 SELECTION OF STUDY SITES:

Two different sites were selected for collection of samples keeping in mind the distribution of the three morphological variants of *Myrica* sp. found in Meghalaya:- Morphotype-1 (please see table 4.1) trees were predominantly found at NEHU permanent campus while morphotype-2 and morphotype-3 trees were found in Nongkrem forest.

Fifty morphotype-1 *Myrica* trees were randomly selected from NEHU permanent campus, located at 25.36° N latitude and 91.53° E longitude at 1431 m above mean sea level. *Myrica* trees were sparsely distributed in a population mostly dominated by pine trees (*Pinus kesiya*). Other trees like *Alnus nepalensis* and *Schima wallichi* were also present in mixed stands.

Second site was located about 35 km (Fig. 3.1) away in Nongkrem forest, located at 25.50° N latitude and 91.88° E longitude at an elevation of 1631 m above mean sea level. A dense population of *Myrica* trees was seen here. Other tree species like *Pinus kesiya, Quercus griffithii, Lithocarpus dealbatus, Alnus nepalensis,* etc. were also found. Fifty morphotype-2 *Myrica* trees were randomly selected from this site. All the eleven morphotype-3 trees found were also selected for the study.

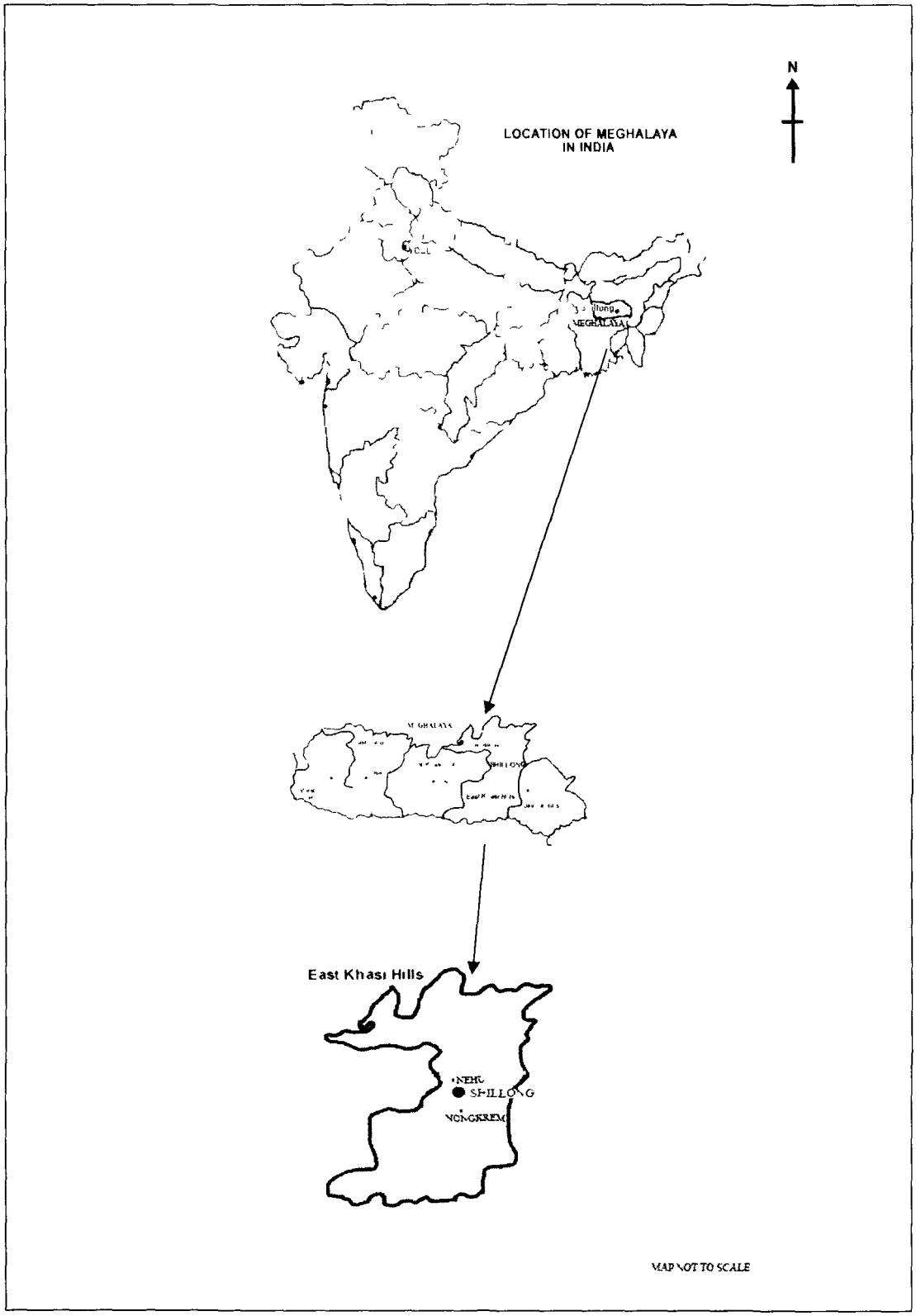


Fig. 3.1: Physical map of India indicating the location of Meghalaya and also the location of NEHU permanent campus and Nongkrem forest therein.

3.2 MORPHOLOGICAL STUDIES:

The different morphotype trees were identified based on morphological features. Only the fruit bearing trees were selected keeping in mind that *Myrica* tree is dioecious in habit. Differences in colour of ripe fruit and leaf morphology between the different morphotype trees were visually keyed out. Fruit and leaf sizes were measured using a centimeter scale. Fruit and leaf sizes were measured in triplicate separately for each of the morphotype trees and their respective mean sizes obtained (Please see Table 4.1).

3.3 COLLECTION OF LEAVES:

Selected stands of *Myrica* trees demarcated from each of these sites were tagged with aluminum plates in numerical order for identification. From each tree, young, uninfected fresh leaves were collected into labeled plastic bags, brought to the laboratory (Fig. 3.2) and stored at -80°C for further studies.

3.4 ISOLATION OF DNA FROM LEAVES:

Isolation of DNA from the leaves was performed using the CTAB method of Dellaporta *et al.* (1983) which was adapted in our laboratory by Chauhan (2000). In order to eliminate chances of isolating extraneous DNA, only young healthy leaves were selected which were surface sterilized thoroughly with 30% hydrogen peroxide (H_2O_2) just before the start of the actual extraction process. Extraction of DNA using older leaves resulted in accumulation of phenolic compounds in the form of slimy substances most often encountered, before pipetting out the aqueous phase at the end of the

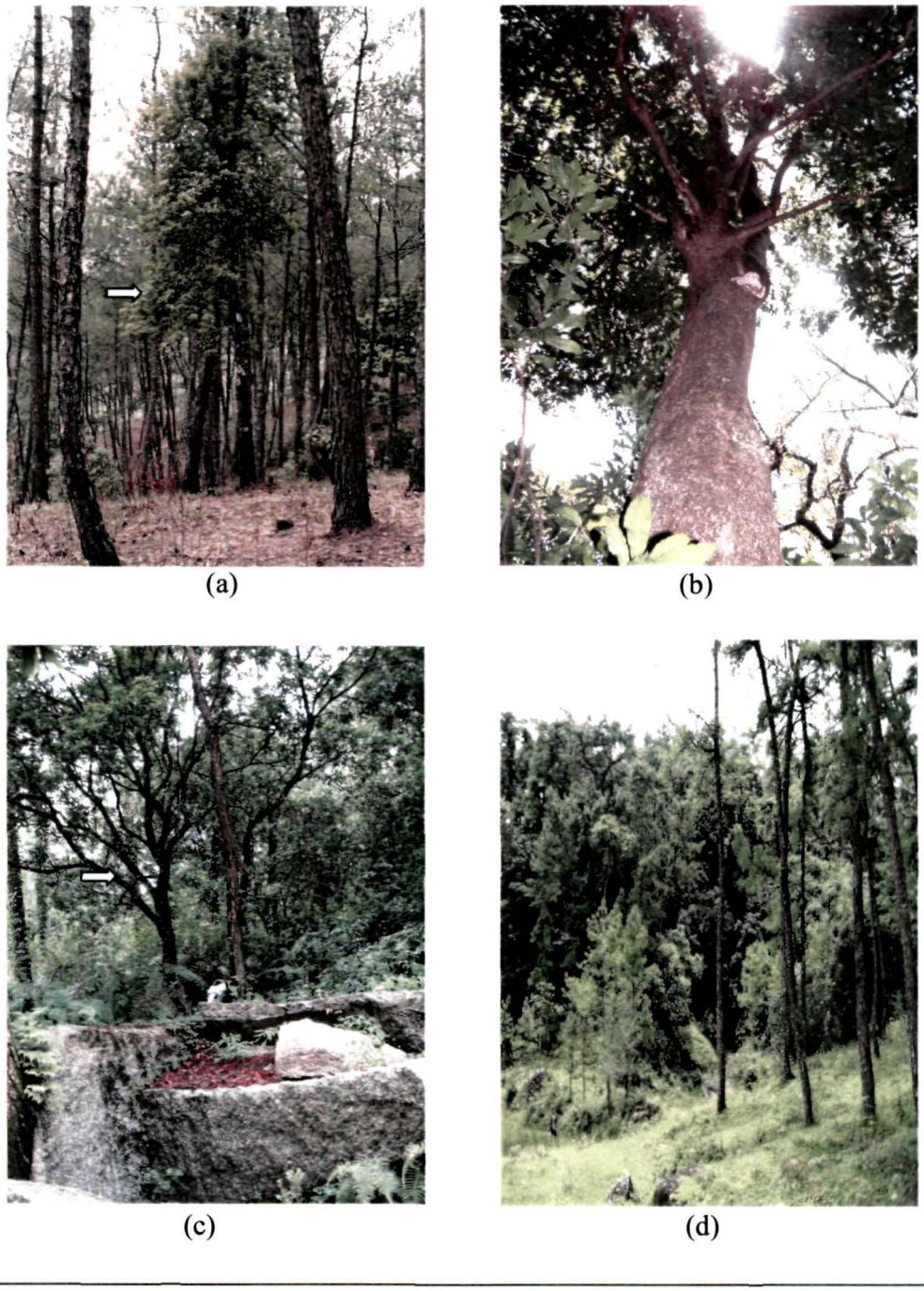


Fig. 3.2: Photographs showing the different collection sites
(a): Morphotype-1 *Myrica* tree at NEHU permanent campus
(b): Morphotype-2 *Myrica* tree at Nongkrem forest
(c): Morphotype-3 *Myrica* tree at Nongkrem forest
(d): A view of the Nongkrem forest

Chloroform: Isoamyl alcohol wash. Therefore, only young leaves were used. The protocol followed is given below:

a) Two gm of leaves, weighed by a Sartorius BP160P weighing balance were first washed with distilled water and then surface sterilized in 30% H_2O_2 for 2-3 minutes followed by repeated washing with double distilled water.

b) Leaves were crushed in liquid nitrogen using mortar and pestle in a cold room.

c) Crushed leaves were transferred into a 1.5 mL microcentrifuge tube containing 1 mL of warm extraction buffer [Tris Base (0.1M), EDTA (0.1M), NaCl (1.4M), CTAB (2% w/v), PVP (1% w/v)].

d) Ten μL of SDS (20% Sodium Dodecyl Sulphate) were added and tube was kept for one hour at 65°C in a water bath.

e) It was centrifuged at 7,500 rpm at room temperature for 5 minutes and the pellet discarded.

f) The supernatant was extracted with equal volume of Chloroform: Isoamyl alcohol (24:1 v/v).

g) It was centrifuged at 13,000 rpm at room temperature for 15 minutes.

h) The aqueous phase was transferred to a fresh tube and the DNA was precipitated with 1 mL of ice cold absolute ethanol and kept at -20°C overnight.

i) It was centrifuged at 13,000 rpm for 30 minutes at 4°C and pellet was washed with 70% alcohol twice for 30 minutes each.

j) The DNA was vacuum dried, dissolved in 15 μL of double distilled water and kept at -20°C for further use.

Two μL of DNA was mixed with 3μL of loading buffer and loaded into a ethidium bromide stained 0.8% agarose gel and electrophoresed at 60V for about 45 minutes along with a λ DNA *Hind* III/ *Eco*R1 double digest marker. The gel was first visualized under a U.V. Transilluminator to check the presence of DNA and then transferred to BIORAD GELDOC 1000 and the photograph taken using MultiAnalyst™ software.

3.5 POLYMERASE CHAIN REACTION (PCR):

Amplification of the isolated DNA was carried out for all the collected samples of *Myrica* using the Polymerase Chain Reaction (PCR). PCR was performed using either Applied Biosystems GeneAmp 9700Gold or Perkin Elmer GeneAmp 2400. A total volume of 25 μL was prepared for each PCR reaction tube. Each reaction mix contained 2.5μl of 10X PCR assay buffer (Bangalore Genei, India), 2.5 μL each of the primer (5 pM), 2.5 μL of $MgCl_2$ (25mM), 0.75 μL of *Taq* polymerase (3 Units/μL) and 2.5 μL each of the dNTP (1.25 mM). A final volume of 25 μL was made up by adding ultra pure water. A total of 35 cycles were run for each reaction. An initial denaturation step of 10 minutes at 94°C preceded 35 cycles of denaturation at 94°C for 1 minute, 1 minute of primer annealing at the appropriate temperature and 1 min of nucleotide extension at 72°C, followed by a post-elongation step of 7 min at 72°C at the end of the cycle.

The annealing temperature for the primers was calculated using the formula;

$$Tm = [4 (G+C) + 2 (A+T)] - 5$$

PCR amplification was performed for the following regions of *Myrica* nuclear DNA (Fig. 3.3);

i) 18S-28S rDNA ITS region

ii) 18S rDNA

The forward primer ITSI PLANT (Navarro *et al.,* 2003) located at the distal part of the 18S rDNA gene and the reverse primer ITSC26A (Wen and Zimmer, 1996) located at the initial part of the 28S rRNA gene were used to amplify the 18S-28S rDNA Internal Transcribed Spacer (ITS) region which yielded a band of approximately 800 bp in length (Table 3.1).

Forward primer SOL23F (personal communication, Pamela Soltis) and reverse primer SOL1769R (personal communication, Pamela Soltis) were used to amplify the conserved nuclear 18S rDNA that produced an approximate expected band size of 1.8 kb (Table 3.2).

3.5.1 Primer stock solution:

The primers used for amplification were procured in lyophilized, desalted and PAGE purified condition from M/s Bangalore Genei, India. A stock solution of 25 μM was prepared by dissolving the primers in ultra pure water. From the stock solution a working solution of 2 μM was prepared by further dilution and stored at -20°C.

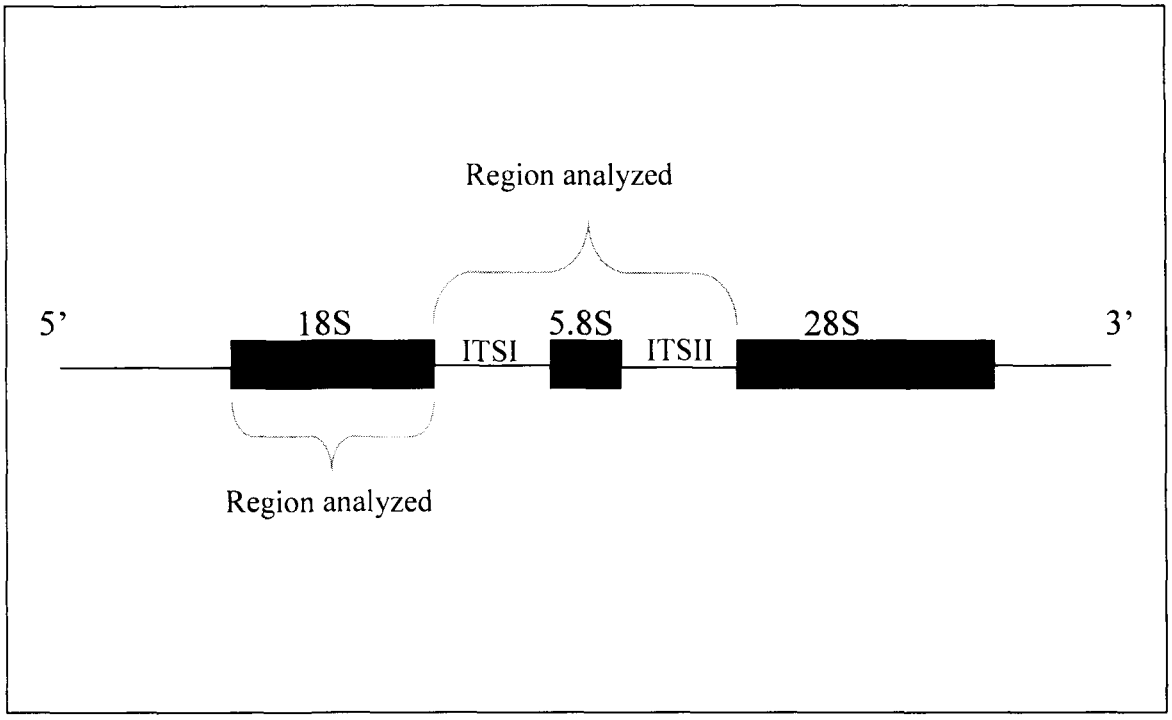


Fig. 3.3: Diagrammatic representation of organization of nuclear 18S-28S *rrn* operon.

TABLE 3.1: PRIMERS USED FOR AMPLIFICATION OF 18S-28S rDNA ITS REGION.

CODE	SEQUENCE	REFERENCE
ITS1-PLANT	5' CGCGAGAAGTCCACTG 3'	Navarro *et al.*, 2003
ITSC26A	5' GTTTCTTTTCCTCCGCT 3'	Wen and Zimmer, 1996

TABLE 3.2: PRIMERS USED FOR AMPLIFICATION OF 18S rDNA.

LAB CODE	SEQUENCE	REFERENCE
SOL23F	5' TTGGTTGATCCTGCCAGTAG 3'	Pamela Soltis, personal communication
SOL1769R	5' CACCTACGGAAACCTTGTT 3'	Pamela Soltis, personal communication

3.5.2 Deoxyribonucleotide solutions:

Deoxyribonucleotide solutions (dNTPs) were procured from M/s Bangalore Genei in 10 mM stock solution. dATP, dCTP, dGTP and dTTP were mixed in equal quantities to prepare a final working solution of 2.5 µM concentration of each and stored at -20°C.

3.5.3 *Taq* polymerase:

Taq DNA polymerase and buffer [10mM Tris HCl (pH 8), 1.5mM KCl and 0.01% gelatin] were procured from M/s Bangalore Genei. Additional $MgCl_2$ (25mM) was also procured. They were stored in a deep freezer at -20°C for further use.

3.6 AGAROSE GEL ELECTROPHORESIS:

For detection of genomic DNA as well as amplified DNA a 0.8% agarose gel was used. For preparation of agarose gel, requisite amount of agarose powder (Hi-Media) was weighed and dissolved in 1X TBE (Tris borate EDTA) buffer. The solution was warmed for about 30 seconds in a microwave oven (IFB) till agarose dissolved completely and a consistent transparent solution obtained. It was allowed to cool for sometime. Then it was poured in the casting tray sealed at the two open ends with cello tape and left for about 15 minutes to harden after ensuring that the comb has been aligned properly. The prepared agarose gels were immersed in an electrophoresis tank which also contained 1X TBE buffer. Two µL of genomic DNA sample or PCR amplicons were mixed with 3 µL of the loading buffer separately (Sambrook *et al.*,

1989), loaded into the gel wells and electrophoresed at 60 V for 1 hour. One of the wells was loaded with 3 µL of λ DNA / *Hind* III – *Eco*RI double digest marker.

3.7 VISUALIZATION OF DNA:

DNA was visualized under a U.V. Transilluminator after staining the gel with ethidium bromide. Photographs of agarose gels were captured using BIORAD Gel Doc 1000 supported with MultiAnalyst™ software (version 1.1, build 34) or KODAK GEL LOGIC 1500 IMAGING SYSTEM. Sizes of the bands of interest were estimated against the standard molecular weight markers using tools which were available in the software. Photographs were saved in the computer hard disc as well as in external storage devices.

3.8 DNA SEQUENCING:

DNA sequencing was performed for the 18S-28S ITS region employing the Single Pass Analysis (SPA) technique for both the forward and reverse strands based on the dideoxy chain termination method of Sanger *et al.* (1977). Amplification of the ITS region with primers ITS1-PLANT and ITSC 26A produced a single band size of approximately 800 bp. Since there were no extra bands present it was considered not necessary to go for gel elution of the band of interest. Twenty µL (~150 ng) of the amplified 18S-28S rDNA operon was directly lyophilized in a 1.5 mL eppendorf tube and sent to M/s Bangalore Genie for sequencing.

Sequencing of the complete 18S rDNA was done employing Primer Walking technique. A band size of approximately 1.8 kb was amplified using two specific primers. One hundred µL of each representative sample was lyophilized in a 1.5 mL microfuge tube, sealed with Parafilm$^{(R)}$ and sent to M/s Bangalore Genei. Fifty µL of 2mM concentration of each of the primers were also lyophilized, sealed with Parafilm® and sent along with the samples. Primer walk reaction was started using the primers which were used for DNA amplification. Sequences were obtained in the form of electropherogram where each of the nitrogen bases was represented with different colours.

3.9 PHYLOGENETIC ANALYSIS BASED ON SEQUENCE DATA OF VARIABLE 18S-28S rDNA ITS REGION AND THE CONSERVED 18S rDNA GENE:

Phylogenetic analyses of the sequence data generated for 18S-28S ITS region sequence as well as the 18S rDNA sequence were done using the online software programmes CLUSTAL W (Version 2.0.8) and PHYLIP (Version 3.66). Each of the sequences of our samples was separately used as a query sequence for BLAST search for related sequences in GenBank (http://www.ncbi.nlm.nih.gov). Retrieved sequences were aligned with our samples using the multiple sequence alignment programme CLUSTAL W. The aligned sequences were then used to construct Neighbour Joining (neighbor.exe), Maximum Parsimony (dnapars.exe) and Maximum Likelihood (dnamlk.exe) phylogenetic trees of the PHYLIP inference package.

3.10 AMPLICON RESTRICTION PATTERN (ARP):

The amplified 18S-28S rDNA ITS region was subjected to restriction digestion using different restriction endonucleases. Initially, computer simulated restriction digestion of a related ITS sequence, retrieved from the internet was performed using the Mac Vector® software. Restriction endonucleases were shortlisted based on the cutting sites and the fragments generated for the purpose of digestion. Only two restriction endonucleases, *Mbo*I and *Sau*96I (Table 3.3), produced differing restriction fragments between the three morphotypes when digested with these enzymes separately. Therefore, restriction digestion analysis of all the samples was performed using these two enzymes. The digestion mixture was prepared in a 1.5 mL tube containing 10 µL of the amplicon, 0.5 µL of the enzyme, 2 µL of the buffer and the remaining was adjusted with distil water to make a final volume of 20 µL. The tube was incubated in a water bath at 37° C for 5 hours or according to the manufacturer's instructions.

3.10.1: Agarose gel electrophoresis of ARP:

Restriction digested amplicons were electrophoresed in a 2% agarose gel of considerable thickness. Twenty µL of restriction product was mixed with 10 µL of loading buffer and was then loaded into the well of the thick gel and electrophoresed for about 6-7 hours at 45 V. A 100 bp molecular weight marker was also loaded in one of the lanes which served as a standard for estimating the band sizes of DNA fragments. Observation and quantification of the restriction fragments was done using BIORAD

TABLE 3.3: LIST OF RESTRICTION ENZYMES USED.

SL. NO.	RESTRICTION ENZYME	CUTTING SITE	SOURCE ORGANISM	BUFFER (INCUBATION TEMPERATURE)
1	*Mbo* I	↓ GATC CTAG↑	*Moraxella bovis*	*B (37°C)
2	*Sau* 96I	G↓GNCC CCNG↑G	*Staphylococcus aureus*	**NEBuffer 4 (37°C)

* Bangalore genei
** New England Biolabs

Gel Doc 1000 with MultiAnalyst software and also using KODAK GEL LOGIC 1500 IMAGING SYSTEM.

3.11 CLUSTER ANALYSIS

Cluster analysis can be defined as the classification of variables on the basis of the similarity of characteristics they possess. The term 'cluster analysis' (Tyron, 1939) also encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. Therefore, the result of cluster analysis is a number of heterogeneous groups with homogenous contents. Cluster analysis can either be agglomerative or divisive. The agglomerative methods begin with each observation being considered as separate clusters and then proceeds to combine them until all observations belong to one cluster. Whereas the divisive methods start with all of the observations in one cluster and then proceeds to split (partition) them into smaller clusters.

In the present study, the banding patterns that emerged from ARP analysis was used to generate cluster dendrogram with the method SAHN (Sequential Agglomerative Hierarchical Nested cluster analysis) using NTSYS software (2.1 version) that was based on Jaccard's coefficient of similarity (Jaccard, 1908). Jaccard's coefficient is based on an algorithm that is used to estimate the level of similarity for multivariate data. Presence or absence of bands for each individual sample was scored for all the individual samples so that a descending pattern of values was obtained.

The Jaccard's coefficient value for each sample was calculated using the formula given below:

Jaccard's coefficient = Nab/ [(Na+Nb) – Nab]

Where, Nab = Number of common bands in both samples

Na = Total number of bands present in the first sample

Nb = Total number of bands present in the second sample

3.12 ESTIMATION OF SOIL NITROGEN:

The conversion of organic nitrogen to ammonia and its subsequent estimation is the basis of Kjeldahl method (Jackson, 1974). A catalyst is required for the digestion. In the digestion stage organic nitrogen is converted into ammonium nitrogen. Free ammonia is liberated from the diluted digest by steam distillation in the presence of excess of alkali. The distillate is collected in a receiver containing excess boric acid indicator. The ammonia is then titrated with standard HCl.

Soil samples were collected from NEHU permanent campus site as well as Nongkrem reserve forest site. From each site ten trees were randomly demarcated from where the soil samples were collected within a periphery of about 1 meter from the base of the tree. Humus and other debris present above the top soil was first cleared off and soil was then collected at a depth of 1 foot from the ground level using a 10 x 3 cm cylindrical soil corer. The collected soil was brought to the laboratory and air-dried at room temperature for two to three days until it was completely desiccated. The soil was then filtered through a 0.2 mm porous net and the weight recorded

Following method was used for the purpose:

1. 1 gm of air dried soil was mixed with 6 mL of Sulphuric acid (H_2SO_4) and one Kjeldahl tablet was taken in a digestion flask.
2. The mixture was digested in a block digester.
3. The digestion was stopped after the colour turned green in the flask and allowed to cool.
4. The mixture was filtered using Whatman No.1 paper and the final volume was adjusted to 50 mL with distilled water in a volumetric flask.
5. Distillation was done in a Kjeldahl distillation set with 10 mL of 40% sodium hydroxide (NaOH).
6. Distillate was collected in a beaker with 5 mL of Boric acid indicator.
7. Distillation was stopped when the pink colour of boric acid indicator turned greenish.
8. The collected distillate was taken out and titrated against M/140 HCl until the colour turned to pink.

The percent soil nitrogen was calculated by using the following formula:

$$N\% = \frac{\text{T- blank X solution volume}}{10^2 \text{ X aliquot volume X sample weight}}$$

3.13 COLLECTION OF NODULES:

Nodules were collected from all the hundred and eleven trees at both the collection sites. Before collecting nodules the upper thin layer of soil just below each tree was first cleared off and the roots containing the nodules were traced to their respective trees in order to eliminate the chances of collecting nodules from a different tree. Since it was not possible to estimate the nitrogenase activity of all the hundred and eleven samples in one day, a period of two months was considered starting from September to October for collection of the nodules keeping in mind that the effect of changing environmental factors was minimal. However, for those nodules which were collected in the early part of the day the ARA was performed on the same day itself. In order to draw a valid comparative inference, the time of collection of nodules till the estimation of ARA was maintained uniformly for all the samples. The ARA values for all the hundred and eleven trees were estimated in triplicate, i.e., a total of 150 nodules were studied for each tree.

3.14 ACETYLENE REDUCTION ASSAY (ARA):

Estimation of nitrogenase activity of the root nodules was done using the Acetylene Reduction Assay (ARA) method described by Stewart *et al.* (1968). Apart from reducing nitrogen, the nitrogenase enzyme can also reduce acetylene (C_2H_2) gas to ethylene (C_2H_4) which is used to determine the efficiency of the nitrogenase enzyme. The quantity of ethylene produced was measured using a Gas Chromatograph (HP 4890D) equipped with flame ionization detectors (FID). As the mixture of injected gases

passed through the column in the carrier gas stream, they got separated depending upon their boiling points. The injection port, oven and detector temperatures were 120°C, 90°C and 175°C respectively. The carrier gas was nitrogen and the flow rates of hydrogen, air and nitrogen were 50, 120 and 10 mL/min respectively. An approximate retention time of 1.40 minutes for ethylene and 2.45 minutes for acetylene was recorded when 1 mL of standard ethylene solution was injected into the gas chromatograph and the ethylene peak in the standard and the samples compared.

The procedure for estimation of Acetylene Reduction Assay (ARA) is given below:

1. Freshly collected nodules were brought to the laboratory and washed thoroughly with distilled water to remove soil debris adhered to it. The nodules were then surface dried on a filter paper.
2. Fifty nodules from each tree were weighed and put in air tight BD Vacutainer™ stoppered vial which was then sealed with parafilm. This was done in triplicate. i.e., a total of 150 nodules from each tree were taken.
3. Using an air tight syringe 1 mL of air in the vial was replaced with 1 mL of acetylene. The vials were incubated for 3 hours at room temperature (29 ± 2°C).
4. At the end of the incubation period 1 mL of the gas mixture was injected at the injection port of the gas chromatograph (HP 4890D) fitted with FID.
5. The values were obtained in a printed form using a printer (hp-3395 integrator) which was connected to the gas chromatograph.

6. The mass of the nodules was again recorded using a Sartorius electronic balance (BP160P) after the incubation was over.

The nitrogenase activity was estimated in terms of nmoles of ethylene produced per milligram fresh weight per hour using the following formula:

$$\text{N.A.} = \frac{\text{nmole } C_2H_4\text{/ unit area in the standard X area of } C_2H_4 \text{ in the sample X vol. of vial}}{\text{Fresh weight X incubation time}}$$

Where, N.A. stands for n moles of ethylene produced/ mg fresh weight/ hour.

RESULTS AND DISCUSSION

CHAPTER 4

RESULTS AND DISCUSSION

The genus *Myrica*, globally represented by about 97 species, belongs to the actinorhizal group of trees in which the root enters into symbiotic association with the actinomycete *Frankia* to form nodules. This symbiotic association is important because this allows the fixation of elemental nitrogen to usable forms which is made available to the plants.

Myrica esculenta Buch.-Ham. ex D. Don, an underutilized plant found growing abundantly in Meghalaya and also other northeastern states including the Himalayas has been reported to be the only species found in India (Haridasan and Rao, 1987). Synonyms of *Myrica esculenta* are *Myrica nagi* Thunb., *Myrica farquhariana* Wall., *Myrica integrifolia* Roxb. and *Myrica sapida* Wall. (Hooker, 1885; Kanjilal and Bor, 1940; Haridasan and Rao, 1987). It was interesting to find that diversity within the species existed in Meghalaya. These differences were observed at the morphological level especially with regard to the fruit size, fruit color, leaf length, leaf serration and tree height (Table 4.1). In the present study three distinct morphotypes were found to be growing in two different locations, NEHU permanent campus and Nongkrem forest, which were separated by a road distance of about 35 km. The morphotype-1 *Myrica* trees were found growing at NEHU permanent campus whereas the morphotype-2 and morphotype-3 *Myrica* trees were found growing at Nongkrem forest (Fig. 4.1). Even

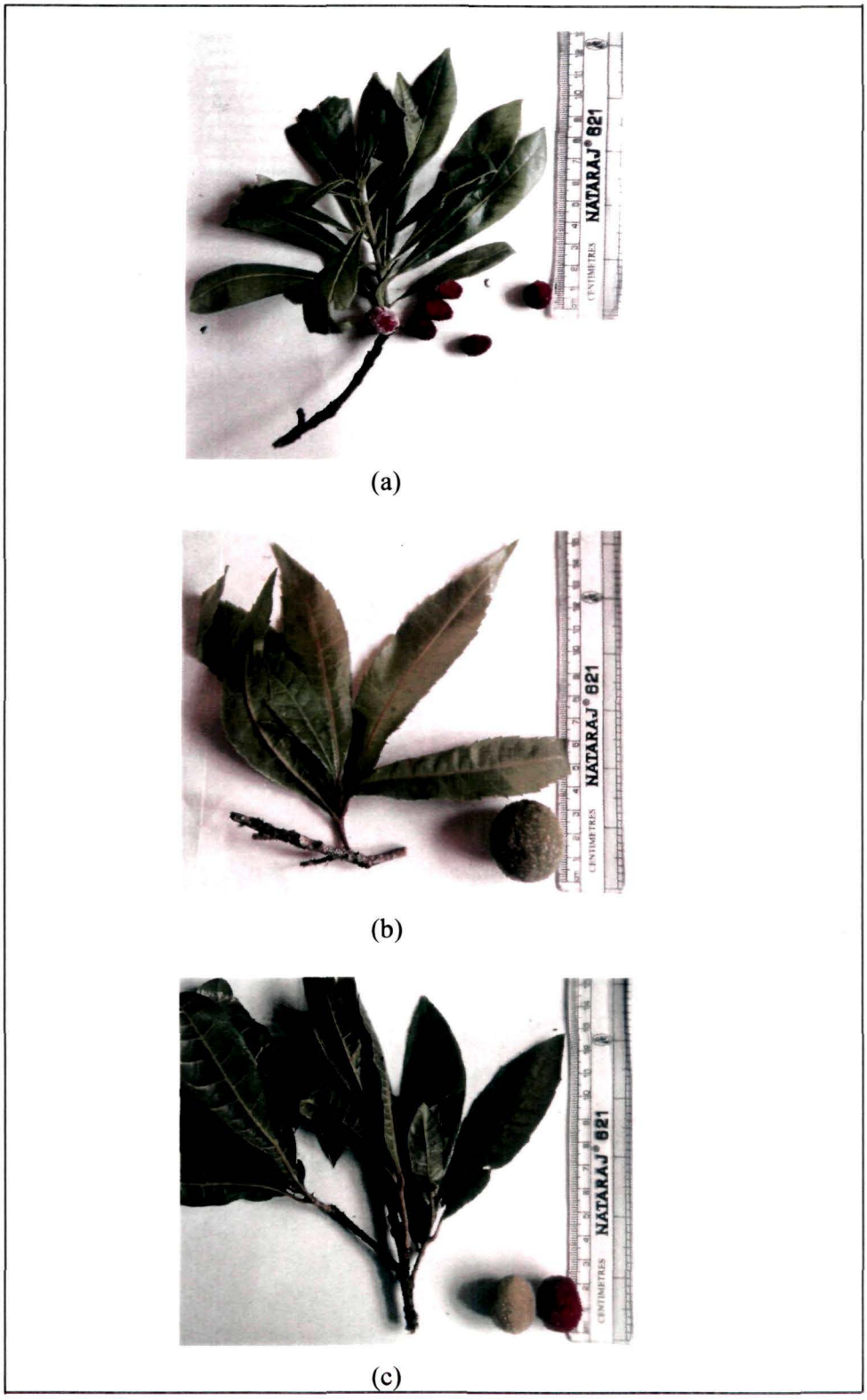


Fig. 4.1: Photographs of leaves and fruits of various morphotypes of *Myrica* trees measured in centimetre scale.

(a): Morphotype-1 *Myrica* leaves and fruits
(b): Morphotype-2 *Myrica* leaves and fruits
(c): Morphotype-3 *Myrica* leaves and fruits

TABLE 4.1: TABLE SHOWING THE DIFFERENT MORPHOLOGICAL DESCRIPTORS FOR THE MORPHOTYPE-1, MORPHOTYPE-2 AND MORPHOTYPE-3 *MYRICA* TREES.

SAMPLE	LEAF EDGES	MEAN LEAF LENGTH (cm)	MEAN FRUIT LENGTH (cm)	TREE HEIGHT (m)	FRUIT COLOUR WHEN RIPE	ALTITUDE (meters asl)
MORPHOTYPE-1 (NEHU Permanent campus)	Entire	12.8 ±0.05	1.3 ±0.05	5-8	Reddish	1431
MORPHOTYPE-2 (Nongkrem Forest)	Sharply serrated	18 ±0.25	3.4 ±0.15	7-15	Yellowish green	1631
MORPHOTYPE-3 (Nongkrem Forest)	Moderate to high serration	13 ±0.15	2.3 ±0.03	7-12	Yellowish green to red	1631

though, reports from literature (Kanjilal and Bor, 1940: Haridasan and Rao, 1987) indicate that these synonyms of *Myrica esculenta* refer to the same plant, confusion still persists among some authors with regard to its nomenclature. Some consider the morphotype-1 plant as *Myrica nagi* and the morphotype-2 plant as *Myrica esculenta* exclusively. Others consider that the different synonyms refer to the same plant. Therefore, the present study was undertaken in order to address the dispute regarding the nomenclature of *Myrica* species growing in Meghalaya.

Two separate regions of DNA, the variable 18S-28S rDNA ITS (Internal Transcribed Spacer) region and the highly conserved 18S rDNA, were studied for the purpose. A separate study on the nitrogenase activity of the root nodules of these trees was also undertaken to screen out trees with high or low nitrogenase activity. The study of the nitrogenase activity was further supplemented by a separate analysis of soil for estimating soil nitrogen status.

I. PHYLOGENETIC STUDY OF THE DIFFERENT MORPHOTYPES BASED ON DNA SEQUENCE DATA.

4.1 DNA ISOLATION:

Approximately 21 kb genomic DNA bands were obtained on electrophoresis in 0.8% agarose gel for all the samples under study [Fig. 4.2 (a-f), Fig. 4.3 (a & b), Fig. 4.4 (a-e) and Fig. 4.5 (a-d)]. An intense 21 kb DNA band was not obtained for all the samples but the yield was sufficient for subsequent amplification reactions.

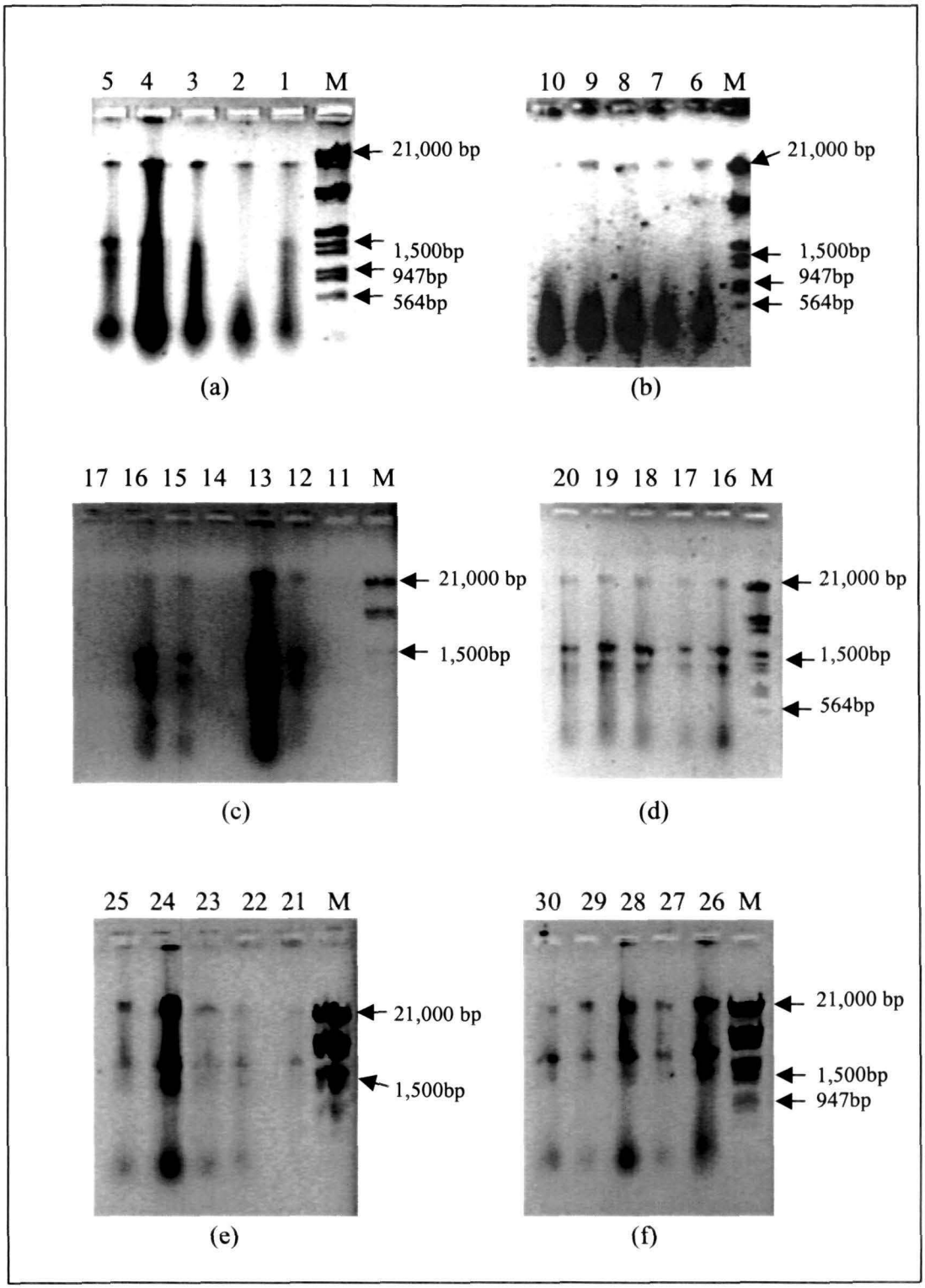


Fig. 4.2 (a-f): Agarose gel photographs of genomic DNA of morphotype-1 *Myrica* trees (Samples 1-30).
(M= λ DNA *Hind* III *Eco*R1 double digest marker)

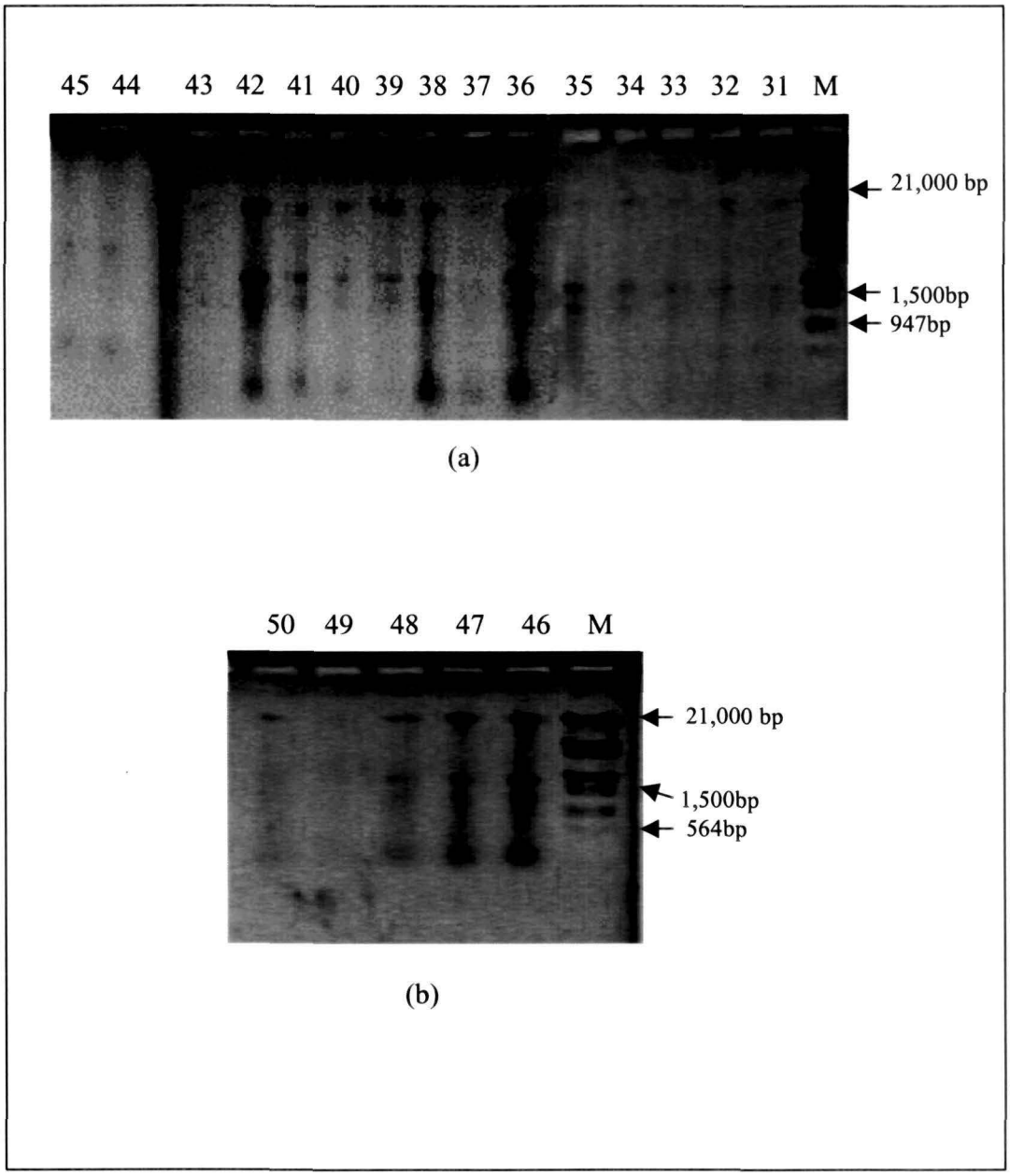


Fig. 4.3 (a & b): Agarose gel photographs of genomic DNA of morphotype-1 *Myrica* trees (Samples 31-50). (M= λ DNA *Hind* III/ *Eco*R1 double digest marker)

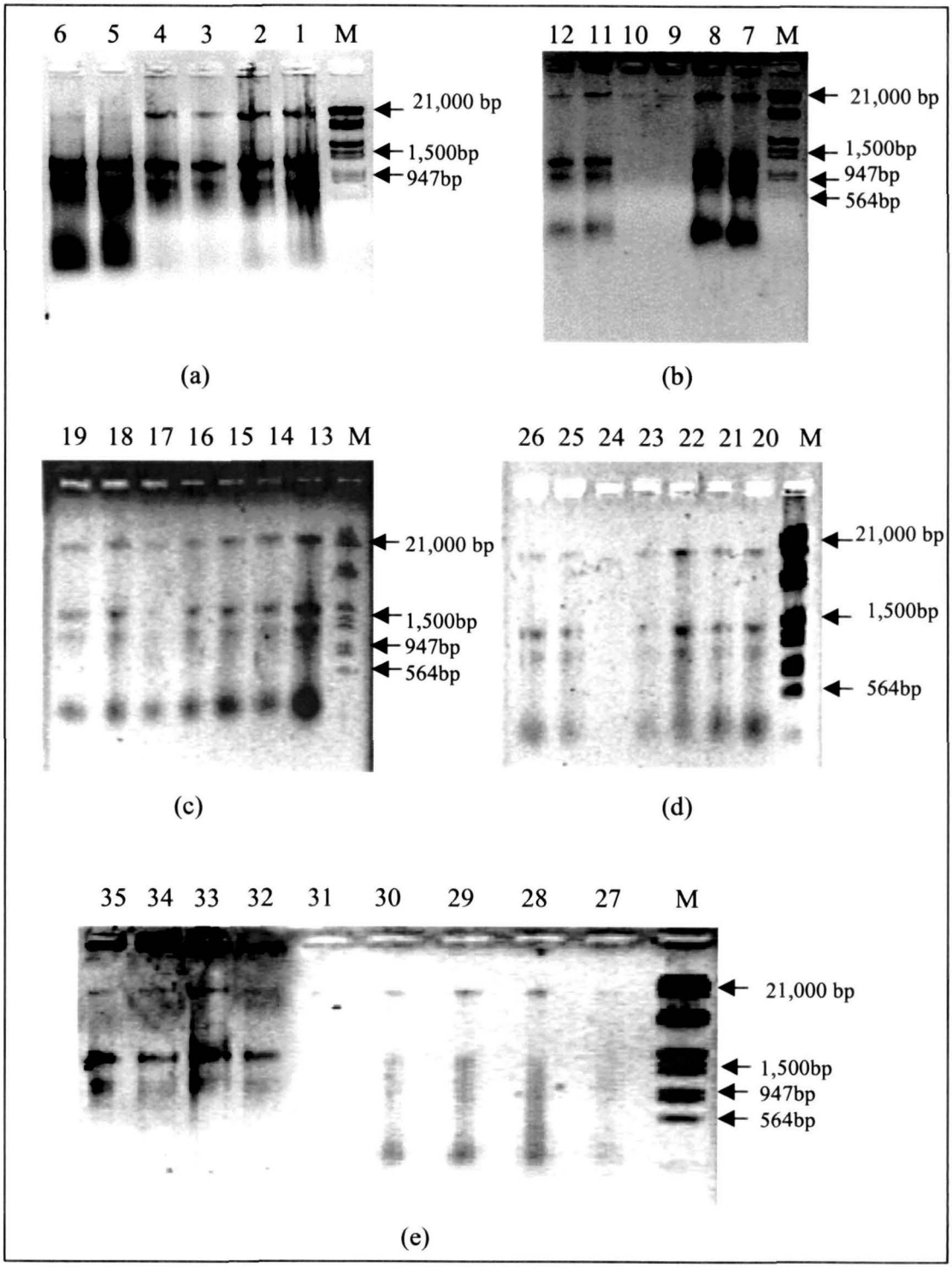


Fig. 4.4 (a-e): Agarose gel photographs of genomic DNA of morphotype-2 *Myrica* trees (Samples 1-35).
(M = λ DNA *Hind* III/ *Eco*R1 double digest marker)

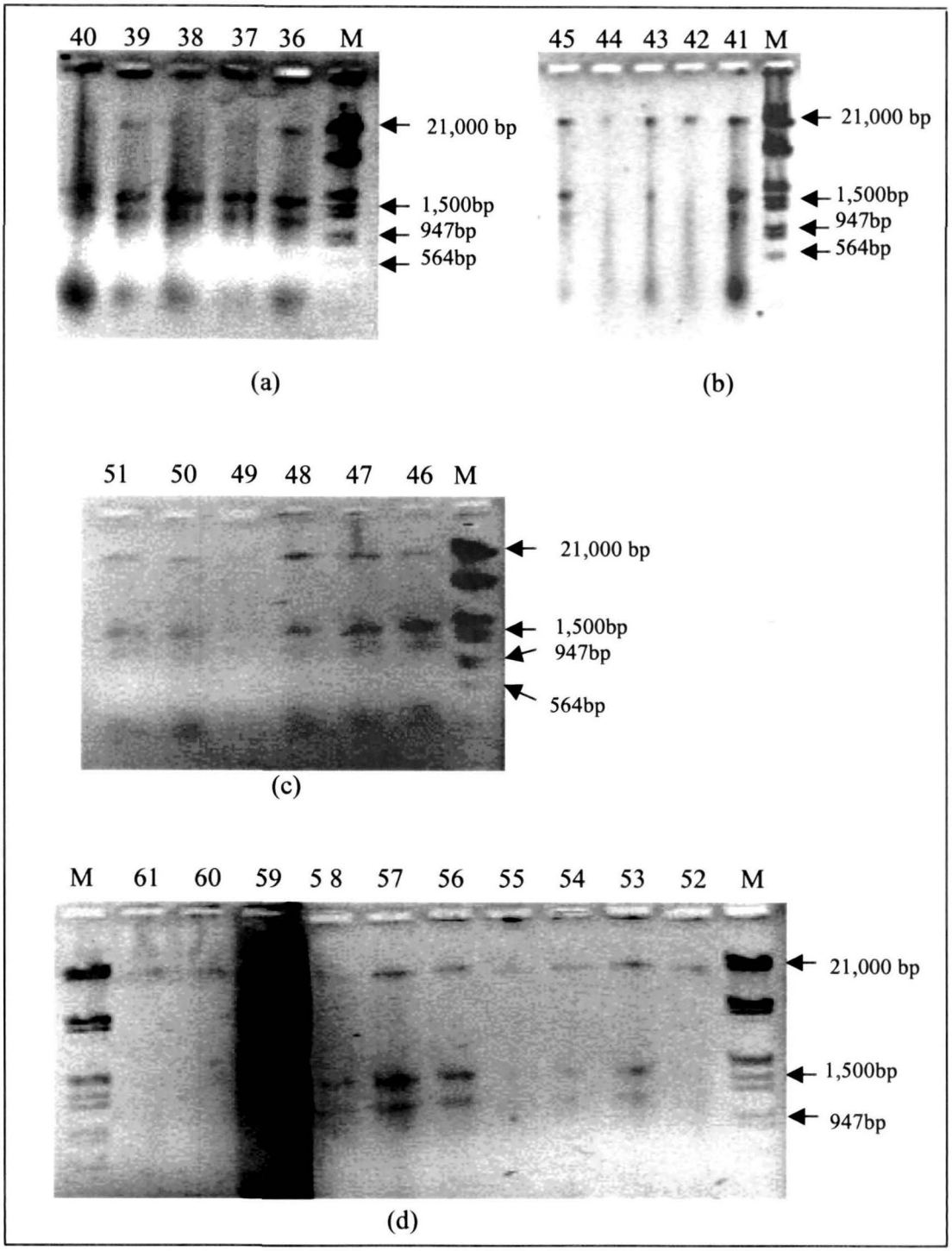


Fig. 4.5 (a-d): Agarose gel photographs of genomic DNA of morphotype-2 (Samples 36-50) and morphotype-3 (Samples 51-61) *Myrica* trees. (M = λ DNA *Hind* III/ *Eco*R1 double digest marker)

4.2 DNA AMPLIFICATION REACTIONS

4.2.1 Amplification of the 18S-28S rDNA ITS region:

The *Myrica* genomic Internal Transcribed Spacer (ITS) region of 18S-28S rDNA was amplified using two universal primers ITS1PLANT (5'CGCGAGAAGTCCACTG 3', Navarro *et al.*, 2003) and ITSC26A (5' GTTTCTTTTCCTCCGCT 3', Wen and Zimmer, 1996).

