The 18S rDNA gene discriminates between red-listed and unexplored ethnomedicinal species of Myristicaceae restricted to humid tropics of India

Article *in* Genetic Resources and Crop Evolution · December 2013

DOI: 10.1007/s10722-013-0060-7

8 authors, including:

Siju Senan

17 PUBLICATIONS **110** CITATIONS

SEE PROFILE

Rahul P R

Centre for Medicinal Plants Research

15 PUBLICATIONS **52** CITATIONS

SEE PROFILE

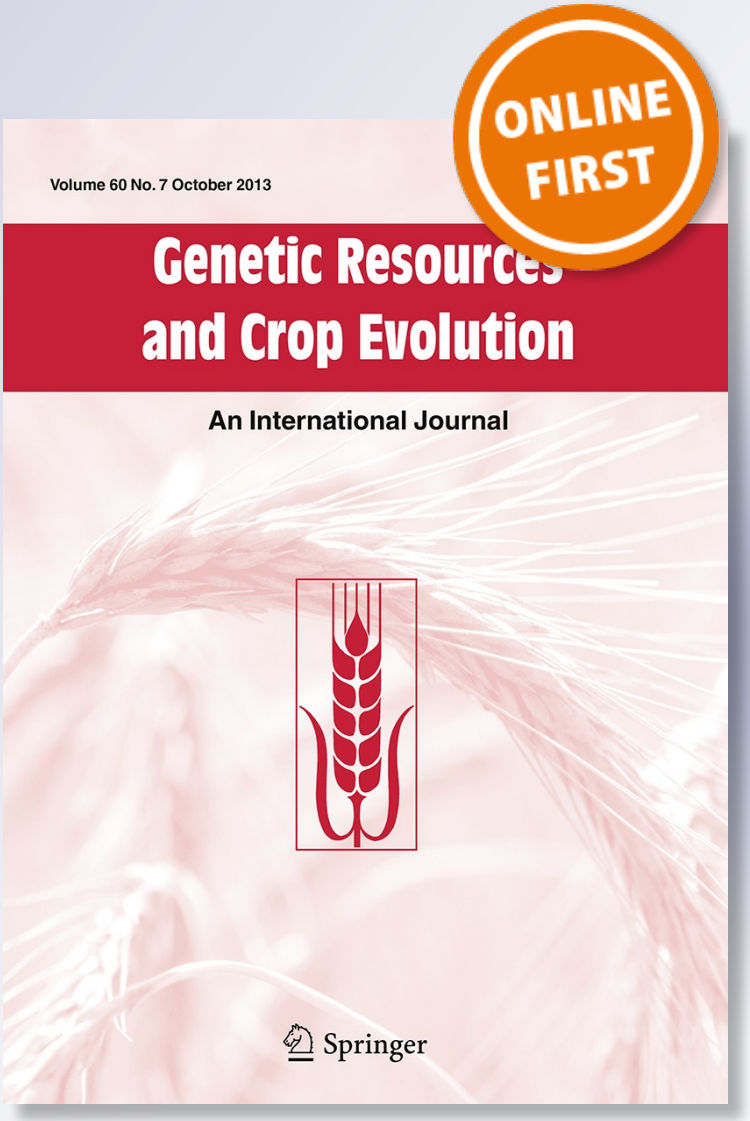
The 18S rDNA gene discriminates between red-listed and unexplored ethnomedicinal species of Myristicaceae restricted to humid tropics of India

T. E. Sheeja, O. B. Rosana, V. P. Swetha, R. S. Shalini, S. Siju, R. Dhanya, P. R. Rahul & B. Krishnamoorthy

Genetic Resources and Crop Evolution
An International Journal

ISSN 0925-9864

Genet Resour Crop Evol
DOI 10.1007/s10722-013-0060-7



Springer

Genet Resour Crop Evol
DOI 10.1007/s10722-013-0060-7

NOTES ON NEGLECTED AND UNDERUTILIZED CROPS

The 18S rDNA gene discriminates between red-listed and unexplored ethnomedicinal species of Myristicaceae restricted to humid tropics of India