The calculated annealing temperatures of forward primer ITSIPLANT and reverse primer ITSC26A were 47°C and 45°C respectively. Normally, in cases where the annealing temperatures of the two primers differed, amplification was done at the lower annealing temperature. However, amplification at annealing temperature of 45°C yielded more than one band including the band of interest. The appearance of extra bands could be the result of nonspecific binding of the primers at sites other than the target sequence region. Therefore, the annealing temperature was increased to 52°C. However, at this temperature no amplification occurred. The best amplification result was obtained at annealing temperature of 49°C where a single band of approximately 800 bp was amplified. Amplifications for the rest of the samples were performed at annealing temperature of 49°C (Fig. 4.6, Fig. 4.7, Fig. 4.8 and Fig. 4.9).

4.2.2 Amplification of the 18S gene:

During the initial optimization a number of annealing temperatures (TM) were tried to obtain the expected 1.8 kb single band. Multiple bands were obtained at annealing temperature of 57°C. On increasing the temperature to 64°C a single band of

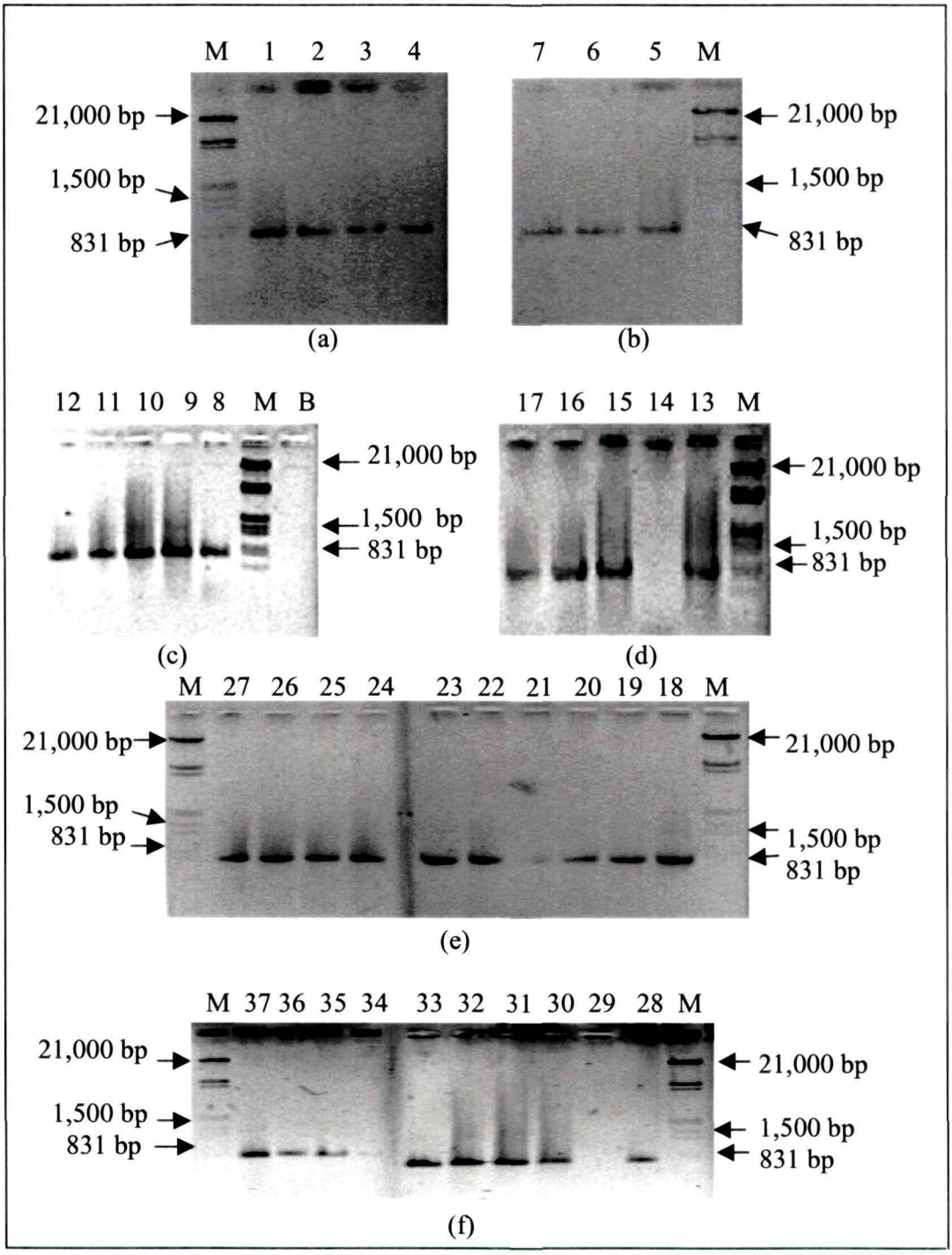


Fig. 4.6 (a-f): PCR amplification of the 18S-28S rDNA ITS region of morphotype-1 *Myrica* trees (Samples 1-37).
(M= λ DNA *Hind* III/ *Eco*R1 double digest marker; B= Blank)

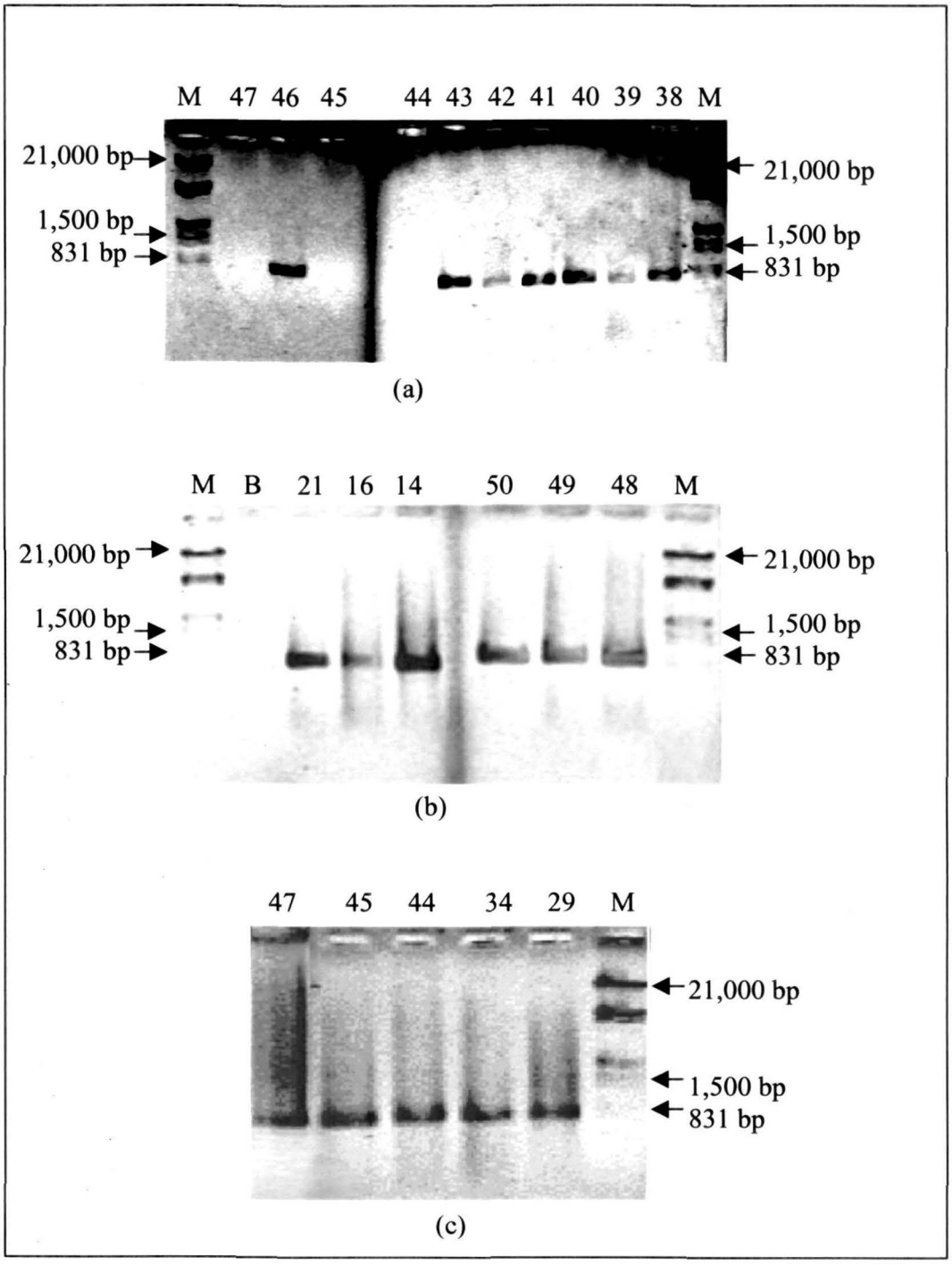


Fig. 4.7 (a-c): PCR amplification of the 18S-28S rDNA ITS region of morphotype-1 *Myrica* trees (Samples 38-50).
(M= λ DNA *Hind* III/ *Eco*R1 double digest marker; B= Blank)

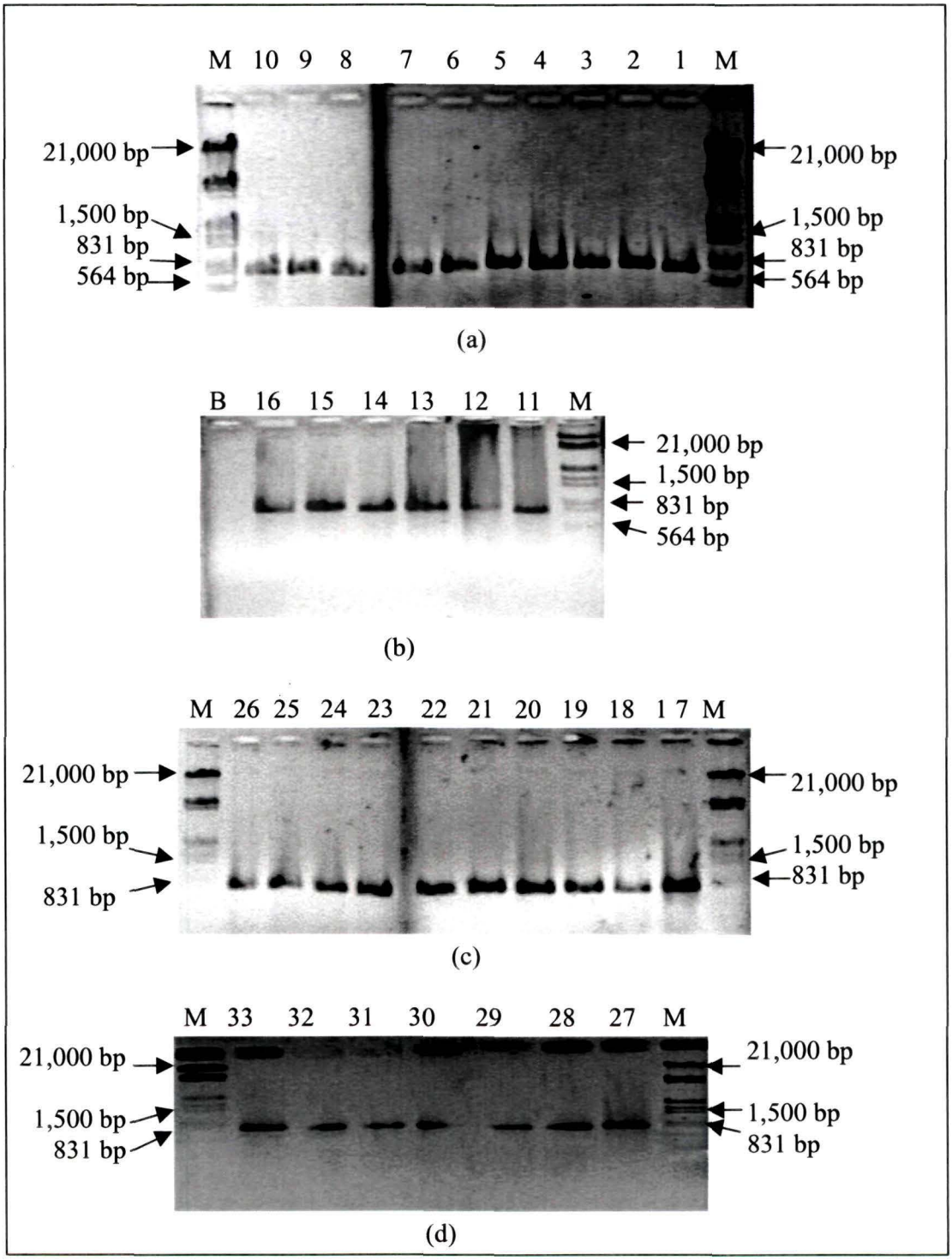


Fig. 4.8 (a-d): PCR amplification of the 18S-28S rDNA ITS region of morphotype-2 *Myrica* trees (Samples 1-33).
(M= λ DNA *Hind* III/ *Eco*R1 double digest marker)

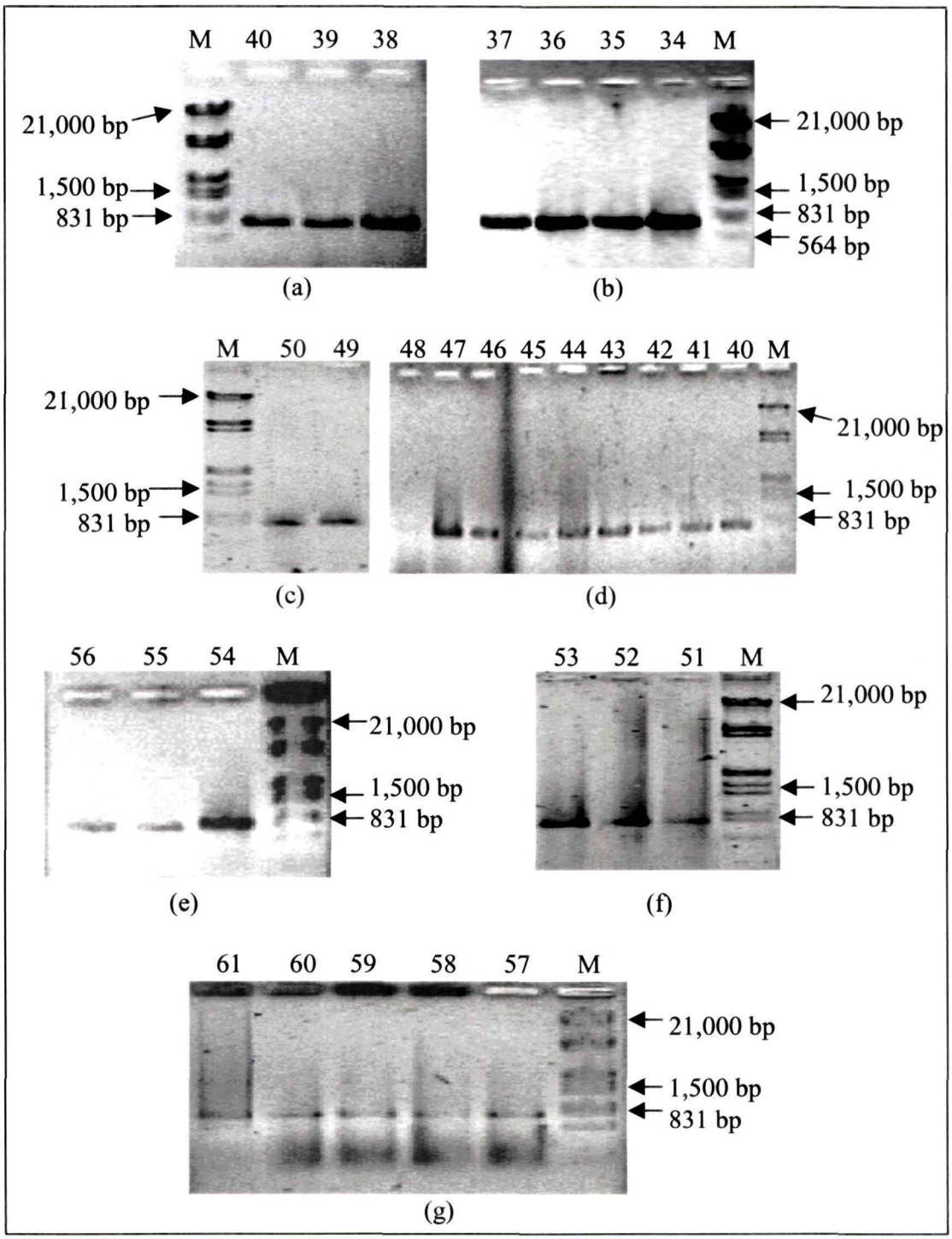


Fig. 4.9 (a-g): PCR amplification of the 18S-28S rDNA ITS region of morphotype-2 (34-50) and morphotype-3 (51-61) *Myrica* trees. (M= λ DNA *Hind* III/ *Eco*R1 double digest marker)

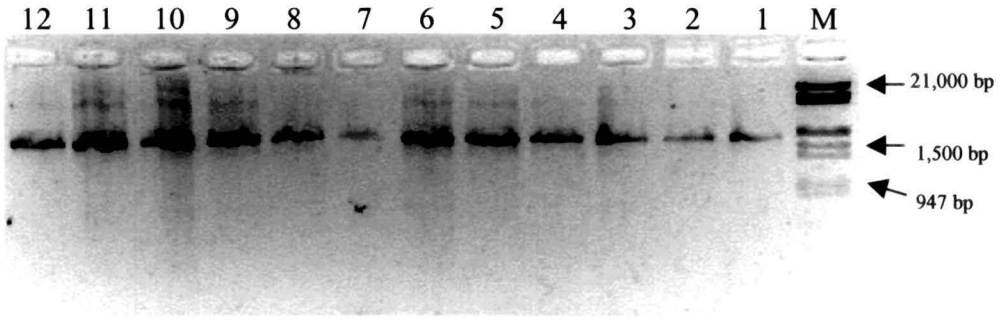
approximately 1.8 kb size could be obtained. However, on electrophoresis of the amplicon in 2% agarose gel, additional faint DNA bands could be detected. At higher annealing temperature of 67°C the band of interest appeared faint. Best results were obtained at annealing temperature of 66.5°C. Photographs of agarose gels showing amplifications at various annealing temperatures are given in Fig. 4.10 and Fig. 4.11.

4.3 DNA SEQUENCING:

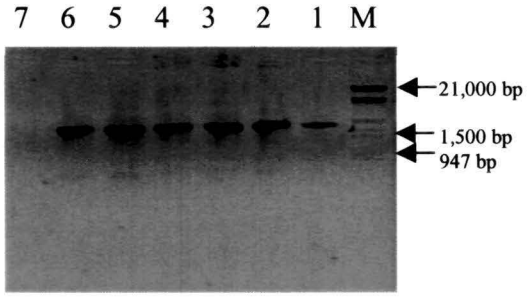
Nucleotide sequencing could not be performed for all the hundred and eleven samples collected across the three morphotypes of *Myrica* trees. Therefore, amplicons of only three samples representing each of the morphotypes from the two sites were sent for sequencing. Sequencing was performed for two different regions of the genomic DNA, the variable 18S-28S ITS region and the conserved 18S rDNA. The amplified DNA was prepared as discussed in section 3.6 of chapter 3 and sent to M/s Bangalore Genei for sequencing. Sequencing was done based on the dideoxy chain termination method of Sanger *et al.* (1977) (Fig. 4.12).

4.3.1 Nucleotide sequencing of the 18S-28S rDNA ITS region:

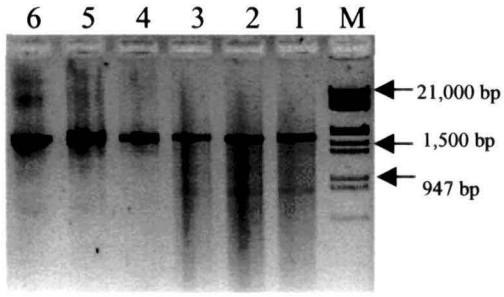
DNA sequencing of the ITS region was done employing the Single Pass Analysis (SPA) method using the same primers as in amplification for both the forward and reverse strands. Sequences were obtained in the form of electropherogram in which each nucleotide base was represented by a different colour. For the Internal Transcribed Spacer (ITS) region a total sequence length of 784 bp, 780 bp and 790 bp were obtained



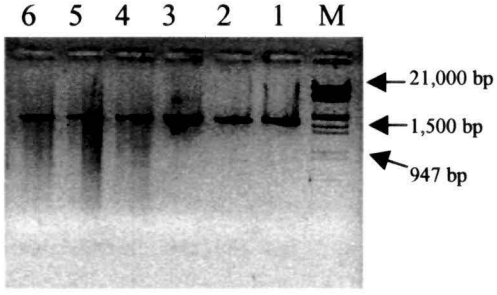
(a): Amplification at TM 57°C {1,2,3,4 –morphotype-1; 5,6,7,8-morphotype-2; 9,10,11,12-morphotype-3}.



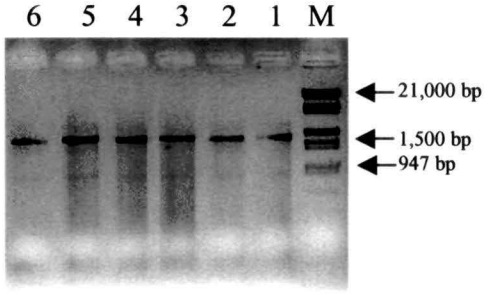
(b): Amplification at TM 60°C (1,2-morphotype-1; 3,4-morphotype-2; 5,6-morphotype-3; 7- negative control)



(c): 1,2,3- Amplification at TM 64°C; 4,5,6- Amplification at TM 57°C (1,4- morphotype-1; 2,5-morphotype-2; 3,6- morphotype-3)

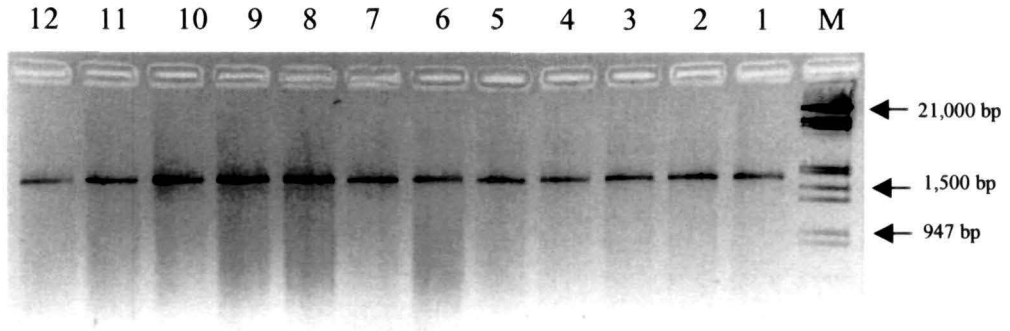


(d): 1,2,3- Amplification at TM 63°C; 4,5,6- Amplification at TM 64°C (1,4- morphotype-1; 2,5- morphotype-2; 3,6- morphotype-3)

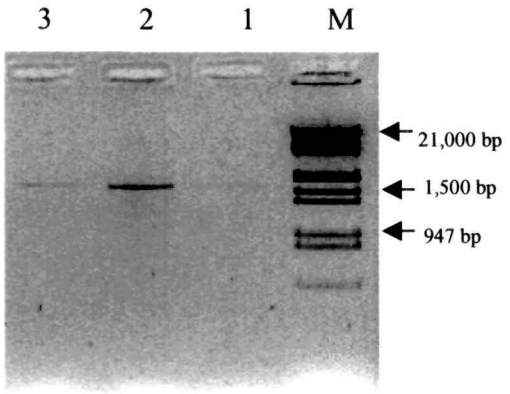


(e): Amplification at TM 65.2°C (1,2-morphotype-1; 3,4-morphotype-2; 5,6-morphotype-3)

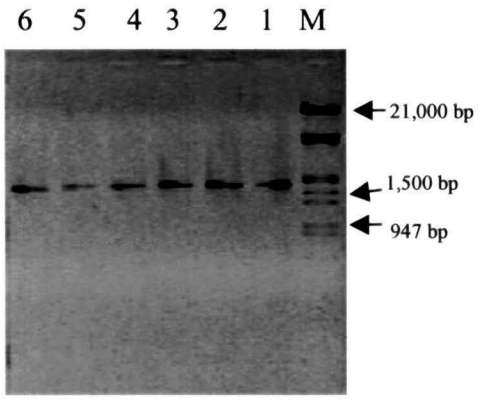
Fig. 4.10 (a-e): Agarose gel photographs showing amplification of the 18S rDNA at various annealing temperatures.
(M = λ DNA *Hind* III/ *Eco*R1 double digest marker)



(a): Amplification at TM 66.5°C (1,2,3,4- morphotype-1;
5,6,7,8- morphotype-2; 9,10,11,12- morphotype-3)



(b): Amplification at TM 67°C
(1 - morphotype-1;
2 - morphotype-2;
3 - morphotype-3)



(c): Amplification at TM 66.5°C
(1,2- morphotype-1;
3,4- morphotype-2;
5,6- morphotype-3)

Fig. 4.11 (a-c): Agarose gel photographs showing amplification of the 18S rDNA at various annealing temperatures.
(M = λ DNA *Hind* III/ *Eco*R1 double digest marker)

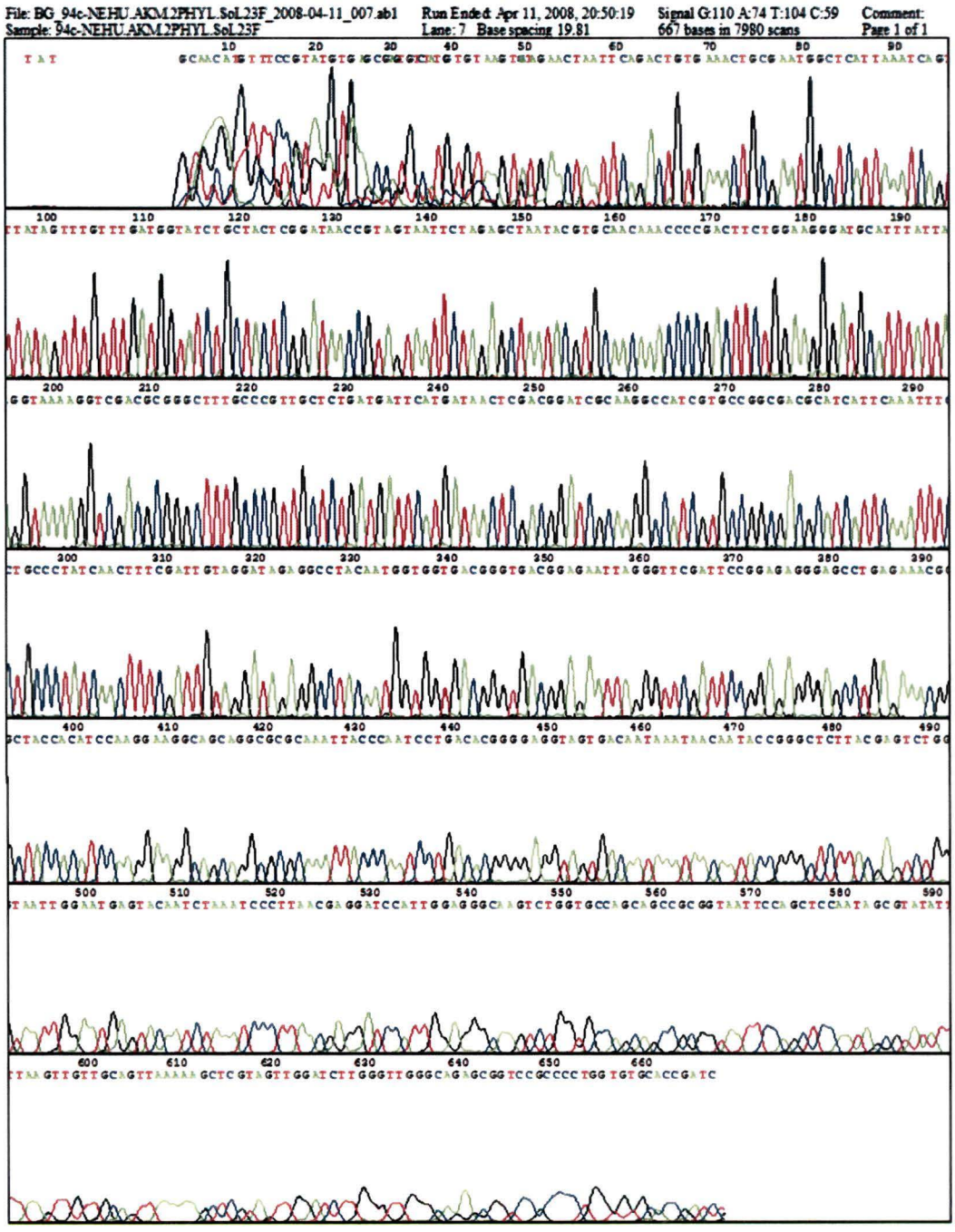


Fig. 4.12: Electrophrrogram showing partial sequence of 18S rDNA of morphotype-2 *Myrica* tree. Each of the four nucleotide bases are represented by different colours.

for samples ME1, ME2 and ME3 respectively (Fig. 4.13). ME1 was the representative sequence sample for the morphotype-1 *Myrica* trees. ME2 was the representative sequence sample for the morphotype-2 *Myrica* trees. ME3 was the representative sequence sample for the morphotype-3 *Myrica* trees.

4.3.1.1 Sequence annotation of the 18S-28S rDNA ITS region:

Sequence annotations of the 18S-28S ITS region of all the three representative samples (ME1, ME2 and ME3) of *Myrica* were done using the 'sequin' programme available at http://www.ncbi.nlm.nih.gov. All the three morphotypes of *Myrica* tree had an ITS1 larger than ITS2. The sizes of ITS1 and ITS2 were 268 bp and 225 bp respectively. The size of ITS1 and ITS2 was similar for all the three morphotypes. The sequence length of 5.8S rDNA gene (160 bp) was similar for all the three morphotypes. However, the annotation positions for different regions in the sequences of all the three morphotypes of *Myrica* trees were different.

Marked differences of about 20 nucleotide bases at the beginning and the end of the sequences were observed among the three morphotypes.

In the morphotype-1 *Myrica* tree sequence (ME1), bases 70-337 constituted ITS1, bases 338-497 constituted the 5.8S rDNA gene and bases 498-722 constituted ITS2.

In the morphotype-2 *Myrica* tree sequence (ME2), bases 67-334 constituted ITS1, bases 335-494 constituted the 5.8S rDNA gene and bases 495-719 constituted ITS2.

CLUSTAL W 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ME1	784	2	ME2	780	93
1	ME1	784	3	ME3	790	95
2	ME2	780	3	ME3	790	97

```
ME2             GGGGGAAGAATGTTGTGGGCGGG--GTCGTAAC---CTGGTTTCCGTAGGTGAACCTGCG 55
ME3             -GGGGATAATGGTTGAGGACGGG--GTCGTAAC---CTGGTTTCCGTAGGTGAACCTGCG 54
ME1             --CGCTCTGTCGATGAGGGCGAGAAGTCGTAACAGGTTACTACTCGTAGGTGAACCTGCG 58
                   *       * ** ** ** *  ********    *  *   ***************

ME2             GAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCG 115
ME3             GAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCG 114
ME1             GAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGTGAACATGTTAATAACTACCG 118
                ************************************* ********************

ME2             GGGGCAGGGGGCGATCAAAAGCCTCCCGTCCCCAAAACGGTTGGGGAGCACGTGCTGTTG 175
ME3             GGGGCAGGGGGCGATCAAAAGCCTCCCGTCCCCAAAACGGTTGGGGAGCACGTGCTGTTG 174
ME1             GGGGCAGGGGGCGATCAAAAGCCTCCTGTCCCCAAAACGGTTGGGGAGCACGTGCTGTTG 178
                ************************** ***********************************

ME2             TCCTGTCGGCCCTCTGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAA 235
ME3             TCCTGTCGGCCCTCTGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAA 234
ME1             TCCTGTCGACCCTCTGGCGCGAACAAGGAACCACACGCACATGTCCCCCCAATCGAACAA 238
                ******** ************ *** ************** *********** *******

ME2             CGAACCCCGGCGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCC-GAAGGCCCC 294
ME3             CGAACCCCGGCGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCC-GAAGGCCCC 293
ME1             CGAACCCCGGCGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCCGATCGCCCC 298
                ************************************************ **  *****

ME2             GGAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACG 354
ME3             GGAAACGGTGTGCCTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACG 353
ME1             GGAAACGGTGTGCGTTGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACG 358
                ************* * ********************************************

ME2             GATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTG 414
ME3             GATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTG 413
ME1             GATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTG 418
                ************************************************************

ME2             CAGAATCCCGCGAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGA 474
ME3             CAGAATCCCGCGAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGA 473
ME1             CAGAATCCCGCGAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGA 478
                ************************************************************
```

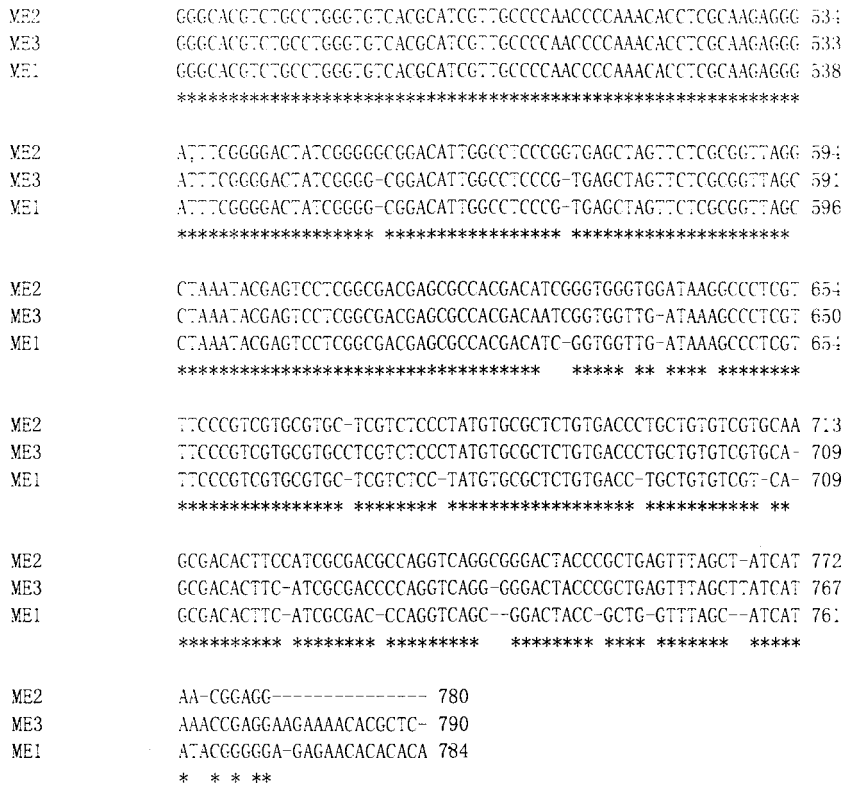


Fig. 4.13: Nucleotide sequence alignment of sample ME1, ME2 and ME3 of the variable 18S-28S rDNA ITS region using the multiple sequence alignment program CLUSTAL W. The asterisk (*) mark indicates the site of nucleotide base showing homology in all the three sequences.

In the morphotype-3 *Myrica* tree sequence (ME3), bases 66-333 constituted ITS1, bases 334-493 constituted the 5.8S rDNA gene and bases 494-718 constituted ITS2.

4.3.1.2 Secondary structure of the 5.8S rDNA:

Secondary structure of the 5.8S rRNA for the morphotype-1, morphotype-2 and morphotype-3 *Myrica* trees was constructed using the GeneBee programme available at http://www.genebee.msu.su/services/rna2_reduced.html. Secondary structure of 5.8S rRNA of Morphotype-1 was different from that of morphotype-2 and morphotype-3 secondary structures that were similar (Fig. 4.14).

4.3.2 Nucleotide sequencing of the 18S rDNA gene:

The nuclear 18S rDNA gene which forms part of the small sub-unit of the eukaryotic ribosome is approximately 1.8 kb in length (Soltis and Soltis, 1998). Sequencing of the 18S rDNA was done using primer walking technique which was based on the dideoxy chain termination method of Sanger *et al.* (1977). Initially, the primers used for amplification of the 18S rDNA gene were used at the start of the primer walk reaction. Total nucleotide sequence lengths of 1721 bp, 1603 bp and 1615 bp were obtained for the representative samples MYR1, MYR2 and MYR3 respectively (Fig. 4.15). MYR1 was the representative sequenced sample for the morphotype-1, MYR2 was the representative sequenced sample for the morphotype-2 and MYR3 was the representative sequenced sample for the morphotype-3 *Myrica* trees respectively.

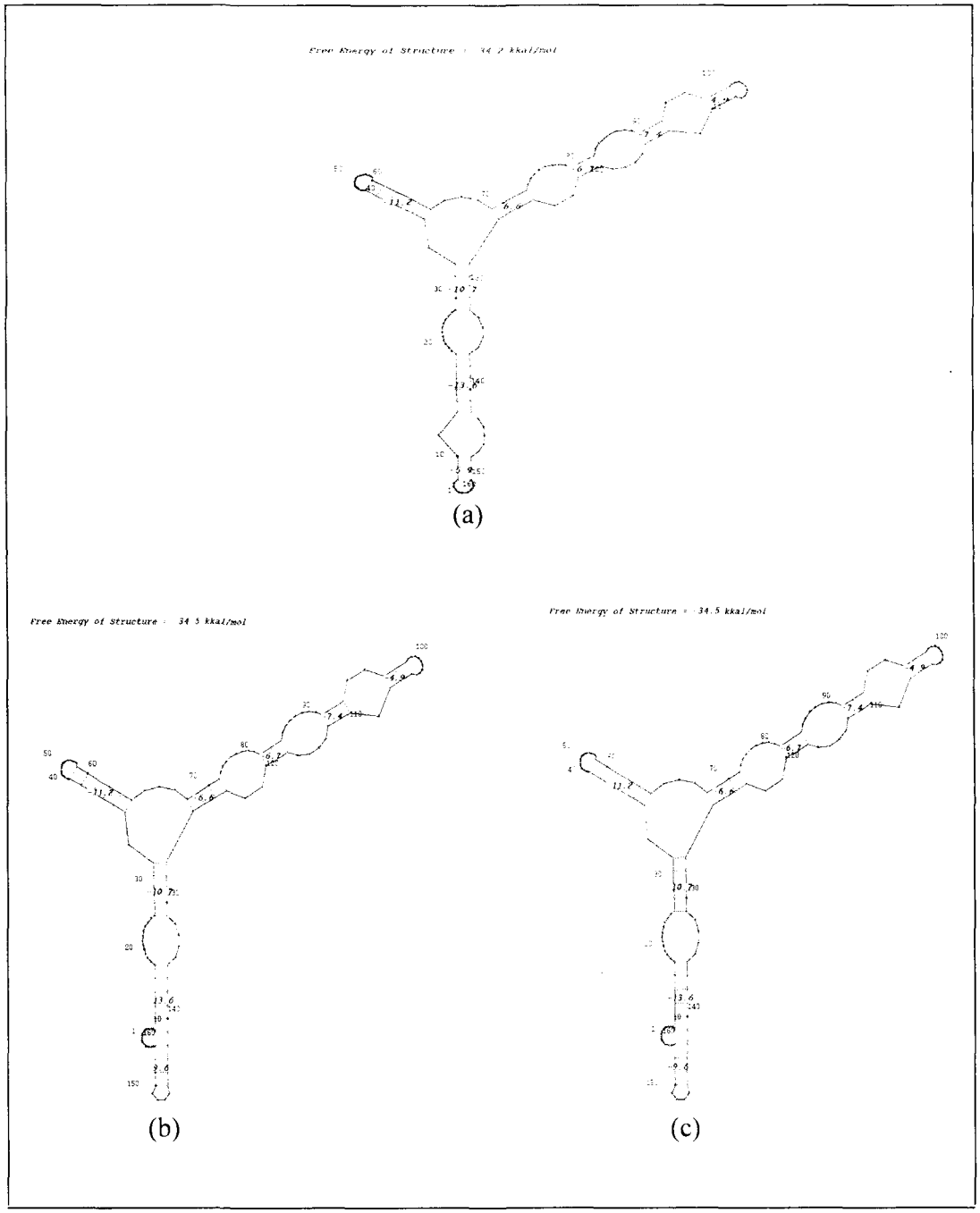


Fig. 4.14 (a) 5.8S mRNA secondary structure of morphotype-1.
(b) 5.8S mRNA secondary structure of morphotype-2
(c) 5.8S mRNA secondary structure of morphotype-3

CLUSTAL W 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	MYR1	1721	2	MYR2	1603	97
1	MYR1	1721	3	MYR3	1615	98
2	MYR2	1603	3	MYR3	1615	98

```
MYR1            GGGGCAAAAGCTATGCGTGTCTCATATGT-AGCCATGCTATCGT-GCCTAAGTGAT-GAA 57
MYR3            ----TGCAAGCGA-GCGTGTC-CGTATGT-GAGCCTGC----AT-GTGTAAGT-AT-GAA 46
MYR2            -----TATGCAACATGTTTC-CGTATGTGAGCCGAGTGTCTATCGTGTAAGTGATAGAA 53
                      * ** *   ** ** * *****    *  *     * *  ***** ** ***

MYR1            CTAATTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTGTTTGATGGT 117
MYR3            CTAATTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTGTTTGATGGT 106
MYR2            CTAATTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTGTTTGATGGT 113
                ************************************************************

MYR1            ATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCGACTT 177
MYR3            ATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCGACTT 166
MYR2            ATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCGACTT 173
                ************************************************************

MYR1            CTGGAAGGGATGCATTTATTAGGTAAAAGGTCGACGCGGGCTCTGCCCGTTGCTCTGATG 237
MYR3            CTGGAAGGGATGCATTTATTAGGTAAAAGGTCGACGCGGGCTTTGCCCGTTGCTCTGATG 226
MYR2            CTGGAAGGGATGCATTTATTAGGTAAAAGGTCGACGCGGGCTTTGCCCGTTGCTCTGATG 233
                ************************************ ****************

MYR1            ATTCATGATAACTCGACGGATCGCAAGGCCATCGTGCCGGCGACGCATCATTCAAATTTC 297
MYR3            ATTCATGATAACTCGACGGATCGCAAGGCCATCGTGCCGGCGACGCATCATTCAAATTTC 286
MYR2            ATTCATGATAACTCGACGGATCGCAAGGCCATCGTGCCGGCGACGCATCATTCAAATTTC 293
                ************************************************************

MYR1            TGCCCTATCAACTTTCGATTGTAGGATAGAGGCCTACAATGGTGGTGACGGGTGACGGAG 357
MYR3            TGCCCTATCAACTTTCGATTGTAGGATAGAGGCCTACAATGGTGGTGACGGGTGACGGAG 346
MYR2            TGCCCTATCAACTTTCGATTGTAGGATAGAGGCCTACAATGGTGGTGACGGGTGACGGAG 353
                ************************************************************

MYR1            AATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCA 417
MYR3            AATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCA 406
MYR2            AATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCA 413
                ************************************************************

MYR1            GCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACC 477
MYR3            GCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACC 466
MYR2            GCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACC 473
                ************************************************************

MYR1            GGGCTCTTACGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCA 537
MYR3            GGGCTCTTACGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCA 526
MYR2            GGGCTCTTACGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCA 533
                ************************************************************
```

```
MYR1        TTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTA 597
MYR3        TTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTA 586
MYR2        TTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTA 593
            ************************************************************

MYR1        AGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTTGGGTTGGGCAGAGCGGTCCGCCCCT 657
MYR3        AGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTTGGGTTGGGCAGAGCGGTCCGCCCCT 646
MYR2        AGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTTGGGTTGGGCAGAGCGGTCCGCCCCT 653
            ************************************************************

MYR1        GGTGTGCACCGATCTGCTCGTCCCTTCTACCGGCGATGCGCTCCTGGCCTTAACTGGCCG 717
MYR3        GGTGTGCACCGATCTGCTCGTCCCTTCTACCGGCGATGCGCTCCTGGCCTTAACTGGCCG 706
MYR2        GGTGTGCACCGATCTGCTCGTCCCTTCTACCGGCGATGCGCTCCTGGCCTTAACTGGCCG 713
            ************************************************************

MYR1        GGTCGTGCCTCCGGTGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCT 777
MYR3        GGTCGTGCCTCCGGTGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCT 766
MYR2        GGTCGTGCCTCCGGTGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCT 773
            ************************************************************

MYR1        CTGTATACATTAGCATGGGATAACATCATAGGATTTCGGTCCTATTGTGTTGGCCTTCGG 837
MYR3        CTGTATACATTAGCATGGGATAACATCATAGGATTTCGGTCCTATTGTGTTGGCCTTCGG 826
MYR2        CTGTATACATTAGCATGGGATAACATCATAGGATTTCGGTCCTATTGTGTTGGCCTTCGG 833
            ************************************************************

MYR1        GATCGGAGTAATGATTAACAGGAACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGA 897
MYR3        GATCGGAGTAATGATTAACAGGAACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGA 886
MYR2        GATCGGAGTAATGATTAACAGGAACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGA 893
            ************************************************************

MYR1        AATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATT 957
MYR3        AATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATT 946
MYR2        AATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATT 953
            ************************************************************

MYR1        AATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATA 1017
MYR3        AATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATA 1006
MYR2        AATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATA 1013
            ************************************************************

MYR1        AACGATGCCGACCAGGGATCGGCGGATGTTGCTTTAAGGACTCCGCCGGCACCTTATGAG 1077
MYR3        AACGATGCCGACCAGGGATCGGCGGATGTTGCTTTAAGGACTCCGCCGGCACCTTATGAG 1066
MYR2        AACGATGCCGACCAGGGATCGGCGGATGTTGCTTTAAGGACTCCGCCGGCACCTTATGAG 1073
            ************************************************************

MYR1        AAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATT 1137
MYR3        AAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATT 1126
MYR2        AAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATT 1133
            ************************************************************

MYR1        GACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAAC 1197
MYR3        GACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAAC 1186
MYR2        GACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAAC 1193
            ************************************************************
```

Fig. 4.15: Nucleotide sequence alignment of sample MYR1, MYR2 and MYR3 of the conserved 18S rDNA using the multiple sequence alignment program CLUSTAL W.
The asterisk (*) mark indicates the site of nucleotide base showing homology in all the three sequences.

4.4 PHYLOGENETIC ANALYSIS

Aligned sequences were used to retrieve related sequences from the GenBank using BLAST (Basic Local Alignment Search Tool) programme which was available at http://www.ncbi.nlm.nih.gov. Related sequences were then aligned with the sample sequences using the basic alignment tool of CLUSTAL W which was available at the site http://www.ebi.ac.uk/Tools/clustalw2/index.html. The generated aligned sequences were saved in Phylip format that served as an input file for PHYLIP (version 3.66) programme for generating phylogenetic trees.

4.4.1 Phylogenetic analysis of 18S-28S Internal Transcribed Spacer (ITS) region:

Phylogenetic analysis of the variable ITS region located between the distal part of 18S gene and the initial part of 28S gene of the ribosomal DNA was performed using different software. *Betula humilis* was selected as the outgroup species in all the phylogenetic analyses (Table 4.2). A bootstrap value of 1000 was considered for constructing different trees. Parsimony, Neighbour Joining and Maximum Likelihood phylogenetic trees are shown in Fig. 4.16, Fig. 4.17 and Fig. 4.18 respectively.

All the three samples clustered together in Neighbour Joining, Parsimony as well as the robust Maximum Likelihood trees scoring 1000.0, 964.9 and 950.0 bootstrap values respectively. Samples ME2 and ME3 clustered together with a bootstrap value of 774.0 in the Neighbour Joining tree and 804.0 in the Maximum Likelihood trees. In the parsimony tree clustering of all three samples in one group with 964.9 bootstrap value

TABLE 4.2: SEQUENCES OF 18S-28S INTERNAL TRANSCRIBED SPACER (ITS) OF *MYRICA, BETULA, COMPTONIA* AND *MORELLA* SP. USED FOR PHYLOGENETIC ANALYSIS

Sl. No.	NAME OF THE GENUS	GENBANK ACCESSION NUMBER	REFERENCE
1.	*Myrica nagi* (ME1)	FJ469992	Present study
2.	*Myrica esculenta*1 (ME2)	FJ469993	Present study
3.	*Myrica esculenta*2 (ME3)	FJ469994	Present study
4.	*Betula humilis*	AJ783643	Forest *et al.*, 2005
5.	*Comptonia peregrina*	AJ626764	Huguet *et al.*, 2005
6.	*Morella californica*	AJ626782	Huguet *et al.*, 2005
7.	*Morella cerifera*	AJ626771	Huguet *et al.*, 2005
8.	*Morella faya*	AJ626777	Huguet *et al.*, 2005
9.	*Morella heterophylla*	AJ626773	Huguet *et al.*, 2005
10.	*Morella rivas-martinezii*	AJ626781	Huguet *et al.*, 2005
11.	*Morella rubra*	AJ626784	Huguet *et al.*, 2005
12.	*Morella spathulata*	AJ626774	Huguet *et al.*, 2005
13.	*Myrica gale*	AJ626768	Huguet *et al.*, 2005
14.	*Morella quercifolia*	AJ626775	Huguet *et al.*, 2005
15.	*Morella nagi*	AJ626783	Huguet *et al.*, 2005
16.	*Morella pensylvanica*	AJ626772	Huguet *et al.*, 2005

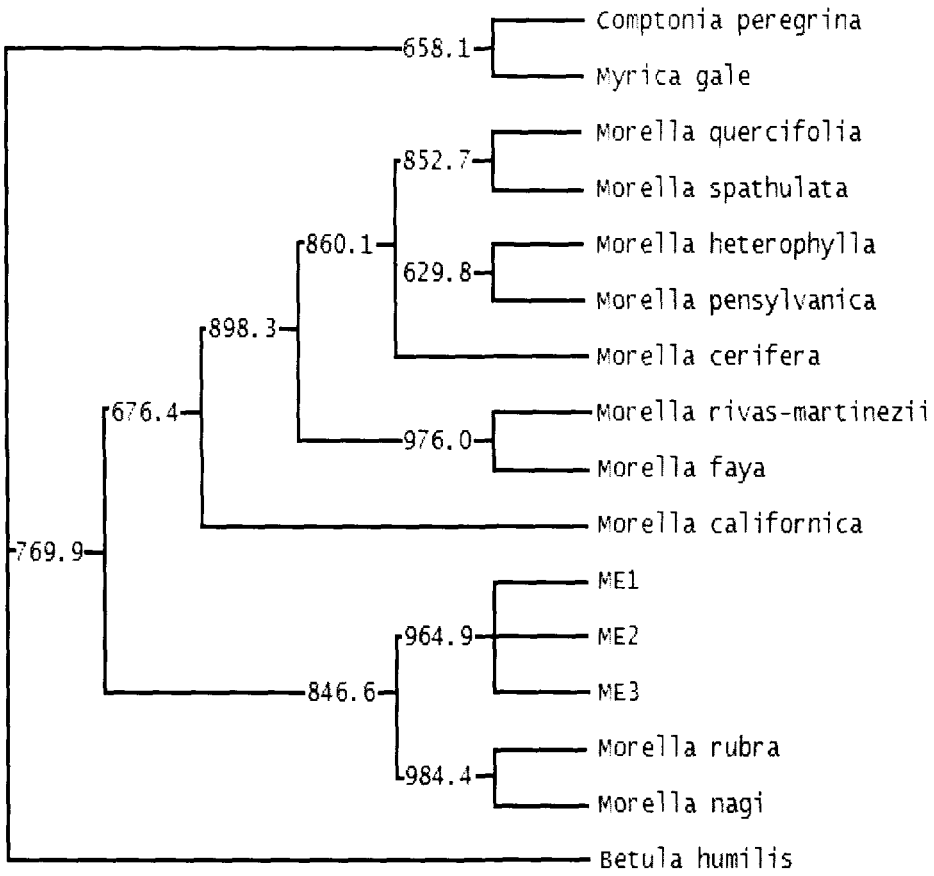


Fig. 4.16: Parsimony tree constructed from aligned sequences of 18S-28S Internal Transcribed Spacer (ITS) region considering 1000 bootstrap value.