T. E. Sheeja · O. B. Rosana · V. P. Swetha · R. S. Shalini · S. Siju · R. Dhanya · P. R. Rahul · B. Krishnamoorthy

Received: 15 February 2013 / Accepted: 7 November 2013

Abstract A potentially diagnostic 18S rDNA (ribosomal DNA) gene was amplified reliably from red-listed ethnomedicinal species of *Myristica* and its wild and related genera. Individuals from nine species of Myristicaceae were utilized for the study. The sequences ranged from 1,767 to 1,794 nucleotide (nt) in length. The GC content (%) varied from 52.77 to 51.04. The frequencies (%) of nt were A (23.31), T (23.82), C (24.48) and G (28.39). The alignment of all sequences produced 195/1,516 variable sites and 1,257/1,516 conserved sites. Total numbers of single nucleotide polymorphism (SNP) sites found in the alignment were 146/1,516. *Knema andamanica* (Warb.) W.J. de Wilde was the most distinct that included 18 variable regions and 15 *InDel* with 27 SNP sites, specific to this species. The identified regions from nine species of *Myristica* and its wild and closely related genera were deposited in the GenBank Database (Accession numbers JN228257-JN228265). Comparison of morphological identifications and phylogenetic analysis indicated that the specimens were correctly assigned on the basis of a short stretch of 18S rDNA (~1,600 bp) making this a potentially useful marker for the rapid molecular assignment of an unknown related species also. Significant sequence homology ranging from 72 to 99 % was observed on comparison with 18S rDNA genes of other plants in the public domain. A comparison of intraspecific data information of nine 18S with that of 73 *matK* and 86 *rbcL* sequences from GenBank revealed that polymorphism, divergence and conservation is higher in 18S locus for Myristicaceae. Hence these markers may be utilized for phylogenetic analysis, evaluation of species richness during ecological surveys or for environmental assessments. These molecular markers are especially important due to the fact that the species studied are mostly vulnerable and red-listed with limited availability in endangered ecological niches.

Keywords Barcoding · Cluster analysis · Endemic · Molecular marker · Myristicaceae · Nuclear DNA

Introduction

The family Myristicaceae is a primitive angiosperm family and is also known as nutmeg family containing about 20 genera and 500 species of evergreen trees found throughout the moist tropical lowlands. The

Electronic supplementary material The online version of this article (doi:10.1007/s10722-013-0060-7) contains supplementary material, which is available to authorized users.

T. E. Sheeja (✉) · O. B. Rosana · V. P. Swetha · R. S. Shalini · S. Siju · P. R. Rahul · B. Krishnamoorthy
Division of Crop Improvement and Biotechnology, Indian Institute of Spices Research, Marikunnu PO, Calicut 673012, Kerala, India
e-mail: sheeja@spices.res.in; teshee@rediffmail.com

R. Dhanya
Central Food Technological Research Institute, Mysore 570 020, India

Published online: 08 December 2013

economically important and cultivated species, *Myristica fragrans* Houtt. or nutmeg belongs to this family and is valued for both its nut and mace which are two valuable spices. Nutmeg plays a major role in economics of countries like Indonesia and Grenada. The present study deals with some important species of Myristicaceae those are included in "The IUCN Red List of Threatened Species" (http://www.iucnredlist.org/details/37093/0 accessed on Feb 11, 2013). These species showed low germination and survival rates (Tambat et al. 2006) and are restricted to Andaman and Nicobar (A&N) islands, India or *Myristica* swamps of Western Ghats, India, that are extensively encroached and destroyed habitats. The A&N islands are extremely vulnerable to environmental perturbations and irreparable damage to forest biodiversity is reported as a consequence of the Tsunami in December 2004 (Porwal et al. 2012). The *Myristica* swamps act as natural flood and erosion control devise and have good watershed value (Chandran and Mesta 2001) and the loss and desiccation of these regions will eventually have far reaching consequences on the animal population. Hence, conservation and protection of species restricted to these regions are serious concerns.