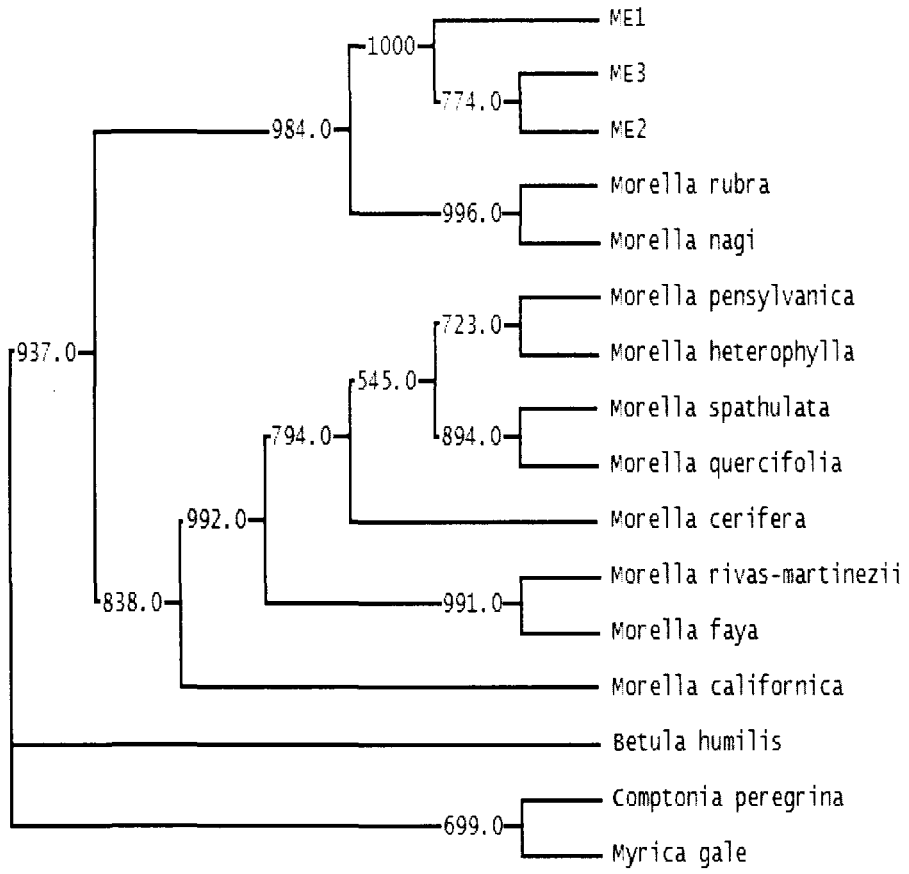


Fig. 4.17: Neighbor joining strict consensus tree constructed using aligned sequences of 18S-28S Internal Transcribed Spacer (ITS) region considering 1000 bootstrap value.

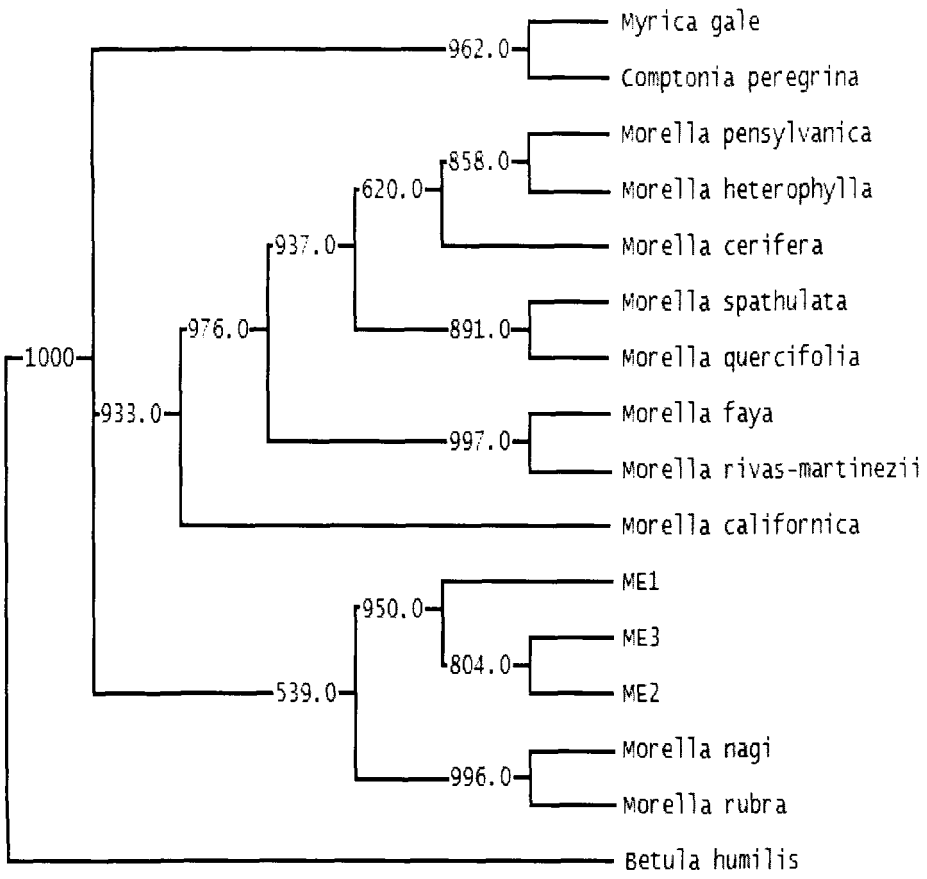


Fig. 4.18: Maximum Likelihood strict consensus tree constructed using aligned sequences of 18S-28S Internal Transcribed Spacer (ITS) region considering 1000 bootstrap value.

was observed. *Morella nagi* and *Morella rubra* clustered together in one group and showed closest affinity to our sample cluster. This clustering pattern showed similarity in Neighbour Joining, Parsimony as well as the Maximum Likelihood tree scoring 984.0, 846.6 and 539.0 bootstrap values. Three major clusters were observed where our samples clustered with *Morella rubra* and *Morella nagi* in one group and the remaining plants, excepting *Betula humilis*, *Comptonia peregrina* and *Myrica gale*, formed the other group. This grouping pattern was similar in the Parsimony, Neighbour Joining as well as the Maximum Likelihood trees. Results obtained from the phylogenetic analyses of the ITS sequences revealed close clustering of all three samples. However, samples ME2 and ME3 tended to cluster together marginally away from sample ME1. Thus samples ME2 (morphotype-2) and ME3 (morphotype-3) appeared closer than the sample ME1 (morphotype-1). This pattern was also observed at the morphological level where the morphotype-1 trees carries with it certain morphological descriptors which differ from the other two morphotypes. Sequence homology between samples ME1 and ME2 was 93%. Sequence homology between samples ME1 and ME3 and was 95%. Sequence homology between samples ME2 and ME3 was 97%. To determine the level of divergence among these samples, sequence data of some angiosperms was mined from data base and compared.

4.4.1.1. ITS sequence homology study of some tree angiosperms:

Sequence homologies were estimated using ITS sequences of different species within the same genus and also between different species belonging to different genera.

ITS sequences of twelve different species of *Morella* were retrieved from the GenBank for the purpose. The ITS sequence homology ranged from 94% to 100 % (Appendix 3) between different species of this genus. Even at the ITS level two cases of 100% homology were observed between species of the same genus. Similarly, ITS sequences of thirteen different species of *Alnus* were retrieved (Table 4.3). The percentages of sequence homology between the different species of *Alnus* too ranged from 94% to 100% (Apppendix 4). Seven ITS sequences belonging to different species of *Betula* were also aligned to check the sequence homology (Table 4.4). Sequence homology ranged from 96% to 100% (Appendix 5). Further, ITS sequence homology was also analysed for four different species of *Coriaria* in which the sequence homology ranged from 93% to 98% (Appendix 6). These results demonstrated that two sequences with homology around 94% may belong to two different species.

In another analysis, sequence homologies of retrieved ITS sequences of different species belonging to different genera were studied. ITS sequences of *Alnus glutinosa, Betula nana, Morella cerifera* and *Coriaria sarmentosa* were used for this analysis. Results obtained from these analyses showed that the ITS sequences varied considerably ranging from 65% to 90% (Appendix 7).

Intraspecific sequence homology test was also performed using ITS sequences retrieved from the GenBank that had been collected from different locations and also from different countries (Table 4.5). Four *Myrica gale* ITS sequences of Canada, Spain, Finland and Belgium were aligned in which the sequence homology ranged from 99 to 100% (Appendix 8). ITS sequences of four *Morella faya* species of Portugal, Spain

TABLE 4.3: SEQUENCES OF 18S-28S INTERNAL TRANSCRIBED SPACER (ITS) OF *ALNUS* SPECIES USED FOR PHYLOGENETIC ANALYSIS

Sl. No.	NAME OF THE GENUS	GENBANK ACCESSION NUMBER	REFERENCE
1.	*Alnus cordata*	AY352306	Chen, Z. and Li, J., 2004
2.	*Alnus glutinosa*	AY352310	Chen, Z. and Li, J., 2004
3.	*Alnus incana*	AY352312	Chen, Z. and Li, J., 2004
4.	*Alnus japonica*	AY352314	Chen, Z. and Li, J., 2004
5.	*Alnus nepalensis*	AY352318	Chen, Z. and Li, J., 2004
6.	*Alnus orientalis*	AY352320	Chen, Z. and Li, J., 2004
7.	*Alnus rubra*	AY352321	Chen, Z. and Li, J., 2004
8.	*Alnus rugosa*	AY352313	Chen, Z. and Li, J., 2004
9.	*Alnus serrulata*	AY352322	Chen, Z. and Li, J., 2004
10.	*Alnus sibirica*	AY352323	Chen, Z. and Li, J., 2004
11.	*Alnus tenuifolia*	AY352327	Chen, Z. and Li, J., 2004
12.	*Alnus trabeculosa*	AY352328	Chen, Z. and Li, J., 2004
13.	*Alnus viridis*	AY352329	Chen, Z. and Li, J., 2004

TABLE 4.4: SEQUENCES OF 18S-28S INTERNAL TRANSCRIBED SPACER (ITS) OF *BETULA* AND *CORIARIA* SPECIES USED FOR PHYLOGENETIC ANALYSIS

Sl. No.	NAME OF THE GENUS	GENBANK ACCESSION NUMBER	REFERENCE
1.	*Betula davurica*	AY352335	Chen, Z. and Li, J. 2004
2.	*Betula lenta*	AY352330	Chen, Z. and Li, J. 2004
3.	*Betula nana*	AY352336	Chen, Z. and Li, J. 2004
4.	*Betula nigra*	AY352331	Chen, Z. and Li, J. 2004
5.	*Betula pendula*	AY352332	Chen, Z. and Li, J. 2004
6.	*Betula populifolia*	AY352333	Chen, Z. and Li, J. 2004
7.	*Betula uber*	AY352334	Chen, Z. and Li, J. 2004
8.	*Coriaria microphylla*	AY091813	Yang *et al.*, 2003
9.	*Coriaria ruscifolia*	AY091815	Yang *et al.*, 2003
10.	*Coriaria sarmentosa*	AY091816	Yang *et al.*, 2003
11.	*Coriaria terminalis*	AY091817	Yang *et al.*, 2003

TABLE 4.5: SEQUENCES OF 18S-28S INTERNAL TRANSCRIBED SPACER (ITS) OF *MORELLA, COMPTONIA* AND *MYRICA* SP. USED FOR PHYLOGENETIC ANALYSIS

Sl. No.	NAME OF THE GENUS	GENBANK ACCESSION NUMBER	REFERENCE
1.	*Comptonia peregrina* USA	AJ626764	Huguet *et al.*, 2005
2.	*Comptonia peregrina* Canada	AJ626765	Huguet *et al.*, 2005
3.	*Morella adenophora* China	DQ501419	Herbert, J., 2007
4.	*Morella adenophora* Taiwan	DQ501420	Herbert, J., 2007
5.	*Morella cerifera* Jamaica	AJ626770	Huguet *et al.*, 2005
6.	*Morella cerifera* USA	AJ626771	Huguet *et al.*, 2005
7.	*Morella faya* Spain: Tenerife, Canary Islands	AJ626776	Huguet *et al.*, 2005
8.	*Morella faya* Spain: La Palma, Canary islands	AJ626777	Huguet *et al.*, 2005
9.	*Morella faya* Spain: Gomera, Canary Islands	AJ626778	Huguet *et al.*, 2005
10.	*Morella faya* Portugal	AJ626779	Huguet *et al.*, 2005
11.	*Morella rivas-martinezii* Spain: El Hierro, Canary Islands	AJ626780	Huguet *et al.*, 2005
12.	*Morella rivas-martinezii* Spain: Gomera, Canary Islands	AJ626781	Huguet *et al.*, 2005
13.	*Myrica gale* Belgium	AJ626766	Huguet *et al.*, 2005
14.	*Myrica gale* Finland	AJ626767	Huguet *et al.*, 2005
15.	*Myrica gale* Spain	AJ626768	Huguet *et al.*, 2005
16.	*Myrica gale* Canada	AJ626769	Huguet *et al.*, 2005
17.	*Morella esculenta* UK	DQ501421	Herbert, J., 2007

Gomera, Spain Tenerife and Spain Lapalma were also aligned in which 100% sequence homology was observed (Appendix 9). *Comptonia peregrina* ITS sequences of Canada and the USA were also aligned in which 100% sequence homology was observed (Appendix 10). Similarly, ITS sequence homology of *Morella rivas-martinezii* of Elhierro, Spain and *Morella rivas-martinezii* of Gomera, Spain revealed 100% sequence homology (Appendix 11). *Morella adenophora* of China and *Morella adenophora* of Taiwan showed 98% sequence homology (Appendix 12). *Morella cerifera* of Jamaican origin and *Morella cerifera* of USA origin also showed 99% sequence homology (Appendix 13). Results obtained from these intraspecific sequence homology tests showed high level of sequence similarity between members of same species that were geographically separated. Therefore, the low level of sequence homology observed between ME1 on one hand and ME2 and ME3 on the other hand is indicative of the possibility of ME1 being a different species belonging to the same genus.

4.4.2 Phylogenetic analysis of the 18S rDNA:

Eight related BLAST searched sequences retrieved from the GenBank were used in conjunction with our 18S rDNA sequences to construct phylogenetic trees (Table 4.6). Of these the most distant, *Berberidopsis corallina,* was used as the out species.

Neighbour joining analysis was performed using SEQBOOT (bootstrap), DNADIST (distance matrix) and CONSENSE (consensus tree programme). One thousand bootstrap values were considered to select the best consensus tree. All three samples MYR1, MYR2 and MYR3 clustered together scoring a bootstrap value of

TABLE 4.6: SEQUENCES OF 18S rDNA USED FOR PHYLOGENETIC ANALYSIS

Sl. No.	NAME OF THE GENUS	GENBANK ACCESSION NUMBER	REFERENCE
1.	*Myrica nagi* (MYR1)	FJ469989	Present study
2.	*Myrica esculenta*1 (MYR2)	FJ569990	Present study
3.	*Myrica esculenta*2 (MYR3)	FJ469991	Present study
4.	*Morella cerifera*	AF206967	Soltis *et al.*, 1999 (*unpublished)
5.	*Juglans nigra*	AF206943	Soltis *et al.*, 1999 (*unpublished)
6.	*Fagus grandifolia*	AF206910	Soltis *et al.*, 1999 (*unpublished)
7.	*Carya glabra*	AF206880	Soltis *et al.*, 1999 (*unpublished)
8.	*Chrysolepis sempervirens*	AF206886	Soltis *et al.*, 1999 (*unpublished)
9.	*Cucurbita pepo*	AF206895	Soltis *et al.*, 1999 (*unpublished)
10.	*Berberidopsis corallina*	AF206866	Soltis *et al.*, 1999 (*unpublished)
11.	*Stachyurus praecox*	AF207025	Soltis *et al.*, 1999 (*unpublished)

*Sequences submitted directly to GenBank.

1000.0. Within the common cluster of our samples the morphotype-2 and morphotype-3 sub-cluster scored a bootstrap value of 731.0. *Morella cerifera* showed the closest affinity to all the three samples scoring a bootstrap value of 970.0. The out group tree, *Berberidopsis coralline*, appeared to be more closely related to *Stachyurus praecox* than any other genus within the group (Fig. 4.19).

In the Parsimony tree (Fig. 4.20) a bootstrap value of 999.6 was obtained for the cluster of all three samples. They were related closest to *Morella cerifera* scoring 920.9 bootstrap value. In both the neighbour joining and parsimony trees, *Juglans nigra* and *Carya glabra* tended to remain in one cluster scoring bootstrap values of 991.0 (Neighbour Joining) and 726.2 (Parsimony). The *Juglans nigra* - *Carya glabra* cluster being closer to *Morella cerifera* - our samples cluster. This clustering pattern was similar for *Fagus grandifolia* - *Chrysolepis sempervirens* cluster scoring 942.0 bootstrap value for Neighbour Joining tree and 941.0 bootstrap value for Parsimony tree. They appeared more distant to our samples than the *Juglans nigra* - *Carya glabra* cluster.

To obtain a robust inference to validate our result the Maximum Likelihood analysis (Fig. 4.21) was performed taking into consideration 1000 bootstrap using the dnamlk.exe of the programme PHYLIP (version 3.66). The out file of this run was used as an input data to generate the most consensus tree (CONSENSE). A bootstrap value of 954.0 was obtained for all the three samples that clustered together showing the closest affinity to *Morella cerifera* scoring a bootstrap value of 507.0. Two major groups were observed where our samples together with *Morella cerifera*, *Carya glabra* and *Juglans nigra* formed one group and the rest of the species formed the other major group.

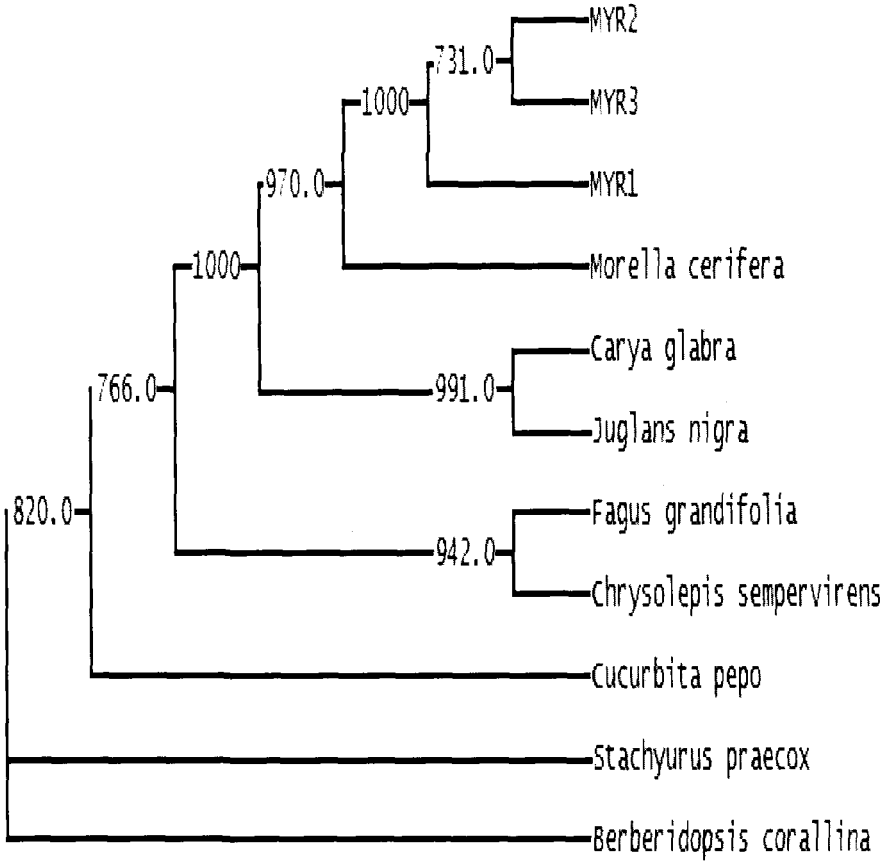


Fig. 4.19: Neighbor Joining strict consensus tree constructed using aligned sequences of 18S rDNA gene considering 1000 bootstrap value. The numbers at each node represents the bootstrap value.

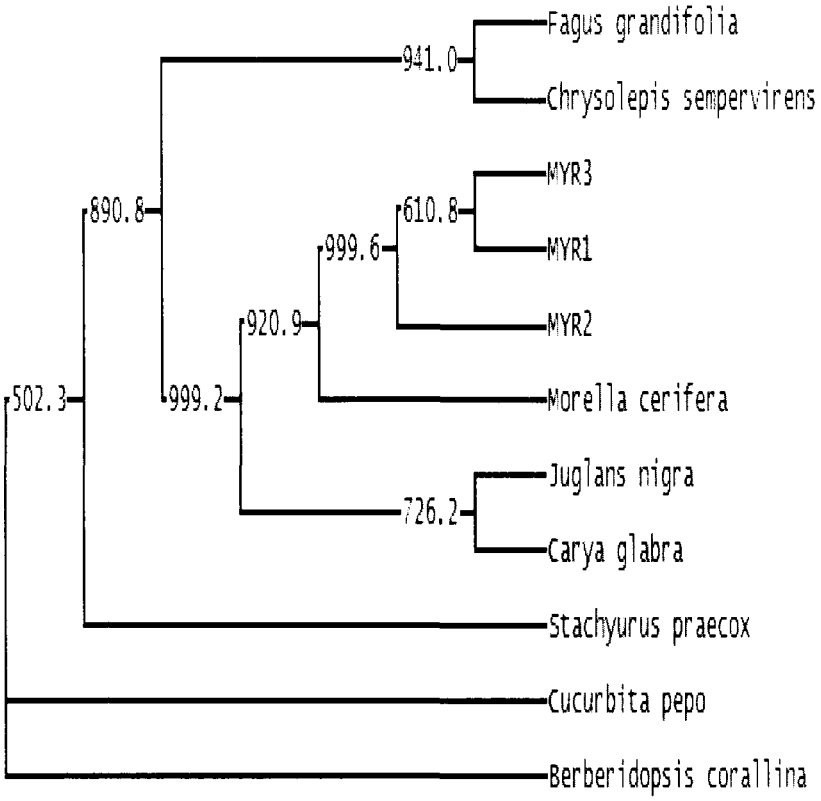


Fig. 4.20: Parsimony strict consensus phylogenetic tree constructed using aligned sequences of 18S rDNA gene considering 1000 bootstrap value. The numbers at each node represents the bootstrap value.

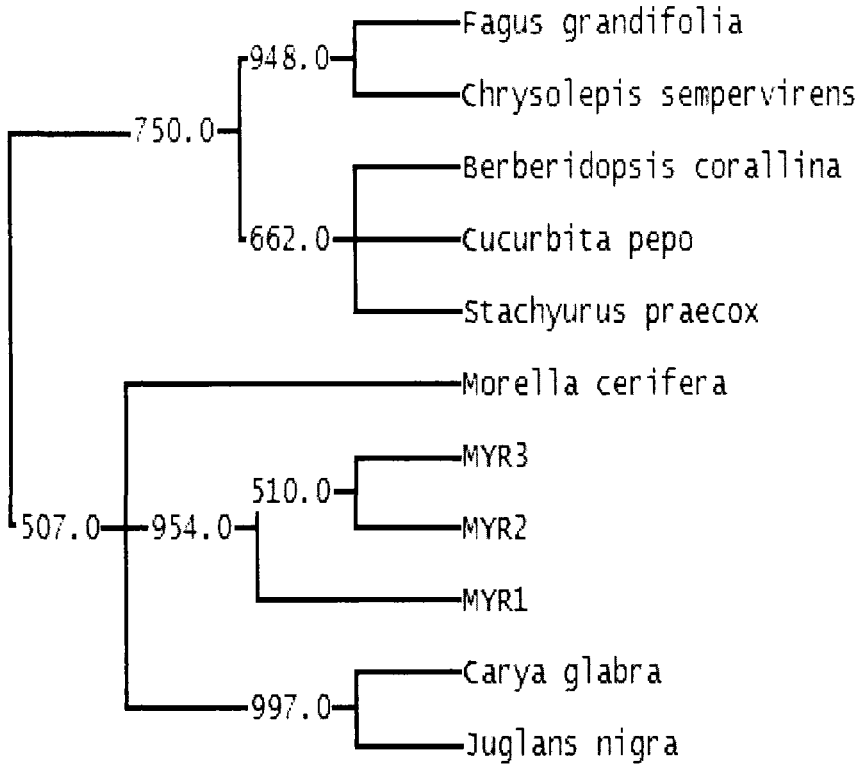


Fig. 4.21: Maximum Likelihood phylogenetic tree constructed using aligned sequences of 18S rDNA gene considering 1000 bootstrap value. The numbers at each node represents the bootstrap value.

Results obtained from these dendrograms showed that all the three samples clustered together in the Parsimony, Neighbour Joining as well as the robust Maximum Likelihood trees. However, within the cluster, there was separation of the morphotype-1 trees with those of the morphotype-2 and the morphotype-3 trees that clustered together.

4.4.2.1. 18S rDNA sequence homology study of some tree angiosperms:

In order to draw comparisons, we retrieved 18S sequences of different species of angiosperms from the GenBank and aligned them to analyze their sequence homology (Table 4.7). The sequence homology of four different species of the genus *Coriaria* ranged from 97% to 99% (Appendix 14). Similarly, sequence alignment between *Datisca glomerata* and *Datisca cannabina* showed 98% homology (Appendix 15).

In a similar analysis we also aligned 18S sequences between species of different genera to check the level of sequence homology. For this purpose the 18S sequences of *Carya glabra, Datisca cannabina, Coriaria ruscifolia, Morella cerifera* and *Juglans nigra* were used. The sequence homology ranged from 96% to 99% (Appendix 16). Sequence homology between samples MYR1 and MYR2 was 97%, MYR1 and MYR3 was 98% and MYR2 and MYR3 was 98%. This comparative study is suggestive of the possibility that morphotype 1 and morphotype 2 trees may belong to two different species of genus *Myrica.*

TABLE 4.7: SEQUENCES OF 18S rDNA OF *CORIARIA, DATISCA, JUGLANS* AND *CARYA* SP. USED FOR PHYLOGENETIC ANALYSIS

Sl. No.	NAME OF THE GENUS	GENBANK ACCESSION NUMBER	REFERENCE
1.	*Coriaria myrtifolia*	AF206891	Soltis *et al.*, 2003
2.	*Coriaria nepalensis*	AY968394	Zhang *et al.*, 2006
3.	*Coriaria ruscifolia*	AY968395	Zhang *et al.*, 2006
4.	*Coriaria sarmentosa*	AY968396	Zhang *et al.*, 2006
5.	*Datisca cannabina*	AF008952	Swensen, S. M., 2004
6.	*Datisca glomerata*	DGU42426	Swensen, S. M., 1996
7.	*Juglans nigra*	AF206943	Soltis *et al.*, 2003
8.	*Carya glabra*	AF206880	Soltis *et al.*, 2003

4.5 AMPLICON RESTRICTION PATTERN (ARP)/ PCR- RESTRICTION FRAGMENT LENGTH PROFILE (PCR-RFLP):

Two restriction enzymes, *Sau*96I and *Mbo*I, produced polymorphic restriction patterns. Restriction digestion was carried out on all the one hundred and eleven samples with the two enzymes separately.

4.5.1 Amplicon Restriction Pattern (ARP) with *Mbo*I:

Restriction enzyme *Mbo*I generated different patterns for the three morphotypes (Fig. 4.22, 4.23, 4.24 & 4.25). Nine different profiles, PM1, PM2, PM3, PM4, PM5, PM6, PM7, PM8 and PM9 were obtained. Five of these profiles (PM1, PM2, PM3, PM4 and PM5) were found only for morphotype-1 trees. The profiles PM6, PM7, PM8 and PM9 were found only for morphotype-2 and morphotype-3 trees (Table 4.8).

Six samples (NPC-1, NPC-3, NPC-4, NPC-6, NPC-7 and NPC-9) were found to have profile PM1 (Fig. 4.26). They showed four fragments of ~700 bp, ~530 bp, ~180 bp and ~90 bp.

Profile PM2 (Fig. 4.27) corresponded to only one sample, NPC-2, where five restriction fragments were obtained including the commonly occurring ~530 bp, ~180 bp and ~90 bp fragments. The two additional bands corresponded to ~730 bp and ~380 bp size along with the remnants of the undigested band of ~800 bp. This undigested band could be because of the presence of more than one copy of the gene in the genome. In one of the copies, the restriction site for *Mbo*I could be absent.

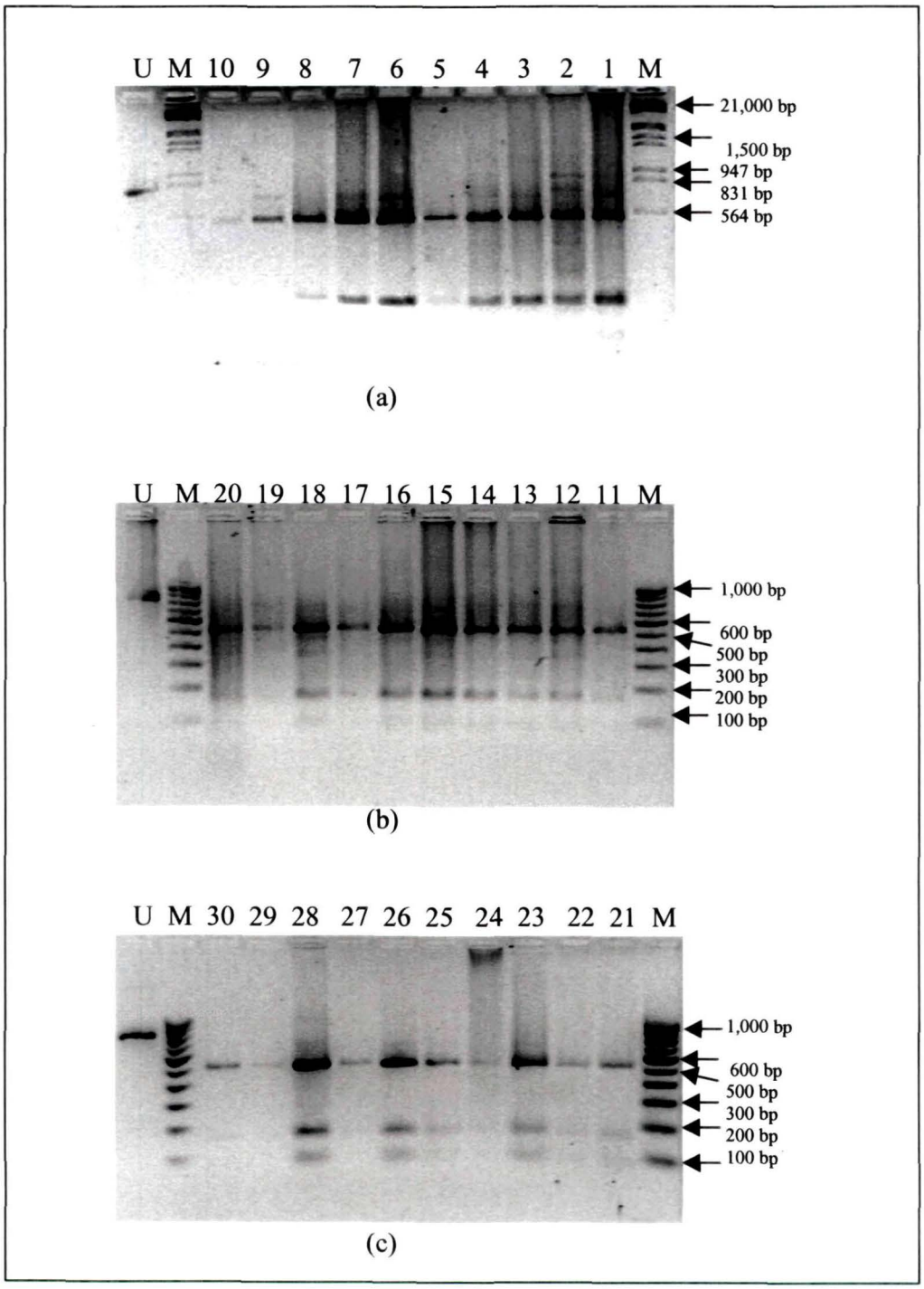


Fig. 4.22 (a-c): Amplicon Restriction Pattern (ARP) of morphotype-1 *Myrica* trees (samples 1-30) with *Mbo*I.
(M= 100 bp ladder; U= undigested DNA)
{Fig a; M= λ DNA *Hind* III/ *Eco*R1 double digest marker}

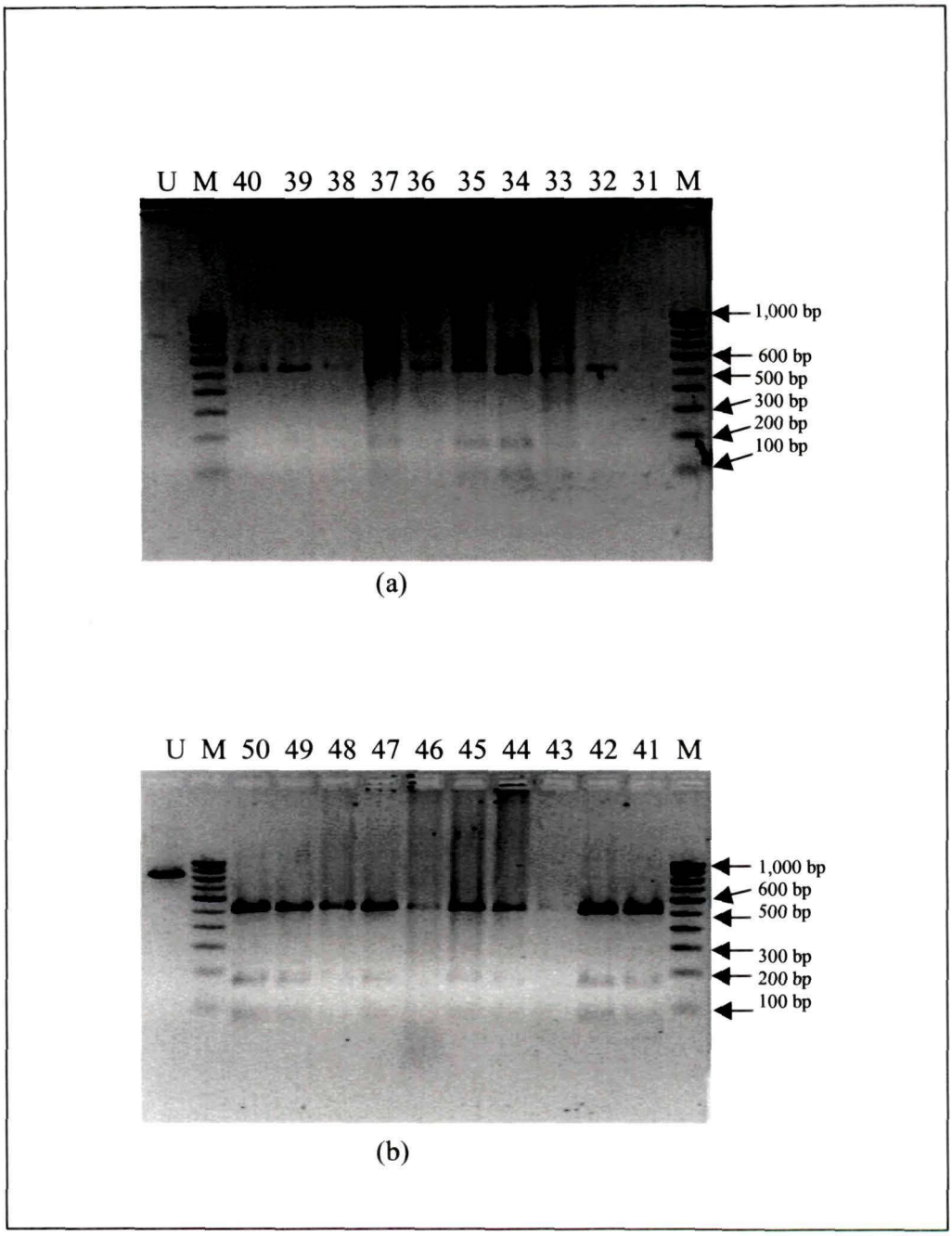


Fig. 4.23 (a & b): Amplicon Restriction Pattern (ARP) of morphotype-1 *Myrica* trees (Samples 31-50) with *Mbo*I.
(M= 100 bp ladder; U=undigested DNA)

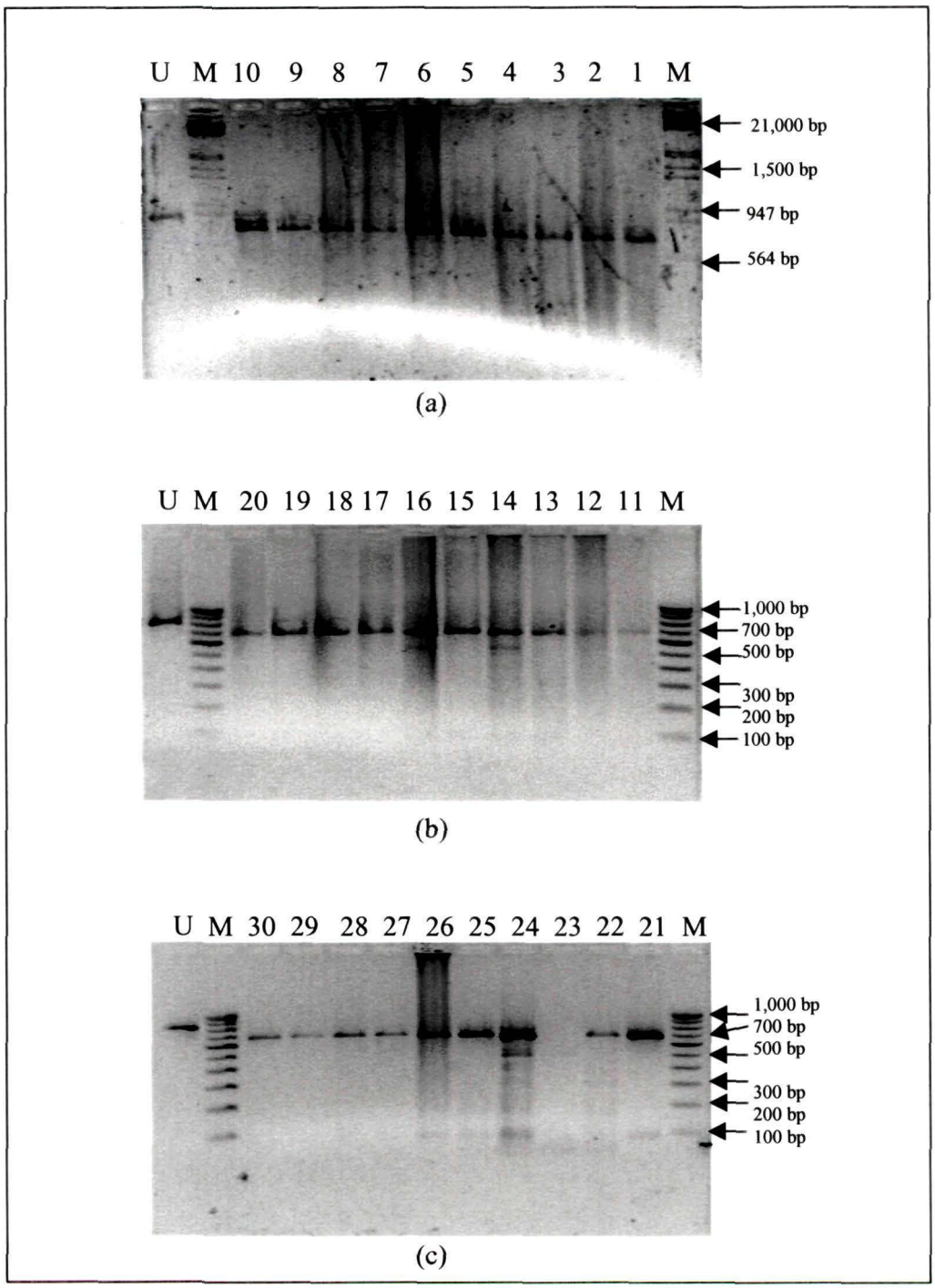


Fig. 24 (a-c): Amplicon Restriction Pattern (ARP) of morphotype-2 *Myrica* trees (Samples 1-30) with *Mbo*I.
(M= 100 bp ladder; U=undigested DNA)
{Fig a; M= λ DNA *Hind* III/ *Eco*R1 double digest marker}

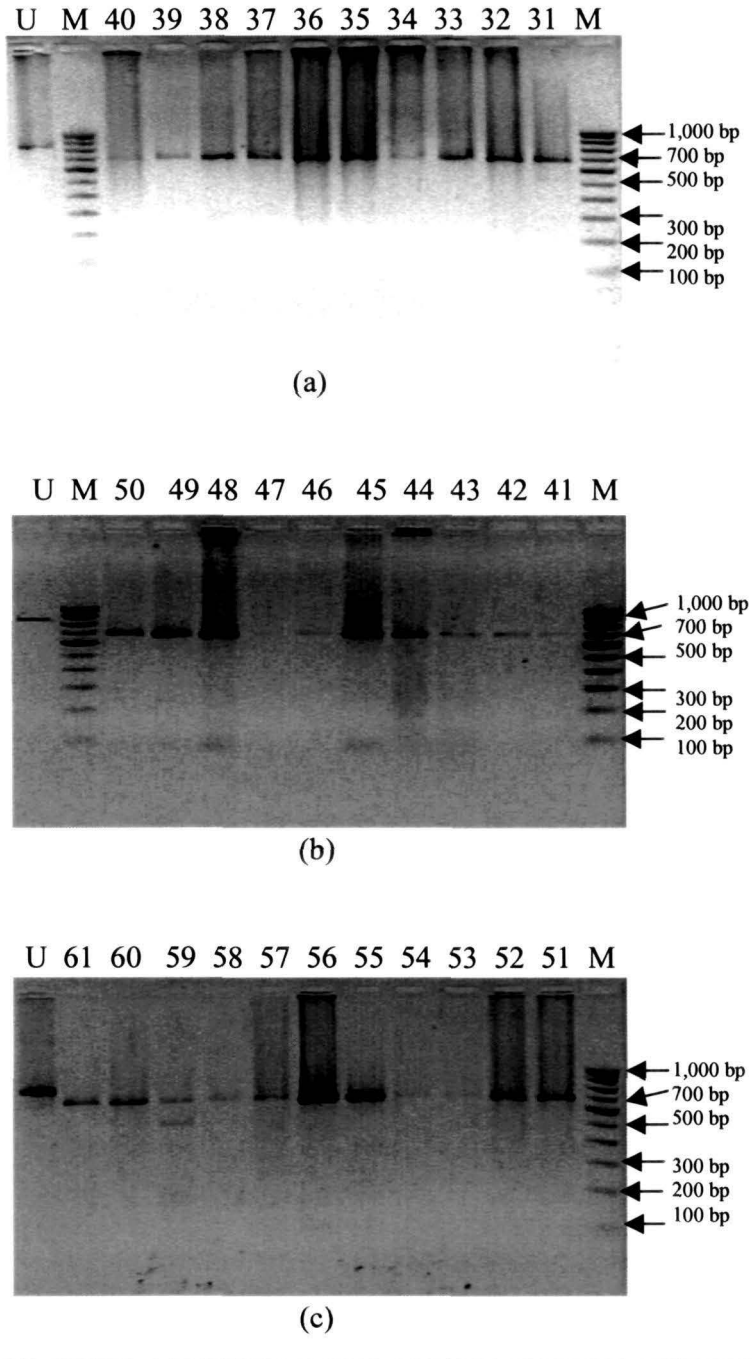


Fig. 25 (a-c): Amplicon Restriction Pattern (ARP) of morphotype-2 (samples 31-50) and morphotype-3 (Samples 51-61) *Myrica* trees with *Mbo*I.
(M= 100 bp ladder; U=Undigested DNA)

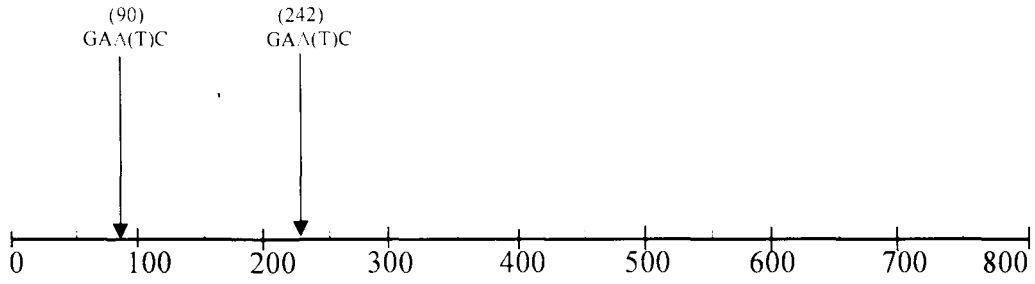


Fig. 4.26: Map of sequence ME1 for profile PM1 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

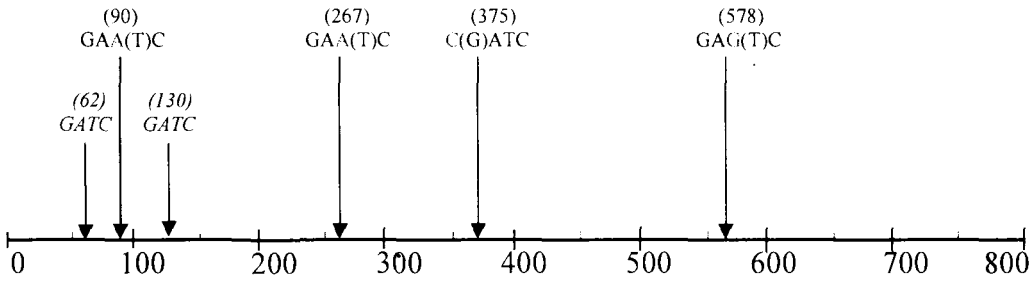


Fig. 4.27: Map of sequence ME1 for profile PM2 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

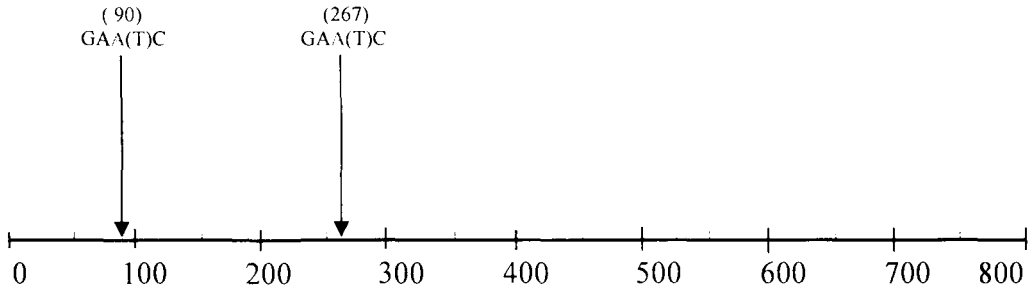


Fig. 4.28: Map of sequence ME1 for profile PM3 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

TABLE 4.8: ARP PROFILES WITH THE RESPECTIVE RESTRICTION FRAGMENTS USING *MBOI*

PROFILES FOR *MboI*	**SAMPLES**	**SIZE OF RESTRICTION FRAGMENTS (bp)**						
		~800	**~730**	**~700**	**~530**	**~380**	**~180**	**~90**
PM1	NPC-1; NPC-3; NPC-4; NPC-6; NPC-7; NPC-9	-	-	+	+	-	+	+
PM2	NPC-2	+	+	-	+	+	+	+
PM3	NPC-5; NPC-8; NPC-10; NPC-11; NPC-13; NPC-14; NPC-17; NPC-20; NPC-21; NPC-22; NPC-23; NPC-24; NPC-25; NPC-27; NPC-29; NPC-30; NPC-31; NPC-32; NPC-33; NPC-34; NPC-35; NPC-36; NPC-37; NPC-38; NPC-39; NPC-40; NPC-41; NPC-42; NPC-43; NPC-44; NPC-45; NPC-46; NPC-47; NPC-48; NPC-49; NPC-50	-	-	-	+	-	+	+
PM4	NPC-12; NPC-15; NPC-18	-	+	-	+	+	+	+
PM5	NPC-16; NPC-19; NPC-26; NPC-28	-	+	-	+	-	+	+
PM6	NRF-1; NRF-2; NRF- 3; NRF-4; NRF-5; NRF-6; NRF-7; NRF-8; NRF-9; NRF-11; NRF-12; NRF-13; NRF-15; NRF-17; NRF-18; NRF-19; NRF-20; NRF-21; NRF-22; NRF-23; NRF-25; NRF-26; NRF-27; NRF-28; NRF-29; NRF-30; NRF-31; NRF-32; NRF-33; NRF-34; NRF-35; NRF-36; NRF-37; NRF-38; NRF-39; NRF-40; NRF-41; NRF-42; NRF-43; NRF-44; NRF-45; NRF-46; NRF-47; NRF-48; NRF-49; NRF-50; NRF-51; NRF-52; NRF-53; NRF-54; NRF-55; NRF-56; NRF-57; NRF-58; NRF-60; NRF-61	-	-	+	-	-	-	+
PM7	NRF-10	+	+	+	-	-	-	+
PM8	NRF-14; NRF-16; NRF-59	-	-	+	+	-	-	+
PM9	NRF-24	-	-	+	+	+	+	+

Thirty six samples constituted the profile PM3 (Fig. 4.28) where three fragments of sizes ~530 bp, ~180 bp and ~90 bp were obtained. These three fragments were also common to all the other samples of NEHU campus trees in addition to additional fragments generated.

Three samples (NPC-12, NPC-15 and NPC-18) had the profile PM4 (Fig. 4.29) generating five restriction fragments of ~730 bp, ~530 bp, ~380 bp, ~180 bp and ~90 bp respectively.

Samples NPC-16, NPC-19, NPC-26 and NPC-28 had the profile PM5 (Fig. 4.30) where four fragments of ~730 bp, ~530 bp, ~180 bp and ~90 bp were observed.

Restriction digestion of morphotype-2 *Myrica* tree samples (NRF-1 to NRF-50) using enzyme *Mbo*I produced four profile patterns designated as PM6, PM7, PM8 and PM9.

Profile PM6 (Fig. 4.31) was the most commonly occurring pattern with two restriction fragments of ~700 bp and ~90 bp respectively. Forty six samples showed this profile.

Profile PM7 (Fig. 4.32) was observed for only one sample, NRF-10, in which an additional ~730 bp fragment and undigested band of ~800 bp were observed. The undigested band could have been the result of the presence of more than one copy of the gene in the genome of *Myrica* where the restriction site for this particular enzyme was absent in one of the copies.

Two samples, NRF-14 and NRF-16 showed the profile PM8 (Fig. 4.33) consisting of three restriction fragments of ~700 bp, ~530 bp and ~90 bp respectively.

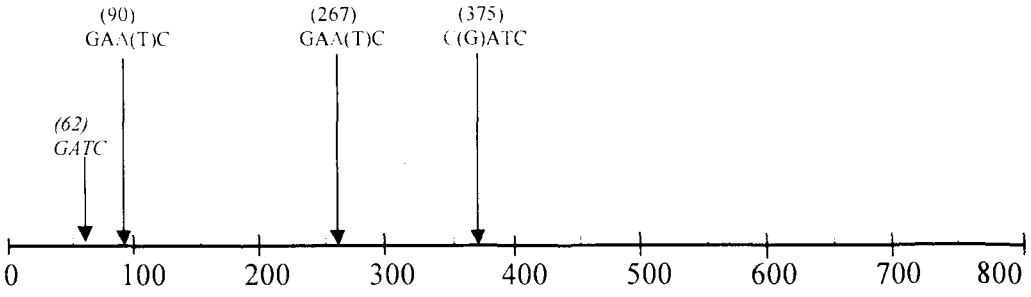


Fig. 4.29: Map of sequence ME1 for profile PM4 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

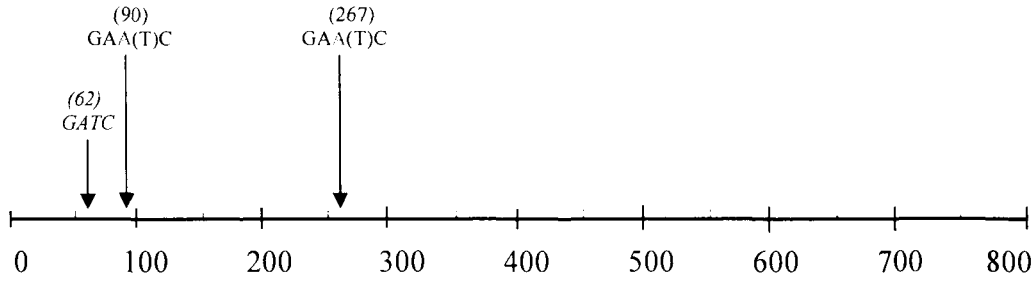


Fig. 4.30: Map of sequence ME1 for profile PM5 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

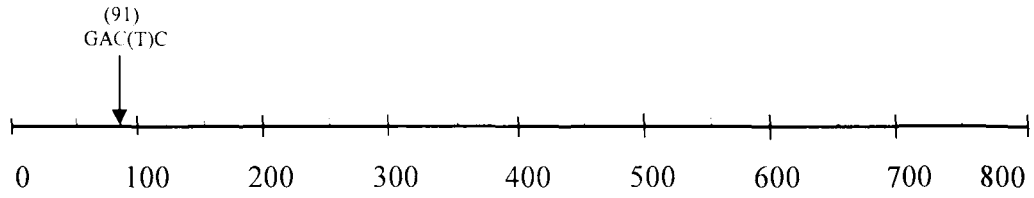


Fig. 4.31: Map of sequence ME2 for profile PM6 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

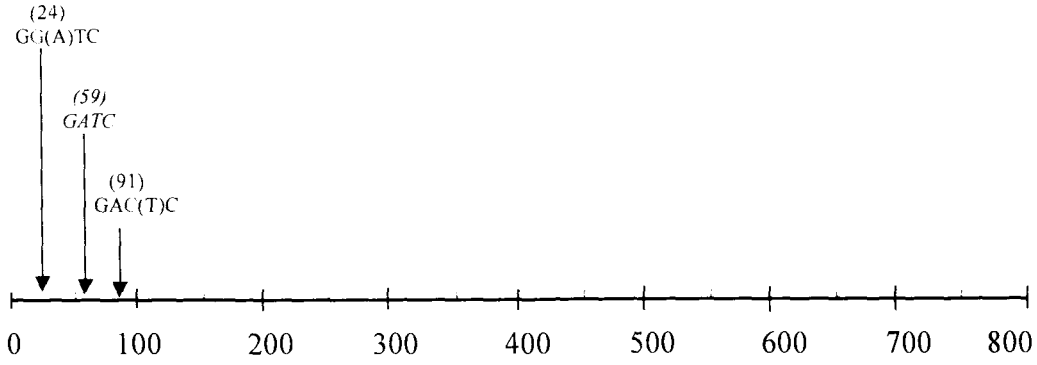


Fig. 4.32: Map of sequence ME2 for profile PM7 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

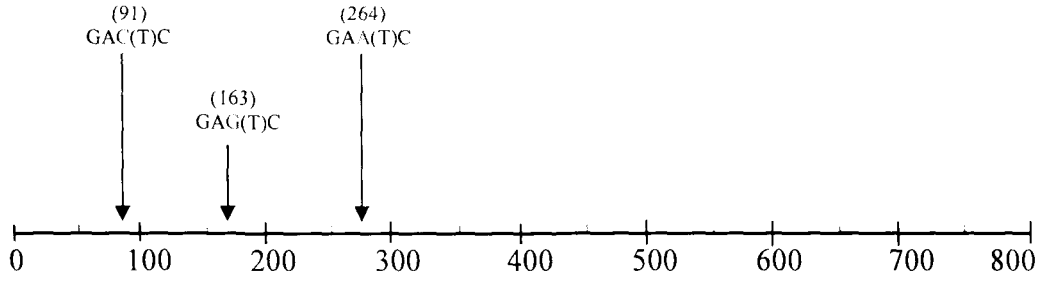


Fig. 4.33: Map of sequence ME2 for profile PM8 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

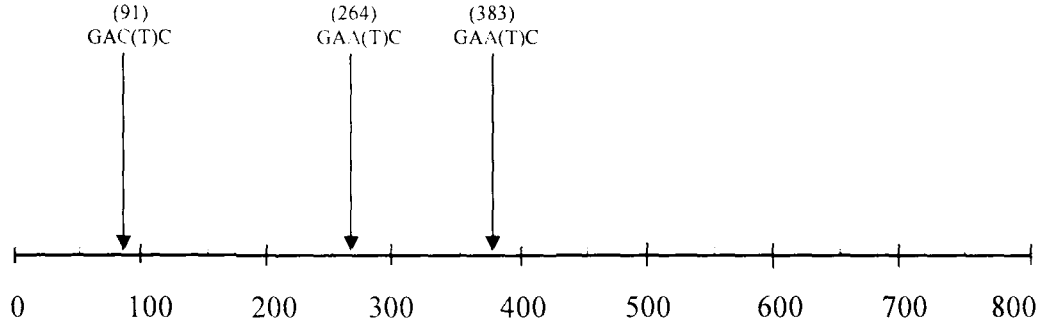
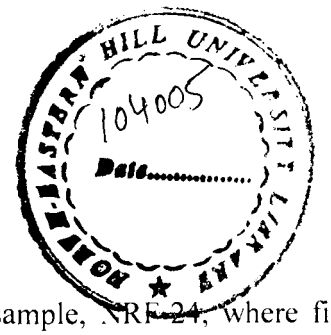


Fig. 4.34: Map of sequence ME2 for profile PM9 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)



Profile PM9 (Fig. 4.34) was found for only one sample, NRF-24, where five restriction fragments of ~700 bp, ~530 bp, ~380 bp, ~180 bp and ~90 bp were observed.

The morphotype-3 trees showed two profiles (PM6 and PM8) similar to those observed for morphotype-2 trees. Except for sample NRF-59 which showed profile PM8 all the other ten samples (NRF-51 to NRF-58, NRF-60 and NRF-61) showed profile PM6.

4.5.1.1 Restriction digestion analysis of sample ME1 using *Mbo*I;

Computer simulated restriction digestion analysis of the sequence ME1 with endonuclease *Mbo*I using the software DS Gene (version 1.1) revealed two cutting sites at position 62 and 130 producing three fragments of 62 bp, 78 bp and 654 bp. Fragments of ~530 bp, ~180 bp and ~90 bp were generated on actual digestion of our samples with *Mbo*I. This was common to all the morphotype-1 *Myrica* trees (Table 4.9 in page 105). Manual examination of the nucleotide sequences for all the three representative sample sequences revealed a number of sites found along the whole stretch where substitution of a single nucleotide base would have provided restriction sites for the enzyme thereby producing the fragments which we actually saw in the gel images. The ITS being highly variable because of its non-coding nature and high copy number, it is possible that single base substitutions (transition/ transversion) at certain positions could have occurred resulting in the patterns as observed in our gel images (Fig. 4.22, 4.23, 4.24 and 4.25 in page 92-95).

Based on the restriction fragment patterns generated using *Mbo*I for trees belonging to both the sites, nine different restriction profiles were obtained. The profiles were designated as PM1, PM2, PM3, PM4, PM5, PM6, PM7, PM8 and PM9.

Six samples (NPC-1, NPC-3, NPC-4, NPC-6, NPC-7 and NPC-9) of profile PM1 (Fig. 4.26 in page 96) showed four restriction fragments sizes of ~700 bp, ~530 bp, ~180 bp and ~90 bp which is explained as given below:

- In the first case transversion of purine base "A" with pyrimidine base "T" at position 90 downstream identified a restriction site which would generate 87 bp and 697 bp fragments.
- In the second case transversion of purine base "A" with pyrimidine base "T" at positions 90 and 242 downstream identified two restriction sites generating fragment sizes of 87 bp, 152 bp and 545 bp.

Profile PM2 (Fig. 4.27 in page 96) corresponded to only one sample, NPC-2, where five restriction fragments were obtained including the commonly occurring ~530 bp, ~180bp and ~90 bp fragments. The two additional bands corresponded to ~730 bp and ~380 bp fragments along with the undigested fragments of ~800 bp. This incomplete digestion could be because of the presence of more than one copy of the gene where the restriction site of *Mbo*I was absent in one of the copies resulting in the undigested amplicon.

- In the first case the presence of a single restriction site at position 62 downstream generated two fragments of 63 bp and 722 bp.

- In the second case transition of purine bases "G" with "C" at position 375 downstream and transversion of purine base "G" with pyrimidine base "T" at position 578 downstream would generate a 374 bp, 201 bp and 209 bp fragments.
- In the third case transversions of purine base "A" with pyrimidine base "T" at position 90 downstream and purine base "A" with pyrimidine base "T" at position 267 downstream would generate three fragments of sizes 87 bp, 177 bp and 520 bp respectively.

Thirty six samples constituted the profile PM3 (Fig. 4.28 in page 96) where three fragments of sizes ~530 bp, ~180 bp and ~90 bp were obtained. These three fragments were also common to all the other samples of the NEHU campus trees in addition to additional fragments generated.

- Transversions of purine base "A" with pyrimidine base "T" at position 90 and purine base "A" with pyrimidine base "T" at position 267 generated three fragments of sizes 87 bp, 177 bp and 520 bp.

Three samples (NPC-12, NPC-15 and NPC-18) constituted the profile PM4 (Fig. 4.29 in page 99) generating five restriction fragments of ~730 bp, ~530 bp, ~380 bp, ~180 bp and ~90 bp size.

- In the first case the presence of a single restriction site at position 62 downstream generated two fragments of 63 bp and 722 bp respectively.
- In the second case transversions of purine base "A" with pyrimidine base "T" at position 90 downstream and purine base "A" with pyrimidine base "T" with at

position 267 downstream would produce three fragments of 87 bp, 177 bp and 520 bp fragments respectively.

- In the third case transversion of purine base "A" with pyrimidine base "C" at position 375 downstream would generate two fragments of 374 bp and 410 bp which together would have merged into the ~380 bp fragment and appeared as one single band of ~380 bp fragment.

Samples NPC-16, NPC-19, NPC-26 and NPC-28 constituted the profile PM5 (Fig. 4.30 in page 99) where four digestion fragments of sizes ~730 bp, ~530 bp, ~180 bp and ~90 bp were observed.

- In the first case the presence of a single restriction site at position 62 downstream generated two fragments of 63 bp and 722 bp which may have appeared as the ~90 bp and and ~730 bp fragments respectively.
- In the second case transversions of purine base "A" with pyrimidine base "T" at position 90 downstream and purine base "A" with pyrimidine base "T" with at position 267 downstream would produce three fragments of 87 bp, 177 bp and 520 bp respectively.

4.5.1.2 Restriction digestion analysis of sample ME2 using *Mbo*I;

Restriction digestion analysis of morphotype-2 *Myrica* tree samples at Nongkrem forest site (NRF-1 to NRF-50) using *Mbo*I produced four profile patterns designated as PM6, PM7, PM8 and PM9. Only 780 bp could be sequenced for the sample ME2.

TABLE 4.9: RESTRICTION FRAGMENTS OF MORPHOTYPE-1 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE APPROXIMATE BAND SIZES USING *MBO*I. (+ = BAND PRESENT; - = BAND ABSENT)

	~800	~730	~700	~530	~380	~180	~90
NPC1	-	-	+	+	-	+	+
NPC2	+	+	-	+	+	+	+
NPC3	-	-	+	+	-	+	+
NPC4	-	-	+	+	-	+	+
NPC5	-	-	-	+	-	+	+
NPC6	-	-	+	+	-	+	+
NPC7	-	-	+	+	-	+	+
NPC8	-	-	-	+	-	+	+
NPC9	-	-	+	+	-	+	+
NPC10	-	-	-	+	-	+	+
NPC11	-	-	-	+	-	+	+
NPC12	-	+	-	+	+	+	+
NPC13	-	-	-	+	-	+	+
NPC14	-	-	-	+	-	+	+
NPC15	-	+	-	+	+	+	+
NPC16	-	+	-	+	-	+	+
NPC17	-	-	-	+	-	+	+
NPC18	-	+	-	+	+	+	+
NPC19	-	+	-	+	-	+	+
NPC20	-	-	-	+	-	+	+
NPC21	-	-	-	+	-	+	+
NPC22	-	-	-	+	-	+	+
NPC23	-	-	-	+	-	+	+
NPC24	-	-	-	+	-	+	+
NPC25	-	-	-	+	-	+	+
NPC26	-	+	-	+	-	+	+
NPC27	-	-	-	+	-	+	+
NPC28	-	+	-	+	-	+	+
NPC29	-	-	-	+	-	+	+
NPC30	-	-	-	+	-	+	+
NPC31	-	-	-	+	-	+	+
NPC32	-	-	-	+	-	+	+
NPC33	-	-	-	+	-	+	+
NPC34	-	-	-	+	-	+	+
NPC35	-	-	-	+	-	+	+
NPC36	-	-	-	+	-	+	+
NPC37	-	-	-	+	-	+	+
NPC38	-	-	-	+	-	+	+
NPC39	-	-	-	+	-	+	+
NPC40	-	-	-	+	-	+	+
NPC41	-	-	-	+	-	+	+
NPC42	-	-	-	+	-	+	+
NPC43	-	-	-	+	-	+	+
NPC44	-	-	-	+	-	+	+
NPC45	-	-	-	+	-	+	+
NPC46	-	-	-	+	-	+	+
NPC47	-	-	-	+	-	+	+
NPC48	-	-	-	+	-	+	+
NPC49	-	-	-	+	-	+	+
NPC50	-	-	-	+	-	+	+

TABLE 4.10: RESTRICTION FRAGMENTS OF MORPHOTYPE-2 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE APPROXIMATE BAND SIZES USING *MBO*I. (+ = BAND PRESENT; - = BAND ABSENT)

	~800	~730	~700	~530	~380	~180	~90
NRF1	-	-	+	-	-	-	+
NRF2	-	-	+	-	-	-	+
NRF3	-	-	+	-	-	-	+
NRF4	-	-	+	-	-	-	+
NRF5	-	-	+	-	-	-	+
NRF6	-	-	+	-	-	-	+
NRF7	-	-	+	-	-	-	+
NRF8	-	-	+	-	-	-	+
NRF9	-	-	+	-	-	-	+
NRF10	+	+	+	-	-	-	+
NRF11	-	-	+	-	-	-	+
NRF12	-	-	+	-	-	-	+
NRF13	-	-	+	-	-	-	+
NRF14	-	-	+	+	-	-	+
NRF15	-	-	+	-	-	-	+
NRF16	-	-	+	+	-	-	+
NRF17	-	-	+	-	-	-	+
NRF18	-	-	+	-	-	-	+
NRF19	-	-	+	-	-	-	+
NRF20	-	-	+	-	-	-	+
NRF21	-	-	+	-	-	-	+
NRF22	-	-	+	-	-	-	+
NRF23	-	-	+	-	-	-	+
NRF24	-	-	+	+	+	+	+
NRF25	-	-	+	-	-	-	+
NRF26	-	-	+	-	-	-	+
NRF27	-	-	+	-	-	-	+
NRF28	-	-	+	-	-	-	+
NRF29	-	-	+	-	-	-	+
NRF30	-	-	+	-	-	-	+
NRF31	-	-	+	-	-	-	+
NRF32	-	-	+	-	-	-	+
NRF33	-	-	+	-	-	-	+
NRF34	-	-	+	-	-	-	+
NRF35	-	-	+	-	-	-	+
NRF36	-	-	+	-	-	-	+
NRF37	-	-	+	-	-	-	+
NRF38	-	-	+	-	-	-	+
NRF39	-	-	+	-	-	-	+
NRF40	-	-	+	-	-	-	+
NRF41	-	-	+	-	-	-	+
NRF42	-	-	+	-	-	-	+
NRF43	-	-	+	-	-	-	+
NRF44	-	-	+	-	-	-	+
NRF45	-	-	+	-	-	-	+
NRF46	-	-	+	-	-	-	+
NRF47	-	-	+	-	-	-	+
NRF48	-	-	+	-	-	-	+
NRF49	-	-	+	-	-	-	+
NRF50	-	-	+	-	-	-	+

TABLE 4.11: RESTRICTION FRAGMENTS OF MORPHOTYPE-3 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE APPROXIMATE BAND SIZES USING *MBO*I. (+ = BAND PRESENT; - = BAND ABSENT)

	~800	~730	~700	~530	~380	~180	~90
NRF51	-	-	+	-	-	-	+
NRF52	-	-	+	-	-	-	+
NRF53	-	-	+	-	-	-	+
NRF54	-	-	+	-	-	-	+
NRF55	-	-	+	-	-	-	+
NRF56	-	-	+	-	-	-	+
NRF57	-	-	+	-	-	-	+
NRF58	-	-	+	-	-	-	+
NRF59	-	-	+	+	-	-	+
NRF60	-	-	+	-	-	-	+
NRF61	-	-	+	-	-	-	+

Computer simulated restriction digestion analysis of the sequence ME2 with *Mbo*I using the software DS Gene (version 1.1) revealed two restriction sites at position 59 and 127 generating three fragments of 59 bp, 68 bp and 653 bp respectively. Agarose gel images from our experiment with the same enzyme revealed two common bands of ~700 bp and ~90 bp fragments in all the samples. However, some samples showed additional fragments in addition to these three common bands (Table 4.10 in page 106).

Profile PM6 (Fig. 4.31 in page 99) was the most commonly occurring pattern with two restriction fragments of ~700 bp and ~90 bp. Forty six samples constituted this profile. Fragments normally ranging from 100 bp to 20 kbp are resolved with 1% agarose gel electrophoresis. Assuming that there was transition at position 91 where base "C" got replaced with "T". This would generate two fragments of sizes 88 bp and 692 bp respectively.

Profile PM7 (Fig. 4.32 in page 100) was shown by only one sample, NRF-10. The undigested band could have been the result of the presence of more than one copy of the gene where the restriction site for this particular enzyme was absent in one of the copies. The presence of additional fragments may be explained as given below;

> Assuming that transversion occurred at position 24 where the pyrimidine base "G" was replaced by "A" three fragments of sizes 22 bp, 37 bp and 721 bp were generated. In this case the 721 bp fragment may have appeared as the ~730 bp band whereas the two smaller fragments of 22 bp and 37 bp may not have been resolved due to their small sizes.

- Assuming that there was transition at position 91 in which pyrimidine base "C" got replaced with "T". This would have generated two fragments of sizes 88 bp and 692 bp respectively.

Two samples, NRF-14 and NRF-16 constituted the profile PM8 (Fig. 4.33 in page 100) consisting of three restriction fragments of ~700 bp, ~530 bp and ~90 bp.

- In the first case a transition of "C" with "T" at position 91 would generate two fragments of sizes 88 bp and 692 bp respectively.
- In the second case transition of "C" with "T" at position 91, transversion of "G" with "T" at position 163 and transversion of "A" with "T" at position 264 in one of the copies of the gene would produce four fragments of nucleotide sizes 89 bp, 72 bp, 101 bp and 519 bp. The 89 bp, 72 bp, 101 bp fragments may have merged together and appeared as the ~90 bp fragment and the 519 bp fragment may have appeared as the ~530 bp band.

Profile PM9 (Fig. 4.34 in page 100) constituted only one sample, NRF-24, where five restriction fragments of ~700 bp, ~530 bp, ~380 bp , ~180 bp and ~90 bp were observed.

- In the first case transition of "C" with "T" at position 91 would generate two fragments of 88 bp and 692 bp respectively.
- In the second case transition of "C" with "T" at position 91 and transversion of "A" with "T" at position 264 would generate three fragments of sizes 88 bp, 173 bp and 519 bp respectively.

- In the third case a single transversion of "A" with "T" at position 383 would produce two fragments of sizes 380 bp and 400 bp which may merge to appear as the ~380 bp band.

4.5.1.3 Restriction digestion analysis of sample ME3 using *Mbo*I;

Computer simulated restriction digestion analysis of sequenced sample ME3 representing the morphotype-3 trees at Nongkrem forest showed two similar profiles (PM6 & PM8) as observed in ME2. Except for sample NRF-59 which showed profile PM8 all the other 10 samples (NRF-51 to NRF-58, NRF-60 and NRF-61) showed profile PM6 (Table 4.11 in page 107).

Profile PM6 (Fig. 4.35) consisted of two restriction fragments of sizes ~90 bp and ~700 bp.

- Transition of "C" with "T" at position 90 would generate two fragments of sizes 87 bp and 703 bp respectively.

Sample NRF-59 constituted the profile PM8 (Fig. 4.36) in which three fragments of ~700 bp, ~530 bp and ~90 bp were observed.

- In the first case transition of "C" with "T" at position 90 would generate two fragments of 87 bp and 703 bp respectively.
- In the second case presence of a restriction site at position 126 and transversion of "A" with "T" at position 238 would generate three fragments of size 126 bp, 109 bp and 555 bp. The 126 bp and 109 bp fragments may merge together into ~90 bp fragment.

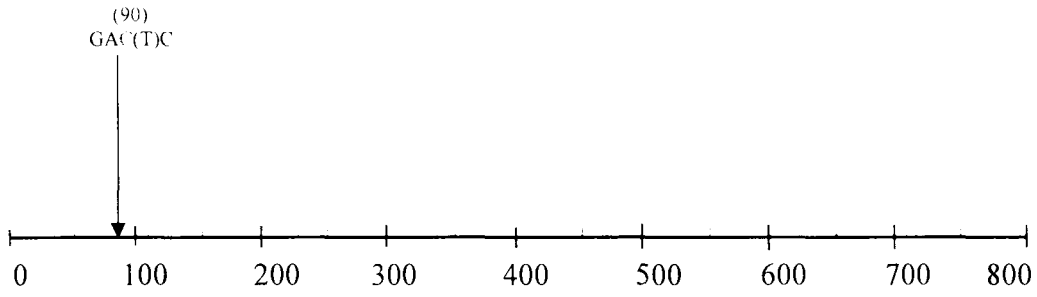


Fig. 4.35: Map of sequence ME3 for profile PM6 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

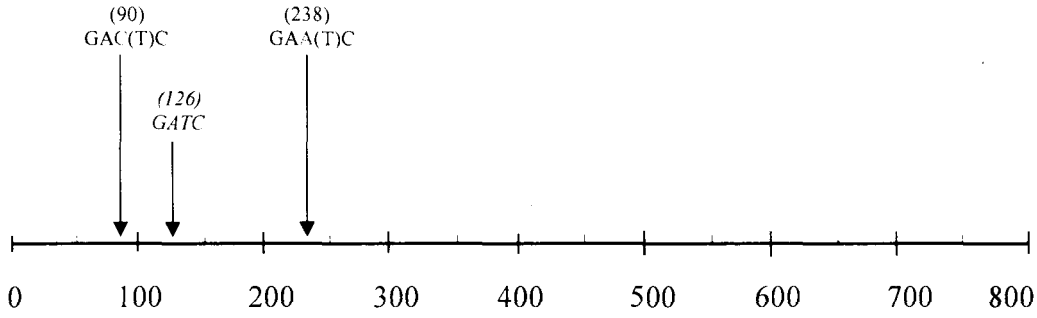


Fig. 4.36: Map of sequence ME3 for profile PM8 showing the possible alternate restriction sites using *Mbo*I. (Distances not to scale)

4.5.2 Amplicon Restriction Patterns (ARPs) with *Sau*96I:

Restriction enzyme *Sau*96I obtained from the bacterium *Staphylococcus aureus* is a five base cutter that cuts the site "GGnCC" slicing between the two purine bases G and G. Altogether seven different profiles were generated when this enzyme was used to digest the amplicons of the sample trees studied. Interestingly, amplicons of morphotype-1 trees at NEHU campus did not undergo complete digestion in all the samples showing incompletely digested bands even where digested fragments were visible (Fig. 4.37, 4.38, 4.39 & 4.40). This could be because of the heterozygous condition or presence of more than one copy where the restriction site for the enzyme *Sau*96I was absent in one of the copies.

Based on the restriction fragments obtained from the restriction digestion of all the samples using *Sau*96I, seven different profiles were detected namely, PS0, PS1, PS2, PS3, PS4, PS5 and PS6 (Table 4.12). Restriction fragment patterns generated produced profiles which were morphotype specific. Profiles PS1 and PS0 were confined to morphotype-1 trees. The other profiles PS2, PS3, PS4, PS5 and PS6 were found only for morphotype-2 and morphotype-3 trees. Based on the morphotype specificity of the restriction digestion fragment profiles, we could differentiate between morphotype-1 trees at NEHU campus from the morphotype-2 and morphotype-3 trees at Nongkrem forest. However, no marked polymorphism in the restriction banding pattern between the sequenced samples ME2 and ME3 of Nongkrem forest was observed.

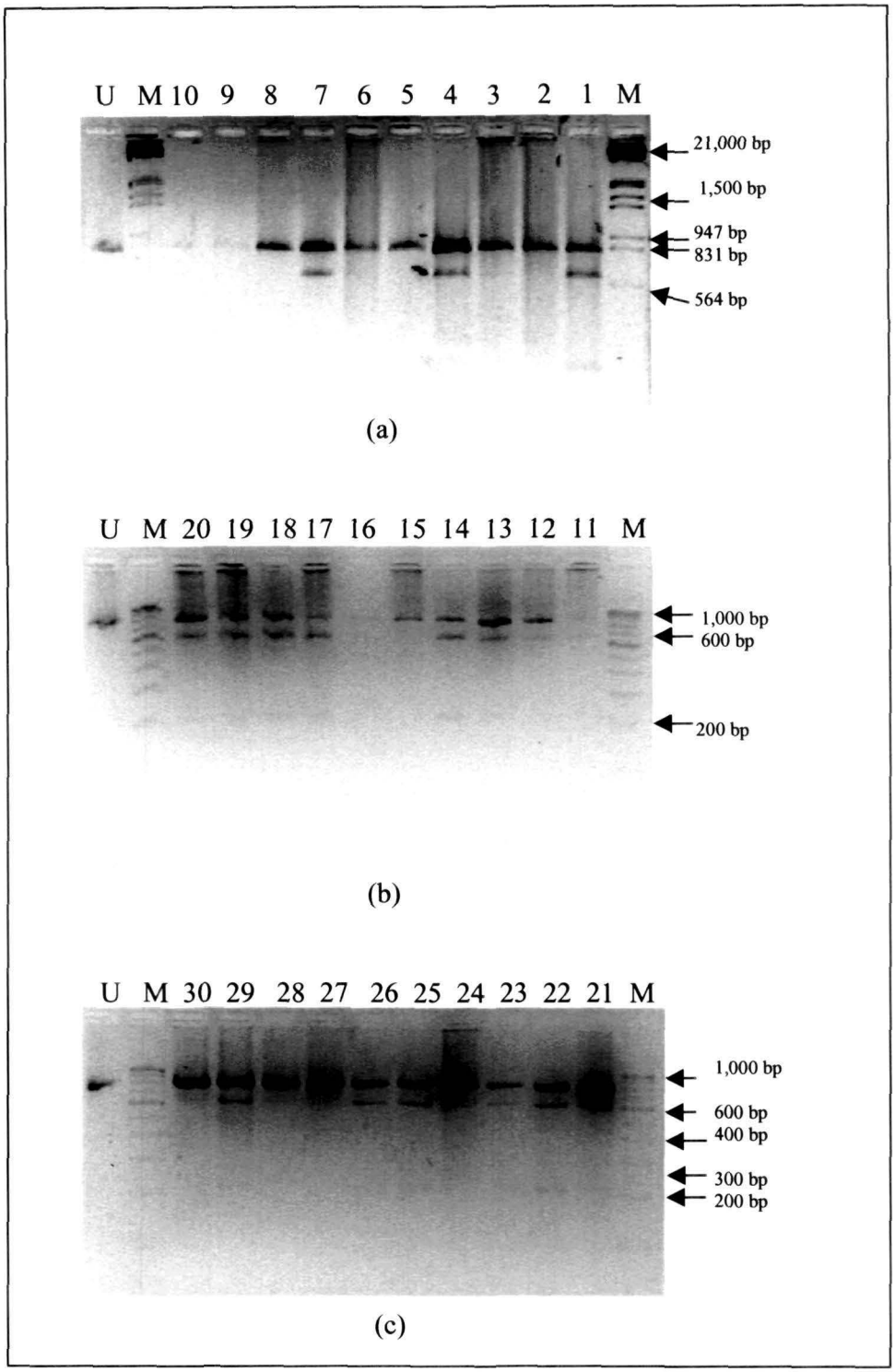


Fig. 4.37 (a-c): Amplicon Restriction Pattern (ARP) of morphotype-1 *Myrica* trees (Samples 1-30) using S*au*96I.
(Fig. b &c; M= 100 bp ladder; U= Undigested DNA)
{Fig a; M= λ DNA *Hind* III/ *Eco*R1 double digest marker}

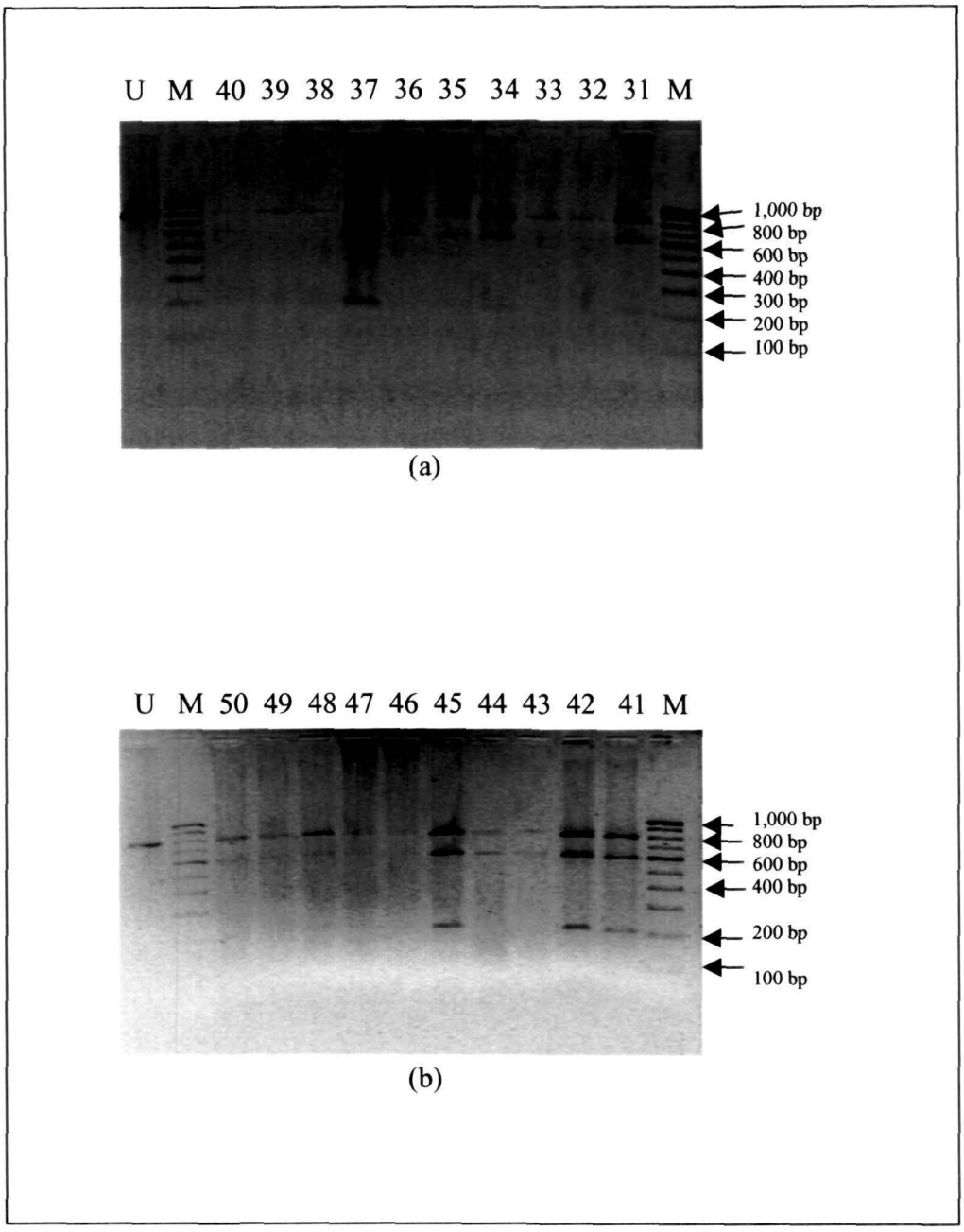


Fig. 4.38 (a&b): Amplicon Restriction Pattern (ARP) of morphotype-1 *Myrica* trees (Samples 31-50) using *Sau*96I. (M= 100 bp ladder; U= Undigested DNA)

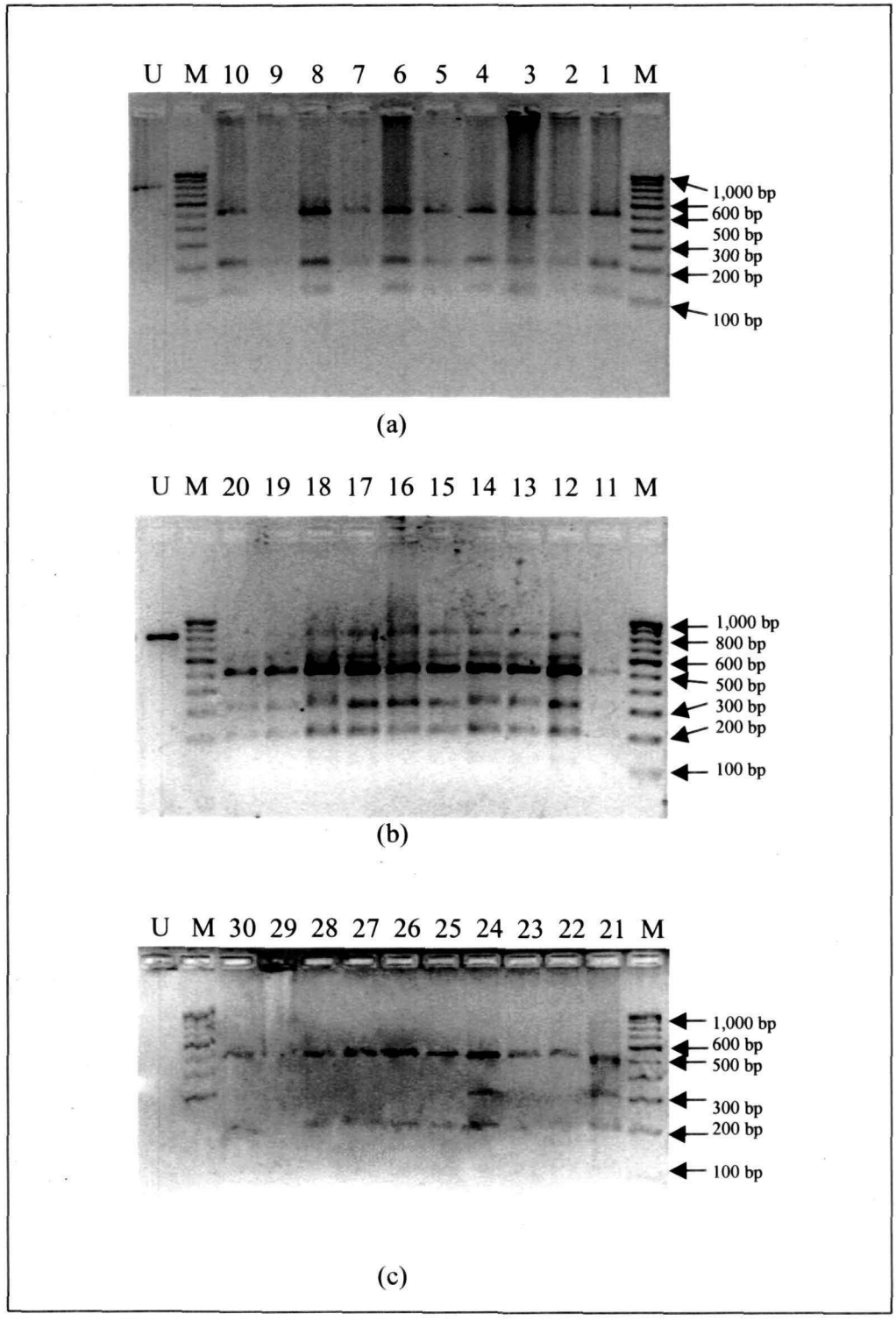


Fig. 4.39 (a-c): Amplicon Restriction Pattern (ARP) of morphotype-2 *Myrica* trees (Samples 1-30) using *Sau*96I.
(M= 100 bp ladder; U= Undigested DNA)

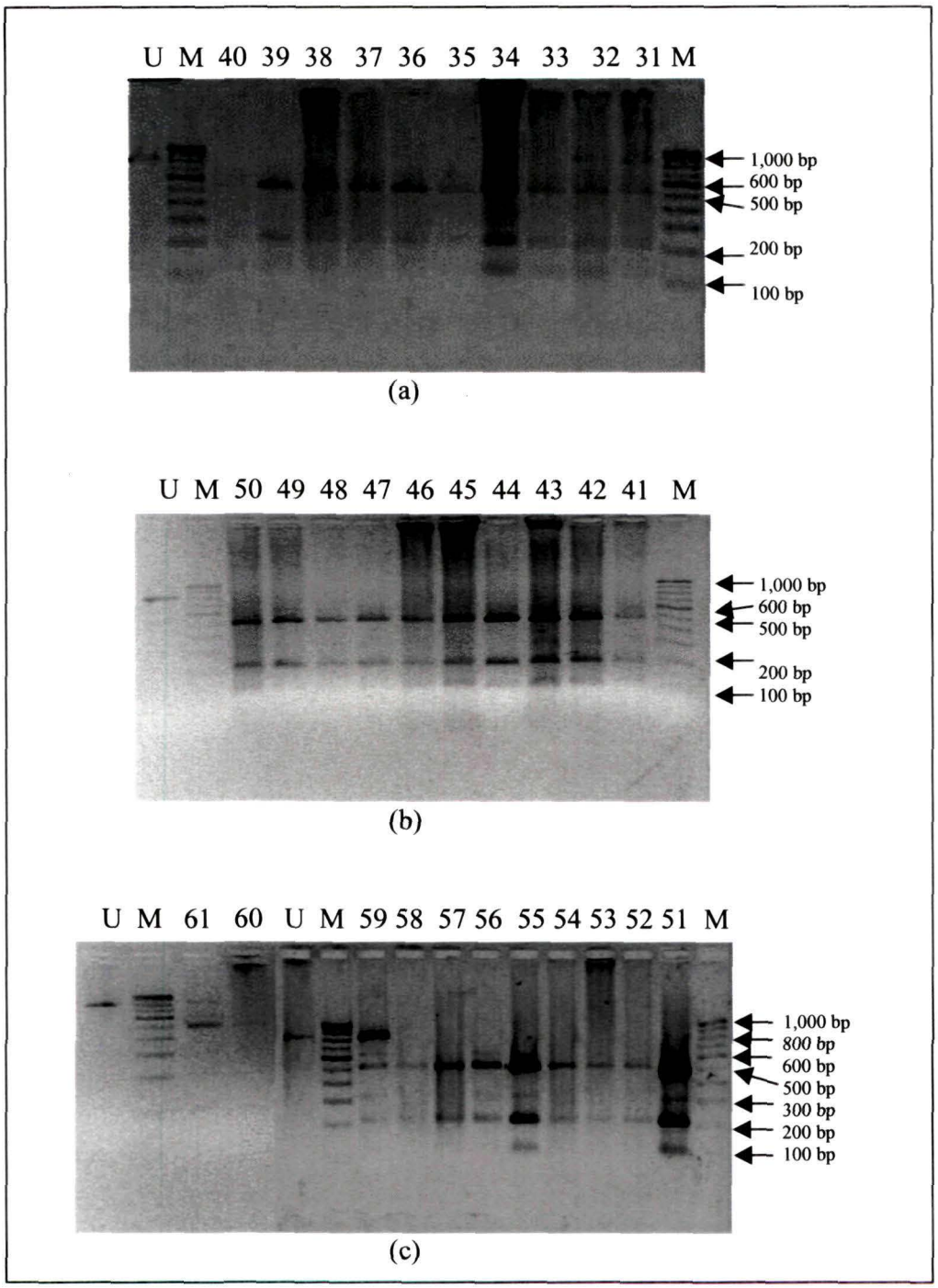


Fig. 4.40 (a-c): Amplicon Restriction Pattern (ARP) of morphotype-2 (Samples 31-50) and morphotype-3 (Samples 51-61) *Myrica* trees using *Sau*96I.
(M= 100 bp ladder; U= Undigested DNA)

TABLE 4.12: ARP PROFILES WITH THEIR RESPECTIVE RESTRICTION FRAGMENTS USING *SAU*96I (ND = NO DIGESTION)

PROFILES FOR *Sau*96I	**SAMPLES**	**SIZE OF RESTRICTION FRAGMENTS (bp)**					
		~800	**~600**	**~500**	**~310**	**~200**	**~100**
PS0 (ND)	NPC-5; NPC-6; NPC-8; NPC-9; NPC-10	+	-	-	-	-	-
PS1	NPC-1; NPC-2; NPC-3; NPC-4; NPC-7; NPC-11; NPC-12; NPC-13; NPC-14; NPC-15; NPC-16; NPC-17; NPC-18; NPC-19; NPC-20; NPC-21; NPC-22; NPC-23; NPC-24; NPC-25; NPC-26; NPC-27; NPC-28; NPC-29; NPC-30; NPC-31; NPC-32; NPC-33; NPC-34; NPC-35; NPC-36; NPC-37; NPC-38; NPC-39; NPC-40; NPC-41; NPC-42; NPC-43; NPC-44; NPC-45; NPC-46; NPC-47; NPC-48; NPC-49; NPC-50	+	+	-	-	+	-
PS2	NRF-1; NRF-2; NRF-3; NRF-4; NRF-5; NRF-6; NRF-7; NRF-8; NRF-9; NRF-10; NRF-11; NRF-22; NRF-23; NRF-25; NRF-26; NRF-27; NRF-27; NRF-29; NRF-30; NRF-33; NRF-35; NRF-36; NRF-37; NRF-38; NRF-39; NRF-40; NRF-41; NRF-44; NRF-45; NRF-46; NRF-47; NRF-48; NRF-52; NRF-53; NRF-54; NRF-57; NRF-58; NRF-60	-	-	+	-	+	+
PS3	NRF-12; NRF-13; NRF-14; NRF-15; NRF-16; NRF-17; NRF-18; NRF-19; NRF-20; NRF-34; NRF-59	+	+	+	+	+	+
PS4	NRF-21; NRF-24; NRF-42; NRF-43; NRF-49; NRF-50; NRF-51; NRF-55;	-	-	+	+	+	+
PS5	NRF-31; NRF-32; NRF-61	+	-	+	-	+	+
PS6	NRF-56	-	+	+	+	+	+

4.5.2.1 Restriction digestion analysis of sample ME1 using *Sau*96I;

The sequence ME1 was a representative sequence for the morphotype-1 *Myrica* trees belonging to NEHU permanent campus. Computer simulated restriction digestion analysis of the sequence ME1 with *Sau*96I using the software DS Gene (version 1.1) failed to locate even one restriction site in the whole stretch of the sequence. However, manual check of the sequence base by base revealed a number of sites where the substitution of one nucleotide base could have produced restriction sites for the enzyme. Agarose gel electrophoresis of samples NPC-1 to NPC-50 revealed that restriction fragments could not be obtained in five samples (NPC-5, NPC-6, NPC-8, NPC-9 and NPC-10). Other samples showed similar restriction fragment pattern revealing two bands of sizes ~600 bp and ~200 bp. The undigested ~800 bp band was always found even where digested bands were seen (Table 4.13). Assuming that transition of "A" to "G" took place at position 606 (Fig. 4.41). This could create a restriction site for the enzyme *Sau*96I generating two restriction fragments of sizes 606 bp and 178 bp. Therefore, two restriction profiles were detected. The profile PS1 included samples where ~800 bp, ~600 bp and ~200 bp fragments were present. Profile PS0 included samples where no digestion was detected.

4.5.2.2 Restriction digestion analysis of sample ME2 using *Sau*96I;

Sequenced sample ME2 was the representative for the morphotype-2 *Myrica* trees (NRF-1 to NRF-50) at Nongkrem forest. Computer simulated restriction digestion analysis of the sequence ME2 with the restriction enzyme *Sau*96I using the software DS

Gene (version 1.1) identified two restriction sites at positions 183 and 289 (Fig. 4.42) generating three restriction fragments of sizes 183 bp, 106 bp and 491 bp (Table 4.14).

Four different profiles namely PS2, PS3, PS4 and PS5 were obtained when we performed restriction digestion of the amplicons of sample trees (NRF-1 to NRF-50) at Nongkrem forest. Thirty two samples constituted Profile PS2 where three restriction fragments of sizes ~510 bp, ~210 bp and ~110 bp were generated. Detection of restriction sites at position 183 and 289 generated three fragments of 183 bp, 106 bp and 491 bp which may have appeared as ~200 bp, ~100 bp and ~500 bp bands respectively in the agarose gel images. Since the resolving power of agarose gels is limited, it is likely that bands of almost similar sizes separated by nucleotide bases less than ~30 bp may have appeared as one single band.

Profile PS3 consisted of ten samples where five restriction fragments, ~600 bp, ~500 bp, ~310 bp, ~200 bp and ~100 bp were detected along with the incompletely digested ~800 bp band. The fragments ~500 bp, ~200 bp and ~100 bp were obtained as was discussed above for profile PS3. As for the other two bands, we assumed that if one of the copies of the gene had a single restriction site at position 183 then two fragments of sizes 183 bp and 598 bp would be produced. The other copy had a single restriction site at position 289 that generated two fragments of sizes 289 bp and 491 bp.

Samples NRF-21, NRF-24, NRF-42, NRF-43, NRF-49 and NRF-50 constituted the profile PS4 in which four restriction fragments of sizes ~500 bp, ~310 bp, ~200 bp and ~100 bp were obtained. The fragments ~500 bp, ~200 bp and ~100 bp corresponded to the restriction sites at position 183 and 289 that generated restriction fragments of

TABLE 4.13: RESTRICTION FRAGMENTS OF MORPHOTYPE-1 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE APPROXIMATE BAND SIZES USING *SAU*96I. (+ = BAND PRESENT; - = BAND ABSENT)

	~800	~600	~500	~310	~200	~100
NPC1	+	+	-	-	+	-
NPC2	+	+	-	-	+	-
NPC3	+	+	-	-	+	-
NPC4	+	+	-	-	+	-
NPC5	+	-	-	-	-	-
NPC6	+	-	-	-	-	-
NPC7	+	+	-	-	+	-
NPC8	+	-	-	-	-	-
NPC9	+	-	-	-	-	-
NPC10	+	-	-	-	-	-
NPC11	+	+	-	-	+	-
NPC12	+	+	-	-	+	-
NPC13	+	+	-	-	+	-
NPC14	+	+	-	-	+	-
NPC15	+	+	-	-	+	-
NPC16	+	+	-	-	+	-
NPC17	+	+	-	-	+	-
NPC18	+	+	-	-	+	-
NPC19	+	+	-	-	+	-
NPC20	+	+	-	-	+	-
NPC21	+	+	-	-	+	-
NPC22	+	+	-	-	+	-
NPC23	+	+	-	-	+	-
NPC24	+	+	-	-	+	-
NPC25	+	+	-	-	+	-
NPC26	+	+	-	-	+	-
NPC27	+	+	-	-	+	-
NPC28	+	+	-	-	+	-
NPC29	+	+	-	-	+	-
NPC30	+	+	-	-	+	-
NPC31	+	+	-	-	+	-
NPC32	+	+	-	-	+	-
NPC33	+	+	-	-	+	-
NPC34	+	+	-	-	+	-
NPC35	+	+	-	-	+	-
NPC36	+	+	-	-	+	-
NPC37	+	+	-	-	+	-
NPC38	+	+	-	-	+	-
NPC39	+	+	-	-	+	-
NPC40	+	+	-	-	+	-
NPC41	+	+	-	-	+	-
NPC42	+	+	-	-	+	-
NPC43	+	+	-	-	+	-
NPC44	+	+	-	-	+	-
NPC45	+	+	-	-	+	-
NPC46	+	+	-	-	+	-
NPC47	+	+	-	-	+	-
NPC48	+	+	-	-	+	-
NPC49	+	+	-	-	+	-
NPC50	+	+	-	-	+	-

TABLE 4.14: RESTRICTION FRAGMENTS OF MORPHOTYPE-2 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE APPROXIMATE BAND SIZES USING *SAU*96I. (+ = BAND PRESENT; - = BAND ABSENT)

	~800	~600	~500	~310	~200	~100
NRF1	-	-	+	-	+	+
NRF2	-	-	+	-	+	+
NPC3	-	-	+	-	+	+
NRF4	-	-	+	-	+	+
NRF5	-	-	+	-	+	+
NRF6	-	-	+	-	+	+
NRF7	-	-	+	-	+	+
NRF8	-	-	+	-	+	+
NRF9	-	-	+	-	+	+
NRF10	-	-	+	-	+	+
NRF11	-	-	+	-	+	+
NRF12	+	+	+	+	+	+
NRF13	+	+	+	+	+	+
NRF14	+	+	+	+	+	+
NRF15	+	+	+	+	+	+
NRF16	+	+	+	+	+	+
NRF17	+	+	+	+	+	+
NRF18	+	+	+	+	+	+
NRF19	+	+	+	+	+	+
NRF20	+	+	+	+	+	+
NRF21	-	-	+	+	+	+
NRF22	-	-	+	-	+	+
NRF23	-	-	+	-	+	+
NRF24	-	-	+	+	+	+
NRF25	-	-	+	-	+	+
NRF26	-	-	+	-	+	+
NRF27	-	-	+	-	+	+
NRF28	-	-	+	-	+	+
NRF29	-	-	+	-	+	+
NRF30	-	-	+	-	+	+
NRF31	+	-	+	-	+	+
NRF32	+	-	+	-	+	+
NRF33	-	-	+	-	+	+
NRF34	+	+	+	+	+	+
NRF35	-	-	+	-	+	+
NRF36	-	-	+	-	+	+
NRF37	-	-	+	-	+	+
NRF38	-	-	+	-	+	+
NRF39	-	-	+	-	+	+
NRF40	-	-	+	-	+	+
NRF41	-	-	+	-	+	+
NRF42	-	-	+	+	+	+
NRF43	-	-	+	+	+	+
NRF44	-	-	+	-	+	+
NRF45	-	-	+	-	+	+
NRF46	-	-	+	-	+	+
NRF47	-	-	+	-	+	-
NRF48	-	-	+	-	+	-
NRF49	-	-	+	-	+	+
NRF50	-	-	+	-	+	-

TABLE 4.15: RESTRICTION FRAGMENTS OF MORPHOTYPE-3 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE APPROXIMATE BAND SIZES USING *SAU*96I. (+ = BAND PRESENT; - = BAND ABSENT)

	~800	~600	~500	~310	~200	~100
NRF 51	-	-	+	+	+	+
NRF 52	-	-	+	-	+	+
NPC 53	-	-	+	-	+	+
NRF 54	-	-	+	-	+	+
NRF 55	-	-	+	+	+	+
NRF 56	-	+	+	+	+	+
NRF 57	-	-	+	-	+	+
NRF 58	-	-	+	-	+	+
NRF 59	+	+	+	+	+	+
NRF60	-	-	+	-	+	+
NRF 61	+	-	+	-	+	+

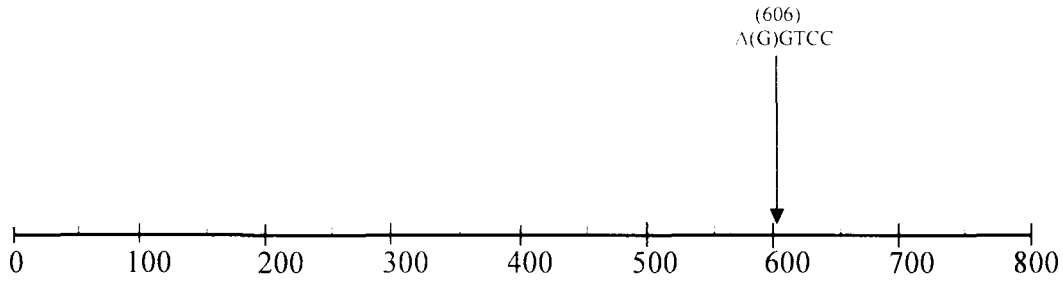


Fig. 4.41: Map of sequence ME1 for profile PS1 showing the possible alternate restriction site using *Sau*96I. (Distances not to scale)

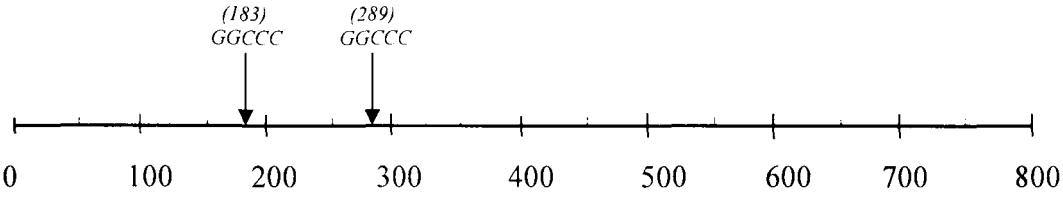


Fig. 4.42: Map of sequence ME2 for profile PS3, PS4, PS5 and PS6 showing the restriction sites using *Sau*96I. (Distances not to scale)

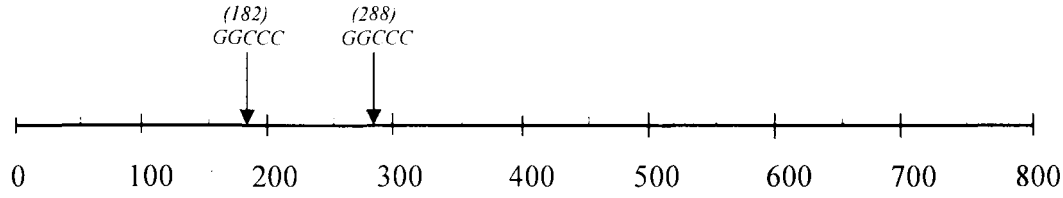


Fig. 4.43: Map of sequence ME3 for profile PS3, PS4, PS5, PS6 and PS7 showing the restriction sites using *Sau*96I. (Distances not to scale)

sizes 491 bp, 183 bp and 106 bp. Presence of a restriction site at position 289 generated two fragments of sizes 289 bp and 491 bp which may have appeared as the ~310 bp and ~500 bp bands respectively.

Profile PS5 consisted of sample NRF-31 and NRF-32 in which an incompletely digested amplicon of ~800 bp band along with digested fragments of sizes ~500 bp, ~200 bp and ~100 bp were detected. The three digested fragments were obtained as a result of the restriction sites available at positions 183 and 289 respectively. The undigested band may have occurred because of the presence of more than one copy in the genome of the plant such that restriction sites for the enzyme *Sau*96I were absent in one of the copies.