The family Myristicaceae has attracted a great deal of interest on account of its ability to produce a series of unusual phenylacylphenols (Wiart 2006), essential oils, terpenes, flavonoids and total free amino acids (Chandran and Mesta 2001; Zachariah et al. 2008), with potential for treatment of various diseases (Patro et al. 2005; Iyer et al. 2009). Many of these are ingredients of tribal medicines and possess very good pharmaceutical properties (Table1). Some of the species like *Myristica amygdalina* Wall. ex Hook. f. et Thomson are used as adulterants of the popular spice, *M. fragrans*. The species under study are good candidates for bioprospecting, but remain underutilized.

Conservation and bioprospecting being two major issues, appropriate identification and discrimination at inter/intra species level and study of phylogenetic and evolutionary relationships is mandatory. Currently, there is much interest in the use of DNA sequences as markers in taxonomic identification and biodiversity surveys augmenting conservation programmes. An approach known as DNA barcoding is widely used for taxon identification (Hebert et al. 2003). 18S rRNA/rDNA sequences have been used as barcodes and for phylogeny reconstruction within eukaryotes, including major groups of plants (Soltis et al. 1997). The rDNA sequencing is more preferred tool than rRNA sequencing due to various problems associated with extraction, stability, secondary structure barriers and sequence ambiguity of RNA (Nickrent and Soltis 1995). Commonly used PCR primers bind to highly conserved regions of the gene and will potentially amplify any 18S homologue, regardless of its origin.

Myristicaceae or nutmeg family is an older group within the angiosperms that contains recently evolved species, obligatory to specialized niches with low amount of genetic divergence (Sauquet et al. 2003; Newmaster et al. 2008). Myristicaceae are an ideal group for testing barcoding loci and identification of a suitable molecular barcode is a challenge (Janovec and Harrison 2002; Newmaster et al. 2008; Sauquet et al. 2003). Myristicaceae have so far received very less attention compared to other related taxa like Annonaceae and Magnoliaceae (Sauquet et al. 2003). A barcoding study (Newmaster et al. 2008) involving 40 nutmegs and seven loci showed that only *trnH-psbA* intergenic spacer had unique sequences for each species and for *matK*, only half of the species sampled had species-specific sequences. This study also suggested that a combination of *trnH-psbA* intergenic spacer and *matK* was the most potential for barcoding. However, the major disadvantages with the *trnH-psbA* and *matK* loci is that in case of the former there can be errors in sequencing and alignment, while the latter is difficult to amplify and sequence (Dong et al. 2012), Hence, in Myristicaceae it is suggested to depend on nuclear genome variations (Newmaster et al. 2008) or an optimal combination of more variable barcoding regions as a better approach for species resolution (Sauquet et al. 2003; Newmaster et al. 2008).

While undertaking the present study, our objectives were to examine the utility of 18S loci as a barcode for species identification and phylogenetic analysis of Myristicaceae and to generate genomic information for the conservation of closely related, vulnerable, rarely studied and medicinally important species.

Materials and methods

The leaf samples of *M. fragrans* Houtt., *M. prainii* King, *M. andamanica* Hook.f., *M. malabarica* Lamk., *M. amygdalina* (Wall. ex Hook. f. et Thomson) Warb.,