4.5.2.3 Restriction digestion analysis of sequence of ME3 using *Sau*96I;

Sequence ME3 was the representative for samples NRF-51 to NRF-61 belonging to morphotype-3. Even though the sample size was small with only eleven trees selected we observed five different profile types within this group. These profiles were PS2, PS3, PS4, PS5 and PS6. Profile PS2 consisted of samples NRF-52, NRF-53, NRF-54, NRF-57, NRF-58 and NRF-60. Only one sample, NRF-59, constituted the profile PS3. Samples NRF-51 and NRF-55 constituted the profile PS4 and sample NRF-61 constituted profile PS5. These profiles (PS2, PS3, PS4 and PS5) exhibited the same banding patterns as those of the morphotype-2 trees (NRF-1 to NRF-50). Sample NRF-56 was the only sample which constituted the profile PS6.

Computer generated restriction digestion analysis of sequenced sample ME3 with *Sau*96I using the software DS Gene (version 1.1) identified two restriction sites at positions 182 and 288 (Fig. 4.43 in page 123) that generated three restriction fragments of 182 bp, 106 bp and 502 bp respectively (Table 4.15 in page 122). Profile PS3 consisted of five bands (~600 bp, ~500 bp, ~310 bp, ~200 bp and ~100 bp) in addition to the undigested ~800 bp band. Considering that one of the copies of the gene had a single restriction site at position 182 then two restriction fragments of 182 bp and 608 bp respectively will be obtained. Similarly, considering that there was a single restriction site at position 288 in one of the copies of the gene then this would generate two restriction fragments of 288 bp and 502 bp which appear as the ~310 bp and ~500 bp bands respectively. Incomplete digestion of the sample amplicon or the absence of restriction sites in one of the copies may have produced the ~800 bp band as observed in the agarose gels.

Four restriction fragments of ~500 bp, ~310 bp, ~200 bp and ~100 bp were obtained for profile PS4. Assuming that there was only one restriction site in one of the copies at position 288, two fragments of 288 bp and 502 bp would be generated which may have appeared as the ~310 bp and ~500 bp band respectively.

Profile PS5 consisted of one sample (NRF-61) where four bands of ~500 bp, ~200 bp, ~100 bp and the undigested ~800 bp were present.

Profile PS6 showed digestion pattern unique to one sample only. Restriction digestion of this sample (NRF-56) produced five restriction fragments of ~600 bp, ~500 bp, ~310 bp, ~200 bp and ~100 bp. The fragments ~500 bp, ~200 bp and ~100 bp are

produced as is discussed for profile PS3. We assumed that presence of a single restriction site at position 182 in one of the copies of the gene would produce two restriction fragments of sizes 182 bp and 608 bp that may have appeared as the ~200 bp and ~600 bp bands respectively. Similarly, considering that a single restriction site was present at position 288 in one of the copies, two restriction fragments of 288 bp and 502 bp would be generated which may have appeared as the ~310 bp and ~500 bp bands respectively.

Restriction patterns of the 18S-28S rDNA ITS region using *Mbo*I for the morphotype-1 trees were different from the patterns for the morphotype-2 and morphotype-3 trees. Restriction pattern generated for the morphotype-3 trees showed basically similar pattern with those of the morphotype-2 trees. The same was observed when the different morphotype samples were digested using the endonuclease enzyme *Sau*96I. Morphotype-1 samples generated similar restriction patterns. However, the restriction patterns generated for the morphotype-2 trees were different from that of morphotype-1 trees. Restriction patterns generated for the morphotype-3 trees were basically the same as obtained for the morphotype-2 trees. Manual checking of the representative ITS sequences of all the three morphotypes base by base identified a number of sites along the stretch of the sequence where the occurrence of a single base substitution (transversion/ transition) could generate the restriction sites for the respective enzymes.

4.6 CLUSTER ANALYSIS:

Cluster dendrogram obtained from the ARP profiles using the enzyme *Mbo*I showed two major clusters, I and II (Fig. 4.44). Samples 1-50 representing the morphotype-1 *Myrica* trees constituted the group I major cluster. Within the cluster I, there were six sub-clusters (I-A, I-B, I-C, I-D, I-E and I-F). Interestingly, the sub-cluster I-B consisted of a lone sample 74 (Sample NRF-24 in Table 4.10) which happened to be a representative of the morphotype-2 trees. Apart from this, no other representatives of the morphotype-2 or morphotype-3 trees existed in the major group I cluster. This sub-cluster I-B appeared to be closest to the sub-cluster I-A than any other sub-cluster within the major cluster I group. It was observed that sample 74 exhibited a unique banding pattern (profile PM9) when digested with the endonuclease enzyme *Mbo*I producing five restriction fragments of ~700 bp, ~530 bp, ~380 bp, ~180 bp and ~90 bp. This banding pattern appeared very close to the banding pattern of profile PM1 where four restriction fragments of ~700 bp, ~530 bp, ~180 bp and ~90 bp were observed. The presence of the additional ~380 bp fragment in the profile PM9 could have been because of the possible occurrence of mutation at position 383 where a transversion took place in which a purine base "A" was replaced with a pyrimidine base "T".

In the group II major cluster four sub-clusters were obtained designated as II-A, II-B, II-C and II-D which exclusively clustered the morphotype-2 and morphotype-3 trees only. Sample 51-100 represented the morphotype-2 trees (NRF-1 to NRF-50) and sample 101-111 represented the morphotype-3 trees (NRF-51 to NRF-61). Sub-cluster II-A represented most of the trees which included both the morphotype-2 and

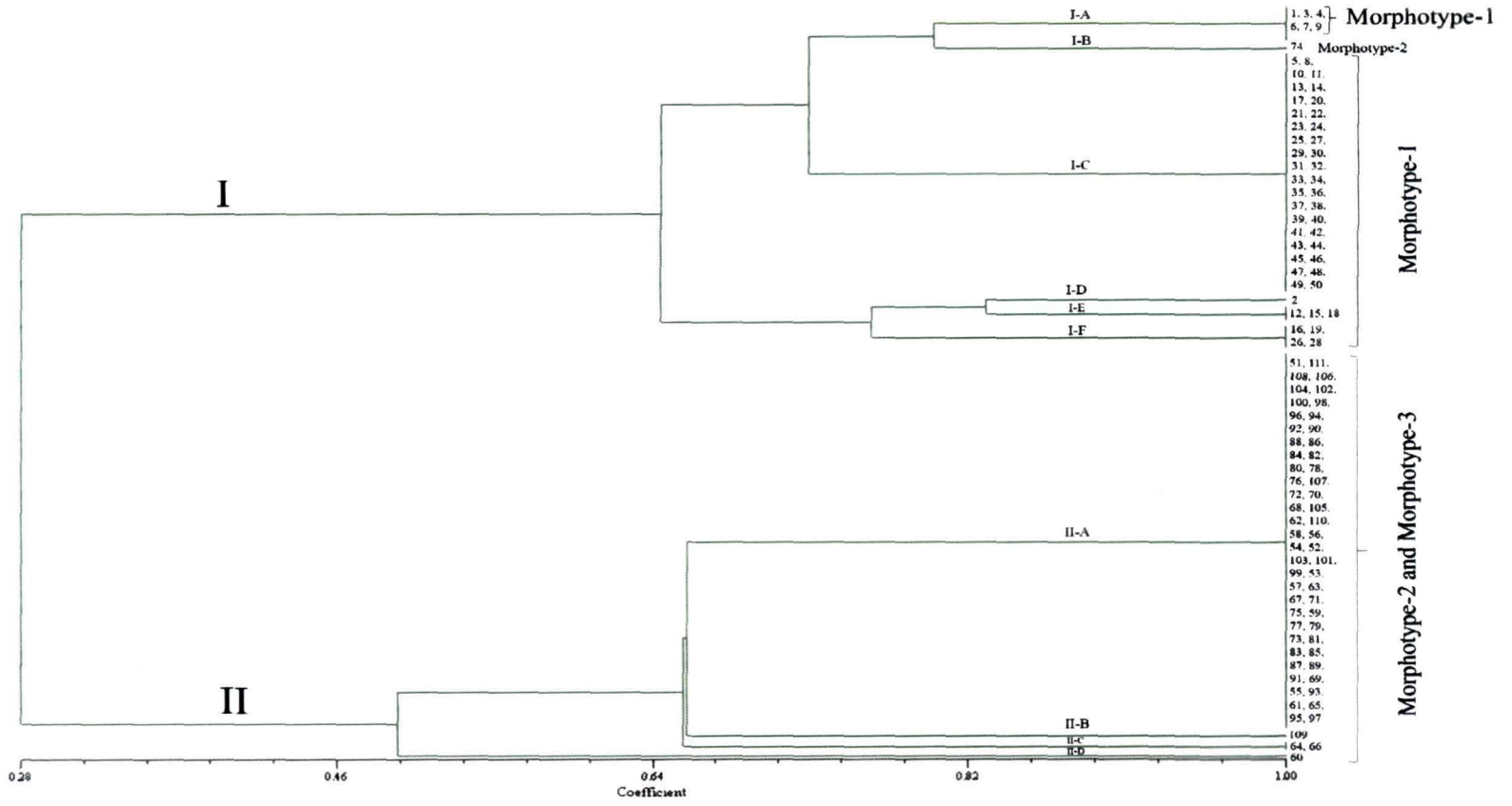


Fig. 4.44: Cluster dendrogram using *Mbo*I: 1 to 50: morphotype-1; 51 to 100: morphotype-2; 101 to 111: morphotype-3.

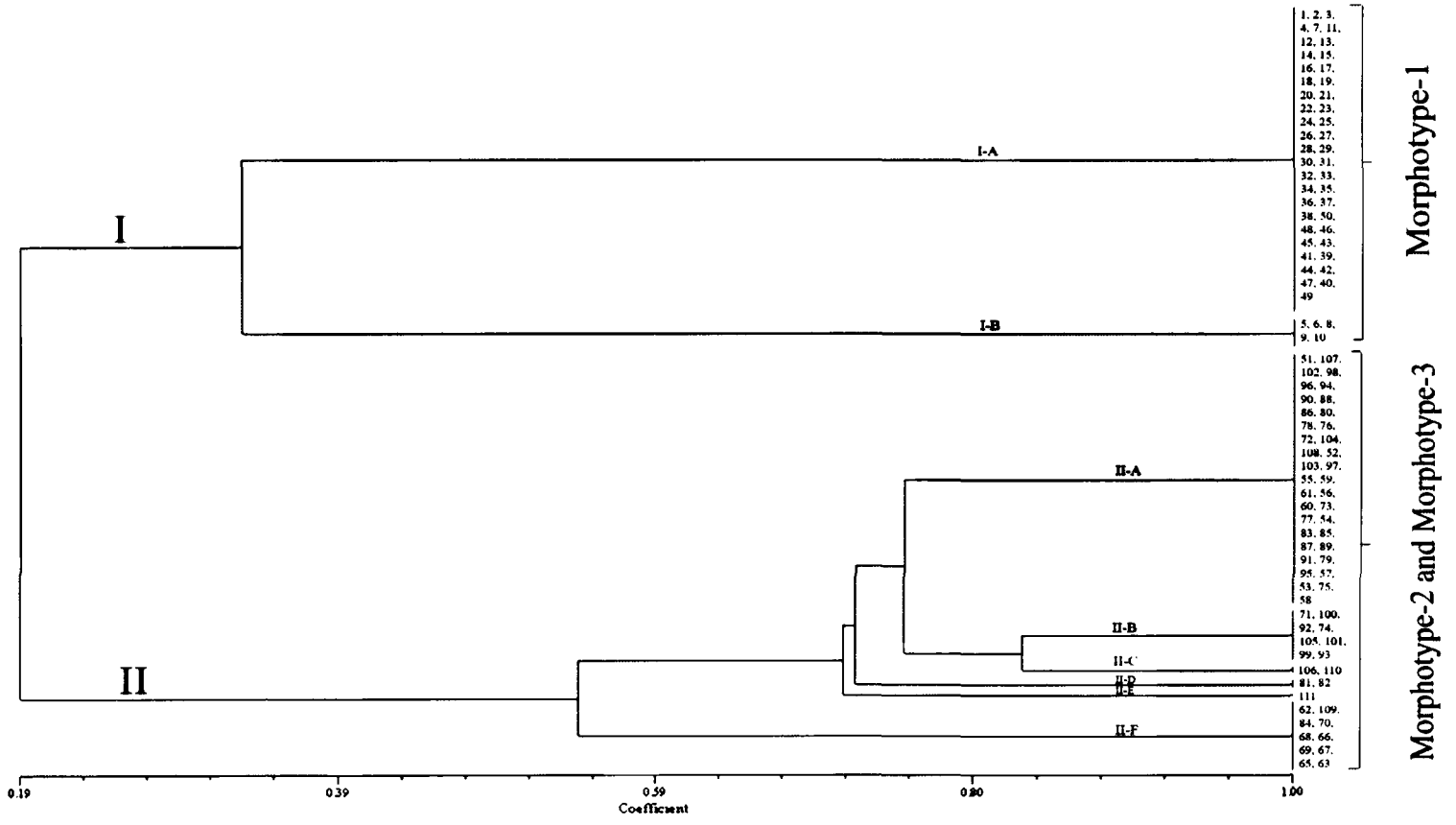


Fig. 4.45: Cluster dendogram using *Sau*96I: 1 to 50: morphotype-1; 51 to 100: morphotype-2; 101 to 111: morphotype-3.

morphotype-3 trees indicating that the trees under this sub-cluster shared one restriction profile. Sub-cluster II-B represented a lone sample, 109 (Sample NRF-59 in Table 4.15) of the morphotype-3 tree whereas sub-cluster II-C represented samples 64 and 66 (Sample NRF-24 and NRF-26 in Table 4.4 and Table 4.6 respectively) of the morphotype-2 trees. Sub-cluster II-D had a lone representative sample 60 (Sample NRF-10 in Table 4.14) of the morphotype-2 tree.

Cluster analysis of the *Sau*96I ARP profiles also revealed two major cluster groups, I and II (Fig. 4.45). All the morphotype-1 trees clustered together in the major cluster group I which was further divided into two sub-clusters, I-A and I-B. The other major cluster group-II consisted of six sub-clusters designated as II-A, II-B, II-C, II-D, II-E and II-F. All the morphotype-2 and morphotype-3 trees clustered in this major cluster group II in mixed occurrence in the different sub-clusters. Morphotype-1 trees stood away from the morphotype-2 and morphotype-3 trees.

4.7 DISCUSSION

The Neighbour Joining, Parsimony and Maximum Likelihood phylogenetic trees for the variable ITS region showed that our samples always clustered together within a common cluster. However, when we considered the clustering pattern within our samples, ME2 (morphotype-2) and ME3 (morphotype-3) appeared to be closer to each other than the sample ME1 (morphotype-1). The clustering patterns in Neighbour Joining and Maximum Likelihood phylogenetic trees obtained for the variable ITS region as well the conserved 18S ribosomal DNA showed almost similar pattern. ME1

definitely seemed to deviate from ME2: ME3 cluster within our common sample cluster. Even at the level of the sequence alignment, sample ME2 and ME3 showed high degree (97%) of sequence homology whereas sequence homology between ME1 and ME2 was 93% and between ME1 and ME3 the sequence homology was 95%. Results obtained from analyses with several ITS sequences of some angiosperm trees also revealed that the sequence homology between species belonging to the same genus was relatively high where the percentage values ranged from 93% to 100%. However, marked divergence in the ITS sequence homology was observed when sequences of species belonging to different genera were aligned where homology as low as 65% was observed. Our hypothesis is that when sequence homology as low as 93% (7% divergence) exists between different species of a genus then the morphotype-1 *Myrica* trees, which shared sequence homology of 93% (7% divergence) with the morphotype-2 may also be considered as a separate species. This proposal is also supported by the morphological differences existing between these two morphotypes. Similar study was conducted by Huguet *et al.* (2005) in which they proposed *Morella cerifera* of Jamaican origin to be treated as a separate species from the USA origin based on the sequence divergence of the ITS region in which sequence divergence of 0.4% was observed between these two geographically isolated species that led to the suggestion that the present Jamaican origin may have sufficiently diverged. ITS divergence values between morphologically distinct species are sometimes less than 1% (Hershkovitz *et al.*, 1999). Even within closely related species there are cases where the ITS divergence is too little to resolve relationships with high statistical confidence (Baldwin *et al.*, 1995; Baldwin

and Markos, 1998). In the present study, intraspecific ITS sequence homology tests revealed 99 to 100% sequence homology between same species that were geographically isolated. This indicated that sequence divergence between plants of the same species even if geographically separated was just 0-1%. However, interspecific ITS divergences within a genus can be highly variable ranging between 0 to 30 % (Susanna *et al.*, 1995; Möller and Cronk, 1997; Pridgeon *et al.*, 1997). Congruently, phylogenetic trees obtained using the 18S gene sequence revealed that the morphotype-1 (MYR1), morphotype-2 (MYR2) and morphotype-3 (MYR3) formed a common cluster in the Neighbour Joining, Parsimony as well as the Maximum Likelihood tests. However, considering within the common cluster of our samples the morphotype-2 and morphotype-3 formed one cluster away from the morphotype-1 as revealed from the Neighbour Joining and the Maximum Likelihood trees. A shift in clustering pattern was observed in the Parsimony tree in which the morphotype-1 and morphotype-3 formed a common cluster, whereas, the morphotype-2 was separated. However, this common cluster between the morphotype-1 (MYR1) and morphotype-3 (MYR3) was not supported by a very high bootstrap value (610.8). Sequence alignment showed that the homology between the between MYR1 and MYR2 was 97%. Even in this case the morphotype-1 showed closer sequence homology to morphotype-3 (98%) than to morphotype-2 (97%) tree. Results obtained from the sequence homology tests as well as from the different phylogenetic trees seemed to indicate that the morphotype-1 *Myrica* trees may be a distinct species from the morphotype-2 and morphotype-3 *Myrica* trees. This was also supported by the cluster dendrograms obtained using the ARP profiles

with the restriction endonuclease enzymes *Mbo*I and *Sau96*I where all the morphotype-1 *Myrica* trees clustered in the major cluster group-I whereas the morphotype-2 and morphotype-3 *Myrica* trees clustered in the major cluster group-II. This is noteworthy because the cluster dendrogram clearly differentiated between the morphotype-1 *Myrica* trees from the other two morphotype trees. This was further supported by the results of the 5.8S rDNA gene secondary structure. Secondary structure of the morphotype-1 differed from the morphotype-2 and morphotype-3 *Myrica* trees in the secondary folding of 5.8S rRNA. The morphotype-2 and morphotype-3 secondary structure showed similarity in folding structure. This result further supports our claim that the morphotype-1 may be different from the morphotype-2 and morphotype-3 *Myrica* trees.

Therefore, based on the results obtained from the phylogenetic tree analysis, sequence alignment tests, 5.8S rRNA secondary structures and also from the ARP cluster dendrograms it may be proposed that the morphotype-1 trees which are referred to as *Myrica nagi* by some authors may retain this nomenclature. Whereas, the morphotype-2 trees preferably known as *Myrica esculenta* may also be proposed to retain this nomenclature. The morphotype-3 tree could be a hybrid of the morphotype-1 and morphotype-2 trees. This is observed from the sequence alignment test of our samples for both the ITS as well as the 18S gene in which the morphotype-1 aligned closer to morphotype-3 (95% and 98% for the ITS and 18S gene respectively) than to morphotype-2 (93% and 97% for the ITS and 18S gene respectively). This is also evident from the ITS parsimony tree where a common cluster (ME1:ME2:ME3) supported by a high bootstrap value 964.9 replicates was obtained. Similarly, the 18S

parsimony tree showed a clustering pattern in which MYR1 (morphotype-1) and MYR3 (morphotype-3) formed a common cluster supported by a bootstrap value of 610.8 replicates within our samples in which MYR2 (morphotype-2) was separated. It may be hypothesized that a common progenitor may have existed during the early times which diverged into the morphotype-1 and the morphotype-3 *Myrica* trees. In due course of time the morphotype-1 may have diverged farther away from the morphotype-3 trees. The morphotype-2 *Myrica* trees may have emerged from the morphotype-3 *Myrica* trees as a result of selection pressure by man for commercial as well as its aesthetic value for its bigger fruit size over the average sized fruit of the morphotype-3 *Myrica* trees. Another hypothesis may also be considered as regards to the existence of the different morphotype *Myrica* trees where a cross between the morphotype-1 and the morphotype-2 *Myrica* tree may have given rise to the F1 progeny. The morphotype-1 *Myrica* trees may have sufficiently diverged away from the morphotype-2 trees. The F1 progeny may have back crossed with the morphotype-2 parent to give rise to the morphotype-3 *Myrica* tree. As such, the morphotype-3 *Myrica* trees may be considered as hybrid of the morphotype-1 and morphotype-2 *Myrica* trees or it may be considered as a variant of the morphotype-2 *Myrica* trees.

Therefore, it may be erroneous to consider *Myrica nagi* and *Myrica esculenta* as synonyms. We propose that these two different species of *Myrica* are found in Meghalaya.

II. MOLECULAR MARKER DEVELOPMENT BASED ON THE ARP PROFILES AND THE ARA VALUES

The ARP profile of each tree was related with its respective Acetylene Reduction Assay (ARA) value in order to look for profiles which would serve as molecular markers for screening out trees belonging to high or low nitrogenase activity. Since the activity of the nitrogenase enzyme of the actinomycete *Frankia* is dependent on the available nitrogen in the surrounding soil environment, estimation of the soil nitrogen content of ten soil samples from each site was also performed in order to draw an inference on the nitrogenase activity and the soil nitrogen availability.

4.8 SOIL NITROGEN ANALYSIS:

The Kjeldahl method of nitrogen estimation was used to determine the nitrogen content of the soil collected from within 1 metre periphery of some selected trees from both the sites. Procedure of nitrogen estimation is discussed in section 3.6 of chapter 3.

Soil nitrogen content in Nongkrem forest recorded a higher value than those collected from NEHU campus (Table 4.16 and 4.17). The soil nitrogen at NEHU permanent campus, where the morphotype-1 trees predominate, ranged from 0.045 % to 0.496 %. At the Nongkrem forest site, where the morphotype-2 and morphotype-3 trees predominate, the soil nitrogen percentage ranged from 0.408 % to 1.036 %. Values obtained from our analysis showed that the soil at NEHU campus exhibited lower

TABLE 4.16: SOIL NITROGEN PERCENTAGE OF TEN REPRESENTATIVE SAMPLES AT NEHU PERMANENT CAMPUS SITE

NEHU PERMANENT CAMPUS			
Sl. No.	MEAN TITRANT VALUES	NITROGEN PERCENTAGE	MEAN PERCENTAGE
1	5.66±0.384	0.258	0.178
2	2.26±0.033	0.088	
3	3.46±0.120	0.148	
4	2.36±0.066	0.093	
5	1.40±0.057	0.045	
6	3.63±0.696	0.156	
7	1.70±0.208	0.060	
8	6.10±0.011	0.280	
9	3.66±0.088	0.158	
10	10.43±0.240	0.496	

TABLE 4.17: SOIL NITROGEN PERCENTAGE OF TEN REPRESENTATIVE SAMPLES AT NONGKREM FOREST SITE

NONGKREM FOREST			
Sl. No.	MEAN TITRANT VALUES	NITROGEN PERCENTAGE	MEAN PERCENTAGE
1	19.5±0.300	0.950	0.628
2	16.33±0.606	0.791	
3	12.33±0.145	0.591	
4	9.36±0.233	0.443	
5	10.76±0.033	0.513	
6	10.66±0.008	0.508	
7	21.23±0.674	1.036	
8	9.83±0.384	0.466	
9	12.16±0.959	0.583	
10	8.66±0.440	0.408	

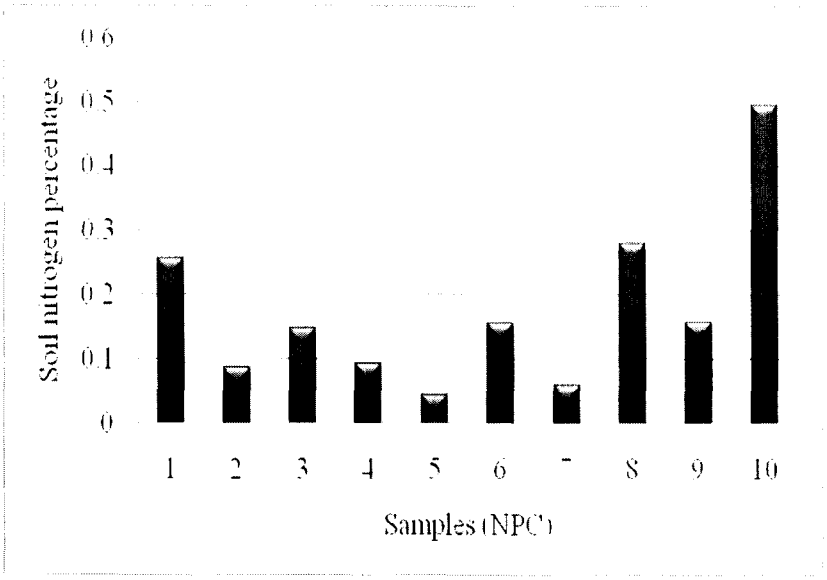


Fig. 4.46: Graphical representation of soil nitrogen content of ten representative samples (1-10) at NEHU campus where x axis indicates percent of nitrogen and y axis indicates samples number.

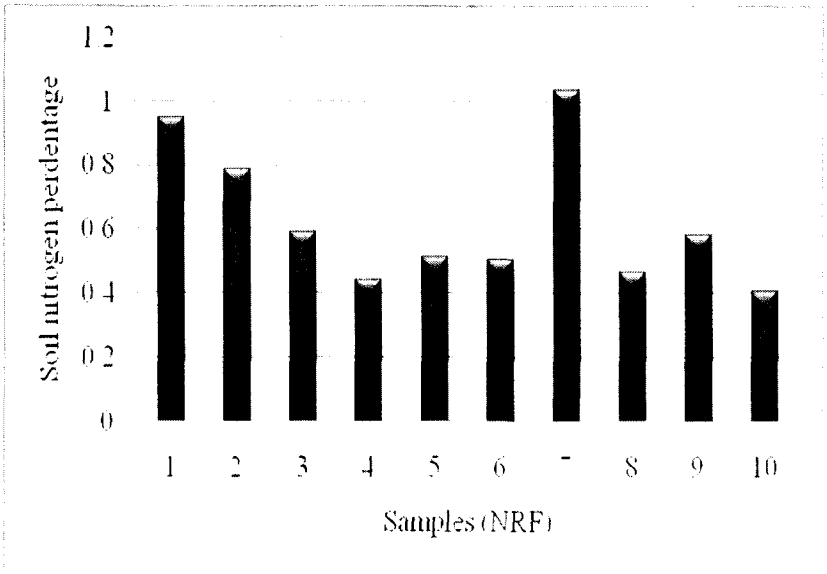


Fig. 4.47: Graphical representation of soil nitrogen content of ten representative samples (1-10) at Nongkrem forest where x axis indicates percent of nitrogen and y axis indicates samples number.

nitrogen content as compared to that of Nongkrem forest soil. Morphotype-1 trees of *Myrica* at NEHU campus were found growing with the major population inhabited by pine trees where the occurrence of other species of trees was sparse. Whereas, dense population of morphotype-2 *Myrica* trees (also morphotype-3 trees) along other angiosperms apart from pine trees prevailed in Nongkrem forest. Apart from the contribution of nitrogen by the nitrogen fixing *Frankia* present in the nodules, a major chunk of nitrogen in soil comes from dead plants and other organic matter where release of nitrogen and other associated organic compounds upon degradation is a natural occurrence. Nitrogen content of soil measured in percent collected at both the sites are represented graphically in Fig. 4.46 and Fig. 4.47.

In order to check whether these mean values obtained were significant or not, students' t-test was performed. The t-test value at 18 degrees of freedom was calculated at 5.625 as against the table value which showed 2.101 at 5% level of significance ($p = 0.05$) and 2.878 at 1% level of significance ($p = 0.01$). The difference was therefore considered statistically significant since the calculated value of 't' was more than the table value at both 1% and 5% levels of significance.

4.9 ESTIMATION OF NITROGENASE ACTIVITY USING ACETYLENE REDUCTION ASSAY:

4.9.1 Collection of nodules:

Estimation of nitrogenase activity was performed on nodules from all the sample trees collected from both the sites following the protocol of Stewart *et al.* (1968).

Procedure for estimaton of nitrogenase activity is discussed in section 3.5 of chapter 3. Care was taken to select only the healthy nodules in order to avoid problems that degenerating nodules would pose in the actual estimation of the nitrogenase activity.

4.9.2 Acetylene Reduction Assay (ARA):

The average nitrogenase activity for morphotype-1 trees (NPC:1 to NPC:50) was recorded at 39.06 n moles C_2H_4 produced/ mg fresh wt/ hr with the values ranging between 19.71 to 56.92 n moles C_2H_4 produced/mg fresh wt/ hr. The mean value obtained for morphotype-2 trees (NRF:1 to NRF:50) stood at 26.91 nmoles C_2H_4 produced/ mg fresh wt/ hr in which the values ranged between 17.82 to 38.48 C_2H_4 produced/ mg fresh wt/ hr. The mean value obtained for the morphotype-3 trees (NRF:51 to NRF:61) was 27.66 C_2H_4 produced/ mg fresh wt/ hr with the ARA values ranging between 15.01 to 46.14 C_2H_4 produced/ mg fresh wt/ hr (Table 4.18, 4.19 and 4.20). Most of the morphotype-1 *Myrica* trees at NEHU campus exhibited an average high nitrogenase activity as compared to those of the morphotype-2 and morphotype-3 *Myrica* trees at Nongkrem forest. One of the reasons for increased nitrogenase activity in the nodules in the morphotype-1 trees could be the poorer nitrogen status of soil from NEHU campus. This is in contrast to those of the morphotype-2 and morphotype-3 trees at Nongkrem forest in which mixed forest stand produced enough decaying organic material providing a rich source of nitrogen to the soil, thereby, lowering the nitrogenase activity of the actinomycete *Frankia* present in the nodules.

TABLE 4.18: AVERAGE NITROGENASE ACTIVITIES OF THE MORPHOTYPE-1 *MYRICA* TREES AT NEHU PERMANENT CAMPUS.

Sl. No.	SAMPLE	AVERAGE NITROGENASE ACTIVITY (n mole C_2H_2 produced / mg fresh wt/ hr)	MEAN ARA VALUE
1	NPC:1	22.12±0.73	
2	NPC:2	45.76±0.84	
3	NPC:3	33.52±0.60	
4	NPC:4	55.76±0.41	
5	NPC:5	55.92±0.46	
6	NPC:6	26.29±0.55	
7	NPC:7	44.50±0.55	
8	NPC:8	54.96±0.81	
9	NPC:9	38.81±0.53	
10	NPC:10	52.06±0.72	
11	NPC:11	49.76±0.29	
12	NPC:12	45.19±0.06	
13	NPC:13	45.64±0.40	
14	NPC:14	41.38±0.68	
15	NPC:15	35.40±0.60	
16	NPC:16	37.84±0.20	
17	NPC:17	31.36±0.53	
18	NPC:18	41.64±0.61	
19	NPC:19	39.01±0.64	
20	NPC:20	26.49±0.76	
21	NPC:21	32.87±0.41	
22	NPC:22	38.70±0.63	
23	NPC:23	37.22±1.13	
24	NPC:24	25.27±0.87	
25	NPC:25	48.04±0.66	39.06
26	NPC:26	38.92±0.44	
27	NPC:27	34.55±0.79	
28	NPC:28	21.10±0.51	
29	NPC:29	32.88±0.87	
30	NPC:30	33.81±0.52	
31	NPC:31	19.71±0.98	
32	NPC:32	38.36±0.96	
33	NPC:33	33.43±0.46	
34	NPC:34	37.18±0.90	
35	NPC:35	51.40±1.12	
36	NPC:36	39.46±0.76	
37	NPC:37	42.05±0.90	
38	NPC:38	39.55±1.15	
39	NPC:39	50.58±0.92	
40	NPC:40	32.66±1.34	
41	NPC:41	56.92±1.06	
42	NPC:42	52.62±0.68	
43	NPC:43	34.84±0.36	
44	NPC:44	56.62±0.76	
45	NPC:45	32.50±1.31	
46	NPC:46	35.49±1.22	
47	NPC:47	35.06±0.99	
48	NPC:48	30.70±0.98	
49	NPC:49	32.43±0.92	
50	NPC:50	34.83±1.00	

Table 4.19: AVERAGE NITROGENASE ACTIVITIES OF THE MORPHOTYPE-2 *MYRICA* TREES AT NONGKREM FOREST.

Sl. No.	SAMPLE	AVERAGE NITROGENASE ACTIVITY (n mole C_2H_2 produced / mg fresh wt/ hr)	MEAN ARA VALUE
1	NRF:1	28.10±0.99	
2	NRF:2	32.06±0.59	
3	NRF:3	17.82±0.91	
4	NRF:4	30.42±0.58	
5	NRF:5	30.58±1.13	
6	NRF:6	25.99±0.51	
7	NRF:7	21.75±0.34	
8	NRF:8	21.98±0.38	
9	NRF:9	33.73±0.30	
10	NRF:10	24.30±1.06	
11	NRF:11	22.54±0.68	
12	NRF:12	21.39±1.08	
13	NRF:13	24.04±0.83	
14	NRF:14	18.62±0.31	
15	NRF:15	30.50±0.88	
16	NRF:16	25.48±0.42	
17	NPC:17	28.55±0.45	
18	NRF:18	25.59±0.25	
19	NRF:19	24.06±0.17	
20	NRF:20	37.17±0.92	
21	NRF:21	23.16±0.69	
22	NRF:22	33.13±0.88	
23	NRF:23	37.64±1.21	
24	NRF:24	31.77±0.28	
25	NRF:25	25.93±0.60	26.91
26	NRF:26	29.93±0.80	
27	NRF:27	28.25±0.68	
28	NRF:28	25.06±0.90	
29	NRF:29	31.10±1.25	
30	NRF:30	26.11±0.81	
31	NRF:31	34.96±0.42	
32	NRF:32	26.24±0.62	
33	NRF:33	38.48±0.36	
34	NRF:34	21.13±0.94	
35	NRF:35	31.12±0.08	
36	NRF:36	19.34±1.17	
37	NRF:37	20.64±0.86	
38	NRF:38	31.29±0.64	
39	NRF:39	19.06±0.63	
40	NRF:40	21.29±0.73	
41	NRF:41	26.35±0.36	
42	NRF:42	34.16±0.19	
43	NRF:43	20.98±0.96	
44	NRF:44	24.03±0.66	
45	NRF:45	25.85±0.71	
46	NRF:46	37.29±0.73	
47	NRF:47	19.11±0.57	
48	NRF:48	34.51±0.71	
49	NRF:49	18.95±0.44	
50	NRF:50	24.06±0.60	

Table 4.20: AVERAGE NITROGENASE ACTIVITIES OF THE MORPHOTYPE-3 *MYRICA* TREES AT NONGKREM FOREST.

Sl. No.	SAMPLE	AVERAGE NITROGENASE ACTIVITY (n mole C_2H_2 produced/ mg fresh wt/ hr)	MEAN ARA VALUE
51	NRF:51	25.90±1.39	27.66
52	NRF:52	23.16±2.17	
53	NRF:53	22.19±0.84	
54	NRF:54	22.53±1.42	
55	NRF:55	19.91±0.22	
56	NRF:56	24.27±0.70	
57	NRF:57	36.08±2.02	
58	NRF:58	46.14±1.97	
59	NRF:59	40.87±2.11	
60	NRF:60	15.01±1.48	
61	NRF:61	28.30±2.54	

However, the nitrogenase activity of trees at both the sites exhibited near about uniformity in their values in their respective sites. Given the wide range of trees selected for this study and also considering the promiscuous nature of the genus *Myrica* towards the actinomycete *Frankia* it was assumed that almost the entire range of *Frankia* strains was represented that was involved in the nitrogen fixation process. Graphical representation of the ARA values for all the morphotypes are given in Fig. 4.48 and Fig. 4.49. Similar study was conducted by Chauhan and Misra (2002) on alder trees.

Students t-test was performed for the mean ARA values to check whether the difference obtained for two comparisons was significant or not. The t-test value between the morphotype-1 and morphotype-2 at 98 degrees of freedom was calculated at 8.01 which was more than the table value (1.960 at 5% level of significance and 2.576 at 1% level of significance). Therefore, the difference in the mean ARA values between the morphotype-1 and morphotype-2 was considered statistically significant. t-test value between the morphotype-1 and morphotype-3 at 59 degrees of freedom was calculated at 3.60 which was more than the table value (1.960 at 5% level of significance and 2.576 at 1% level of significance). Therefore, the difference in the mean ARA values between the morphotype-1 and morphotype-3 was considered statistically significant. t-test value between the morphotype-2 and morphotype-3 at 59 degrees of freedom was calculated at 0.25 which was less than the table value (1.960 at 5% level of significance and 2.576 at 1% level of significance). Therefore, the difference in the mean ARA values between the morphotype-2 and morphotype-3 trees was not considered significant.

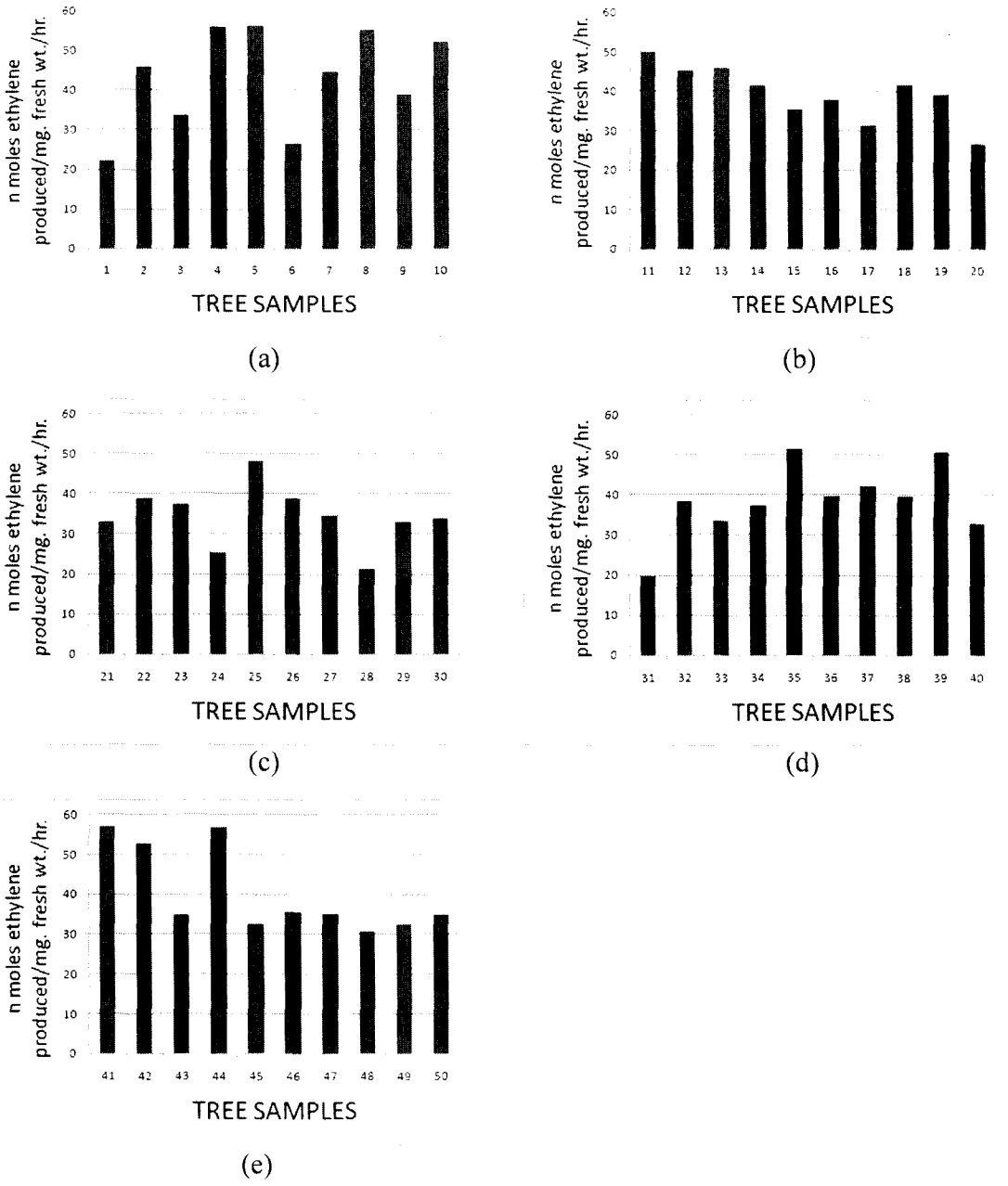


Fig. 4.48 (a-e): Graphical representation of the ARA values of morphotype-1 *Myrica* trees at NEHU permanent campus. The x axis represents the nitrogenase activity given in n mole C_2H_2 produced/ mg fresh wt/ hr. The y axis represents the samples in numerical order.

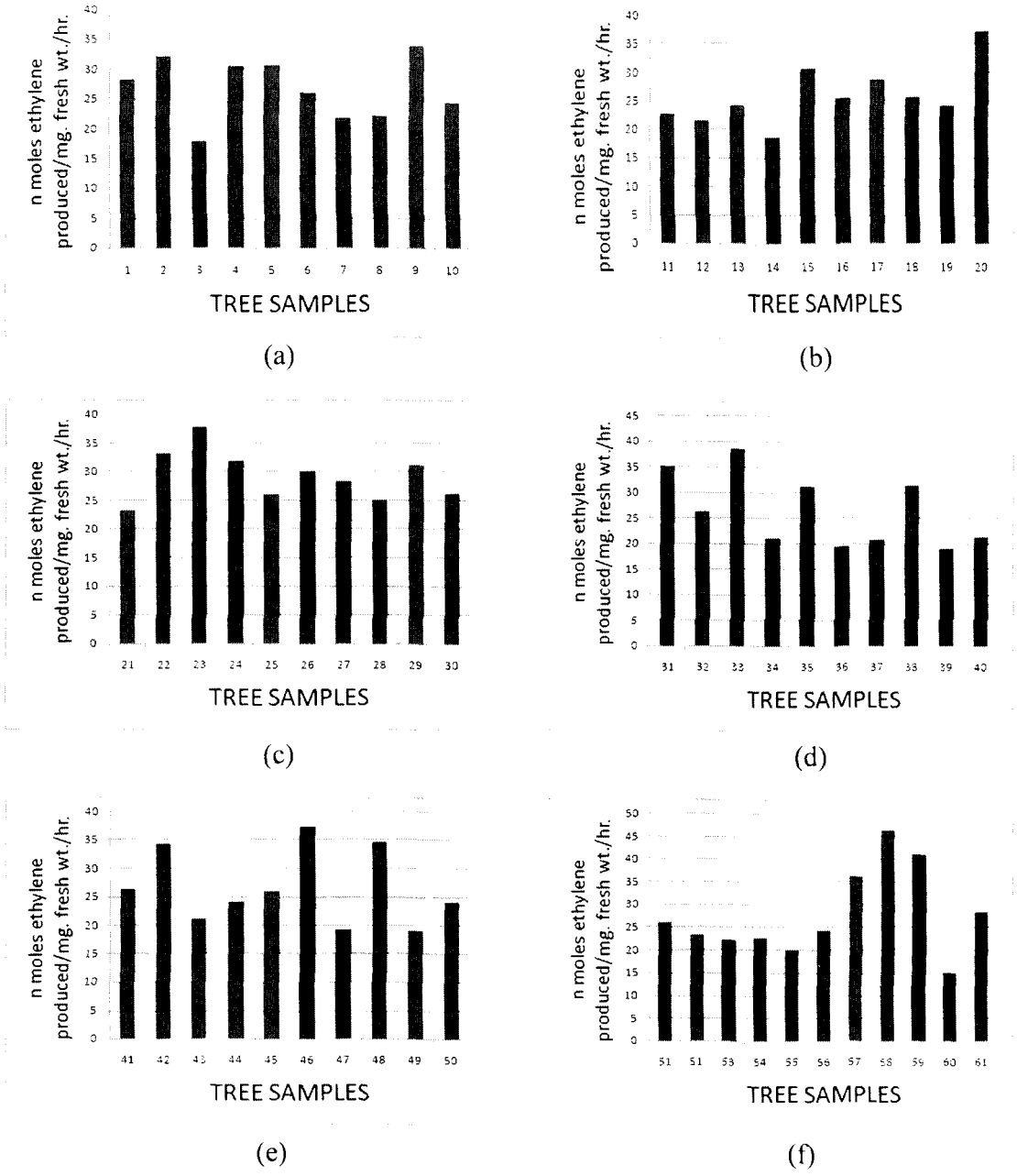


Fig. 4.49 (a-f): Graphical representation of the ARA values of morphotype-2 (a-e) and morphotype-3 *Myrica* trees (f) at Nongkrem forest. The x axis represents the nitrogenase activity given in n mole C_2H_2 produced/ mg fresh wt/ hr. The y axis represents the samples in numerical order.

4.10 RELATION BETWEEN THE ACETYLENE REDUCTION ASSAY AND ARP/ PCR-RFLP PROFILES:

In order to determine the relationship between the Acetylene Reduction Assay (ARA) and the restriction digestion profiles, if any, we tried to relate the ARA values of each sample against the restriction profiles. A frequency class distribution (0-5, 6-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60) of the samples in regard to their nitrogenase activity was prepared to make the comparisons easier. It was observed that most of the morphotype-1 trees fell in the range of 31-35 n moles C_2H_4 produced/ mg fresh wt/ hr frequency class (Table 4.21). Whereas, the morphotype-2 and morphotype-3 trees at Nongkrem forest fell in the range of 21-25 n moles C_2H_4 produced/ mg fresh wt/ hr frequency class (Table 4.22 and 4.23). We considered 30 n moles ethylene produced/ mg fresh wt/ hr as the standard for assigning the trees to high or low nitrogenase activity groups. Morphotype-1 trees at NEHU campus were found to belong to the high nitrogen fixing group. The distribution of sample trees to its respective frequency class group are represented graphically in Fig. 4.50, Fig. 4.51 and Fig. 4.52.

Comparison of the ARP pattern with the ARA values revealed that many of the restriction profiles were morphotype restricted, meaning the profiles which were obtained for one particular morphotype were not present in the other, except for the morphotype-3 where the restriction fragment profiles were similar to those of the morphotype-2 trees. Moreover, many of the high and low nitrogenase activity values were spread across the profiles. Profiles obtained for the enzyme *Mbo*I did not

TABLE 4.21: NUMBER OF MORPHOTYPE-1 *MYRICA* TREES BELONGING TO EACH FREQUENCY CLASS BASED ON ARA VALUES

MORPHOTYPE-1 (NEHU PERMANENT CAMPUS)		
Sl. No.	**FREQUENCY CLASS**	**NUMBER OF SAMPLES**
1	0-5	0
2	6-10	0
3	11-15	0
4	16-20	1
5	21-25	3
6	26-30	3
7	31-35	15
8	36-40	10
9	41-45	7
10	46-50	3
11	51-55	6
12	56-60	2

TABLE 4.22: NUMBER OF MORPHOTYPE-2 *MYRICA* TREES BELONGING TO EACH FREQUENCY CLASS BASED ON ARA VALUES

MORPHOTYPE-2 (NONGKREM FOREST)		
Sl. No.	**FREQUENCY CLASS**	**NUMBER OF SAMPLES**
1	0-5	0
2	6-10	0
3	11-15	0
4	16-20	8
5	21-25	18
6	26-30	10
7	31-35	10
8	36-40	4
9	41-45	0
10	46-50	0
11	51-55	0
12	56-60	0

TABLE 4.23: NUMBER OF MORPHOTYPE-3 *MYRICA* TREES BELONGING TO EACH FREQUENCY CLASS BASED ON ARA VALUES

MORPHOTYPE-3 (NONGKREM FOREST)		
Sl. No.	**FREQUENCY CLASS**	**NUMBER OF SAMPLES**
1	0-5	0
2	6-10	0
3	11-15	1
4	16-20	1
5	21-25	5
6	26-30	1
7	31-35	0
8	36-40	2
9	41-45	0
10	46-50	1
11	51-55	0
12	56-60	0

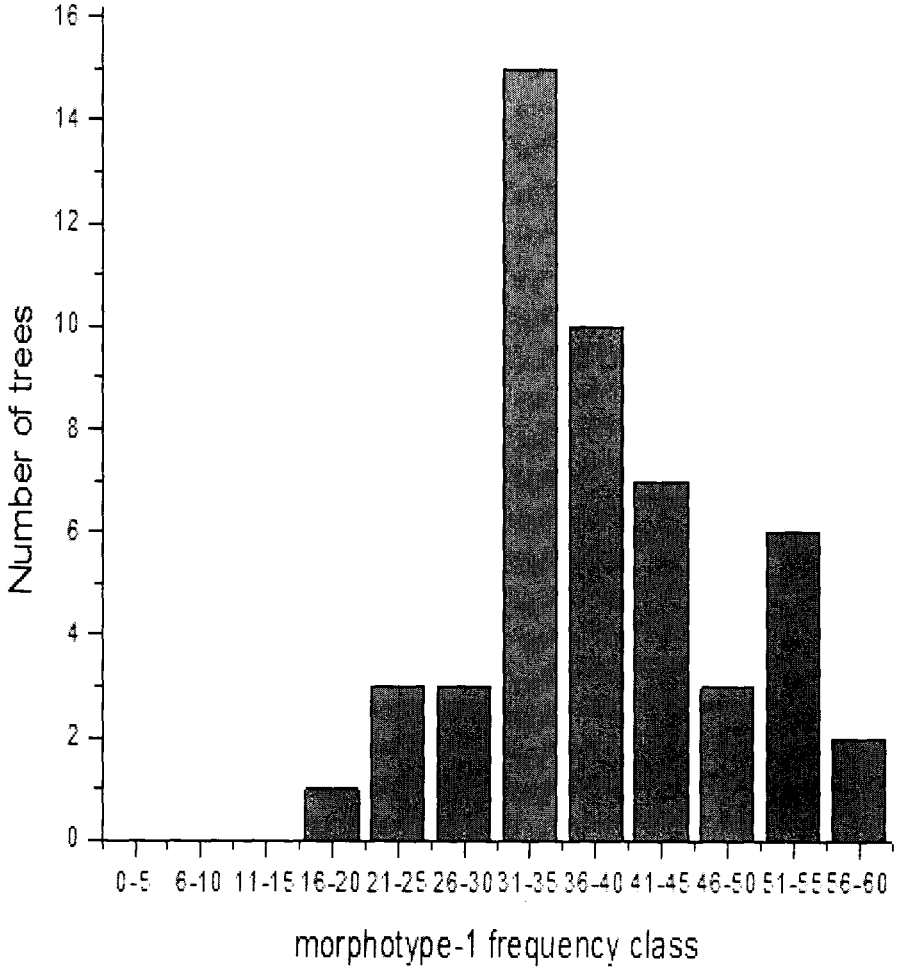


Fig. 4.50: Graphical representation of frequency class distribution of morphotype-1 *Myrica* trees at NEHU permanent campus based on the ARA values.

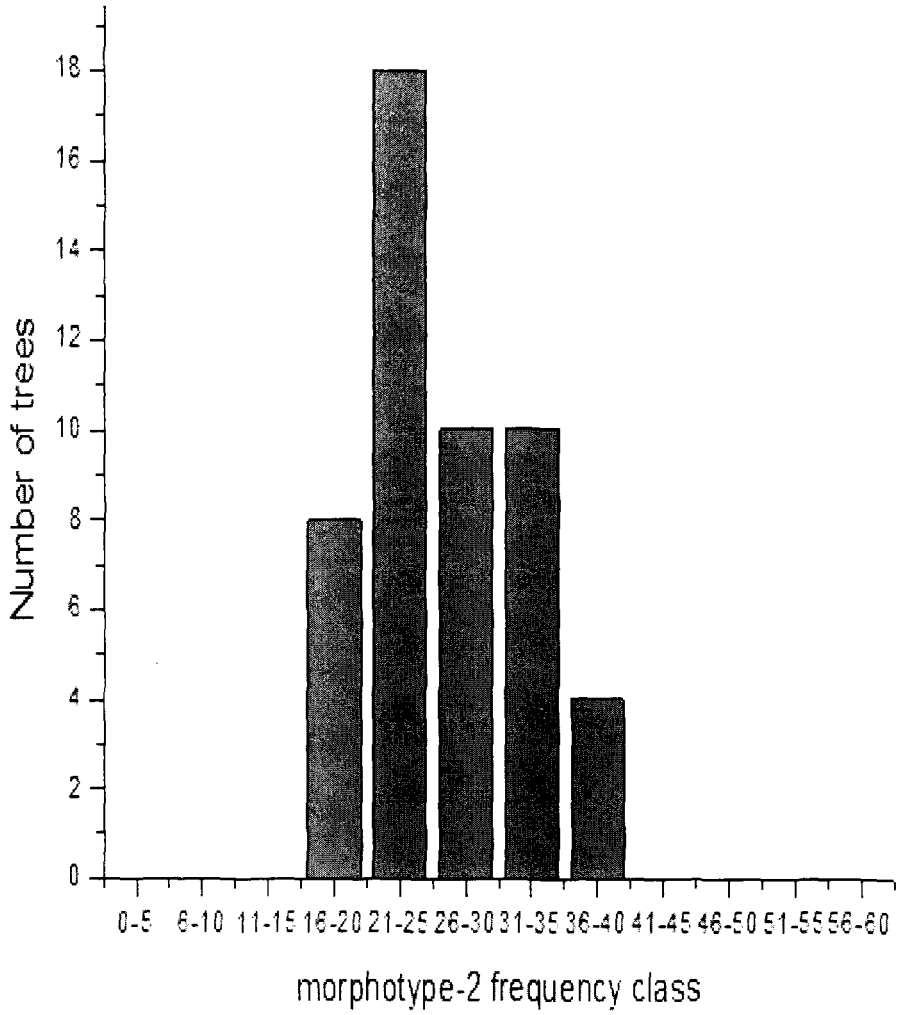


Fig. 4.51: Graphical representation of frequency class distribution of morphotype-2 *Myrica* trees at Nongkrem forest based on the ARA values.

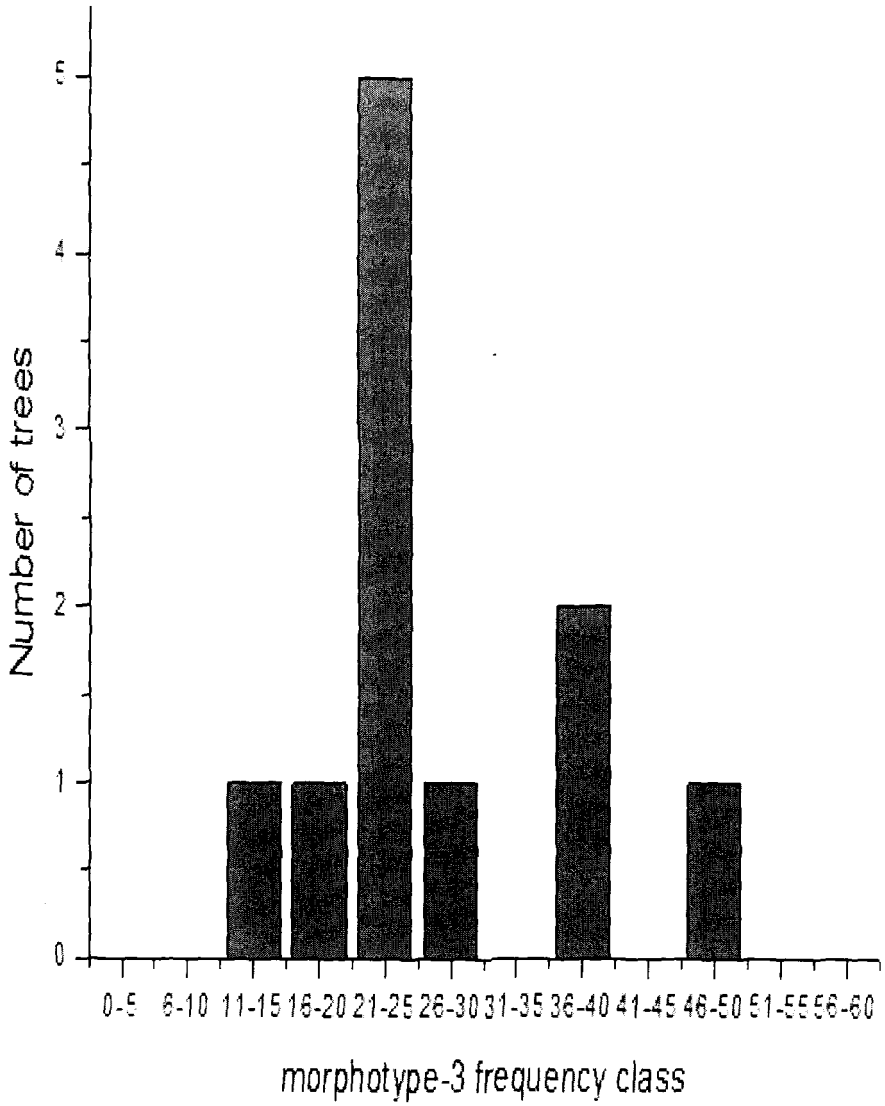


Fig. 4.52: Graphical representation of frequency class distribution of morphotype-3 *Myrica* trees at Nongkrem forest based on the ARA values.

necessarily reflect the same trees with those profiles obtained for the enzyme *Sau*96I and vice versa. Since restriction profiles generated by either of the enzymes in isolation did not reveal any association with nitrogenase activity, we resorted to the multi-site marker approach described by Verghese and Misra (2000) and Chauhan and Misra (2002). It was interesting to note that the profiles PM3 (*Mbo*I) and PS1 (*Sau*96I) (assigned to the same sample tree) denoted both the highest and lowest nitrogenase activity values for samples NPC-1 to NPC-50. The same was observed for the Nongkrem forest samples, NRF-1 to NRF-61, where the profiles PM6 and PS2 (assigned to the same sample) denoted both the highest and lowest nitrogenase activity value. Apart from this, these profiles were spread across samples with varied range of nitrogenase activity. A unique profile combination PM9:PS4 was detected for a single sample (NRF-24) of the morphotype-2 trees in which the value of nitrogenase activity was recorded at 31.77 n mole ethylene produced/ mg fresh wt/ hr which was just about the standard value (Table 4.25).

However, an interesting observation was made in samples NPC-5, NPC-8 and NPC-10 of the morphotype-1 trees with relatively high nitrogenase activity and sharing the profile combination PM3:PS0 (Table 4.24). This profile combination was not found in any tree with low nitrogenase activity. Therefore, it may be considered as a potential molecular marker to screen trees with relatively high nitrogenase activity. The profile combination PM6:PS4 constituting the samples NRF:21, NRF:43, NRF:49, NRF:50 and NRF:55 recorded nitrogenase activity ranging from 18.95 to 24.06 n moles of ethylene produced/ mg fresh wt/ hr which was well below the standard 30 n moles ethylene

TABLE 4.24: ARA VALUES OF MORPHOTYPE-1 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE ARP PROFILES.

SAMPLE	AVERAGE NITROGENASE ACTIVITY (n mole C_2H_2 produced / mg fresh wt/ hr)	PROFILES PCR-RFLP/ ARP *Mbo*I	PROFILES PCR-RFLP/ ARP *Sau*96I
NPC:1	22.12	PM1	PS1
NPC:2	45.76	PM2	PS1
NPC:3	33.52	PM1	PS1
NPC:4	55.76	PM1	PS1
NPC:5	55.92	PM3	PS0
NPC:6	26.29	PM1	PS0
NPC:7	44.50	PM1	PS1
NPC:8	54.96	PM3	PS0
NPC:9	38.81	PM1	PS0
NPC:10	52.06	PM3	PS0
NPC:11	49.76	PM3	PS1
NPC:12	45.19	PM4	PS1
NPC:13	45.64	PM3	PS1
NPC:14	41.38	PM3	PS1
NPC:15	35.40	PM4	PS1
NPC:16	37.84	PM5	PS1
NPC:17	31.36	PM3	PS1
NPC:18	41.64	PM4	PS1
NPC:19	39.01	PM5	PS1
NPC:20	26.49	PM3	PS1
NPC:21	32.87	PM3	PS1
NPC:22	38.70	PM3	PS1
NPC:23	37.22	PM3	PS1
NPC:24	25.27	PM3	PS1
NPC:25	48.04	PM3	PS1
NPC:26	38.92	PM5	PS1
NPC:27	34.55	PM3	PS1
NPC:28	21.10	PM5	PS1
NPC:29	32.88	PM3	PS1
NPC:30	33.81	PM3	PS1
NPC:31	19.71	PM3	PS1
NPC:32	38.36	PM3	PS1
NPC:33	33.43	PM3	PS1
NPC:34	37.18	PM3	PS1
NPC:35	51.40	PM3	PS1
NPC:36	39.46	PM3	PS1
NPC:37	42.05	PM3	PS1
NPC:38	39.55	PM3	PS1
NPC:39	50.58	PM3	PS1
NPC:40	32.66	PM3	PS1
NPC:41	56.92	PM3	PS1
NPC:42	52.62	PM3	PS1
NPC:43	34.84	PM3	PS1
NPC:44	56.62	PM3	PS1
NPC:45	32.50	PM3	PS1
NPC:46	35.49	PM3	PS1
NPC:47	35.06	PM3	PS1
NPC:48	30.70	PM3	PS1
NPC:49	32.43	PM3	PS1
NPC:50	34.83	PM3	PS1

TABLE 4.25: ARA VALUES OF MORPHOTYPE-2 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE ARP PROFILES.

SAMPLE	AVERAGE NITROGENASE ACTIVITY (n mole C_2H_4 produced / mg fresh wt/ hr)	PROFILES PCR-RFLP/ ARP *Mbo*I	PROFILES PCR-RFLP/ ARP *Sau*96I
NRF:1	28.10	PM6	PS2
NRF:2	32.06	PM6	PS2
NRF:3	17.82	PM6	PS2
NRF:4	30.42	PM6	PS2
NRF:5	30.58	PM6	PS2
NRF:6	25.99	PM6	PS2
NRF:7	21.75	PM6	PS2
NRF:8	21.98	PM6	PS2
NRF:9	33.73	PM6	PS2
NRF:10	24.30	PM7	PS2
NRF:11	22.54	PM6	PS3
NRF:12	21.39	PM6	PS3
NRF:13	24.04	PM6	PS3
NRF:14	18.62	PM8	PS3
NRF:15	30.50	PM6	PS3
NRF:16	25.48	PM8	PS3
NPC:17	28.55	PM6	PS3
NRF:18	25.59	PM6	PS3
NRF:19	24.06	PM6	PS3
NRF:20	37.17	PM6	PS3
NRF:21	23.16	PM6	PS4
NRF:22	33.13	PM6	PS2
NRF:23	37.64	PM6	PS2
NRF:24	31.77	PM9	PS4
NRF:25	25.93	PM6	PS2
NRF:26	29.93	PM6	PS2
NRF:27	28.25	PM6	PS2
NRF:28	25.06	PM6	PS2
NRF:29	31.10	PM6	PS2
NRF:30	26.11	PM6	PS2
NRF:31	34.96	PM6	PS5
NRF:32	26.24	PM6	PS5
NRF:33	38.48	PM6	PS2
NRF:34	21.13	PM6	PS3
NRF:35	31.12	PM6	PS2
NRF:36	19.34	PM6	PS2
NRF:37	20.64	PM6	PS2
NRF:38	31.29	PM6	PS2
NRF:39	19.06	PM6	PS2
NRF:40	21.29	PM6	PS2
NRF:41	26.35	PM6	PS2
NRF:42	34.16	PM6	PS4
NRF:43	20.98	PM6	PS4
NRF:44	24.03	PM6	PS2
NRF:45	25.85	PM6	PS2
NRF:46	37.29	PM6	PS2
NRF:47	19.11	PM6	PS2
NRF:48	34.51	PM6	PS2
NRF:49	18.95	PM6	PS4
NRF:50	24.06	PM6	PS4

TABLE 4.26: ARA VALUES OF MORPHOTYPE-3 *MYRICA* TREES ASSIGNED TO ITS RESPECTIVE ARP PROFILES.

SAMPLE	AVERAGE NITROGENASE ACTIVITY (n mole C_2H_2 produced / mg fresh wt/ hr)	PROFILES PCR-RFLP/ ARP *Mbo*I	PROFILES PCR-RFLP/ ARP *Sau*96I
NRF:51	25.90	PM6	PS4
NRF:52	23.16	PM6	PS2
NRF:53	22.19	PM6	PS2
NRF:54	22.53	PM6	PS2
NRF:55	19.91	PM6	PS4
NRF:56	24.27	PM6	PS6
NRF:57	36.08	PM6	PS2
NRF:58	46.14	PM6	PS2
NRF:59	40.87	PM8	PS3
NRF:60	15.01	PM6	PS2
NRF:61	28.30	PM6	PS5

produced/ mg fresh wt/ hr. Sample NRF:42 also shared the same profile combination but the nitrogenase activity was recorded a little above the standard value at 34.16 n moles of ethylene produced/ mg fresh wt/ hr. Nonetheless, the profile combination PM6:PS4 could be considered as a molecular marker to weed out *Myrica* trees with low nitrogenase activity. This profile combination was not found in any tree with high nitrogenase activity.

4.11 DISCUSSION:

It was interesting to note that the nitrogenase activity estimated by ARA (Stewart *et al.*, 1968) recorded an average high value for morphotype-1 trees growing at NEHU permanent campus compared to morphotype-2 and morphotype-3 trees at Nongkrem reserve forest. The mean soil nitrogen content of soil samples collected at NEHU campus recorded less value (0.178%) when compared with those collected at Nongkrem forest where the mean soil nitrogen percentage was recorded to be higher (0.628%). Nitrogen fixation in the symbiotic actinorhizal plants is brought about by the oxygen-sensitive nitrogenase enzyme present in the microsymbiont *Frankia.* The activity of the nitrogenase enzyme is controlled by *nif* genes located in the *Frankia* depending on the availability or non-availability of nitrogen in its surrounding. In times of nitrogen stress the *nif* A activates transcription of the rest of the *nif* genes. If there is sufficient amount of nitrogen present, another gene *nif* L, is activated which inhibits the *nif* A activity resulting in the inhibition of the nitrogenase enzyme. The switch on/ off mechanism of nitrogenase enzyme is affected in this way. In this context, the studies of Dixon and

Wheeler (1986) in *Klebsiella* needs special mention. Their results revealed that when ammonia was present at low concentration, the nitrogen-fixing system was switched on because the *nif* genes were sensitive to the concentration of combined nitrogen. Mutation in the *nif* L gene allowed the expression of the *nif* genes even at higher concentrations of nitrogen. On the other hand, at high combined nitrogen concentrations the organism repressed the necessary genes required to assimilate ammonia, an otherwise energy-demanding process. Therefore, the low soil nitrogen content coupled with high nitrogenase activity of the morphotype-1 trees at NEHU campus and high nitrogen soil content coupled with low nitrogenase activity of the morphotype-2 and morphotype-3 trees at Nongkrem forest can be explained on the basis of the available nitrogen concentration in the surrounding soil which in turn directly influences the nitrogenase activity of the actinomycete *Frankia* present in the root nodules of the *Myrica* tree. The mean soil nitrogen percentage at Nongkrem forest was found to be more than the NEHU campus soil.

In order to develop molecular markers for screening out *Myrica* trees with regard to high or low nitrogenase activity, a table was prepared in which the Amplicon Restriction Patterns (ARP) were assigned to respective Acetylene Reduction Assay (ARA) values. Based on the results obtained from the comparisons we could develop molecular marker to identify samples with relatively high or low nitrogenase activity. ARP profile PM3:PS0 could serve as a molecular marker for screening *Myrica* trees with high nitrogenase activity. Propagation of superior genotypes based on the molecular marker assisted selection would be helpful in selection of *Myrica* trees with

high nitrogenase activity that could be used to replenish the nitrogen deficient soil especially in the agro-forestry sector. On the other hand molecular marker (PM6:PS4) obtained for trees (NRF:43, NRF:49, NRF:50 and NRF:55) with low nitrogenase activity could be used in weeding out inferior genotypes. This would also serve the purpose of cross checking before the superior genotypes are considered for further propagation.

Thus, in the present investigation we were able to develop molecular markers to select *Myrica* trees supporting high nitrogenase activity. At the same time we could also develop molecular marker to weed out *Myrica* trees with low nitrogenase activity.

CONCLUSION

CHAPTER 5

CONCLUSION

1. Based on the molecular data, the phylogenetic relationship between the different morphotypes of *Myrica* trees of Meghalaya could be established. Comparative study of sequence homology revealed a high level of divergence between trees of morphotype-1 on one hand and trees of morphotypes 2 and 3 on the other.

   It is proposed that morphotype-1 trees be classified as *Myrica nagi* (or *Morella nagi* as proposed by Wilbur, 1994) and trees of morphotypes-2 and 3 be classified as varieties of *Myrica esculenta* (or *Morella esculenta*)

2. We attempted developing PCR-RFLP/ ARP markers using 18S-28S ITS for possible screening of trees with higher nitrogenase activity.

   Multisite marker PM3:PS0 may be used for selecting seedlings likely to support higher nitrogenase activity. Multisite marker PM6:PS4 on the other hand may be used for weeding out seedlings likely to support low nitrogenase activity.

REFERENCES

REFERENCES

Abbe, E. C. (1963). The male flowers and inflorescence of the Myricaceae. *American Journal of Botany.* **50**: 632.

Adams, C. D. (1972). Flowering plants of Jamaica. University of the West Indies, Mona, Jamaica.

Akopyanz, N., Bukanov, N., Westblom, T. U. and Berg, D. E. (1992). PCR-based RFLP analysis of DNA diversity in the gastric pathogen *Helicobacter pyroli*. *Nucleic Acids Research.* **20**: 6221-6225.

Arora, R. K. (1985). Genetic resources of less known cultivated food plants, NBPGR Sci. Monogr. No. 9, New Delhi.

Arora, R. K. and Pandey, A. (1996). Wild edible plants of India: Diversity, conservation and use, National Bureau of Plant Genetic Resources, New Delhi.

Bagsara, O. and Hansen, J. (1997). Review of the PCR technique. *In situ PCR techniques.* pp1-16.

Baird, J. R. (1968). A taxonomic revision of the plant family Myricaceae of North America, North of Mexico. Ph.D. thesis, The University of North Carolina.

Bakker, F. T., Olsen, J. L. and Stam, W. T. (1995). Evolution of nuclear rDNA ITS sequences in the *Cladophora albida/sericea* clade (Chlorophyta). *Journal of Molecular Evolution.* **40**: 640-651.

Baldwin, B. G. (1992). Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the compositae. *Molecular Phylogenetics and Evolution.* **1**: 3-16.

Baldwin, B. G. and Markos, S. (1998). Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Molecular Phylogenetics and Evolution.* **10**: 449-463.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S. and Donoghue, M. J. (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden.* **82**: 247-277.

Barua, U. and Srivastava, R. (2003). *Myrica esculenta* Buch.-Hum. –A lesser known wild edible fruit of Meghalaya. *Journal of the North Eastern Council* (October-December). **23** (4): 17-20.

Bhatt, I. D., Rawal, R. S., Dhar, U. (2000). Improvement in seed germination of *Myrica esculenta* Buch.-Ham. ex D. Don - a high value tree species of Kumaun Himalaya, India. *Seed Science and Technology.* **28**: 597-605.

Bor, N. L. (1953). Manual of Indian Forest Botany. Geoffrey Cumberiege: Oxford University Press: p112.

Bosco, M., Fernandez, M. P., Simonet P., Materassi, R. and Normand, P. (1992). Evidence that some *Frankia* sp. Strains are able to cross boundaries between *Alnus* and *Elaeagnus* host specificity groups. *Appl. Environ. Microbial.* **58**: 1569-1576.

Bronstein, A. J. (1997). Myricaceae. In: Flora of North America Editorial Committee (Eds.). Flora of North America North of Mexico. Volume 3, Magnoliophyta: Magnoliidae and Hamamelidae. New York: Oxford University Press: pp429-434.

Brummitt, R. K. (1999). Proposals to conserve or reject. Report of the committee for spermatophyte. *Taxon.* **48**: 367.

Burr, B., Evola, S. V., Burr, F. A. and Beckmann, J. S. (1983). The application of restriction fragment length polymorphisms to plant breeding. In: Setlow, J. K., Hollaender, A. (eds.). *Genetic engineering principles and methods*. Plenum Press. New York London. **5**: 45-59.

Bult, C. J., Sweere, J. A. and Zimmer, E. A. (1995). Cryptic sequence simplicity, nucleotide composition bias, and molecular coevolution in the large subunit of ribosomal DNA in plants: Implications for phylogenetic analyses. *Annals of the Missouri Botanical Garden.* **82**: 235-246.

Caetano-Anolles, G., Bassam, B. J. and Gresshoff, P. M. (1991). DNA Amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/Technol.* **9**: 553-557.

Campbell, B. C., Steffen-Campbell, J. D. and Werren, J. H. (1993). Phylogeny of *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and 28S rDNA sequences. *Insect Molecular Biology* **2**: 225-237.

Carranza, S., Giribet, G., Ribera, C., Baguña, J. and Riutort, M. (1996). Evidence that two types of 18S rDNA coexist in the genome of *Dugesia (Schmidtea)*

mediterranea (Platyhelminthes, Turbellaria, Tricladid). *Molecular Biology and Evolution*. **13**: 824-832.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qui, Y. -L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Systma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedren, M., Gaut, B. S., Jansen, R. K., Kim, K, -J., Wimpee, C. F., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q. -Y., Plunkett, G. M., Soltis, P. S., Swensen, S. M., Williams, S. E., Gadek, P. A., Quinn, C. J., Equiarte, L. E., Golenberg, E., Learn, JR., G. H., Graham, S. W., Barrett, S. C. H., Dayanandan, S. and Albert, V. A. (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden*. **80**: 528-580.

Chauhan, V. S. (2000). Generation of molecular signatures for *Alnus nepalensis* genotypes with high nitrogenase activity in symbiotic association with *Frankia*. Ph.D. thesis, North Eastern Hill University, Shillong, India.

Chauhan, V. S. and Misra, A. K. (2002). Development of molecular markers for early screening of *Alnus nepalensis* (D. Don) genotypes for the nitrogenase activity of actinorhizal root nodules. *Molecular Genetics and Genomics*. **267**: 303-312.

Chen, H. B., Martin, J. M., Lanvin, M. and Talbert, L. E. (1994). Genetic diversity in hard red spring wheat based on Sequence-Tagged-Site PCR markers. *Crop Science*. **34**: 1628-1632.

Chevalier, A. (1901). Monographie des Myricaceae; anatomie et histology, organographies, classification et description des espèces, distribution géographique. *Mém. Soc. Natl. Sci. Nat. Math. Cherbourg.* **32**: 85-340.

Chourey, M. S. (1974). A study of the Myricaceae from Eocene sediments of Southeastern North America. *Palaeontographica.* **146**: 88-153.

Clegg, M. T. (1993). Chloroplast gene sequences and the study of plant evolution. *Proceedings of the National Academy of Sciences.* U.S.A. **90**: 363-367.

Coleman, A. W., Suarez, A. and Goff, L. J. (1994). Molecular delineation of species and syngens in volvocean green algae (Chlorophyta). *Journal of Molecular Evolution.* **30**: 80-90.

Crane, P. R. (1989). Early fossil history and evolution of the Betulaeae. In: Evaluation, Systematics and Fossil history of the Hamamelidae, Vo. 2. Higher Hamamelidae (Eds Crane, P. R., Blackmore, S.) Systematics Association special volume NO. 40B. 87-116. Clarendon Press, Oxford.

Cronquist, A. (1978). The evolution and classification of flowering plants. Second edition. Allen Press, Lawrence, Kansas.

Cronquist, A. (1981). An integrated system of classification of flowering plants. Columbia University Press, New York, NY. ISBN 0-231-03880-1.

Cronquist, A. (1988). The evolution and classification of flowering plants, 2nd edn. The New York Botanical Garden, New York.

Dahlgren, M. R. T. (1980). A revised system of classification of the angiosperms. *Botanical Journal of the Linnean Society.* **80**: 91-124.

Dellaporta, S. L., Wood, J. and Hicks, J. B. (1983). A plant DNA minipreparation: version II. *Plant Molecular Biology Reporter*. **1** (4): 19-21.

dePamphilis, C. W., Young, N. D. and Wolfe, A. D. (1997). Evolution of plastid gene *rps*2 in a lineage of hemiparasitic and holoparasitic plants: many losses of photosynthesis and complex patterns of rate variation. *Proceedings of the National Academy of Sciences* U.S.A. **94**: 7367-7372.

Dhyani, P. P. and Dhar, U. (1994). *Myrica esculenta* Box Myrtle Kaiphal. A promising underexploited tree crop of the Himalaya. HIMAVIKAS occasional Publication. Bishen Singh Mahendra Pal Singh. **3**: 3-8.

Dixon, R. O. D. and Wheeler, C. T. (1986). Nitrogen fixation in plants. Blackie, Glasgow and London. USA: Chapman and Hall, New York.

Eckenrode, V. K., Arnold, J. and Meagher, R. B. (1985). Comparison of the nucleotide sequence of soybean 18S rRNA with the sequences of other small-subunit rRNAs. *Journal of Molecular Evolution*. **21**: 259-269.

Elrich, H. A. (1989). PCR Technology: Principles and applications for DNA amplification. (Stockton, New York).

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*. **17**: 368-376.

Fritz, G. N., Conn, J., Cockburn, A. and Seawright, J. (1994). Sequence analysis of the ribosomal internal transcribed spacer 2 from populations of *Anopheles nuneztovari* (Diptera: Culidae). *Molecular Biology and Evolution*. **11**: 406-416.

Ghareyazie, B., Huang, N., Second, G. and Khush, G. S. (1995). Classification of rice germplasm I. Analysis using ALP and PCR based RFLP. *Theoretical Applied Genetics*. **91**: 218-227.

Gielly, I. and Taberlet, P. (1996). A phylogeny of the European Gentians inferred from chloroplast *trnL* (UAA) intron sequences. *Botanical Journal of the Linnean Society*. **120**: 57-75.

Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J. and Pace, N. R. (1988). Evolutionary relationships among cyanobacteria and green chloroplasts. *Journal of Bacteriology*. **170**: 3584-3592.

Gladkova, A. N. (1962). Fragments of the history of the Myricaceae family. *Pollen and Spores*. **4**: 345.

Gomes, E. A., Kasuya, M. C. M., de Barros, E. G., Borges, A. C. and Araújo, E. F. (2002). Polymorphism in the internal transcribed spacer (ITS) of ribosomal DNA of 26 isolates of ectomycorrhizal fungi. *Genetics and Molecular Biology*. **25** (4): 477-483.

Gusella, J. F., Keys, C., Varsanyi-Breiner, A., Kao, F-T., Jones, C., Puck, T. T. and Housman, D. (1980). *Proceedings of the National Academy of Sciences USA*. **77**: 2829-2833.

Hamby, K. R. and Zimmer, E. A. (1992). Ribosomal RNA as a phylogenetic tool in plant systematics. In *Molecular Systematics of Plants*, eds. P. Soltis and J. Doyle, pp. 50-91. Chapman & Hall, New York.

Hadrys, H., Balick, M. and Schierwater., B. (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology.* **1**: 55–63

Haridasan, K. and Rao, R. R. (1987). Forest flora of Meghalaya. Volume II. Caprifoliaceae to Salicaceae. Bishen Singh Mahendra Pal Singh. **2**: 851-852.

Harris, S. A. and Ingram, R. (1991). Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. *Taxon.* **40**: 393-412.

Harrison, R. G. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution.* **4**: 6-11.

Hershkovitz, M. A., Zimmer, E. A. and Hahn, W. J. (1999). Ribosomal DNA sequences and angiosperm systematic. In *Molecular systematic and plant evolution* (1999). Eds. P. M. Hollingsworth, R. M. Bateman and R. J. Gornall. Taylor & Francis, London. pp268-326.

Hillis, D. M. (1996). Inferring complex phylogenies. *Nature.* **383**: 130.

Hooker, J. D. (1885). Flora of British India. Vol. 5. Reeve, London.

Horn, G. T., Richards, B. and Klinger, K. W. (1989). Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nucleic Acids Research.* **17**: 2140.

Huguet, V., Gouy, M., Normand, P., Zimpfer, J. F. and Fernandez, M. P. (2005). Molecular phylogeny of Myricaceae: a reexamination of host-symbiont specificity. *Molecular Phylogenies and Evolution.* **34**: 557-568.

Index Kewensis (1977-1985 Rep.). Vol.II and supplements 1-15. Otto Koeltz Science Publ., Germany.

Jaccard, P. (1908). Nouvelles researches sur la distribution florate- *Bull. Soc. Vaudoise Sci. Nat.* **44**: 223-270.

Jackson, M. L. (1974). Soil chemical analysis-advanced course, 2nd edn. Pub. by the author, Department of Soils, University of Wisconsin, Madison.

Jeffreys, A. J., Wilson, V., Newmann, R. and Keyte, J. (1988). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Research.* **16** (23): 10953-10971.

Johnson, L. A. and Soltis, D. E. (1995). Phylogenetic inference in Saxifragraceae sensu strict and *Gilia* (Polemoniaceae) using *matK* sequences. *Annals of the Missouri Botanical Garden.* **82**: 149-175.

Johnson, L. A. S. and Wilson, K. L. (1989). Casuarinaceae: a synopsis. In: Evaluation, Systematics and Fossil history of the Hamamelidae, Vol. 2. Higher Hamamelidae (Eds Crane, P.R., Blackmore, S.), Systematics Association special volume NO. 40B. 87-116. Clarendon Press, Oxford.

Joshi, S. P., Ranjekar, P. K. and Gupta, V. S. (1999). Molecular markers in plant genome analysis (review article). *Current Science.* **77** (2): 204-308.

Kanjilal, U. N. and Bor, N. L. (1940). Flora of Assam. Nyctaginaceae to Cycadaceae. **4**: 302-303.

Karp, A., Seberg, O. and Buiatti, M. (1996). Molecular techniques in the assessment of botanical diversity. *Annals of Botany.* **78**: 143-149.

Khan, A., Myrold, D. D. and Misra, A. K. (2007). Distribution of *Frankia* genotypes occupying *Alnus nepalensis* nodules with respect to altitude and soil characteristics in the Sikkim Himalayas. *Physiologia Plantarum*. 130: 364-371.

Kim, K, -J. and Jansen, R. K. (1995). *ndhF* sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Science U.S.A.* **92**: 10379-10383.

Kirtikar, K. R. and Basu, B. D. (1984 Rep.). Indian medicinal plants, Vol. III. Bishen Singh Mahendra Pal Singh, Dehradun.

Kogan, S. C., Doherty, M. and Gitschier, J. (1987). An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. Application to hemophilia A. *New England Journal of Medicine*. **317**: 985-990.

Lakobashvili, R. and Lapidot, A. (1999). Low temperature cycled PCR protocol for Klenow fragment of DNA polymerase I in the presence of proline. *Nucleic Acids Research*. **27** (6): 1566-1568.

Lashermes, P., Combes, M. C., Trouslot, P. and Charrier, A. (1997). Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical Applied Genetics*. **94**: 947-955.

MacDonald, A. D. (1977). Myricaceae: Floral hypothesis for *Myrica gale* and *Comptonia*. *Canadian Journal of Botany*. **55**: 2636-2651.

MacDonald, A. D. (1989). The morphology and relationships of the Myricaceae. In: Crane, P.R., Blackmore, S. (Eds). Evolution, systematics and fossil history of the

Hamamelidae. Vol. 2: Higher Hamamelidae. *Systematics Association Special Volume No. 40B*. Clarendon Press, Oxford. pp147-165.

Maggia, L. and Bousquet, J. (1994). Molecular phylogeny of the actinorhizal Hamamelidae and relationships with host promiscuity towards *Frankia*. *Molecular Ecology*. **3**: 459-464.

Manhart, J. R. (1995). Chloroplast 16S rDNA sequences and phylogenetic relationships of fern allies and ferns. *American Fern Journal*. **85**: 182-192.

Messing, J., Carlson, J., Hagen, G., Rubenstein, I. and Oleson, A. (1984). Cloning and sequencing of the ribosomal RNA genes in maize: the 17S region. *DNA*. **3**: 31-40.

Misra, A. K. and Verghese, S. K. (2004). Some aspects of actinorhizal symbiosis. *Indian Journal of Microbiology*. **44** (4): 235-246.

Mitton, J. B. (1994). Molecular approaches to population biology. *Annual Review of Ecology and Systematics*. **25**: 45-69.

Möller, M. and Cronk, Q. C. B. (1997). Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany*. **84**: 956-965.

Mullis, K. B. and Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*. **155**: 335-350.

Murai, S. (1964). Phytotaxonomical and geobotanical studies on so-called genus *Alnus* in Japan (II). Comparative studies on all species, including shrubby species. *Bull. Gov. For. Exp. Sta. (Tokyo, Japan)*. **154**: 20-72.

Nadot, S., Bajon, R. and Lejeune, B. (1994). The chloroplast gene *rps4* as a tool for the study of poaceae phylogeny. *Plant Systematics and Evolution.* **191**: 27-38.

Nairn, C. J. and Ferl, R. J. (1988). The complete nucleotide sequence of the small-subunit ribosomal RNA coding region for the cycad *Zamia pumila:* phylogenetic implications: *Journal of Molecular Evolution.* **27**: 133-141.

Navarro, E., Bousquet, J., Moiroud, A., Munive, A., Piou, D. and Normand, P. (2003). Molecular phylogeny of *Alnus* (Betulaceae), inferred from nuclear ribosomal DNA ITS sequences. *Plant and Soil.* **254**: 207-217.

Nickrent, D. L. and Franchina, C. R. (1990). Phylogenetic relationships of the Santalales and relatives. *Journal of Molecular Evolution.* **31**: 294-301.

Nickrent, D. L. and Soltis, D. E. (1995). A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbcL* sequences. *Annals of the Missouri Botanical Garden.* **82**: 208-234.

Normand P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., Evtushenko, L. and Misra, A. K. (1996). Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. *International Journal of Systematic Bacteriology.* **46** (1): 1-9.

Ochman, H., Gerber, A. S. and Hartl, D. L. (1988). Genetic Applications of an Inverse Polymerase Chain Reaction. *Genetics.* **120** (3): 621-623.

Olmos, J., Paniagua, J and Contreras, R. (2000). Molecular identification of *Dunaliea* sp. utilizing the 18S rDNA gene. *Letters in Applied Microbiology.* **30**: 80-84.

Osmaston, A. E. (1987 Rep.). Forest flora for Kumaon. Bishen Singh Mahendra Pal Bishen Singh, Dehradun.

Pace, N. R., Olsen, G. J. and Woese, C. R. (1986). Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell.* **45**: 325-326.

Palmer, J. D. (1985a). Evolution of chloroplast and mitochondrial DNA in plants and algae. In *Molecular Evolutionary Genetics*, ed. R. J. MacLntyre. 131-240. Plenum Press, New York.

Palmer, J. D. (1985b). Comparative organization of chloroplast genomes. *Annual Review of Genetics*. **19**: 325-354.

Palmer, J. D. (1986). Isolation and structural analysis of chloroplast DNA. *Methods in Enzymlogy*. **188**: 167-186.

Palmer, J. D. (1992). Mitochondrial DNA in plant systematic: applications and limitations. In Molecular Systematics of Plants. Eds. Soltis, P.S. and Soltis, D. E. and Doyle, J. J. Chapman & Hall, New York. 36-49.

Palmer, J. D. and Hebron, L. A. (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*. **28**: 87-97.

Palmer, J. D., Jansen, R. K., Michaels, H. J., Chase, M. W. and Manhart, J. R. (1988a). Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Garden*. **75**: 1180-1206.

Pandey, A. K., Wen, J. and Pathak, M. K. (2004). Relationships in Indian Araliaceae as inferred from sequences of Internal Transcribed Spacer (ITS) regions of nuclear Ribosomal DNA. *Rheedea*. **14**: 1-8.

Para-O. C. (2001). Lectotypification and epitypification of *Morella cerifera* (L.) Small (Myricaceae). *Caldasia*. **23**: 135-137.

Pareek, O. P., Sharma, S. and Arora, A. K. (1998). Underutilised edible fruits and nuts: An inventory of genetic resources in their regions of diversity. New Delhi IPGRI.

Perry, D. J., Isabel, N. and Bousquet, J. (1999). Sequence-Tagged-Sites (STS) markers of arbitrary genes: the amount and nature of variation revealed in Norway spruce. *Heredity*. **83**: 239-248.

Pridgeon, A. M., Bateman, R. M., Cox, A. V., Hapeman, J. R. and Chase, M. W. (1997). Phylogenetics of subtribe Orchidinae (Orchidoideae, Orchidaceae) based on nuclear ITS sequences. 1. Intergeneric relationships and polyphyly of *Orchis sensu lato*. *Lindleyana*. **12**: 89-109.

Raven, P. H. and Axelrod, D. I. (1974). Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden*. **61**: 539-673.

Raven, P. H., Evert, R. F. and Eichorn, S. E. (1986). Biology of plants. (Fourth edition). Worth Publishers, Inc., New York.

Rastogi, R. P. and Mehrotra, B. N. (1991). Compendium of Indian medicinal plants (Vol. I, 1960-1969). CDRI, Lucknow and PID, New Delhi.

Rieseberg, L. H. and Brunsfeld, S. J. (1992). Molecular evidence and plant introgression. In *Molecular Systematics of Plants*., (Soltis, P. S., Soltis, D. E. and Doyle, J. J. eds.). Chapman & Hall, New York, pp151-176.

Rieseberg, L. H. and Soltis, D. E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*. **5**: 65-84.

Ritland, K. and Clegg, M. T. (1987). Evolutionary analysis of plant DNA sequences. *American Naturalist*. **130**: S74-S100.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for constructing phylogenetic trees. *Molecular Biology and Evolution.* **4:** 406-425.

Sambrook, J., Friysch, E. F. and Maniatis, T. (1989). Molecular cloning laboratory manual. Cold Spring Harbor Laboratory Press, USA.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences U.S.A.* **74**: 5463- 5467.

Savard, L. and Lalonde. M. (1991). Sequence of 18S rDNA of actinorhizal *Alnus glutinosa* (Betulaceae). *Plant Molecular Biology*. **16**: 725-728.

Savard, L., Michaud, M. and Bousquet, J. (1993). Genetic diversity and phylogenetic relationships between Birches and Alders using ITS, 18S rRNA and *rbc*L gene sequences. *Molecular Phylogenetics and Evolution.* **2** (2): 112-118.

Schlötterer, C., Hauser, M.-T., Von Haesleler, A. and Tautz, D. (1994). Comparative evolutionary analysis of rDNA ITS region in *Drosophila. Molecular Biology and Evolution.* **11**: 513-522.

Schlotterer, C. and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acid Research.* **20**: 211-215.

Schmidt-Puchta. W., Kiitemeier. G., Gonther. I., Haas. B. and Slinger. H. L. (1989). Cloning and sequence analysis of the 18S ribosomal RNA gene of tomato and

secondary structure model for the 18 S rRNA of angiosperms. *Molecular and General Genetics.* **219**: 17-25.

Schwintzer, C. R. and Tjepkema, J. D. (1990). The biology of *Frankia* and Actinorhizal plants. Academic press, California, USA.

Shi, S., Chang, H. T., Chen, Y., Qu, L. and Wen, J. (1998). Phylogeny of the Hamamelidaceae based on the ITS sequences of nuclear ribosomal DNA. *Biochemical Systematics and Ecology.* **26**: 55-69.

Simonet, P., Grosjean, M. C., Misra, A. K., Nazaret, S., Cournoyer, B. and Normand, P. (1991). *Frankia* genus-specific characterization by polymerase chain reaction. *Appl. Environ Microbiol.* **57**: 3278-3286.

Smith, J. (1977). Vascular Plant Families. Mad River Press, CA.

Soltis, D. E. and Kuzoff, R. K. (1995). Discordance between molecular and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution.* **49**: 727-742.

Soltis, D. E., Soltis, P. S., Nickrent, D. L., Johnson, L. A., Hahn, W. J., Hoot, S. B., Sweere, J. A., Kuzoff, R. K., Kron, K. A., Chase, M. W., Swensen, S. M., Zimmer, E. A., Chaw, S. M., Gillespie, L. J., Kress, W. J. and Sytsma, K. J. (1997). Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missourie Botanical Garden.* **84**: 1-49.

Soltis, P. S. and Soltis, D. E. (1998). Choosing an approach and an appropriate gene for phylogenetic analysis. In *Molecular Systematics of Plants II. DNA sequencing* (Edited by Douglas E. Soltis, Pamela S. Soltis and Jeff J. Doyle). 1-42.

Soltis, P. S. and Soltis, D. E. (1998). Molecular evolution of 18S rDNA in angiosperms: Implications for character weighting in phylogenetic analysis. In *Molecular Systematics of Plants II. DNA sequencing* (Edited by Douglas E. Soltis, Pamela S. Soltis and Jeff J. Doyle). 188-210.

Soltis, P. S. and Soltis, D. E. and Doyle, J. J. (1992). Molecular systematic of plants. Chapman & Hall, New York.

Sporne, K. R. (1975). The morphology of angiosperms. St. Martin's Press, New York.

Stewart, W. P. D., Fritzgerald, G. P. and Burris, R. H. (1968). Acetylene reduction in nitrogen fixing blue-green algae. *Arch. Microbiol.* **62**: 336-348.

Suh, Y., Thien, L. B., Reeve, H. E. and Zimmer, E. A. (1993). Molecular evolution and phylogenetic implications of Internal Transcribed Spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany.* **80** (9): 1042-1055.

Susanna, A., Nuria Garcia, J. and Soltis, P. S. (1995). Phylogenetic relationships in the tribe Cardueae (Asteraceae) based on ITS sequences. *American Journal of Botany.* **82**: 1056-1068.

Swensen, S. M. and Mullin, B. C. (1997). Phylogenetic relationships among actinorhizal plants. The impact of molecular systematics and implications for the evolution of actinorhizal symbioses. *Physiologia Plantarum.* **99**: 565-573.

Swofford, D. L. (1997). PAUP*: Phylogenetic analysis using parsimony, Ver. 4.0 Sinauer Assoc., Sunderland, Massachusetts.

Swofford, D. L., Olsen, G. J., Waddell, P. J. and Hillis, D. M. (1996). Phylogenetic inference. In *Molecular Systematics*, eds D.M. Hillis, C. Moritz and B. K. Mable, 2nd edn.. 407-514. Sinauer Assoc., Sunderland, Massachusetts.

Takaiwa, F., Oono, K. and Sugiura, M. (1984). The complete nucleotide sequence of a rice 17S rRNA gene. *Nucleic Acids Research*. **12**: 5441-5448.

Takaiwa, F., Ooono, K. and Sugiura, M. (1985). Nucleotide sequence of the 17S-25S spacer region from rice rDNA. *Plant Molecular Biology*. **4**: 355-365.

Takhtajan, A. (1969). Flowering plants: Origin and dispersal. Oliver and Boyd, Edinburgh. p310.

Takhtajan, A. (1980). Outline of the classification of flowering plants (Magnoliophyta). *Bot. Rev.* (Lancaster) **46**: 225-348.

Tanaka, T. (1976). Tanaka's cyclopedia of edible plants of the world. Keigaku Publishing Company, Tokyo, Japan.

Tanaka, Y., Dyer, T. A. and Brownlee, G. G. (1980). An improved direct RNA sequence method: its application to *Vicia fabia* 5.8S ribosomal RNA. *Nucleic Acids Research*. **8**: 1259-1272.

Thompson, W. F. and Flavell, R. B. (1988). DNAase I sensitivity of ribosomal RNA genes in chromatin and nucleolar dominance in wheat. *Journal of Molecular Biology*. **204**: 535-548.

Thorne, R. T. (1992). Classification and geography of the flowering plants. *Bot. Rev.* (Lancaster). **58**: 225-359.

Tingey, S. V., Rafalski, J. A. and Williams, J. G. K. (1993). Application of RAPD technology to plant breeding (ed. Neff, M.). ASHS Publishers, Minnesota. pp3-8.

Tjepkema, J. D., Schwintzer, C. R. and Benson, D.R. (1986). Physiology of actinorhizal nodules. *Annu. Rev. Plant Physiol.* **37**: 209-232.

Triglia, T., Peterson, M. G. and Kemp, D. J. (1988). A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Research.* **16**: 8186.

Tsumura, Y., Suyama, Y., Yoshimura, K., Shirato, N. and Mukai, Y. (1997). Sequence-Tagged-Sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evaluation as molecular markers in conifers. *Theoretical Applied Genetics.* **94**: 764-772.

Tyron, R. C. (1939). Cluster analysis. Ann. Arbor, MI: Edward Brothers.

Urbach, E., Robertson, D. L. and Chisholm, S. W. (1992). Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature.* **355**: 267-270.

Van de Peer, Y., Chapelle, S. and De Wachter, R. (1996). A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Research.* **24**: 3381-3391.

Varghese, R., Chauhan, V. S. and Misra, A. K. (2003). Hypervariable spacer regions are good sites for developing specific PCR-RFLP markers and PCR primers for screening actinorhizal symbionts. *J. Biosci.* **28**: 437-442.

Verghese, S. K. and Misra, A. K. (2000). PCR-RFLP based screening of *Frankia* in alder nodules having different levels of nitrogenase activity. *Symbiosis*: **28**: 337-350.

Vilgalys, R. and Sun, B. L. (1994). Ancient and recent patterns of geographic speciation in the oyster mushroom *Pleurotus* revealed by phylogenetic analysis of ribosomal DNA sequences. *Proceedings of the National Academy of Sciences* U.S.A. **91**: 4599-4603.

Vogler, A. P. and DeSalle, R. (1994). Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis*. *Molecular Biology and Evolution*. **11**: 393-405.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van der Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*. **23** (21): 4407-4414.

.Watt, G. (1891). A dictionary of the economic products of India. Vol. 5. Cosmo Publications, Delhi.

Weakley, A. S. (2000). Flora of the Carolinas and Virginia: Working draft of 15 May 2000. Chapel Hill, NC: Nature Conservancy, Southeast Regional Office, Southern Conservation Science Department.

Welsh, J. and McClelland, M. (1991). Genomic fingerprints produced by PCR with consensus tRNA gene primers. *Nucleic Acids Research*. **19:** 861-866.

Wen, J. and Zimmer, E. A. (1996). Phylogeny and biogeography of *Panax* L. (the Ginseng Genus, Araliaceae): References from ITS sequences of nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution.* **6** (2): 167-177.

White, T. J., Bruns, T., Lee, S. and Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR protocols: A guide to Methods and Amplifications.* (M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White eds.). Academic Press, San Diego, California. pp315-322

Wilbur, R. L. (1994). The Myricaceae of the United States and Canada: *Genera, subgenera and series*. SIDA **16** (1): 93-107.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research.* **18**: 6531-6535.

Woese, C. R. (1987). Bacterial evolution. *Microbiology Review*. **51**: 221-271.

Wolters, J. and Erdmann, V. A. (1986). Cladistic analysis of ssRNA and 16SrRNA secondary and primary structure- the evolution of eukaryotes and their rlation to Archaebacteria. *Journal of Molecular Evolution*. **24**: 152-166.

Zabeau, M. (1993). European Patent Application. Publication No. 0534858A1.

Zeven, A. C. and Wet, J. M. J. de (1982). Dictionary of cultivated plants and their regions of diversity- excluding most ornamentals, forest trees and lower plants, Centre for Agricultural Publishing and Documentation, (PUDOC), Wageningen.

Zimmer, E. A., Hamby, R. K., Arnold, M. L., LeBlanc, D. A., Theriot, E. E. (1989). Ribosomal RNA phylogenies and flowering plant evolution. Hierarchy of Life,

(Fernholm, B., Bremer, K. and Jörnvall, J. eds.). *Elsevier Science*. Amsterdam. pp205-214.