Table 1 Details of Myristicaceae analysed for the 18S rDNA loci

Species	Germplasm Accession no.	Status	Distribution	Important medicinal and other uses	References
M. fragrans Houtt.	378	Widely cultivated	Native of eastern islands of Moluccas. Introduced into most of the tropical countries including Philippines, most West Indian Islands, Tropical America and Pacific Islands. Cultivated in Indonesia, Sri Lanka, Grenada, India, Guatemala, Mexico and Nicaragua. In India it is grown in Assam, Tamilnadu, Kerala and other southern states. It requires a hot and moist climate with a rainfall of 150-300 cm per annum. It does not thrive above an altitude of 750 m.	Popular spice used for the flavouring of bakery products, sauces, pickles and meat products. The oleoresins and essential oils from nutmeg are used in the food, beverage cosmetic and pharmaceutical industries. In traditional Indian medicine it is used to cure fever, asthma, heart disease, digestive disorders, kidney diseases and lymphatic ailments. Seed butter is used as industrial lubricant.	http://zipcodezoo.com/Key/Plantae/Myristica_Genus.asp Conley (2002) Krishnamoorthy and Rema (2001)
M. prainii King	392	Endemic to Andaman and Nicobar Islands in India, status not assessed, Wild	Indonesia, Papua New Guinea, China, Philippines, Thailand and Andaman and Nicobar Islands in India.	Pharmacological properties unexplored	–
M. andamanica Hook.f.	393	Endemic, Vulnerable, Wild	Restricted to Andaman and Nicobar islands, India	For cure of sickness, stopping bleeding, wound healing, stimulant, carminative, effective in digestive, dehydration and skin disorders	Sharief (2007) Arunachalam and Subhashini (2011)
M. malabarica Lamk.	394	Endemic, Vulnerable, Wild	Western Ghats, India	High in t-caryophyllene, antioxidant, hepatoprotective, febrifuge, expectorant, febrifuge, wound healing, aphrodisiac. Seed butter is good for rheumatism, ulcer, myalgia, sprains and sores, used as rootstock for cultivated nutmeg	Zachariah et al. (2008), Manjunatha et al. (2011)
Myristica amygdalina Wall. ex Hook. f. et Thomson	395	Status not assessed, Wild	Native to South East Asia. Found in the dense forests on mountain slopes and in ravines, sparse forests of Bangladesh, India, China, Laos, Myanmar, Thailand and Vietnam.	Used as spice/adulterant of nutmeg, Inhibits biosynthesis of prostaglandins and leukotrienes and useful in inflammation and allergy with little side effects	Flora of China (2008) 7:99–101
K. andamanica (Warb.) W.J. de Wilde	396	Endemic, status not assessed, Wild	Distributed in Andaman and Nicobar Islands	Throat pain and cough	Sharief (2007)

Table 1 continued

Species	Germplasm Accession no.	Status	Distribution	Important medicinal and other uses	References
M. beddomei King	397	Endemic, status not assessed, Wild	Western Ghats in India	Used as spice/adulterant of nutmeg, high in t-caryophyllene, anti-inflammatory and antioxidant properties, check diarrhea, cough, bronchitis, used as rootstock for cultivated nutmeg	Zachariah et al. (2008)
Myristica sp.	398	–		Western Ghats, India	–
Gymnocranthera canarica (King)Warb.	399	Endemic, status not assessed, Wild	Western Ghats in India	Pharmacological properties unexplored	–

K. andamanica (Warb.) W.J. de Wilde, *M. beddomei* King, *Myristica* sp., *Gymnocranthera canarica* (King) Warb. were collected from the germplasm repository, which houses 482 different accessions of Myristicaceae collected from different locations in India. The details on sample collection, protocol for DNA isolation, PCR amplification of partial 18S rDNA loci and sequence analysis is as per Sheeja et al. (2008, 2013b). In brief, three sets of forward and reverse primers were designed for the 18S rDNA region based on the sequences already available in NCBI (GenBank accession No AF206968) and products were cloned individually into TOPO TA vector (Invitrogen Inc, USA), transformed and sequenced. For each sequence the length and proportion of GC contents were estimated, sequence alignment and trimming were done using BioEdit tool (Hall 1999). Homology searches were performed within GenBank's existing Myristicaceae (taxid: 22274) 18S sequences using basic local alignment search tool (BLASTN; Atschul et al. 1997) algorithm of BioEdit LocalBlast. A final alignment of 1,516 bases of the 18S rDNA gene was used in the analyses out of a total of approximately 1,750 bases using ClustalW2 programme (Larkin et al. 2007). The polymorphic sites, mutations, insertion deletions events (*InDel*), sequence divergence and conservation threshold were identified using DNAsp.5 (Librado and Rozas 2009) and MEGA5 (Tamura et al. 2007).