APPENDICES

APPENDIX 1

BUFFERS USED FOR EXTRACTION OF DNA AND AGAROSE GEL ELECTROPHORESIS

1. DNA EXTRACTION BUFFER (pH 8.0)

COMPONENT	AMOUNT/100ML
TRIS BASE (0.1 M/ 100 mM)	50 ml
EDTA (0.1 M/ 100 mM)	20 ml
NaCl (1.4 M)	8.18 gm
CTAB (2%)	2 gm (wt/ vol)
PVP (1%)	1 gm (wt/ vol)

CTAB : Cetyl Trimethyl Ammonium Bromide

PVP : Polyvinypyrollidone

2. 5X TBE BUFFER (pH 8.0)

COMPONENT	AMOUNT/1000ML
TRIS BASE	54.0 gm
BORIC ACID	27.50 gm
EDTA (0.5 M)	20 ml

3. TYPE III LOADING BUFFER (6X)

Bromophenol blue	0.25% (wt/ vol)
Xylene Cyanol FF	0.25% (wt/ vol)
Glycerol	30% in water (wt/ vol)

APPENDIX 2

BUFFERS USED FOR PCR AND RESTRICTION ANALYSIS

1. **10X PCR BUFFER** (pH 8.3 at 25°C)

Tris-HCl	100 mM
KCl	500 mM
$MgCl_2$	15 mM
Gelatin	0.01% (wt/ vol)

2. **BUFFER B [BANGALORE GENEI]** (pH 8.0)

Tris-HCl	10 mM
NaCl	100 mM
$MgCl_2$	10 mM
2-mercaptoethanol	5 mM

3. **NEBUFFER 4 [NEW ENGLAND BIOLABS]** (pH 7.9)

Tris-acetate	20 mM
Potassium acetate	50 mM
Magnesium acetate	10 mM
Dithiothreitol	1 mM

APPENDIX 3

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	Morellanagi	762	2	Morellaadenophora	804	98
1	Morellanagi	762	3	Morellarubra	796	97
1	Morellanagi	762	4	Morellaesculenta	803	97
1	Morellanagi	762	5	Morellaheterophylla	762	98
1	Morellanagi	762	6	Morellapensylvanica	762	98
1	Morellanagi	762	7	Morellacerifera	762	98
1	Morellanagi	762	8	Morellacalifornica	762	97
1	Morellanagi	762	9	Morellarivas-martinez	762	97
1	Morellanagi	762	10	Morellafaya	762	97
1	Morellanagi	762	11	Morellaspathulata	762	97
1	Morellanagi	762	12	Myricagale	761	96
2	Morellaadenophora	804	3	Morellarubra	796	98
2	Morellaadenophora	804	4	Morellaesculenta	803	98
2	Morellaadenophora	804	5	Morellaheterophylla	762	96
2	Morellaadenophora	804	6	Morellapensylvanica	762	96
2	Morellaadenophora	804	7	Morellacerifera	762	96
2	Morellaadenophora	804	8	Morellacalifornica	762	96
2	Morellaadenophora	804	9	Morellarivas-martinez	762	96
2	Morellaadenophora	804	10	Morellafaya	762	96
2	Morellaadenophora	804	11	Morellaspathulata	762	96
2	Morellaadenophora	804	12	Myricagale	761	95
3	Morellarubra	796	4	Morellaesculenta	803	97
3	Morellarubra	796	5	Morellaheterophylla	762	96
3	Morellarubra	796	6	Morellapensylvanica	762	96
3	Morellarubra	796	7	Morellacerifera	762	96
3	Morellarubra	796	8	Morellacalifornica	762	96
3	Morellarubra	796	9	Morellarivas-martinez	762	96
3	Morellarubra	796	10	Morellafaya	762	96
3	Morellarubra	796	11	Morellaspathulata	762	96
3	Morellarubra	796	12	Myricagale	761	95
4	Morellaesculenta	803	5	Morellaheterophylla	762	96
4	Morellaesculenta	803	6	Morellapensylvanica	762	96
4	Morellaesculenta	803	7	Morellacerifera	762	96
4	Morellaesculenta	803	8	Morellacalifornica	762	95
4	Morellaesculenta	803	9	Morellarivas-martinez	762	95
4	Morellaesculenta	803	10	Morellafaya	762	95
4	Morellaesculenta	803	11	Morellaspathulata	762	95
4	Morellaesculenta	803	12	Myricagale	761	94
5	Morellaheterophylla	762	6	Morellapensylvanica	762	100
5	Morellaheterophylla	762	7	Morellacerifera	762	99
5	Morellaheterophylla	762	8	Morellacalifornica	762	98
5	Morellaheterophylla	762	9	Morellarivas-martinez	762	99
5	Morellaheterophylla	762	10	Morellafaya	762	99
5	Morellaheterophylla	762	11	Morellaspathulata	762	99
5	Morellaheterophylla	762	12	Myricagale	761	96
6	Morellapensylvanica	762	7	Morellacerifera	762	99
6	Morellapensylvanica	762	8	Morellacalifornica	762	98

6	Morellapensylvanica	762	9	Morellarivas-martinez	762	99
6	Morellapensylvanica	762	10	Morellafaya	762	99
6	Morellapensylvanica	762	11	Morellaspathulata	762	99
6	Morellapensylvanica	762	12	Myricagale	761	96
7	Morellacerifera	762	8	Morellacalifornica	762	98
7	Morellacerifera	762	9	Morellarivas-martinez	762	99
7	Morellacerifera	762	10	Morellafaya	762	99
7	Morellacerifera	762	11	Morellaspathulata	762	99
7	Morellacerifera	762	12	Myricagale	761	96
8	Morellacalifornica	762	9	Morellarivas-martinez	762	98
8	Morellacalifornica	762	10	Morellafaya	762	98
8	Morellacalifornica	762	11	Morellaspathulata	762	98
8	Morellacalifornica	762	12	Myricagale	761	96
9	Morellarivas-martinez	762	10	Morellafaya	762	100
9	Morellarivas-martinez	762	11	Morellaspathulata	762	98
9	Morellarivas-martinez	762	12	Myricagale	761	96
10	Morellafaya	762	11	Morellaspathulata	762	98
10	Morellafaya	762	12	Myricagale	761	96
11	Morellaspathulata	762	12	Myricagale	761	96

```
morellaheterophylla         --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
morellapensylvanica         --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
morellaspathulata           --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
morellacerifera             --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
morellarivas-martinez       --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
morellafaya                 --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
morellacalifornica          --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
myricagale                  --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
morellaadenophora           GGCCGCGGGAATTCGATTGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGG 50
morellarubra                GGCCGCGGGAATTCGATTGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGG 50
morellaesculenta            GGCCGCGGGAATTCGATTGGAAGTTAAAGTCGTAACAAGGTTTCCGTAGG 50
morellanagi                 --------CATTTAGA--GGAAGGAGAAGTCGTAACAAGGTTTCCGTAGG 40
                                     * ** **  *****   **********************

morellaheterophylla         TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
morellapensylvanica         TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
morellaspathulata           TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
morellacerifera             TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
morellarivas-martinez       TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
morellafaya                 TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
morellacalifornica          TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
myricagale                  TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
morellaadenophora           TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 100
morellarubra                TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 100
morellaesculenta            TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGT 100
morellanagi                 TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 90
                            **************************************************
```

```
morellaheterophylla       GAACATGTTAATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCGGTC 140
morellapensylvanica       GAACATGTTAATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCGGTC 140
morellaspathulata         GAACATGTTAATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCGGTC 140
morellacerifera           GAACATGTTAATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCGGTC 140
morellarivas-martinez     GAACATGTTAATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCGGTC 140
morellafaya               GAACATGTTAATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCGGTC 140
morellacalifornica        GAACATGTTAATAACTTTCGGGGGCAGGGGGCGATCAAAAGCCTCCTGTC 140
myricagale                GAACATGTTAATAACTACCGGGGGCGGGGGGCGATCAAAAGCCTCCCGTC 140
morellaadenophora         GAACATGTTAATAACTACCGGGGGCAGAGGGCGATCAAAAGCCTCCCGTC 150
morellarubra              GAACATGTTAATAACTACTGGGGGCAGGGGGCGATCAAAAGCCTCCCGTC 150
morellaesculenta          GAACATGTTAATAACTACCGGGGGCGGGGGGCGATCAAAAGCCTCCTGTC 150
morellanagi               GAACATGTTAATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCCGTC 140
                          ***************   ****** * ****************** ***

morellaheterophylla       CCCAAAACGGTCGGGGAGCATGTGCTGTTGTCCCGTCGGCCCTCGGGCGC 190
morellapensylvanica       CCCAAAACGGTCGGGGAGCATGTGCTGTTGTCCCGTCGGCCCTCGGGCGC 190
morellaspathulata         CCCAAAACGGTCGGGGAGCATGTGCTGTTGTCCCGTCGGCCCTCGGGCGC 190
morellacerifera           CCCAAAACGGTCGGGGAGCATGTGCTGTTGTCCCGTCGGCCCTCGGGCGC 190
morellarivas-martinez     CCCAAAACGGTTGGGGAGCATGTGCTGTTGTCCCGTCGGCCCTCGGGCGC 190
morellafaya               CCCAAAACGGTTGGGGAGCATGTGCTGTTGTCCCGTCGGCCCTCGGGCGC 190
morellacalifornica        CCCAAAACGGTTGGGGAGCATGTGCTGTTATCCCGTCGGCCCTCGGGCGC 190
myricagale                CCCAAAACGGTTGGGGAGCATGTGCTGTTAACCTGTCGGCCCTCGGGTGC 190
morellaadenophora         CCCAAAACGGTTGGGGAGCCCGTGCTGTTGTCCTGTTGGCCCTCTGGCGC 200
morellarubra              CCCAAAACGGTTGGGGAGCACGTGCTGTTGTCCTGTCGGCCCTCTGGTGC 200
morellaesculenta          CCCAAAACGGTTGGGGAGCACGTGCTGTTGTCCTGTCGGCCCTCTGGCGC 200
morellanagi               CCCAAAACGGTTGGGGAGCACGTGCTGTTGTCCTGTCGGCCCTCTGGCGC 190
                          ********** *******  ********  ** ** ******* ** **

morellaheterophylla       GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
morellapensylvanica       GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
morellaspathulata         GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
morellacerifera           GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
morellarivas-martinez     GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
morellafaya               GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
morellacalifornica        GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
myricagale                GGACGGCAAAC-ACATGTACGTGTCCCCCCAGCCGAACAACGAACCCCGG 239
morellaadenophora         GGACAGGGAACCACATGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 250
morellarubra              GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 250
morellaesculenta          GGACAAGGAACCACACGCACGTGTCCCCCCAATCAAACAACGAACCCCGG 250
morellanagi               GGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
                          ****  * *** *** * *************  * ***************

morellaheterophylla       CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCG 290
morellapensylvanica       CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCG 290
morellaspathulata         CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCG 290
morellacerifera           CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCG 290
morellarivas-martinez     CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCG 290
morellafaya               CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCG 290
morellacalifornica        CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCTGATTGCCCCG 290
myricagale                CGCGGACTGCGCCAAGGAATTTCAACAAAAGAGTGCCTCTGATCGCCCCG 289
morellaadenophora         CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGAAGGCCCCG 300
morellarubra              CGCGGACTGCGCCGAGGAACTTCAACAAAAGAGTGCCTCCGAAGGCCCCG 300
morellaesculenta          CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCG 300
morellanagi               CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGAAGGCCCCG 290
                          ************* ***** ******************** **  ******
```

```
morellaheterophylla     GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 340
morellapensylvanica     GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 340
morellaspathulata       GAAACGGTGTGCGTCGGATGGGACGTCTTGACTGTTATACAAAACGACTC 340
morellacerifera         GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 340
morellarivas-martinez   GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 340
morellafaya             GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 340
morellacalifornica      GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 340
myricagale              GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 339
morellaadenophora       GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 350
morellarubra            GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 350
morellaesculenta        GAAACGGTGTGCGTTGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 350
morellanagi             GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 340
                        ********** ** ********************************

morellaheterophylla     TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 390
morellapensylvanica     TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 390
morellaspathulata       TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 390
morellacerifera         TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 390
morellarivas-martinez   TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 390
morellafaya             TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 390
morellacalifornica      TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 390
myricagale              TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 389
morellaadenophora       TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 400
morellarubra            TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 400
morellaesculenta        TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 400
morellanagi             TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 390
                        **************************************************

morellaheterophylla     CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 440
morellapensylvanica     CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 440
morellaspathulata       CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 440
morellacerifera         CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 440
morellarivas-martinez   CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 440
morellafaya             CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 440
morellacalifornica      CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 440
myricagale              CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 439
morellaadenophora       CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 450
morellarubra            CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 450
morellaesculenta        CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 450
morellanagi             CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 440
                        **************************************************

morellaheterophylla     GCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTGCCTGGGTGTC 490
morellapensylvanica     GCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTGCCTGGGTGTC 490
morellaspathulata       GCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTGCCTGGGTGTC 490
morellacerifera         GCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTGCCTGGGTGTC 490
morellarivas-martinez   GCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTGCCTGGGTGTC 490
morellafaya             GCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTGCCTGGGTGTC 490
morellacalifornica      GCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCCTGGGTGTC 490
myricagale              GCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCCTGGGTGTC 489
morellaadenophora       GCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCCTGGGTGTC 500
morellarubra            GCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCCTGGGTGTC 500
morellaesculenta        GCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCCTGGGTGTC 500
morellanagi             GCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCCTGGGTGTC 490
                        ****************** ******************************
```

```
morellaheterophylla       ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTGGGGAC 540
morellapensylvanica       ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTGGGGAC 540
morellaspathulata         ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAA 540
morellacerifera           ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
morellarivas-martinez     ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
morellafaya               ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
morellacalifornica        ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
myricagale                ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTGGGGAC 539
morellaadenophora         ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGATTTCGGGGAC 550
morellarubra              ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGATTTCGGGGAC 550
morellaesculenta          ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGATTTCGGGGAC 550
morellanagi               ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGATTTCGGGGAC 540
                          **************************************** ** *****

morellaheterophylla       TATCGGGGCGGACATTGGCCTCCCGTGAGCCAGTTCTCGCGGTTAGCCTA 590
morellapensylvanica       TATCGGGGCGGACATTGGCCTCCCGTGAGCCAGTTCTCGCGGTTAGCCTA 590
morellaspathulata         TATCGGGGCGGACATTGGCCTCCCGTGAGCCAGTTCTCGCGGTTAGCCTA 590
morellacerifera           TATCGGGGCGGACATTGGCCTCCCGTGAGCCAGTTCTCGCGGTTAGCCTA 590
morellarivas-martinez     TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 590
morellafaya               TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 590
morellacalifornica        TATGGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 590
myricagale                TATCGGGGCGGACATTGGCCTCCCGTGAGGTAGTTCTCGCGGTTAGCCTA 589
morellaadenophora         TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 600
morellarubra              TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 600
morellaesculenta          TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 600
morellanagi               TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 590
                          *** ************************  ********************

morellaheterophylla       AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 640
morellapensylvanica       AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 640
morellaspathulata         AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 640
morellacerifera           AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 640
morellarivas-martinez     AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAGC 640
morellafaya               AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAGC 640
morellacalifornica        AATACGAGTCCTCGGCGACGAGCGCCATGACAATCGGTGGTTGATAAAGC 640
myricagale                AATACGAGTCCTCGGCGATGAGCGCCACGACAATCGGTGGTTGATAAAGC 639
morellaadenophora         AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 650
morellarubra              AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 650
morellaesculenta          AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 650
morellanagi               AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 640
                          ****************** ******** ***************** *****

morellaheterophylla       CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGCGTTCTCTGTGACC 690
morellapensylvanica       CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGCGTTCTCTGTGACC 690
morellaspathulata         CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGCGTTCTCTGTGACC 690
morellacerifera           CCTCGTTTCCCGTCGTGCGTGCTTCGTCTCCCTATGCGTTCTCTGTGACC 690
morellarivas-martinez     CCTCGTTTCCCGTCGTGGGTGCGTCGTCTCCATATGCGTTCTCTGTGACC 690
morellafaya               CCTCGTTTCCCGTCGTGGGTGCGTCGTCTCCATATGCGTTCTCTGTGACC 690
morellacalifornica        CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGTGTTCTCTGTGACC 690
myricagale                CCTCGTTTCCAGTCGTGCGCGCCTCGTTGCCCTACGTGTGCTCCGTGACC 689
morellaadenophora         CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGTGCGCTCTGTGACC 700
morellarubra              CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGTGCGCTCTGTGACC 700
morellaesculenta          CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGTGCGCTCTGTGACC 700
morellanagi               CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGTGCGCTCTGTGACC 690
                          ********** ****** * ** ****  ** ** * *  *** ******
```

```
morellaheterophylla       CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 740
morellapensylvanica       CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 740
morellaspathulata         CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 740
morellacerifera           CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 740
morellarivas-martinez     CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 740
morellafaya               CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 740
morellacalifornica        CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 740
myricagale                CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 739
morellaadenophora         CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 750
morellarubra              CTGCCGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 750
morellaesculenta          CTGCTGTGTCGTTCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 750
morellanagi               CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTACAGGCGG 740
                          **** ******* ****************************   *  **

morellaheterophylla       ACTACCCGCTGAGTTTAAGCAT---------------------------- 762
morellapensylvanica       ACTACCCGCTGAGTTTAAGCAT---------------------------- 762
morellaspathulata         ACTACCCGCTGAGTTTAAGCAT---------------------------- 762
morellacerifera           ACTACCCGCTGAGTTTAAGCAT---------------------------- 762
morellarivas-martinez     ACTACCCGCTGAGTTTAAGCAT---------------------------- 762
morellafaya               ACTACCCGCTGAGTTTAAGCAT---------------------------- 762
morellacalifornica        ACTACCCGCTGAGTTTAAGCAT---------------------------- 762
myricagale                ACTACCCGCTGAGTTTAAGCAT---------------------------- 761
morellaadenophora         ACTACCCGCTGAGTTTAAGCATATCAATAAGCGGAAGGAA-TCACTAGTG 799
morellarubra              ACTACCCGCTGAGTTTAAGCGTATCAATAAGCGGAAGGAAATCACT---- 796
morellaesculenta          ACTACCCGCTGAGTTTAAGCATATCAATAAGCGGAAGGAAATCACTAGTG 800
morellanagi               ACTACCCGCTGAGTTTAAGCAT---------------------------- 762
                          ******************** *

morellaheterophylla       -----
morellapensylvanica       -----
morellaspathulata         -----
morellacerifera           -----
morellarivas-martinez     -----
morellafaya               -----
morellacalifornica        -----
myricagale                -----
morellaadenophora         TATTC 804
morellarubra              -----
morellaesculenta          TAT-- 803
morellanagi               -----
```

APPENDIX 4

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	a. viridis	608	2	a. trabeculosa	539	96
1	a. viridis	608	3	a. tenuifolia	608	94
1	a. viridis	608	4	a. sibirica	608	97
1	a. viridis	608	5	a. serrulata	608	96
1	a. viridis	608	6	a. rubra	499	96
1	a. viridis	608	7	a. orientalis	609	96
1	a. viridis	608	8	a. nepalensis	539	96
1	a. viridis	608	9	a. japonica	609	96
1	a. viridis	608	10	a. incana	607	96
1	a. viridis	608	11	a. rugosa	529	96
1	a. viridis	608	12	a. cordata	608	97
1	a. viridis	608	13	a. glutinosa	608	97
2	a. trabeculosa	539	3	a. tenuifolia	608	94
2	a. trabeculosa	539	4	a. sibirica	608	98
2	a. trabeculosa	539	5	a. serrulata	608	98
2	a. trabeculosa	539	6	a. rubra	499	98
2	a. trabeculosa	539	7	a. orientalis	609	99
2	a. trabeculosa	539	8	a. nepalensis	539	98
2	a. trabeculosa	539	9	a. japonica	609	98
2	a. trabeculosa	539	10	a. incana	607	97
2	a. trabeculosa	539	11	a. rugosa	529	97
2	a. trabeculosa	539	12	a. cordata	608	99
2	a. trabeculosa	539	13	a. glutinosa	608	98
3	a. tenuifolia	608	4	a. sibirica	608	96
3	a. tenuifolia	608	5	a. serrulata	608	94
3	a. tenuifolia	608	6	a. rubra	499	96
3	a. tenuifolia	608	7	a. orientalis	609	94
3	a. tenuifolia	608	8	a. nepalensis	539	94
3	a. tenuifolia	608	9	a. japonica	609	94
3	a. tenuifolia	608	10	a. incana	607	95
3	a. tenuifolia	608	11	a. rugosa	529	95
3	a. tenuifolia	608	12	a. cordata	608	95
3	a. tenuifolia	608	13	a. glutinosa	608	96
4	a. sibirica	608	5	a. serrulata	608	98
4	a. sibirica	608	6	a. rubra	499	99
4	a. sibirica	608	7	a. orientalis	609	98
4	a. sibirica	608	8	a. nepalensis	539	98
4	a. sibirica	608	9	a. japonica	609	98
4	a. sibirica	608	10	a. incana	607	99
4	a. sibirica	608	11	a. rugosa	529	99
4	a. sibirica	608	12	a. cordata	608	99
4	a. sibirica	608	13	a. glutinosa	608	100
5	a. serrulata	608	6	a. rubra	499	98
5	a. serrulata	608	7	a. orientalis	609	98
5	a. serrulata	608	8	a. nepalensis	539	98
5	a. serrulata	608	9	a. japonica	609	99
5	a. serrulata	608	10	a. incana	607	98

5	a. serrulata	608	11	a. rugosa	529	98
5	a. serrulata	608	12	a. cordata	608	99
5	a. serrulata	608	13	a. glutinosa	608	98
6	a. rubra	499	7	a. orientalis	609	98
6	a. rubra	499	8	a. nepalensis	539	97
6	a. rubra	499	9	a. japonica	609	98
6	a. rubra	499	10	a. incana	607	99
6	a. rubra	499	11	a. rugosa	529	99
6	a. rubra	499	12	a. cordata	608	98
6	a. rubra	499	13	a. glutinosa	608	99
7	a. orientalis	609	8	a. nepalensis	539	98
7	a. orientalis	609	9	a. japonica	609	98
7	a. orientalis	609	10	a. incana	607	98
7	a. orientalis	609	11	a. rugosa	529	98
7	a. orientalis	609	12	a. cordata	608	99
7	a. orientalis	609	13	a. glutinosa	608	98
8	a. nepalensis	539	9	a. japonica	609	98
8	a. nepalensis	539	10	a. incana	607	97
8	a. nepalensis	539	11	a. rugosa	529	98
8	a. nepalensis	539	12	a. cordata	608	99
8	a. nepalensis	539	13	a. glutinosa	608	98
9	a. japonica	609	10	a. incana	607	98
9	a. japonica	609	11	a. rugosa	529	98
9	a. japonica	609	12	a. cordata	608	99
9	a. japonica	609	13	a. glutinosa	608	98
10	a. incana	607	11	a. rugosa	529	99
10	a. incana	607	12	a. cordata	608	98
10	a. incana	607	13	a. glutinosa	608	99
11	a. rugosa	529	12	a. cordata	608	98
11	a. rugosa	529	13	a. glutinosa	608	99
12	a. cordata	608	13	a. glutinosa	608	99

```
a. trabeculosa    ------------------------------------------------------------
a. nepalensis     ------------------------------------------------------------
a. orientalis     TCGAAACCTGCCCAGCAGAACGACCCGCGAACCTGTCACAACAACTGGGGGCGGGGGGCG 60
a. cordata        TCGAAACCTGCCCAGCAGAACGACCCGCGAACCTGTCACAACAACTGGGGGCGGGGGGCG 60
a. serrulata      TCGAAACCTGCCCAGCAGAACGACCCGCGAACCCGTCACAACAACTGGGGGCGGGGGGCG 60
a. japonica       TCGAAACCTGCCCAGCAGAACGACCCGCGAACCCGTCACAACAACTGGGGGCGGGGGGCG 60
a. viridis        TCGAAACCTGCCCAGCAGAACGACCCGCGAACCTGTCACAACAACTGGGGGCGGGGGGCG 60
a. sibirica       TCGAAACCTGCCCAGCAGAACGACCCGCGAACCTGTCACAACAACTGGGGGCGGGGGGCG 60
a. glutinosa      TCGAAACCTGCCCAGCAGAACGACCCGCGAACCTGTCACAACAACTGGGGGCGGGGGGCG 60
a. rubra          ------------------------------------------------------------
a. rugosa         ------------------------------------------------------------
a. incana         TCGAAACCTGCCCAGCAGAACGACCCGCGAACCTGTCACAACAACTGGGGGCGGGGGGCG 60
a. temnifolia     TCGAAACCTGCCCAGCAGAACGACCCGCGAACCTGTCACAACAACTGGGGGCGGGGGGCG 60
```

```
a. trabeculosa  ----------CCCGCCCCCGACCGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 51
a. nepalensis   ----------CCCGCCCCCGAACGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 51
a. orientalis   ATCTCGTGCCCCGCCCTCGAACGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 120
a. cordata      ATCTCGCGCCCCGCCCTCGAACGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 120
a. serrulata    ATCTCGTGCCCCGCCCTCGAACGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 120
a. japonica     ATCTCGCGCCCCGCCCTCGAACGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 120
a. viridis      ATCTCGCACCCCGCCCTCGAACGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 120
a. sibirica     ATCTCGCGCCCCGCCCTCGAACGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 120
a. glutinosa    ATCTCGCGCCCCGCCCTCGAACGGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 120
a. rubra        ----------------------------------------------------CGAACAACGTA 11
a. rugosa       ------------------GAACGCAGGGAGACACTCGTGCCTTCCTGCCGAACAACGTA 41
a. incana       ATCTCGCGCCCCCGCCCTCGAACGGCAGGAAGACACTCGTGCCTTCCTGCCGAACAACGTA 120
a. tenuifolia   ATCTCGCNCCCCGCCCTCCAACGGNAGGGAGACACTCGTGCNTTCCTGCCGAACAACGTA 120
                                                          **********

a. trabeculosa  CCCCGGCGCGGTCCGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 111
a. nepalensis   CCCCGGCGCGGTCCGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCACGGAGA 111
a. orientalis   CCCCGGCGCGGTCCGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 180
a. cordata      CCCCGGCGCGGTCCGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 180
a. serrulata    CCCCGGCGCGGTCCGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 180
a. japonica     CCCCGGCGCGGTCCGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 180
a. viridis      CCCCGGCGCGGTCTGCGCCCAAGGAACATTAACGAAAGAGTGCCTCCGGTAGCCTCGGAAA 180
a. sibirica     CCCCGGCGCGGTCCGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 180
a. glutinosa    CCCCGGCGCGGTCCGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 180
a. rubra        CCCCGGCGCGGTCTGCGCCCAAGGAACATGAACGAAAGAGTGCCTCTGGTCGCCTCGGAAA 71
a. rugosa       CCCCGGCCCGGTCTGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 101
a. incana       CCCCGGCGCGGTCTGCGCCCAAGGAACATGAACGAAAGAGTGCCTCCGGTCGCCTCGGAAA 180
a. tenuifolia   CCCCNGNGCGGTCTGCGCCCAAGGAACATGAACNAAAGAATGCCTCCGGTCGCCTCGNAAA 180
                **** *  ***** ************** *** ***** ****** *** *** ** * *

a. trabeculosa  CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 171
a. nepalensis   CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCGTAACGACTCTCGGCAACGGATAT 171
a. orientalis   CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCGTAACGACTCTCGGCAACGGATAT 240
a. cordata      CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
a. serrulata    CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
a. japonica     CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
a. viridis      CGCTGTGCTTGCCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
a. sibirica     CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
a. glutinosa    CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
a. rubra        CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 131
a. rugosa       CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 161
a. incana       CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
a. tenuifolia   CGCTGCGCGCACCGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
                ***** **   ************************ ************************
```

```
a. trabeculosa    CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 231
a. nepalensis     CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 231
a. orientalis     CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
a. cordata        CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
a. serrulata      CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
a. japonica       CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
a. viridis        CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
a. sibirica       CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
a. glutinosa      CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
a. rubra          CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 191
a. rugosa         CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 221
a. incana         CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
a. tenuifolia     CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
                  ************************************************************

a. trabeculosa    TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 291
a. nepalensis     TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 291
a. orientalis     TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
a. cordata        TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
a. serrulata      TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
a. japonica       TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
a. viridis        TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
a. sibirica       TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
a. glutinosa      TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
a. rubra          TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 251
a. rugosa         TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 281
a. incana         TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
a. tenuifolia     TCCCGCGAATCATNGAGTCTTTGAACGCAAGTTGCGCCNGAAGCCACCTGGCCGAGGGCA 360
                  ************ ************************ **********************

a. trabeculosa    CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-CGCAAAGAGGCGGT 350
a. nepalensis     CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGT 350
a. orientalis     CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCCCTGCAAAGAGGCGGT 420
a. cordata        CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGT 419
a. serrulata      CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGC 419
a. japonica       CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGC 419
a. viridis        CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGAT 419
a. sibirica       CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGT 419
a. glutinosa      CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGT 419
a. rubra          CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGT 310
a. rugosa         CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGT 340
a. incana         CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGACGCGGT 419
a. tenuifolia     CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCC-TGCAAAGAGGCGGT 419
                  ********************************************  ******* ***

a. trabeculosa    GGGGGCACGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 409
a. nepalensis     GGGGGCACGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGATGGCCTAAAA 409
a. orientalis     GGGGGCACGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 479
a. cordata        GGGGGCACGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 478
a. serrulata      GGGGGCGCGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 478
a. japonica       GGGGGCGCGCGGGGGCGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 479
a. viridis        GGGGGCATGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 478
a. sibirica       GGGGGCATGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 478
a. glutinosa      GGGGGCATGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 478
a. rubra          GGGGGCATGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 369
a. rugosa         GGGGGCATGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAAAA 399
a. incana         GGGGGCATGCGGGG-CGGACATTGGCCTCCCGTGGGCTGATGCCCGCGGCTGGCCTAAAA 478
a. tenuifolia     GGGGGCATGCGGGG-CGGACATTGGCCTCCCGTGGGNTGATGCNTGCGGNTGGCCTAAAA 478
```

```
                ******  ****** ******************** ******  **** **********

a. trabeculosa  ACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 469
a. nepalensis   ACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 469
a. orientalis   ACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 539
a. cordata      ACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 538
a. serrulata    ACGAGTCCTCGGCGACGATCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 538
a. japonica     ACGAGTCCTCGGCGACGATCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 539
a. viridis      ACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 538
a. sibirica     ACGAGTCCTCGGCGACGATCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 538
a. glutinosa    ACGAGTCCTCGGCGACGATCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 538
a. rubra        ACGAGTCCTCGGCGACGATCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 429
a. rugosa       ACGAGTCCTCGGCGACGATCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 459
a. incana       ACGAGTCCTCGGCGACGATCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 538
a. tenuifolia   ACGAGTCNTCGGNGACGATCGCCACGACAATCGGTGGTTGACAAACCCTCGTGACCCGTC 538
                ******* **** ***** ************************** ***********

a. trabeculosa  GTGCGCGCATCGTCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 529
a. nepalensis   GTGCGCGCATCGTCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 529
a. orientalis   GTGCGCGCATCGTCGCTCAACGCGTGCTCTTACGACCCTGTCGCGTCGCGCTCGCGACGC 599
a. cordata      GTGCGCGCATCGTCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 598
a. serrulata    GTGCGCGCACCGTCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 598
a. japonica     GTGCGCGCACCGTCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGGCGC 599
a. viridis      GTGCGTGCATCGTCGCTCAATGTGTGCTCCTTTGACCCTGTCGTGTCGCGCTCGCGACGC 598
a. sibirica     GTGCGCGCATCGCCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 598
a. glutinosa    GTGCGCGCATCGCCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 598
a. rubra        GTGCGCGCATCGCCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 489
a. rugosa       GTGCGCGCATCGCCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 519
a. incana       GTGCGCGCATCGCCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGACGC 598
a. tenuifolia   GTGCGCGCATNGCCGCTCAACGCGTGCTNTTTTGACCCTGTCGGGTTGCGCTNGNGACGC 598
                ***** ***  * ******* * *****  *  ********** ** ***** * * ***

a. trabeculosa  TTCCAACGCG 539
a. nepalensis   TTCCAACGCG 539
a. orientalis   TTCCAACGCG 609
a. cordata      TTCCAACGCG 608
a. serrulata    TTCCAACGCG 608
a. japonica     TTCCAACGCG 609
a. viridis      TTCCAACGCG 608
a. sibirica     TTCCAACGCG 608
a. glutinosa    TTCCAACGCG 608
a. rubra        TTCCAACGCG 499
a. rugosa       TTCCAACGCG 529
a. incana       T-CCAACGCG 607
a. tenuifolia   TTCCAACGCG 608
                * ********
```

APPENDIX 5

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	b. nana	606	2	b. davurica	576	98
1	b. nana	606	3	b. uber	599	98
1	b. nana	606	4	b. populifolia	606	99
1	b. nana	606	5	b. pendula	606	99
1	b. nana	606	6	b. nigra	606	96
1	b. nana	606	7	b. lenta	606	98
2	b. davurica	576	3	b. uber	599	97
2	b. davurica	576	4	b. populifolia	606	98
2	b. davurica	576	5	b. pendula	606	98
2	b. davurica	576	6	b. nigra	606	97
2	b. davurica	576	7	b. lenta	606	97
3	b. uber	599	4	b. populifolia	606	98
3	b. uber	599	5	b. pendula	606	97
3	b. uber	599	6	b. nigra	606	96
3	b. uber	599	7	b. lenta	606	100
4	b. populifolia	606	5	b. pendula	606	99
4	b. populifolia	606	6	b. nigra	606	96
4	b. populifolia	606	7	b. lenta	606	98
5	b. pendula	606	6	b. nigra	606	97
5	b. pendula	606	7	b. lenta	606	98
6	b. nigra	606	7	b. lenta	606	96

```
b. uber          TCGAAACCTGCCCAGCAGAACGACCCGTGAACATGTTGAAACAACTGGGGGCGGGGGGCG 60
b. lenta         TCGAAACCTGCCCAGCAGAACGACCCGTGAACATGTTGAAACAACTGGGGGCGGGGGGCG 60
b. nana          TCGAAACCTGCCCAGCAGAACGACCCGTGAACCTGTTGAAACAACTGGGGGTGKGGGGCG 60
b. pendula       TCGAAACCTGCCCAGCAGAACGACCCGTGAACCTGTTGAAACAACTGGGGGTGTGGGGCG 60
b. populifolia   TCGAAACCTGCCCAGCAGAACGACCCGTGAACCTGTTGAAACAACTGGGGGNGGGGGGCG 60
b. davurica      TCGAAACCTGCCCAGCAGAACGACCCGTGAACCTGTTGAAACAACTGGGGGTGGGGGGCG 60
b. nigra         TCGAAACCTGCCCAGCAGAACGACCCGTGAACCTGTTGAAACAACTGGGGGTGGGGGGCG 60
                 ******************************** ****************** * ******

b. uber          ATCTCGCCCCCGTGCCCCCGAACGGCAGGGAGACACTCGTGCATCCCTGCCGAACAACGAA 120
b. lenta         ATCTCGCCCCCGTGCCCCCGAACGGCAGGGAGACACTCGTGCATCCCTGCCGAACAACGAA 120
b. nana          ATCTCGCCCCCTTGCCCCCGAACGGTAGGGAGACACTTGTGCATCCCTGCCGAACAACGAA 120
b. pendula       ATCTCGCCCCCTTGCCCCCGAACGGTAGGGAGACACTTGTGCATCCCTGCCGAACAACGAA 120
b. populifolia   ATCTCGCCCCCTTGTCCCCGAACGGTAGGGAGACACTTGTGCATCCCTGCCGAACAACGAA 120
b. davurica      ATCTCGCCCCCTTGCCCCCGAACGGTAGGGAGACACTTGTGCATCCCTGCCGAACAACGAA 120
b. nigra         ATCTCGCCCCCGTGCCTCCGAACGGTAGGGAGACACTTGTGCATCCCTGCTGAACAACGAA 120
                 ********** ** * ******** *********** ************ **********
```

```
b. uber          CCCCGGCGCGGTCTGCGCCAAGGAACTTTAACGAAAGAGTGCCTCCGGCCGTCTCGGAAA 180
b. lenta         CCCCGGCGCGGTCTGCGCCAAGGAACTTTAACGAAAGAGTGCCTCCGGCCGTCTCGGAAA 180
b. nana          CCCCGGCGCGGTCYGCGCCAAGGAACTTTAACGAAAGAGTGCCTCCGGCCGCCTCGGAAA 180
b. pendula       CCCCGGCGCGGTCCGCGCCAAGGAACTTTAACGAAAGAGTGCCTCCGGCCGCCTCGGAAA 180
b. populifolia   CCCCGGCGCGGTCCGCGCCAAGGAACTTTAACGAAAGAGTGCCTCCGGCCGCCTCGGAAA 180
b. davurica      CCCCGGCGCGGTCTGCGCCAAGGAACTTTAACGAAAGAGTGCCTCCCGCCGCCTCGGAAA 180
b. nigra         CCCCGGCGCGGTCTGCGCCAAGGAACTTTAACGAAAGAGTGCCTCCGGCCGCCTCGGAAA 180
                 ************ ******************************** **** ********

b. uber          CGGTGTGCGTGCGGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
b. lenta         CGGTGTGCGTGCGGGAGGCGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
b. nana          CGGTGTGCGTGCGGGAGGTGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
b. pendula       CGGTGTGCGTGCGGGAGGTGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
b. populifolia   CGGTGTGCGTGCGGGAGGTGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
b. davurica      CGGTGTGCGTGCGGGAGGTGAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
b. nigra         CGGTGTGCGTGCAGGAAAACAATCTTGTCTAGAACCATAACGACTCTCGGCAACGGATAT 240
                 ************ ***    ************************************

b. uber          CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
b. lenta         CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
b. nana          CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
b. pendula       CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
b. populifolia   CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
b. davurica      CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
b. nigra         CTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAA 300
                 ************************************************************

b. uber          TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
b. lenta         TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
b. nana          TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
b. pendula       TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
b. populifolia   TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
b. davurica      TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
b. nigra         TCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGGGCA 360
                 ************************************************************

b. uber          CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCTCCTTGCAAAGGGACGAGG 420
b. lenta         CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCTCCTTGCAAAGGGACGAGG 420
b. nana          CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCTCCTTGCAAAGGGACGAGG 420
b. pendula       CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCTCCTTGCAAAGGGACGAGG 420
b. populifolia   CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCTCCTTGCAAAGGGACGAGG 420
b. davurica      CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCTCCTTGTAAAGGGACGAGG 420
b. nigra         CGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCTCCTTGCAAAGGGACGAGA 420
                 ********************************************** ***********

b. uber          GGGCCTGTGGGGCAGAAATTGGCCTCCCGTGAGCTCATGCATGCGGTTGGCCTAAAAGCG 480
b. lenta         GGGCCTGTGGGGCAGAAATTGGCCTCCCGTGAGCTCATGCATGCGGTTGGCCTAAAAGCG 480
b. nana          GGGCCTGTGGGGCAGAAATTGGCCTCCCGTGAGCTCATGCATGCGGTTGGCCTAAAAGCG 480
b. pendula       GGGCCTGTGGGGCAGAAATTGGCCTCCCGTGAGCTCATGCATGCGGTTGGCCTAAAAGCG 480
b. populifolia   GGGCCTGTGGGGCAGAAATTGGCCTCCCGTGAGCTCATGCATGCGGTTGGCCTAAAAGCG 480
b. davurica      GGGCCCGTGGGGCAGAAATTGGCCTCCCGTGAGCTCATGCATGCGGTTGGCCTAAAAGCG 480
b. nigra         GGGCCAGTGGGGTAGAAATTGGCCTCCCGTGAGCTCATGCATGCGGTTGGCCTAAAAGCG 480
                 ***** ****** ************************************************
```

```
b.uber          AGTCCTCGGCGACGCGCGCCACGACAATCGGTGGTTGACAAACCCTCGTGTCCCGTCGTG 540
b.lenta         AGTCCTCGGCGACGCGCGCCACGACAATCGGTGGTTGACAAACCCTCGTGTCCCGTCGTG 540
b.nana          AGTCCTCGGCGACGCGCGCCACGACAATCGGTGGTTGTCAAACCCTCGTGTCCCGTCGTG 540
b.pendula       AGTCCTCGGCGACGCGCGCCACGACAATCGGTGGTTGTCAAACCCTCGTGTCCCGTCGTG 540
b.populifolia   AGTCCTCGGCGACGCGCGCCACGACAATCGGTGGTTGTCAAACCCTCGTGTCCCGTCGTG 540
b.davurica      AGTCCTCGGCGACGCGCGCCACGACAATCGGTGGTTGTCAAACCCTCGTGTCCCGTCGTG 540
b.nigra         AGTCCTCGGCGACGCGCGCCACGACAATCGGTGGTTGACAAACCCTCGTGTCCCGTCGTG 540
                ************************************ ********************

b.uber          CGTGCCGCGTCGCTCATCGTGTGCTCTTTGACCCTGCTGTGTCGCGCTAGCGACGCTTC- 599
b.lenta         CGTGCCGCGTCGCTCATCGTGTGCTCTTTGACCCTGCTGTGTCGCGCTAGCGACGCTTCC 600
b.nana          CGTGCCGCGTCGCTCATCGTGTGCTCCTTGACCCTGCTGTGTCGCGCTAGCGACGCTTCC 600
b.pendula       CGTGCCGCGTCGCTCATCGTGTGCTCCTTGACCCTGTTGTGTCGCGCTAGCGACGCTTCC 600
b.populifolia   CGTGCCGCGTCGCTCATCGTGTGCTCCTTGACCCTGCTGTGTCGCGCTAGCGACGCTTCC 600
b.davurica      CGTGACGCGTCGCTCATCGTGTGCTCCTTGACCCTG---------------------------- 576
b.nigra         CGTGCCGCGTTGCTCATCGTGTGCTCTTTGACCCTGTTGTGTCGCGCTAGCGATGCTTCC 600
                **** ***** *************** *********

b.uber          ------
b.lenta         AACGCG 606
b.nana          AACGCG 606
b.pendula       AACGCG 606
b.populifolia   AACGCG 606
b.davurica      ------
b.nigra         AATGCG 606
```

APPENDIX 6

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	C. terminalis	601	2	C. sarmentosa	599	94
1	C. terminalis	601	3	C. ruscifolia	605	94
1	C. terminalis	601	4	C. microphylla	607	93
2	C. sarmentosa	599	3	C. ruscifolia	605	96
2	C. sarmentosa	599	4	C. microphylla	607	96
3	C. ruscifolia	605	4	C. microphylla	607	98

```
C. ruscifolia     TTGTCGATGCCTGCACAGCAGAACGACCCGCGAACGAGTTTTCAATCGTAATGGGAGAAC 60
C. microphylla    TTGTCGATGCCTGCACAGCAGAACGACCCGCGAACGAGTTTTCAATTGTAATGGGAGAAC 60
C. sarmentosa     -------TGCCTGCAAAGCAGAACGACCCGCGAACGAGTTTTCAATTCTAACGGGAGAAC 53
C. terminalis     -------TGCCTGCACAGCAGAACGACCCGCGAACGAGTTTTCAATTCTAACGGGAGA-C 52
                         ******** ******************************  *** ****** *

C. ruscifolia     GGGGCCTGCGAGGGCCT-CGCACCTCCCACGGGAGAGCCGTCCCCTCGCGGGACGTCTAC 119
C. microphylla    GGGGCCTGCGAGGGCCT-CGCACCTCCCACGGGAGAGCCGTCCCCTCGCGGGACGTCTAC 119
C. sarmentosa     GGGGCCTGCGAGGGCCT-CGCACCTCCCGCGGGAGAGCCGTCCCCTCGCGGGACGTCTGC 112
C. terminalis     GGGGCCTGCGAGGGCCTTCGCGCTTCCGCCGGGAGAGCCGTCCCATCGCGGGACGTCTCC 112
                  ***************** *** * ***  ************** ************* *

C. ruscifolia     CCCGAGAACAACGAACCCCGACGCAATCCGCGTCAAGGAAACTGTACAAGCGACTTCGCC 179
C. microphylla    CCCGAGAACAACGAACCCCGACGCAATCCGCGTCAAGGAAACTGTACAAGCGACTTCGCC 179
C. sarmentosa     CCCGAGAACAACGAACCCCGACGCAATCCGCGTCAAGGAAACTGTACGAGCGACTTCGCC 172
C. terminalis     CCCGAGAACAACGAACCCCGGCGCAATCTGCGTCAAGGAAACTGTACAAGCGATTCCGCC 172
                  ******************* ******* ***************** ***** * ****

C. ruscifolia     GTGCCCGCCCCGGAGACGGCGAGCGGACATGGCGTGTTTCGTCGCGTGATCACAATCAC- 238
C. microphylla    ATGCCCGCCCCGGAGACGGCGAGCGGACATGGCGCGTTCCGTCGCGTGATCACAATCAC- 238
C. sarmentosa     GTGCCCGCCCCGGAGACGGCGAGCGGACATGGCATGCATCGTCGCGTGATCACAATCAC- 231
C. terminalis     GTGCCTGCCCCGGAGACGGCGAGTGGACACGGCATGTGTCGTCAAGTGATCACAATCACC 232
                   **** ***************** ***** ***  *   ****  **************

C. ruscifolia     AACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCG 298
C. microphylla    AACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCG 298
C. sarmentosa     AACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCG 291
C. terminalis     AACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCG 292
                  ************************************************************

C. ruscifolia     ATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCC 358
C. microphylla    ATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCC 358
C. sarmentosa     ATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCC 351
C. terminalis     ATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCC 352
                    ************************************************************
```

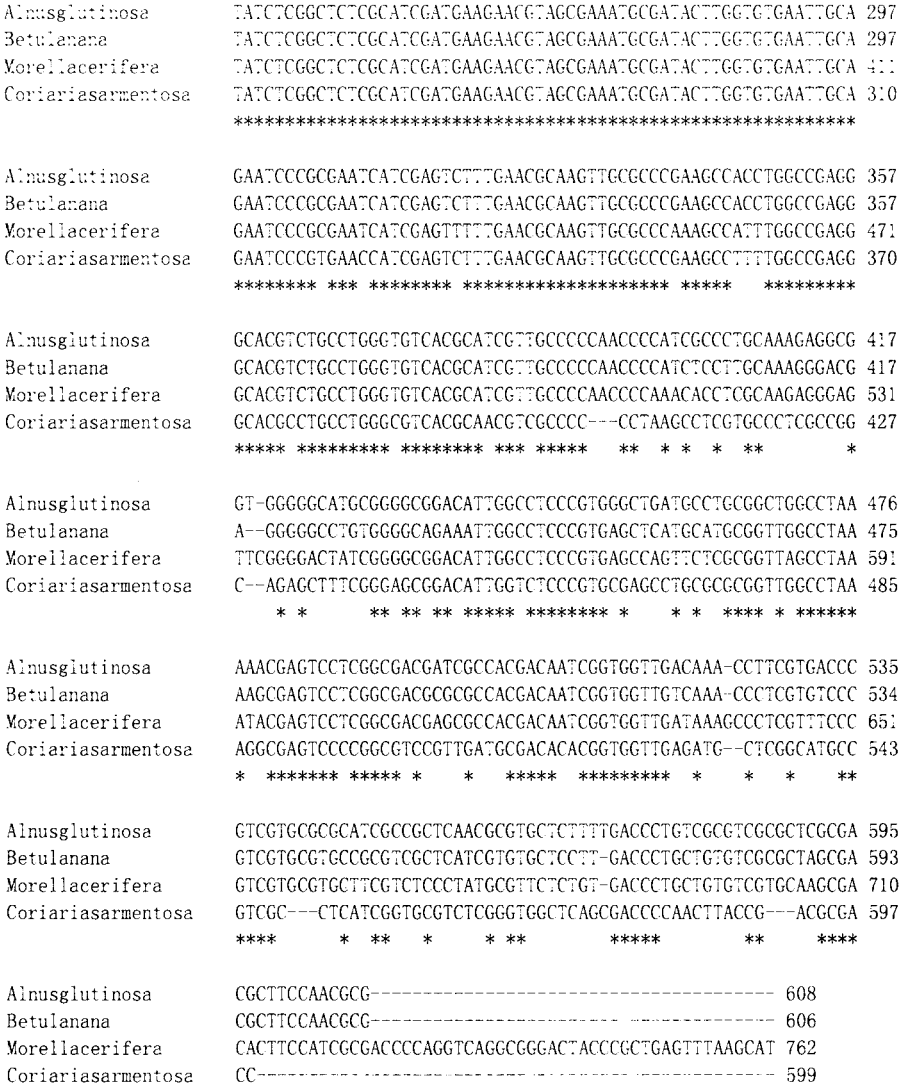
```
C.ruscifolia      -GAAGCCTTTCGGCCGAGGGCACGCCTGCCTGGGCGTCACGCAACGTCGCCCCCCTAAGC 417
C.microphylla     CGAAGCCTTTCGGCCGAGGGCACGCCTGCCTGGGCGTCACGCAACGTCGCCCCCCTAAGC 418
C.sarmentosa      CGAAGCCTTTTGGCCGAGGGCACGCCTGCCTGGGCGTCACGCAACGTCGCCCCCCTAAGC 411
C.terminalis      CGAAGCCTTCTGGCTGAGGGCACGCCTGCCTGGGCGTCACGCAACGTCGCCCCCCCAAGC 412
                   ********  *** ****************************************** ****

C.ruscifolia      CTCATGCCCTCGTCGGCGGAGCTTTTGGGAGCGGACATTGGTCTCCCGTGCGTACCTGCG 477
C.microphylla     CTCGTGCCCCCGTCGGCGGAGCTTTTGGGAGCGGACATTGGTCTCCCGTGCGTACCTGCG 478
C.sarmentosa      CTCGTGCCCTCGCCGGCAGAGCTTTCGGGAGCGGACATTGGTCTCCCGTGCGAGCCTGCG 471
C.terminalis      CTCATGCCCTCGTCGGCGGAGCTTCGGGGAGCGGACATTGGTCTCCCGTGCGACTCTGCG 472
                  *** ***** ** **** ******  *************************   *****

C.ruscifolia      CGCGGTTGGCCTAAAGGCGAGTCCCCGGCGTCCGTTGATGCGACACACGGTGGTTGAGAT 537
C.microphylla     CGCGGTTGGCCTAAAGGCGAGTCCCCGGCGTCCGTTGGTGCGACACACGGTGGTTGAGAT 538
C.sarmentosa      CGCGGTTGGCCTAAAGGCGAGTCCCCGGCGTCCGTTGATGCGACACACGGTGGTTGAGAT 531
C.terminalis      CGCGGTTGGCCTAAAGGCGGGTCCCCGGCGTCTGTTGATGCGACACACGGTGGTTGAGAT 532
                  ****************** ************ **** **********************

C.ruscifolia      GCTCGGCATGCCGTCGCCTCATCGGTGCGTCTCCGGT-GGCTCAGCGACCCCAACTTACC 596
C.microphylla     GCTCGGCATGCCGTCGCCTCATCGGTGCGTCTCCGGTTGGCTCAGCGACCCCAACTTACC 598
C.sarmentosa      GCTCGGCATGCCGTCGCCTCATCGGTGCGTCTCGGGT-GGCTCAGCGACCCCAACTTACC 590
C.terminalis      GCTCGGCATGCCGTCGCCTCATCGGTGCGTCTCCGGTTGGCTCAGCGACCCCAACTTACC 592
                  ********************************** *** **********************

C.ruscifolia      GACGCGACC 605
C.microphylla     GACGCGACC 607
C.sarmentosa      GACGCGACC 599
C.terminalis      GACGCGACC 601
                  *********
```

APPENDIX 7

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	Alnusglutinosa	608	2	Betulanana	606	90
1	Alnusglutinosa	608	3	Morellacerifera	762	84
1	Alnusglutinosa	608	4	Coriariasarmentosa	599	65
2	Betulanana	606	3	Morellacerifera	762	85
2	Betulanana	606	4	Coriariasarmentosa	599	68
3	Morellacerifera	762	4	Coriariasarmentosa	599	67

```
Alnusglutinosa          ------------------------------------------------------------
Betulanana              ------------------------------------------------------------
Morellacerifera         CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
Coriariasarmentosa      ------------------------------------------------------------

Alnusglutinosa          --TCGAAACCTGCCCAGCAGAACGACCCGCGAACCTGTCAC-AA--CAACTGGGGGCGGG 55
Betulanana              --TCGAAACCTGCCCAGCAGAACGACCCGTGAACCTGTTGA-AA--CAACTGGGGGTGXG 55
Morellacerifera         TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAA--CTACCGGGGCCAGG 118
Coriariasarmentosa      ------TGCCTGCAAAGCAGAACGACCCGCGAACGAGTTTTCAATTCTAACGGGAGAACG 54
                               *****  ************** ****  **    **  * *  *** *   *

Alnusglutinosa          GGGCGATC--------TCGCGCCCC-------------------------------------- 72
Betulanana              GGGCGATC--------TCGCCCCTT-------------------------------------- 72
Morellacerifera         GGGCGATCAAAAGCCTCCGGTCCCCAAAACGGTCGGGGAGCATGTGCTGTTGTCCCGTCG 178
Coriariasarmentosa      GGGCCTGCGAGGGCC-TCGCACCTC-------------------------------------- 78
                        ****   *         **  **

Alnusglutinosa          GCCCTCGAACG---GCAGGGAGACACTCGTGCCT-----TCCTGCCGAACAACGTACCCC 124
Betulanana              GCCCCCGAACG---GTAGGGAGACACTTGTGCAT-----CCCTGCCGAACAACGAACCCC 124
Morellacerifera         GCCCTCGGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCC 238
Coriariasarmentosa      --CCGCGGGAG-----AGCCGTCCCCTCGCGGGACGTCTGCCCCGAGAACAACGAACCCC 131
                          ** **   *     **     * *  *           **    ******** *****

Alnusglutinosa          GGCGCGGTCCGCGCCAAGGAACATGAACGAAAGAGTGCCTC-CGGTCGCCTCGGAAACGC 183
Betulanana              GGCGCGGTCYGCGCCAAGGAACTTTAACGAAAGAGTGCCTC-CGGCCGCCTCGGAAACGG 183
Morellacerifera         GGCGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTC-CGATCGCCCCGGAAACGG 297
Coriariasarmentosa      GACGCAATCCGCGTCAAGGAAACTGTACGAGCGACTTCGCCGTGCCCGCCCCGGAGACGC 191
                        * ***   * *** *******  *  ** *  ** * *  *  *  **** **** ***

Alnusglutinosa          TGCGCGCACCGGAGGCGAATCTT-----GTCTAGAAC-CATAACGACTCTCGGCAACGGA 237
Betulanana              TGTGCGTGCGGGAGGTGAATCTT-----GTCTAGAAC-CATAACGACTCTCGGCAACGGA 237
Morellacerifera         TGTGCGT-CGGTTGGGACGTCTT-----GACTGTTATACAAAACGACTCTCGGCAACGGA 351
Coriariasarmentosa      CGAGCGGACATGGCATGCATCGTCGCGTGATCACAAT-CACAACGACTCTCGGCAACGGA 250
                         * ***  *          ** *     *      *  ** ******************
```


Alnusglutinosa TATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCA 297
Betulanana TATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCA 297
Morellacerifera TATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCA 411
Coriariasarmentosa TATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCA 310
Alnusglutinosa GAATCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGG 357
Betulanana GAATCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACCTGGCCGAGG 357
Morellacerifera GAATCCCGCGAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCATTTGGCCGAGG 471
Coriariasarmentosa GAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCTTTTGGCCGAGG 370
Alnusglutinosa GCACGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCGCCCTGCAAAGAGGCG 417
Betulanana GCACGTCTGCCTGGGTGTCACGCATCGTTGCCCCCAACCCCATCTCCTTGCAAAGGGACG 417
Morellacerifera GCACGTCTGCCTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAG 531
Coriariasarmentosa GCACGCCTGCCTGGGCGTCACGCAACGTCGCCCC---CCTAAGCCTCGTGCCCTCGCCGG 427
Alnusglutinosa GT-GGGGGCATGCGGGGCGGACATTGGCCTCCCGTGGGCTGATGCCTGCGGCTGGCCTAA 476
Betulanana A--GGGGGCCTGTGGGGCAGAAATTGGCCTCCCGTGAGCTCATGCATGCGGTTGGCCTAA 475
Morellacerifera TTCGGGGACTATCGGGGCGGACATTGGCCTCCCGTGAGCCAGTTCTCGCGGTTAGCCTAA 591
Coriariasarmentosa C--AGAGCTTTCGGGAGCGGACATTGGTCTCCCGTGCGAGCCTGCGCGCGGTTGGCCTAA 485
Alnusglutinosa AAACGAGTCCTCGGCGACGATCGCCACGACAATCGGTGGTTGACAAA-CCTTCGTGACCC 535
Betulanana AAGCGAGTCCTCGGCGACGCGCGCCACGACAATCGGTGGTTGTCAAA-CCCTCGTGTCCC 534
Morellacerifera ATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGCCCTCGTTTCCC 651
Coriariasarmentosa AGGCGAGTCCCCGGCGTCCGTTGATGCGACACACGGTGGTTGAGATG--CTCGGCATGCC 543
Alnusglutinosa GTCGTGCGCGCATCGCCGCTCAACGCGTGCTCTTTTGACCCTGTCGCGTCGCGCTCGCGA 595
Betulanana GTCGTGCGTGCCGCGTCGCTCATCGTGTGCTCCTT-GACCCTGCTGTGTCGCGCTAGCGA 593
Morellacerifera GTCGTGCGTGCTTCGTCTCCCTATGCGTTCTCTGT-GACCCTGCTGTGTCGTGCAAGCGA 710
Coriariasarmentosa GTCGC---CTCATCGGTGCGTCTCGGGTGGCTCAGCGACCCCAACTTACCG---ACGCGA 597
Alnusglutinosa CGCTTCCAACGCG---------- 608
Betulanana CGCTTCCAACGCG---------- 606
Morellacerifera CACTTCCATCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
Coriariasarmentosa CC---------- 599