Phylogenetic analysis for 18S was performed using different methods such as Bayesian analysis, maximum parsimony, and maximum likelihood and the consensus tree was identified. Bayesian analysis was performed in MrBayes version 3.1 (Ronquist and Huelsenbeck 2003) with two searches run simultaneously for at least two million generations. Flat Dirichlet priors were used for the gamma shape parameter and the proportion of invariable sites. Three heated chains (temperature 0.2) and one cold chain were used in each search. The parameter was then fixed for a bootstrap analysis with 10,000 replicates. Tracer version 1.5 (Rambaut and Drummond 2009) was used to evaluate mixing and convergence, and to estimate appropriate burn-in period. Maximum likelihood analysis was performed using GARLI version 2.0 (Zwickl 2006) with two replicates used to estimate model parameters. These parameters were then fixed for a bootstrap analysis with 10,000 replicates. The majority-rule consensus of the bootstrap replicate trees was calculated using consense and seqboot package in the Phylip (Felsenstein 1993). When consense is run the majority rule consensus tree will result that retains the relationships found in majority of the trees and allows bootstrapping on different methods. *Magnolia virginiana* Linn. and *Magnolia acuminata* Linn. with Gen Bank accession nos. GU476443.1 and D29776.1, were designated as outgroup for 18S rDNA sequences, since Magnoliaceae have been shown to be the sister group to Myristicaceae (Sauquet et al. 2003).

For identifying the ability of 18S region to discriminate Myristicaceae species in comparison

with other markers like *matK* and *rbcL*, 76 Myristicaceae *matK* sequences: http://www.ncbi.nlm.nih.gov/nuccore/?term=txid22274%5BOrganism%3Aexp%5D+matk and 94 Myristicaceae *rbcL* sequences: http://www.ncbi.nlm.nih.gov/nuccore/?term=txid22274%5BOrganism%3Aexp%5D+rbcL. were used. Phylogenetic analysis for *mat*K and *rbcL* was performed with maximum likelihood method using GARLI version 2.0 (Zwickl 2006) with two replicates used to estimate model parameters. These parameters were then fixed for a bootstrap analysis with 10,000 replicates. *Magnolia virginiana* (GQ248153.2) and *Magnolia virginiana* (AB021049.1|) were employed as outgroup for *matK and rbcL*. Out of the 76 Myristicaceae *matK* sequences, JQ626536.1, EU090506.1, EU669473.1 and out of the 94 Myristicaceae *rbcL* sequences AF127617.1, AJ235539.1, JF738496.1, JF738507.1, JF738661.1, JF738509.1, AY299808 have been excluded for alignment correction due to lesser length. The measure of polymorphism and diversity in intra specific data information was estimated using DNAsp.5 Nucleotide diversity Pi (π) (Nei 1987 equation 10.5). Nucleotide polymorphism with Jukes and Cantor correction, π (JC) (Lynch and Crease 1990, equations 1–2). Theta (per site) from Eta (η), the total number of mutations (Watterson 1975, equation 1.4a, but on base pair basis; Nei 1987, equation 10.3). Theta values will not be reported in some cases where codons might differ by multiple changes. Estimation of nucleotide diversity (and of divergence) by the Jukes and Cantor (1969) correction was performed using the simplification indicated in Nei and Miller 1990 (equation 25). Nucleotide divergence, (average proportion of nucleotide differences between species), K (Nei 1987, equation 10.20); K (JC), average number of nucleotide substitutions per site between species with Jukes and Cantor correction and (Nei 1987, equation 10.20). Theta (per sequence) from the number of *InDel* events (Watterson 1975, equation 1.4a). DNAsp.5 will estimate conservation threshold (CT) parameters from given data according to the nucleotide variation. Conservation (C), the average conservation of the data was calculated, from the observed number of polymorphic/variable sites (S) in the data and the number of polymorphic sites (S) was estimated from nucleotide diversity (assuming mutation-drift equilibrium).CT parameters are, respectively, the minimum length and the minimum conservation value required to identify conserved regions. Here, the C was measured as the proportion of conserved sites in the alignment region (Tajima 1983, eq. A3). Based on intraspecific polymorphism, divergence and conservation threshold of 18S, *matK* and *rbcL* sequences, the ability to discriminate Myristicaceae species have been studied.

Results and discussion

Taxonomy and Nomenclature

1. *M. fragrans* Houtt.

Syn.: *M. officinalis* Linn. f.; *M. moschata* Thunb.; *M. aromatica* Lamk.