APPENDIX 8

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	Myricagalecanada	761	2	Myricagalespain	761	99
1	Myricagalecanada	761	3	Myricagalefinland	761	99
1	Myricagalecanada	761	4	Myricagalebelgium	761	99
2	Myricagalespain	761	3	Myricagalefinland	761	100
2	Myricagalespain	761	4	Myricagalebelgium	761	100
3	Myricagalefinland	761	4	Myricagalebelgium	761	100

```
Myricagalefinland    CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
Myricagalebelgium    CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
Myricagalespain      CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
Myricagalecanada     CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
                     ************************************************************

Myricagalefinland    TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCGGGGG 120
Myricagalebelgium    TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCGGGGG 120
Myricagalespain      TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCGGGGG 120
Myricagalecanada     TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCGGGGG 120
                     ************************************************************

Myricagalefinland    GCGATCAAAAGCCTCCCGTCCCCAAAACGGTTGGGGAGCATGTGCTGTTAACCTGTCGGC 180
Myricagalebelgium    GCGATCAAAAGCCTCCCGTCCCCAAAACGGTTGGGGAGCATGTGCTGTTAACCTGTCGGC 180
Myricagalespain      GCGATCAAAAGCCTCCCGTCCCCAAAACGGTTGGGGAGCATGTGCTGTTAACCTGTCGGC 180
Myricagalecanada     GCGATCAAAAGCCTCCCGTCCCCAAAACGGTTGGGGAGCATGTGCTGTTAACCTGTCGGC 180
                     ************************************************************

Myricagalefinland    CCTCGGGTGCGGACGGGAAACACATGTACGTGTCCCCCCAGCCGAACAACGAACCCCGGC 240
Myricagalebelgium    CCTCGGGTGCGGACGGGAAACACATGTACGTGTCCCCCCAGCCGAACAACGAACCCCGGC 240
Myricagalespain      CCTCGGGTGCGGACGGGAAACACATGTACGTGTCCCCCCAGCCGAACAACGAACCCCGGC 240
Myricagalecanada     CCTCGGGTGCGGACGGGAAACACATGTACGTGTCCCCCCAGCCGAACAACGAACCCCGGC 240
                     ************************************************************

Myricagalefinland    GCGGACTGCGCCAAGGAATTTCAACAAAAGAGTGCCTCTGATCGCCCCGGAAACGGTGTG 300
Myricagalebelgium    GCGGACTGCGCCAAGGAATTTCAACAAAAGAGTGCCTCTGATCGCCCCGGAAACGGTGTG 300
Myricagalespain      GCGGACTGCGCCAAGGAATTTCAACAAAAGAGTGCCTCTGATCGCCCCGGAAACGGTGTG 300
Myricagalecanada     GCGGACTGCGCCAAGGAATTTCAACAAAAGAGTGCCTCTGATCGCCCCGGAAACGGTGTG 300
                     ************************************************************

Myricagalefinland    CGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGCT 360
Myricagalebelgium    CGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGCT 360
Myricagalespain      CGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGCT 360
Myricagalecanada     CGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGCT 360
                     ************************************************************
```

```
Myricagalefinland    CTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGCG 420
Myricagalebelgium    CTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGCG 420
Myricagalespain      CTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGCG 420
Myricagalecanada     CTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGCG 420
                     ************************************************************

Myricagalefinland    AATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGC 480
Myricagalebelgium    AATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGC 480
Myricagalespain      AATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGC 480
Myricagalecanada     AATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGC 480
                     ************************************************************

Myricagalefinland    CTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTGGGGACT 540
Myricagalebelgium    CTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTGGGGACT 540
Myricagalespain      CTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTGGGGACT 540
Myricagalecanada     CTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTGGGGACT 540
                     ************************************************************

Myricagalefinland    ATCGGGGCGGACATTGGCCTCCCGTGAGGTAGTTCTCGCGGTTAGCCTAAATACGAGTCC 600
Myricagalebelgium    ATCGGGGCGGACATTGGCCTCCCGTGAGGTAGTTCTCGCGGTTAGCCTAAATACGAGTCC 600
Myricagalespain      ATCGGGGCGGACATTGGCCTCCCGTGAGGTAGTTCTCGCGGTTAGCCTAAATACGAGTCC 600
Myricagalecanada     ATCGGGGCGGACATTGGCCTCCCGTGAGGTAGTTCTCGCGGTTAGCCTAAATACGAGTCC 600
                     ************************************************************

Myricagalefinland    TCGGCGACGAGCGCCACAACAATCGGTGGTTGATAAAGCCCTCGTTTCCAGTCGTGCGCG 660
Myricagalebelgium    TCGGCGACGAGCGCCACAACAATCGGTGGTTGATAAAGCCCTCGTTTCCAGTCGTGCGCG 660
Myricagalespain      TCGGCGACGAGCGCCACAACAATCGGTGGTTGATAAAGCCCTCGTTTCCAGTCGTGCGCG 660
Myricagalecanada     TCGGCGATGAGCGCCACGACAATCGGTGGTTGATAAAGCCCTCGTTTCCAGTCGTGCGCG 660
                     ******* ********* ******************************************

Myricagalefinland    CCTCATCGCCCTAAGTGTGCTCCGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATCG 720
Myricagalebelgium    CCTCATCGCCCTAAGTGTGCTCCGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATCG 720
Myricagalespain      CCTCATCGCCCTAAGTGTGCTCCGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATCG 720
Myricagalecanada     CCTCGTTGCCCTACGTGTGCTCCGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATCG 720
                     **** * ****** ********************************************

Myricagalefinland    CGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 761
Myricagalebelgium    CGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 761
Myricagalespain      CGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 761
Myricagalecanada     CGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 761
                     *******************************************
```

APPENDIX 9

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	M. fayaportugal	762	2	M. fayaspaingomera	762	100
1	M. fayaportugal	762	3	M. fayaspainlapalma	762	100
1	M. fayaportugal	762	4	M. fayaspaintenerife	762	100
2	M. fayaspaingomera	762	3	M. fayaspainlapalma	762	100
2	M. fayaspaingomera	762	4	M. fayaspaintenerife	762	100
3	M. fayaspainlapalma	762	4	M. fayaspaintenerife	762	100

```
M. fayaportugal         CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
M. fayaspaingomera      CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
M. fayaspaintenerife    CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
M. fayaspainlapalma     CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
                        ************************************************************

M. fayaportugal         TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCAGGGG 120
M. fayaspaingomera      TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCAGGGG 120
M. fayaspaintenerife    TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCAGGGG 120
M. fayaspainlapalma     TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCAGGGG 120
                        ************************************************************

M. fayaportugal         GCGATCAAAAGCCTCCGGTCCCCAAAACGGTTGGGGAGCATGTGCTGTTGTCCCGTCGGC 180
M. fayaspaingomera      GCGATCAAAAGCCTCCGGTCCCCAAAACGGTTGGGGAGCATGTGCTGTTGTCCCGTCGGC 180
M. fayaspaintenerife    GCGATCAAAAGCCTCCGGTCCCCAAAACGGTTGGGGAGCATGTGCTGTTGTCCCGTCGGC 180
M. fayaspainlapalma     GCGATCAAAAGCCTCCGGTCCCCAAAACGGTTGGGGAGCATGTGCTGTTGTCCCGTCGGC 180
                        ************************************************************

M. fayaportugal         CCTCGGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
M. fayaspaingomera      CCTCGGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
M. fayaspaintenerife    CCTCGGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
M. fayaspainlapalma     CCTCGGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
                        ************************************************************

M. fayaportugal         CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCGGAAACGGTGT 300
M. fayaspaingomera      CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCGGAAACGGTGT 300
M. fayaspaintenerife    CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCGGAAACGGTGT 300
M. fayaspainlapalma     CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCGGAAACGGTGT 300
                        ************************************************************

M. fayaportugal         GCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGC 360
M. fayaspaingomera      GCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGC 360
M. fayaspaintenerife    GCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGC 360
M. fayaspainlapalma     GCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGC 360
                        ************************************************************
```

```
M. fayaportugal         TCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGC 420
M. fayaspaingomera      TCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGC 420
M. fayaspaintenerife    TCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGC 420
M. fayaspainlapalma     TCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGC 420
                        ************************************************************

M. fayaportugal         GAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTG 480
M. fayaspaingomera      GAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTG 480
M. fayaspaintenerife    GAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTG 480
M. fayaspainlapalma     GAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTG 480
                        ************************************************************

M. fayaportugal         CCTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
M. fayaspaingomera      CCTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
M. fayaspaintenerife    CCTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
M. fayaspainlapalma     CCTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
                        ************************************************************

M. fayaportugal         TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
M. fayaspaingomera      TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
M. fayaspaintenerife    TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
M. fayaspainlapalma     TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
                        ************************************************************

M. fayaportugal         CTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAGCCCTCGTTTCCCGTCGTGGGT 660
M. fayaspaingomera      CTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAGCCCTCGTTTCCCGTCGTGGGT 660
M. fayaspaintenerife    CTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAGCCCTCGTTTCCCGTCGTGGGT 660
M. fayaspainlapalma     CTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAGCCCTCGTTTCCCGTCGTGGGT 660
                        ************************************************************

M. fayaportugal         GCGTCGTCTCCATATGCGTTCTCTGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATC 720
M. fayaspaingomera      GCGTCGTCTCCATATGCGTTCTCTGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATC 720
M. fayaspaintenerife    GCGTCGTCTCCATATGCGTTCTCTGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATC 720
M. fayaspainlapalma     GCGTCGTCTCCATATGCGTTCTCTGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATC 720
                        ************************************************************

M. fayaportugal         GCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
M. fayaspaingomera      GCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
M. fayaspaintenerife    GCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
M. fayaspainlapalma     GCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
                        *******************************************
```

APPENDIX 10

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	C.Peregrinacanada	762	2	C.PeregrinaUSA	762	100

```
C.Peregrinacanada    CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
C.PeregrinaUSA       CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
                     ************************************************************

C.Peregrinacanada    TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCGGGCG 120
C.PeregrinaUSA       TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCGGGCG 120
                     ************************************************************

C.Peregrinacanada    GCGATCAAAAGCCTCCCGTCCCCAAAACGGTTGGGGAGCATGTGCCGTTACCCCGTCGGC 180
C.PeregrinaUSA       GCGATCAAAAGCCTCCCGTCCCCAAAACGGTTGGGGAGCATGTGCCGTTACCCCGTCGGC 180
                     ************************************************************

C.Peregrinacanada    CCTCGGGCGCGGACGGGAAACACAAGCGCGTGCCCCCCAGCCGAACAACGAACCCCGGCG 240
C.PeregrinaUSA       CCTCGGGCGCGGACGGGAAACACAAGCGCGTGCCCCCCAGCCGAACAACGAACCCCGGCG 240
                     ************************************************************

C.Peregrinacanada    CGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGGTCGCCCCGGAAACGGTGTGC 300
C.PeregrinaUSA       CGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGGTCGCCCCGGAAACGGTGTGC 300
                     ************************************************************

C.Peregrinacanada    GTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGCTC 360
C.PeregrinaUSA       GTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGCTC 360
                     ************************************************************

C.Peregrinacanada    TCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGCGA 420
C.PeregrinaUSA       TCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGCGA 420
                     ************************************************************

C.Peregrinacanada    ATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCC 480
C.PeregrinaUSA       ATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCC 480
                     ************************************************************

C.Peregrinacanada    TGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTCGGGGACT 540
C.PeregrinaUSA       TGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTTCGGGGACT 540
                     ************************************************************

C.Peregrinacanada    AATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
C.PeregrinaUSA       AATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
                     ************************************************************
```

```
C.PeregrinaCanada    CTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGCCCTCGTTTCCCGTCGTGCGC 660
C.PeregrinaUSA       CTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGCCCTCGTTTCCCGTCGTGCGC 660
                     ************************************************************

C.PeregrinaCanada    GCCTTGTCGCCCTATGTGCGCTCCGTGACCCTGCTGTGTCGTGCAAGCGACGCTTCCATC 720
C.PeregrinaUSA       GCCTTGTCGCCCTATGTGCGCTCCGTGACCCTGCTGTGTCGTGCAAGCGACGCTTCCATC 720
                     ************************************************************

C.PeregrinaCanada    GCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
C.PeregrinaUSA       GCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
                     *******************************************
```

APPENDIX 11

CLUSTAL W 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	M.r-martineziispainelhierro	762	2	M.r-martineziispaingomera	762	100

```
M.r-martineziispainelhierro     CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCG 50
M.r-martineziispaingomera       CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCG 50
                                **************************************************

M.r-martineziispainelhierro     GAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTA 100
M.r-martineziispaingomera       GAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTA 100
                                **************************************************

M.r-martineziispainelhierro     ATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCGGTCCCCAAAACGG 150
M.r-martineziispaingomera       ATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCGGTCCCCAAAACGG 150
                                **************************************************

M.r-martineziispainelhierro     TTGGGGAGCATGTGCTGTTGTCCCGTCGGCCCTCGGGCGCGGACAGGGAA 200
M.r-martineziispaingomera       TTGGGGAGCATGTGCTGTTGTCCCGTCGGCCCTCGGGCGCGGACAGGGAA 200
                                **************************************************

M.r-martineziispainelhierro     CCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGGCGCGGACTGC 250
M.r-martineziispaingomera       CCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGGCGCGGACTGC 250
                                **************************************************

M.r-martineziispainelhierro     GCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCGGAAACGGTGT 300
M.r-martineziispaingomera       GCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCGGAAACGGTGT 300
                                **************************************************

M.r-martineziispainelhierro     GCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGG 350
M.r-martineziispaingomera       GCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGG 350
                                **************************************************

M.r-martineziispainelhierro     ATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGG 400
M.r-martineziispaingomera       ATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGG 400
                                **************************************************

M.r-martineziispainelhierro     TGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAACGCAAGTTGCG 450
M.r-martineziispaingomera       TGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAACGCAAGTTGCG 450
                                **************************************************

M.r-martineziispainelhierro     CCCAAAGCCATTTGGCCGAGGGCACGTCTGCCTGGGTGTCACGCATCGTT 500
M.r-martineziispaingomera       CCCAAAGCCATTTGGCCGAGGGCACGTCTGCCTGGGTGTCACGCATCGTT 500
                                **************************************************
```

```
M.r-martineziispainelhierro     GCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGACTATCGGGGCG 550
M.r-martineziispaingomera       GCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGACTATCGGGGCG 550
                                **************************************************

M.r-martineziispainelhierro     GACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
M.r-martineziispaingomera       GACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
                                **************************************************

M.r-martineziispainelhierro     CTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAGCCCTCGTTTCC 650
M.r-martineziispaingomera       CTCGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAGCCCTCGTTTCC 650
                                **************************************************

M.r-martineziispainelhierro     CGTCGTGGGTGCGTCGTCTCCATATGCGTTCTCTGTGACCCTGCTGTGTC 700
M.r-martineziispaingomera       CGTCGTGGGTGCGTCGTCTCCATATGCGTTCTCTGTGACCCTGCTGTGTC 700
                                **************************************************

M.r-martineziispainelhierro     GTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGGACTACCCGCT 750
M.r-martineziispaingomera       GTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGGACTACCCGCT 750
                                **************************************************

M.r-martineziispainelhierro     GAGTTTAAGCAT 762
M.r-martineziispaingomera       GAGTTTAAGCAT 762
                                ************
```

APPENDIX 12

CLUSTAL W 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	Seq3	Name	Len(nt)	Score
1	Morellaadenophorachina	804	2	Morellaadenophorataiwan	799	98

```
Morellaadenophorachina      GGCCGCCGGGAATTCGATTGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGG 50
Morellaadenophorataiwan     GGCCGCGGGAATTCGATTGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGG 50
                            **************************************************

Morellaadenophorachina      TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 100
Morellaadenophorataiwan     TGAACCTGCGGAAGGATCATTGTCGAAACCTGCCCAGCAGAACGACCCGC 100
                            **************************************************

Morellaadenophorachina      GAACATGTTAATAACTACCGGGGGCAGAGGGCGATCAAAAGCCTCCCGTC 150
Morellaadenophorataiwan     GAACATGTTAATAACTACCGGGGGCAGGGGGCGATCAAAAGCCTCCCGTC 150
                            ************************** **********************

Morellaadenophorachina      CCCAAAACGGTTGGGGAGCCCGTGCTGTTGTCCTGTTGGCCCTCTGGCGC 200
Morellaadenophorataiwan     CCCAAAACGGTTGGGGAGCACGTGCTGTTGTCCTGTCGGCCCTCTGGCGC 200
                            ****************** *************** *************

Morellaadenophorachina      GGACAGGGAACCACATGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 250
Morellaadenophorataiwan     GGACAGGGAACCACACGCACGTGTCCCCCCAACCAAACAACGAACCCCGG 250
                            ************** ***************** ***************

Morellaadenophorachina      CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGAAGGCCCCG 300
Morellaadenophorataiwan     CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGAAGGCCCCG 300
                            **************************************************

Morellaadenophorachina      GAAACGGTGTGCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTC 350
Morellaadenophorataiwan     GAAACGGTGTGCGTCGGTTGGGACATCTTGACTGTTATACAAAACGACTC 350
                            ************************ ************************

Morellaadenophorachina      TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 400
Morellaadenophorataiwan     TCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATG 400
                            **************************************************

Morellaadenophorachina      CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 450
Morellaadenophorataiwan     CGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAAC 450
                            **************************************************

Morellaadenophorachina      GCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCCTGGGTGTC 500
Morellaadenophorataiwan     GCAAGTTGCGCCCAAAGCCGTTTGGCCGAGGGCACGTCTGCCTGGGTGTC 500
                            **************************************************
```

```
Morellaadenophorachina      ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGATTTCGGGAC 550
Morellaadenophorataiwan     ACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGATTTTGGGAC 549
                            ************************************************ **

Morellaadenophorachina      TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 600
Morellaadenophorataiwan     TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTA 599
                            **************************************************

Morellaadenophorachina      AATACGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGC 650
Morellaadenophorataiwan     AATTCGAGTCCTCGGCGACGAGCGCCACGACAATCGGCGGTTGATAAAGC 649
                            *** ********************************** ************

Morellaadenophorachina      CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGTGCGCTCTGTGACC 700
Morellaadenophorataiwan     CCTCGTTTCCCGTCGTGCGTGCCTCGTCTCCCTATGTGCGCTCTGTGACC 699
                            **************************************************

Morellaadenophorachina      CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 750
Morellaadenophorataiwan     CTGCTGTGTCGTGCAAGCGACACTTCCATCGCGACCCCAGGTCAGGCGGG 749
                            **************************************************

Morellaadenophorachina      ACTACCCGCTGAGTTTAAGCATATCAATAAGCGGAAGGAA-TCACTAGTG 799
Morellaadenophorataiwan     ACTACCCGCTGAGTTTAAGCATATCAATAAGCGGAAGGAAATCACTAGTG 799
                            ************************************* *********

Morellaadenophorachina      TATTC 804
Morellaadenophorataiwan     ----- 
```

APPENDIX 13

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	M.ceriferajamaica	762	2	M.ceriferaUSA	762	99

```
M.ceriferajamaica   CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
M.ceriferaUSA       CATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCAT 60
                    ************************************************************

M.ceriferajamaica   TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCAGGGG 120
M.ceriferaUSA       TGTCGAAACCTGCCCAGCAGAACGACCCGCGAACATGTTAATAACTACCGGGGGCAGGGG 120
                    ************************************************************

M.ceriferajamaica   GCGATCAAAAGCCTCCGGTCCCCAAAACGGTCGGGGAGCATGTGCTGTTGTCCCGTCGGC 180
M.ceriferaUSA       GCGATCAAAAGCCTCCGGTCCCCAAAACGGTCGGGGAGCATGTGCTGTTGTCCCGTCGGC 180
                    ************************************************************

M.ceriferajamaica   CCTCGGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
M.ceriferaUSA       CCTCGGGCGCGGACAGGGAACCACACGCACGTGTCCCCCCAACCGAACAACGAACCCCGG 240
                    ************************************************************

M.ceriferajamaica   CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCGGAAACGGTGT 300
M.ceriferaUSA       CGCGGACTGCGCCAAGGAACTTCAACAAAAGAGTGCCTCCGATCGCCCCGGAAACGGTGT 300
                    ************************************************************

M.ceriferajamaica   GCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGC 360
M.ceriferaUSA       GCGTCGGTTGGGACGTCTTGACTGTTATACAAAACGACTCTCGGCAACGGATATCTCGGC 360
                    ************************************************************

M.ceriferajamaica   TCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGC 420
M.ceriferaUSA       TCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGC 420
                    ************************************************************

M.ceriferajamaica   GAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTG 480
M.ceriferaUSA       GAATCATCGAGTTTTTGAACGCAAGTTGCGCCCAAAGCCATTTGGCCGAGGGCACGTCTG 480
                    ************************************************************

M.ceriferajamaica   CCTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
M.ceriferaUSA       CCTGGGTGTCACGCATCGTTGCCCCAACCCCAAACACCTCGCAAGAGGGAGTTCGGGGAC 540
                    ************************************************************

M.ceriferajamaica   TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
M.ceriferaUSA       TATCGGGGCGGACATTGGCCTCCCGTGAGCTAGTTCTCGCGGTTAGCCTAAATACGAGTC 600
                    ****************************** ****************************
```

```
M. ceriferajamaica   CTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGCCCTCGTTTCCCGTCGTGCGT 660
M. ceriferaUSA       CTCGGCGACGAGCGCCACGACAATCGGTGGTTGATAAAGCCCTCGTTTCCCGTCGTGCGT 660
                     ************************************************************

M. ceriferajamaica   GCCTCGTCTCCATATGCGTTCTCTGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATC 720
M. ceriferaUSA       GCTTCGTCTCCCTATGCGTTCTCTGTGACCCTGCTGTGTCGTGCAAGCGACACTTCCATC 720
                     ** ******** ************************************************

M. ceriferajamaica   GCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
M. ceriferaUSA       GCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCAT 762
                     *******************************************
```

APPENDIX 14

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	C. sarmentosa	1785	2	C. ruscifolia	1781	98
1	C. sarmentosa	1785	3	C. nepalensis	1785	99
1	C. sarmentosa	1785	4	C. myrtifolia	1737	99
2	C. ruscifolia	1781	3	C. nepalensis	1785	97
2	C. ruscifolia	1781	4	C. myrtifolia	1737	97
3	C. nepalensis	1785	4	C. myrtifolia	1737	99

```
C. nepalensis     TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTGCAAGTATGAACTAATTCAGACTGT 60
C. myrtifolia     -CATATGCTTGTCTCAAAGATTAAGCCATGCATGTGCAAGTATGAACTAATTCAGACTGT 59
C. sarmentosa     TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTGCAAGTATGAACTAATTCAGACTGT 60
C. ruscifolia     TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTGCAAGTATGAACTAATTCAGACTGT 60
                   ************************************************************

C. nepalensis     GAAACTGCCAATGGCTCATTAAATCAGTTATAGTTTGTTTGATGGTATCTGCTACTCGGA 120
C. myrtifolia     GAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTGTTTGATGGTATCTGCTACTCGGA 119
C. sarmentosa     GAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTGTTTGATGGTATCTGCTACTCGGA 120
C. ruscifolia     GAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTGTTTGATGGTATCTGCTACTCGGA 120
                  ************************************************************

C. nepalensis     TAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCAACTTCTGGAAGGGATGCA 180
C. myrtifolia     TAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCAACTTCTGGAAGGGATGCA 179
C. sarmentosa     TAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCAACTTCTGGAAGGGACGCA 180
C. ruscifolia     TAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCAACTTCTGGAAGGGAYGCA 180
                  ************************************************************ ***

C. nepalensis     TTTATTAGATAAAAGGTCGACGCGGGCCTAGCTCGTTGCTCTGATGATTCATGATAACTC 240
C. myrtifolia     TTTATTAGATAAAAGGTCGACGCGGGCCTAGCCCGTTGCTCTGATGATTCATGATAACTC 239
C. sarmentosa     TTTATTAGATAAAAGGTCGACGCGGGCCTAGCCCGTTGCTCTGATGATTCATGATAACTC 240
C. ruscifolia     TTTATTAGATAAAAGGTCGACGC----CTAGCCCGTTGCTCTGATGATTCATGATAACTC 236
                  ********************    ***** **************************

C. nepalensis     GACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATATCTGCCCTATCAACTT 300
C. myrtifolia     GACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAACTT 299
C. sarmentosa     GACGGATCGCACGGCCTTCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAACTT 300
C. ruscifolia     GACGGATCGCACGGCCTTCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAACTT 296
                  *************** ************************* ***************

C. nepalensis     TCGATGGTAGGATAGTGGCCTACTATGGTGGTGACGGGTGACGGAGAATTAGGGTTCGAT 360
C. myrtifolia     TCGATGGTAGGATAGTGGCCTACTATGGTGGTGACGGGTGACGGAGAATTAGGGTTCGAT 359
C. sarmentosa     TCGATGGTAGGATAGTGGCCTACTATGGTGGTGACGGGTGACGGAGAATTAGGGTTCGAT 360
C. ruscifolia     TCGATGGTAGGATAGTGGCCTACTATGGTGGTGACGGGTGACGGAGAATTAGGGTTCGAT 356
                  ************************************************************
```

```
C.nepalensis    TCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAAT 420
C.myrtifolia    TCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAAT 419
C.sarmentosa    TCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAAT 420
C.ruscifolia    TCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAAT 416
                ************************************************************

C.nepalensis    TACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTTTGAGTC 480
C.myrtifolia    TACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTTTGAGTC 479
C.sarmentosa    TACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTTTGAGTC 480
C.ruscifolia    TACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTTTGAGTC 476
                ************************************************************

C.nepalensis    TGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCT 540
C.myrtifolia    TGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCT 539
C.sarmentosa    TGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCT 540
C.ruscifolia    TGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCT 536
                ************************************************************

C.nepalensis    GGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAA 600
C.myrtifolia    GGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAA 599
C.sarmentosa    GGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAA 600
C.ruscifolia    GGTGCCAGCAGCCGCGGTARTTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAA 596
                ****************** ****************************************

C.nepalensis    AAAGCTCGTAGTTGGACCTTGGGTCGGGCCGATCGGTCCGCCAATGGTGTGCACCGGTCT 660
C.myrtifolia    AAAGCTCGTAGTTGGACCTTGGGTCGGGCCGATCGGTCCGCCAATGGTGTGCACCGGTCT 659
C.sarmentosa    AAAGCTCGTAGTTGGACCTTGGGTCGGGTCGATCGGTCCGCCAATGGTGTGCACCGGTTT 660
C.ruscifolia    AAAGCTCGTAGTTGGACCTTGGGTYGGGTCGATCGGTCCGCCWATGGTGWGCACCGGTTK 656
                ********************** *** ************ ****** ********

C.nepalensis    GCTCGTCCCTTCTGCCGGCGATGCGCTCCTGTCCTTAACTGGCCGGGTCGTGCCTCCGGC 720
C.myrtifolia    GCTCGTCCCTTCTGCCGGCGATGCGCTCCTGTCCTTAACTGGCCGGGTCGTGCCTCCGGC 719
C.sarmentosa    GCTCGTCCCTTCTGCCGGCGATGCGCTCCTGTCCTTAACTGGCCGGGTCGTGCCTCCGGC 720
C.ruscifolia    GCTCGTCCCTTCTGCCGGCGATGCGCTCCTGKCCTTAACTGGCCGGGTCGTGCCTCCGGC 716
                ******************************** ****************************

C.nepalensis    GCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAGCA 780
C.myrtifolia    GCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGCATATATTAGCA 779
C.sarmentosa    GCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGCATACATTAGCA 780
C.ruscifolia    GCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAGCA 776
                ********************************************** *** *******

C.nepalensis    TGGGATAACATCATAGGATTTCGATCCTATTCTGTTGGCCTTCGGGATCGGAGTAATGAT 840
C.myrtifolia    TGGGATAACATCATAGGATTTCGATCCTATTCTGTTGGCCTTCGGGATCGGAGTAATGAT 839
C.sarmentosa    TGGGATAACATCATAGGATTTCGATCCTATTCTGTTGGCCTTCGGGATCGGAGTAATGAT 840
C.ruscifolia    TGGGATAACATCATAGGATTTCGATCCTATTCTGTTGGCCTTCGGGATCGGAGTAATGAT 836
                ************************************************************

C.nepalensis    TAATAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATTTAT 900
C.myrtifolia    TAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATTTAT 899
C.sarmentosa    TAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATTTAT 900
C.ruscifolia    TAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATTTAT 896
                *** ********************************************************
```

```
C. nepalensis    GAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAG 960
C. myrtifolia    GAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAG 959
C. sarmentosa    GAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAG 960
C. ruscifolia    GAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAG 956
                 ************************************************************

C. nepalensis    TTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGACCAG 1020
C. myrtifolia    TTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGACCAG 1019
C. sarmentosa    TTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGACCAG 1020
C. ruscifolia    TTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGACCAG 1016
                 ************************************************************

C. nepalensis    GGATCGGCGGATGTTACTTTAAGGACTCCGCCGGCACCTTATGAGAAATCAAAGTCTTTG 1080
C. myrtifolia    GGATCGGCGGATGTTACTTTAAGGACTCCGCCGGCACCTTATGAGAAATCAAAGTCTTTG 1079
C. sarmentosa    GGATCGGCGGATGTTACTTTAAGGACTCCGCCGGCACCTTATGAGAAATCAAAGTCTTTG 1080
C. ruscifolia    GGATYGGCGGATGTTRCTTTAAGGACTCCGCCRGCACCTTATGAGAAATCAAAGTCTTTG 1076
                 **** ********** *************** ****************************

C. nepalensis    GGTTCCGGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA 1140
C. myrtifolia    GGTTCCGGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA 1139
C. sarmentosa    GGTTCCGGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA 1140
C. ruscifolia    GGTTCCGGGGGGGAGTATGGTCGCAANNNNNNNNNNNNNNNNAATTGACGGAAGGGCNCCN 1136
                 ************************              ************** **

C. nepalensis    CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGAC 1200
C. myrtifolia    CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGAC 1199
C. sarmentosa    CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGAC 1200
C. ruscifolia    CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGAC 1196
                 ************************************************************

C. nepalensis    ATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATGGCC 1260
C. myrtifolia    ATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATGGCC 1259
C. sarmentosa    ATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATGGCC 1260
C. ruscifolia    ATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATGGCC 1256
                 ************************************************************

C. nepalensis    GTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAGCCT 1320
C. myrtifolia    GTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAGCCT 1319
C. sarmentosa    GTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAGCCT 1320
C. ruscifolia    GTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAGCCT 1316
                 ************************************************************

C. nepalensis    GCTAACTAGCTATGCGGAGGTACCCCTCCGCGGCCAGCTTCTTAGAGGGACTATGGCCTT 1380
C. myrtifolia    GCTAACTAGCTATGCGGAGGTACCCCTCCGCGGCCAGCTTCTTAGAGGGACTATGGCCTT 1379
C. sarmentosa    GCTAACTAGCTATGCGGAGGTACCCCTCCGCGGCCAGCTTCTTAGAGGGACTATGGCCTT 1380
C. ruscifolia    GCTAACTAGCTATGCGGAGGWAYCCCTCCGCGGCCAGCTTCTTAGAGGGACTATGGCCTT 1376
                 ******************* * ***********************************

C. nepalensis    TTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCG 1440
C. myrtifolia    TTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCG 1439
C. sarmentosa    TTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCG 1440
C. ruscifolia    TTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCG 1436
                 ********************************** ***********************
```

```
C. nepalensis    CACGCGCGCTACACTGATGTATTCAACGAGTCTATAGCCTTGGCCGACAGGCCCGGGTAA 1500
C. myrtifolia    CACGCGCGCTACACTGATGTATTCAACGAGTCTATAGCCTTGGCCGACAGGCCCGGGTAA 1499
C. sarmentosa    CACGCGCGCTACACTGATGTATTCAACGAGTCTATAGCCTTGGCCGACAGGCCCGGGTAA 1500
C. ruscifolia    CACGCGCGCTACACTGATGTATTCAACGAGTCTATAGCCTTGGCCGACAGGCCCGGGTAA 1496
                 ************************************************************

C. nepalensis    TCTTTGAAGTTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAGGAA 1560
C. myrtifolia    TCTTTGAAGTTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAGGAA 1559
C. sarmentosa    TCTTTGAAGTTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAGGAA 1560
C. ruscifolia    TCTTTGAAGTTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAGGAA 1556
                 ************************************************************

C. nepalensis    TTCCTAGTAAGCGCGAGTCATCAGCTCGTGCTGACTACGTCCCTGCCCTTTGTACACACC 1620
C. myrtifolia    TTCCTAGTAAGCGCGAGTCATCAGCTCGTGCTGACTACGTCCCTGCCCTTTGTACACACC 1619
C. sarmentosa    TTCCTAGTAAGCGCGAGTCATCAGCTCGTGCTGACTACGTCCCTGCCCTTTGTACACACC 1620
C. ruscifolia    TTCCTAGTAAGCGCGAGTCATCAGCTCGTGCTGACTACGTCCCTGCCCTTTGTACACACC 1616
                 ************************************************************

C. nepalensis    GCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGTGTTCGGATCGCGGCGACGTGGGC 1680
C. myrtifolia    GCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGTGTTCGGATCGCGGCGACGTGGGC 1679
C. sarmentosa    GCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGTGTTCGGATCGCGGCGACGTGGGC 1680
C. ruscifolia    GCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGTGTTCGGATCGCGGCGACGTGGGC 1676
                 ************************************************************

C. nepalensis    GGTTCGCTGCCGGCGACGTCGCGAGAAGTCCACTGAACCTTATCATTTAGAGGAAGGAGA 1740
C. myrtifolia    GGTTCGCTGCCGGCGACGTCGCGAGAAGTCCACTGAACCTTATCATTTAGAGGAAGGA-- 1737
C. sarmentosa    GGTTCGCTGCCGGCGACGTCGCGAGAAGTCCACTGAACCTTATCATTTAGAGGAAGGAGA 1740
C. ruscifolia    GGTTCGCTGCCGGCGACGTCGCGAGAAGTCCACTGAACCTTATCATTTAGAGGAAGGAGA 1736
                 ************************************************************

C. nepalensis    AGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTG 1785
C. myrtifolia    ---------------------------------------------
C. sarmentosa    AGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTG 1785
C. ruscifolia    AGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTG 1781
```

APPENDIX 15

CLUSTAL 2.0.8 multiple sequence alignment

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	D.glomerata	1808	2	D.cannabina	1803	98

```
D.glomerata     TACTTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTGT 60
D.cannabina     TACTTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTG- 59
                ************************************************************

D.glomerata     AAGTATGAACTAGTTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTG 120
D.cannabina     AAGTATGAACTATTTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTG 119
                ************ ********************************************

D.glomerata     TTTGATGGTATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAA 180
D.cannabina     TTTGATGGTATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAA 179
                ************************************************************

D.glomerata     CCCCGACTTTTGGAAGGGATGCATTTATTAGATAAAAGGTCGACGCAGGCTCTGCCCGTT 240
D.cannabina     CCCCGACTTCTGGAAGGGATGCATTTATTAGATAAAAGGTCGACGCAGGCTCTGCTCGTT 239
                ********* ********************************************** ****

D.glomerata     GCTCTGATGATTCATGATAACTCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCAT 300
D.cannabina     GCTCTGATGATTCATGATAACTCGACGGATCGCACGGCCTTCGTGCCGGCGACGCATCAT 299
                ************************************* ********************

D.glomerata     TCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACTATGGTGGTGACGG 360
D.cannabina     TCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACTATGGTGGTGACGG 359
                ************************************************************

D.glomerata     GTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCA 420
D.cannabina     GTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCA 419
                ************************************************************

D.glomerata     AGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAAT 480
D.cannabina     AGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAAT 479
                ************************************************************

D.glomerata     AACAATACCGGGCTCTTAGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACG 540
D.cannabina     AACAATACCGGGCTCTTGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACG 539
                ********************************************************

D.glomerata     AGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCG 600
D.cannabina     AGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCG 599
                ************************************************************
```

```
D. glomerata    TATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGACCTTGGGTTGGGTCAATCGGT 660
D. cannabina    TATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGAACCTTGGGTTGGGTCAATCGGT 659
                ************************************ **********************

D. glomerata    CCGCCTATGGTGTGCACCGGTTGGCTCGTCCCTTCGACCGGCGATGCGCTCCTGGCCTTA 720
D. cannabina    CCGCCACTGGTGTGCACCGGTTGGCTCGTCCCTAATACCGGCGATGCGCTCCTGGCCTTA 719
                *****  *************************   ***********************

D. glomerata    ACTGGCCGGGTCGTGCCTCCGGTACTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAG 780
D. cannabina    ACTGGCCGGGTCGTGCCTCCGGTGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAG 779
                ********************** ************************************

D. glomerata    CCTACGCTCTGTATACATTAGCATGGGATAACATCATAGGATTTCGATCCTATTCTGTTG 840
D. cannabina    CCTACGCTCTGTATACATTAGCATGGGATAACACTACTGGATTTCGATCCTATTCTGTTG 839
                ********************************  *  **********************

D. glomerata    GCCTTCGGGATCGGAGTAATGATTAATAGGGACAGTCGGGGGCATTCGTATTTCATAGTC 900
D. cannabina    GCCTTCGGGATCGGAGTAATGATTAATAGGGACAGTCGGGGGCATTCGTATTTCATAGTC 899
                ************************************************************

D. glomerata    AGAGGTGAAATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATG 960
D. cannabina    AGAGGTGAAATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATG 959
                ************************************************************

D. glomerata    TTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCT 1020
D. cannabina    TTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCT 1019
                ************************************************************

D. glomerata    CAACCATAAACGATGCCGACCAGGGATCGGCGGATGTTACTTTTAGGACACCGCCGGCAC 1080
D. cannabina    CAACCATAAACGATGCCGACCAGGGATCGGCGGATGTTACTTTTAGGACACCGCCGGCAC 1079
                ************************************************************

D. glomerata    CTTATGAGAAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTA 1140
D. cannabina    CTTATGAGAAATCAAAGTCTTTGGGTTCCGGGGGGAG-ATGGTCGCAAGGCTGAAACTTA 1138
                ************************************ **********************

D. glomerata    AAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACAC 1200
D. cannabina    AAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCATGCGGCTTAATTTGACTCAACAC 1198
                ************************************ **********************

D. glomerata    GGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGAT 1260
D. cannabina    GGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGAT 1258
                ************************************************************

D. glomerata    TCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCC 1320
D. cannabina    TCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCC 1318
                ************************************************************

D. glomerata    GTTAACGAACGAGACCTCAGCCTGCTAACTAGCTATGCGGAGGTGACCCTCCGCGGCCAG 1380
D. cannabina    GTTAACGAACGAGACCTCAGCCTGCTAACTAGCTATGCGGAGGTGACCCTCCGCGGCCAG 1378
                ************************************************************

D. glomerata    CTTCTTAGAGGGACTATGGCCGCTTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTGTG 1440
D. cannabina    CTTCTTAGAGGGACTATGGCCGCTTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTGTG 1438
                ************************************************************
```

```
D. glomerata    ATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGATGTATTCAACGAGTATATAA 1500
D. cannabina    ATGCCCTTAGATGTTCTGGGC-GCACGCGCGCTACACTGATGTATTCAACGAGTTTATAA 1497
                ********************* ****************************** *****

D. glomerata    CCTTGGCCGACAGGCCCGGGAAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGC 1560
D. cannabina    CCTTGGCCGACAGGCC-GGGAAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGC 1556
                *************** ****************************************

D. glomerata    AATTGTTGGTCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTA 1620
D. cannabina    AATTGTTGGTCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTA 1616
                ************************************************************

D. glomerata    CGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGT 1680
D. cannabina    CGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGT 1676
                ************************************************************

D. glomerata    GTTCGGATCGCGGCGACGTGGGCGGTTCGCTGCCCGCGACGTCGCGAGAAGTCCACTGAA 1740
D. cannabina    GTTCGGATCGCGGCGACGTGGGCGGTTCG-TGCCCGCGACGTTGCGAGAAGTCCACTGAA 1735
                **************************** ************ *****************

D. glomerata    CCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAG 1800
D. cannabina    CCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAG 1795
                ************************************************************

D. glomerata    GATCATTG 1808
D. cannabina    GATCATTG 1803
                ********
```

APPENDIX 16

CLUSTAL 2.0.8 multiple sequence

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	Datiscacannabina	1803	2	Coriariaruscifolia	1785	97
1	Datiscacannabina	1803	3	Caryaglabra	1739	97
1	Datiscacannabina	1803	4	Juglansnigra	1738	96
1	Datiscacannabina	1803	5	Morellacerifera	1752	96
2	Coriariaruscifolia	1785	3	Caryaglabra	1739	97
2	Coriariaruscifolia	1785	4	Juglansnigra	1738	97
2	Coriariaruscifolia	1785	5	Morellacerifera	1752	97
3	Caryaglabra	1739	4	Juglansnigra	1738	99
3	Caryaglabra	1739	5	Morellacerifera	1752	99
4	Juglansnigra	1738	5	Morellacerifera	1752	98

```
Caryaglabra          ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTGT 37
Juglansnigra         ---------------------TAGTCATATGCTTGTCTCNAAGATTAAGCCATGCATGTGT 40
Morellacerifera      ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTGT 37
Coriariaruscifolia   ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTGC 37
Datiscacannabina     TACTTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTG- 59
                                           ********** ********** ********** *****

Caryaglabra          AAGTATGAACTAATTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTG 97
Juglansnigra         AAGTATGAACTAATTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTG 100
Morellacerifera      AAGTATGAACTAATTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTG 97
Coriariaruscifolia   AAGTATGAACTAATTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTG 97
Datiscacannabina     AAGTATGAACTATTTCAGACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTG 119
                     *********** ********************************************

Caryaglabra          TTTGATGGTATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAA 157
Juglansnigra         TTTGATGGTATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAA 160
Morellacerifera      TTTGATGGTATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAA 157
Coriariaruscifolia   TTTGATGGTATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAA 157
Datiscacannabina     TTTGATGGTATCTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAA 179
                     ************************************************************

Caryaglabra          CCCCGACTTCTGGAAGGGATGCATTTATTAGATAAAAGGTCGACGCGGGCTT-TGCCCGT 216
Juglansnigra         CCCCGACTTCTGGAAGGGATGCATTTATTAGATAAAAGGTCGACGCGGGCTT-TGCCCGT 219
Morellacerifera      CCCCGACTTCTGGAAGGGATGCATTTATTAGGTAAAAGGTCGACGCGGGCTTCTGCCCGT 217
Coriariaruscifolia   CCCCAACTTCTGGAAGGGATGCATTTATTAGATAAAAGGTCGACGCGGGCCT-AGCTCGT 216
Datiscacannabina     CCCCGACTTCTGGAAGGGATGCATTTATTAGATAAAAGGTCGACGCAGGCTC-TGCTCGT 238
                     **** ************************** ********** ***    ** ***
```

```
Caryaglabra        TGCTCTGATGATTCATGATAACTCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCA 276
Juglansnigra       TGCTCTGATGATTCATGATAACTCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCA 279
Morellacerifera    TGCTCTGATGATTCATGATAACTCGACGGATCGCAAGGCCATCGTGCCGGCGACGCATCA 277
Coriariaruscifolia TGCTCTGATGATTCATGATAACTCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCA 276
Datiscacannabina   TGCTCTGATGATTCATGATAACTCGACGGATCGCACGGCCTTCGTGCCGGCGACGCATCA 298
                   ********************************** **** ******************

Caryaglabra        TTCAAATTTCTGCCCTATCAACTTTCGATTGTAGGATAGAGGCCTACAATGGTGGTGACG 336
Juglansnigra       TTCAAATTTCTGCCCTATCAACTTTCGATTGTAGGATAGAGGCCTACAATGGTGGTGACG 339
Morellacerifera    TTCAAATTTCTGCCCTATCAACTTTCGATTGTAGGATAGAGGCCTACAATGGTGGTGACG 337
Coriariaruscifolia TTCAAATATCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACTATGGTGGTGACG 336
Datiscacannabina   TTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACTATGGTGGTGACG 358
                   ******* ********************** ********* ******* ************

Caryaglabra        GGTAACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCC 396
Juglansnigra       GGTAACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCC 399
Morellacerifera    GGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCC 397
Coriariaruscifolia GGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCC 396
Datiscacannabina   GGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCC 418
                   *** ************************************************************

Caryaglabra        AAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAA 456
Juglansnigra       AAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAA 459
Morellacerifera    AAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAA 457
Coriariaruscifolia AAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAA 456
Datiscacannabina   AAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAA 478
                   ************************************************************

Caryaglabra        TAACAATACCGGGCTCTTACGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAA 516
Juglansnigra       TAACAATACCGGGCTCTTACGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAA 519
Morellacerifera    TAACAATACCGGGCTCTTACGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAA 517
Coriariaruscifolia TAACAATACCGGGCTCTT-TGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAA 515
Datiscacannabina   TAACAATACCGGGCTCTT-TGAGTCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAA 537
                   ****************** ****************************************

Caryaglabra        CGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAG 576
Juglansnigra       CGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAG 579
Morellacerifera    CGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAG 577
Coriariaruscifolia CGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAG 575
Datiscacannabina   CGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAG 597
                   ************************************************************

Caryaglabra        CGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTTGGGTTGGGCAGAGCG 636
Juglansnigra       CGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTTGGGTTGGGCAGAGCG 639
Morellacerifera    CGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATCTTGGGTTGGGCAGAGCG 637
Coriariaruscifolia CGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGACCTTGGGTCGGGCCGATCG 635
Datiscacannabina   CGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGAACCTTGGGTTGGGTCAATCG 657
                   ************************************* * ******* ***   * **
```

```
Caryaglabra          GTCCGCCCCTGGTGTGCACCGGTCTGCTCGTCCCTTCTACCGGCGATGCGCTCCTGGCCT 696
Juglansnigra         GTCCGCCCCTGGTGTGCACCGGTCTGCTCGTCCCTTCTACCGGCGATGCGCTCCTGGCCT 699
Morellacerifera      GTCCGCCCCTGGTGTGCACCGATCTGCTCGTCCCTTCTACCGGCGATGCGCTCCTGGCCT 697
Coriariaruscifolia   GTCCGCCAATGGTGTGCACCGGTCTGCTCGTCCCTTCTGCCGGCGATGCGCTCCTGTCCT 695
Datiscacannabina     GTCCGCCACTGGTGTGCACCGGTTGGCTCGTCCCTAATACCGGCGATGCGCTCCTGGCCT 717
                     ******* *********** *  **********  * ***************** ***


Caryaglabra          TAACTGGCCGGGTCGTGCCTCCGGTGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCA 756
Juglansnigra         TAACTGGCCGGGTCGTGCCTCCGGTGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCA 759
Morellacerifera      TAACTGGCCGGGTCGTGCCTCCGGTGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCA 757
Coriariaruscifolia   TAACTGGCCGGGTCGTGCCTCCGGCGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCA 755
Datiscacannabina     TAACTGGCCGGGTCGTGCCTCCGGTGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCA 777
                     ********************** ***********************************


Caryaglabra          AGCCTACGCTCTGTATACATTAGCATGGGATAACATCATAGGATTTCGGTCCTATTGTGT 816
Juglansnigra         AGCCTACGCTCTGTATACATTAGCATGGGATAACATCATAGGATTTCGGTCCTATTGTGT 819
Morellacerifera      AGCCTACGCTCTGTATACATTAGCATGGGATAACATCATAGGATTTCGGTCCTATTGTGT 817
Coriariaruscifolia   AGCCTACGCTCTGTATACATTAGCATGGGATAACATCATAGGATTTCGATCCTATTCTGT 815
Datiscacannabina     AGCCTACGCTCTGTATACATTAGCATGGGATAACACTACTGGATTTCGATCCTATTCTGT 837
                     ***********************************  *  ******** ******* ***


Caryaglabra          TGGCCTTCGGGATCGGAGTAATGATTAACAGGAACAGTCGGGGGCATTCGTATTTCATAG 876
Juglansnigra         TGGCCTTCGGGATCGGAGTAATGATTAACAGGAACAGTCGGGGGCATTCGTATTTCATAG 879
Morellacerifera      TGGCCTTCGGGATCGGAGTAATGATTAACAGGAACAGTCGGGGGCATTCGTATTTCATAG 877
Coriariaruscifolia   TGGCCTTCGGGATCGGAGTAATGATTAATAGGGACAGTCGGGGGCATTCGTATTTCATAG 875
Datiscacannabina     TGGCCTTCGGGATCGGAGTAATGATTAATAGGGACAGTCGGGGGCATTCGTATTTCATAG 897
                     ***************************** *** ***************************


Caryaglabra          TCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGA 936
Juglansnigra         TCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGA 939
Morellacerifera      TCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGA 937
Coriariaruscifolia   TCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGA 935
Datiscacannabina     TCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGA 957
                     ************************************************************


Caryaglabra          TGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGT 996
Juglansnigra         TGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGT 999
Morellacerifera      TGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGT 997
Coriariaruscifolia   TGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGT 995
Datiscacannabina     TGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGT 1017
                     ************************************************************


Caryaglabra          CTCAACCATAAACGATGCCGACCAGGGATCGGCGGATGTTGCTTTAAGGACTCCGCCGGC 1056
Juglansnigra         CTCAACCATAAACGATGCCGACCAGGGATCGGCGGATGTTGCTTTAAGGACTCCGCCGGC 1059
Morellacerifera      CTCAACCATAAACGATGCCGACCAGGGATCGGCGGATGTTGCTTTAAGGACTCCGCCGGC 1057
Coriariaruscifolia   CTCAACCATAAACGATGCCGACCAGGGATCGGCGGATGTTACTTTAAGGACTCCGCCGGC 1055
Datiscacannabina     CTCAACCATAAACGATGCCGACCAGGGATCGGCGGATGTTACTTTTAGGACACCGCCGGC 1077
                     ************************************** **** ***** ********
```

```
Caryaglabra         ACCTTATGAGAAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACT 1116
Juglansnigra        ACCTTATGAGAAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACT 1119
Morellacerifera     ACCTTATGAGAAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACT 1117
Coriariaruscifolia  ACCTTATGAGAAATCAAAGTCTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACT 1115
Datiscacannabina    ACCTTATGAGAAATCAAAGTCTTTGGGTTCCGGGGGGAG-ATGGTCGCAAGGCTGAAACT 1136
                    ************************************ ********************

Caryaglabra         TAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAAC 1176
Juglansnigra        TAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAAC 1179
Morellacerifera     TAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAAC 1177
Coriariaruscifolia  TAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAAC 1175
Datiscacannabina    TAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCATGCGGCTTAATTTGACTCAAC 1196
                    ************************************ ********************

Caryaglabra         ACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTG 1236
Juglansnigra        ACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTG 1239
Morellacerifera     ACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTG 1237
Coriariaruscifolia  ACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTG 1235
Datiscacannabina    ACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTG 1256
                    ************************************************************

Caryaglabra         ATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATT 1296
Juglansnigra        ATTCTATGGGTGGTGGTNCATGNCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATT 1299
Morellacerifera     ATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATT 1297
Coriariaruscifolia  ATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATT 1295
Datiscacannabina    ATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATT 1316
                    ***************** **** ************************************

Caryaglabra         CCGTTAACGAACGAGACCTCAGCCTGCTAACTAGCTATGCGGAGGTGACCTTCCGCGGCC 1356
Juglansnigra        CCGTTAACGAACGAGACCTCAGCCTGCTAACTAGCTATGCGGAGGTGACCTTCCGCGGCC 1359
Morellacerifera     CCGTTAACGAACGAGACCTCAGCCTGCTAACTAGCTATNCGGAGGTGACCTTCCGCGGCC 1357
Coriariaruscifolia  CCGTTAACGAACGAGACCTCAGCCTGCTAACTAGCTATGCGGAGGTACCCCTCCGCGGCC 1355
Datiscacannabina    CCGTTAACGAACGAGACCTCAGCCTGCTAACTAGCTATGCGGAGGTGACCCTCCGCGGCC 1376
                    ************************************ ******* ** *********

Caryaglabra         AGCTTCTTAGAGGGACTATGGCCGCTTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTG 1416
Juglansnigra        AGCTTCTTAGAGGGACTATGGCCGCTTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTG 1419
Morellacerifera     AGCTTCTTAGAGGGACTATGGCCGCTTAGGCCAAGGAAGTTTGAGNCAATAACAGGTCTG 1417
Coriariaruscifolia  AGCTTCTTAGAGGGACTATGGCCTTTTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTG 1415
Datiscacannabina    AGCTTCTTAGAGGGACTATGGCCGCTTAGGCCAAGGAAGTTTGAGGCAATAACAGGTCTG 1436
                    ********************** ******************** **************

Caryaglabra         TGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGATGTATTCAACGAGTTTAT 1476
Juglansnigra        TGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGATGTATTCAACGAGTTTAT 1479
Morellacerifera     TGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGATGTATTCAACGAGTTTAT 1477
Coriariaruscifolia  TGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGATGTATTCAACGAGTCTAT 1475
Datiscacannabina    TGATGCCCTTAGATGTTCTGGGC-GCACGCGCGCTACACTGATGTATTCAACGAGTTTAT 1495
                    ********************** ******************************** ***
```

```
Caryaglabra         AGCCTTGGCCGACAGGCCCGGGTAATCTTTGAAATTTCATCGTGATGGGGATAGATCATT 1536
Juglansnigra        AGCCTTGGCCGACAGGCCCGGGTAATCTTTGAAATTTCATCGTGATGGGGATAGATCATT 1539
Morellacerifera     AGCCTTGGCCGACAGGCCCGGGTAATCTTTGAAATTTCATCGTGATGGGGATAGATCATT 1537
Coriariaruscifolia  AGCCTTGGCCGACAGGCCCGGGTAATCTTTGAAGTTTCATCGTGATGGGGATAGATCATT 1535
Datiscacannabina    AACCTTGGCCGACAGGCC-GGGAAATCTTTGAAATTTCATCGTGATGGGGATAGATCATT 1554
                    * *************** *** ********* ************************

Caryaglabra         GCAATTGTTGGTCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGAC 1596
Juglansnigra        GCAATTGTTGGTCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGAC 1599
Morellacerifera     GCAATTGTTGGTCTTAAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGAC 1597
Coriariaruscifolia  GCAATTGTTGGTCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGTGCTGAC 1595
Datiscacannabina    GCAATTGTTGGTCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGAC 1614
                    ********** ********************************** * ****

Caryaglabra         TACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAA 1656
Juglansnigra        TACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAA 1659
Morellacerifera     TACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAA 1657
Coriariaruscifolia  TACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAA 1655
Datiscacannabina    TACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAA 1674
                    ************************************************************

Caryaglabra         GTGTTCGGATCGCGGCGATGTGGGCGGTCCGCTGCCGGCAACGTCGCGAGAAGTCCACTG 1716
Juglansnigra        GTGTTCGGATCGAGGCGATGTGGGCGGTTCGCTGCCGGCAACGTCGCGAGAAGTCCACTG 1719
Morellacerifera     GTGTTCGGATCGAGGCGATGTGGGCGGTTCGCTGCCGGCAACGTTTTGAGAAGTCCACTG 1717
Coriariaruscifolia  GTGTTCGGATCGCGGCGACGTGGGCGGTTCGCTGCCGGCGACGTCGCGAGAAGTCCACTG 1715
Datiscacannabina    GTGTTCGGATCGCGGCGACGTGGGCGGTTCG-TGCCCGCGACGTTGCGAGAAGTCCACTG 1733
                    ************ ***** ********* ** **** ** ****   *************

Caryaglabra         AACCTTATCATTTAGAGGAAGGA---------------------------------- 1739
Juglansnigra        NACCTTATCATNTAGNGGA------------------------------------- 1738
Morellacerifera     AACCTTATCATTTAGAGGAAGGAGAAGTCGTAACA---------------------- 1752
Coriariaruscifolia  AACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGA 1775
Datiscacannabina    AACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGA 1793
                     ********** *** ***

Caryaglabra         ----------
Juglansnigra        ----------
Morellacerifera     ----------
Coriariaruscifolia  AGGATCATTG 1785
Datiscacannabina    AGGATCATTG 1803
```