Vernacular names: Jathi kai, Japatri (Malayalam-Kerala); Jajikai (Kannada, Karnataka); Jaiphal (Hindi); Japatri (Bengali-West Bengal); Jotri (Marathi-Maharashtra); Jayapatri (Gujarathi-Gujarat)

Brief description

A lofty tree; branches slender, glabrous, leaves coriaceous, 0.64–1.3 cm elliptic—oblong or lanceolate acuminate glaucous beneath, base acute, pale yellow brown, paler with red brown nerves beneath, nerves about 8 pairs slender, petiole 0.64–1.3 cm. Flowers brancteolate males in lax slender supra axillary racemes. Male racemes 2.5–5.1 cm. Flowers 0.64 cm long, ellipsoid or urceolate, nodding; bracteole a scale under the glabrous perianth; anthers 9–12, connate in a cylindrical stipitate column. Fruit ovoid, subglobose or pyriform, the pericarp yellow, the arillus red and much laciniate.

2. *Horsfieldia prainii* (King) Warburg

Syn.: *Myristica prainii* King; *Endocomia macrocoma* (Miquel) W.J. de Wilde subsp. *prainii* (King) W.J. de Wilde; *Horsfieldia longipedunculata* Hu; *H. pandurifolia* Hu

Vernacular name: Kumpang lunau, Nyera kapok (Malay-Indonesia), YunNanFengChuiNan (Mandarin Chinese), Lal Jaiphal (Andamans)

Brief description

A tall straight stemmed tree with a high crown and slender branches, bark dark grey, smooth, cut reddish-brown with thin pinkish juice. Leaves 15.24–30.48 cm long, 7.6–12.7 cm broad, elliptic-oblong to broadly elliptic, acute, base broad and is somewhat rounded, lateral nerves 15–18 pairs. Flowers small in lax much-branched panicles. Fruit ovoid 3.8 cm long, with a thick pericarp, the seed red laciniated.

3. *M. andamanica* Hook.f.

Syn.:? *M. elliptica* Kurtz

Vernacular name: Oro (Jarawa-Andaman and Nicobar)

Brief description

A slender handsome tree) with slender horizontal branches and often with curved stilt like roots at the base, bark blackish green cut dark red with blood red juice. Leaves 20.3–38.1 cm long and 7.6–17.8 cm broad, oblong to elliptic-lanceolate, dark green and glossy above, silvery or coppery beneath, lateral nerves 12–15 pairs, rather distant and wavy; petiole strong, 2.5–3.8 cm long. Flowers few, in the leaf axils, small urceolate-globular, whitish. Fruit about the shape of a hen's egg or larger; pointed, pericarp thick, brown; seed blood red, slashed.

4. *M. malabarica* Lamk.

Syn.:? *M. tomentosa*, Grah.; *M. dactyloides,* Wall.; *M. notha*, Wall.

Vernacular names: Kattujathika (Malayalam-Kerala); Patthiri (Tamil–Tamil Nadu); Adavijajikaya (Telugu-Andhra Pradesh); Kannagi (Kannada-Karnataka)

Brief description

A moderately sized tree reaching 15.3 m in height and a diameter of 0.46 m. Bark greenish-black, smooth; wood yellowish brown tinged with grey, moderately hard. Leaves 4–8 by 3.8–10.2 cm linear-oblong or elliptic-lanceolate subacute glaucous beneath, nerves 8–14 pair, male flower in subcymose panicles bracteolate, peduncles and pedicels slender, perianth globose, anthers 10–15. Ovary sessile, ovoid globose, pointed, densely rufous-tomentose; stigma large, sessile. Fruit elongate, oblong, pointed, densely rufous-tomentose, 5.1–6.4 cm by 2.5–3.2 cm. Seed oblong, obtuse, slightly flattened on one side; testa shining; aril yellow, irregularly lobed and laciniate, extending to the apex of the seed.

5. *Horsfieldia amygdalina* (Wall. ex Hook. f. et Thomson) Warb.

Syn.: *H. prunoides* C. Y. Wu; *H. tonkinensis* Lecomte; *Myristica amygdalina* Wall. ex Hook. f. et Thomson

Vernacular names: Dieng-soh-jodao (Khasi-Meghalaya); Bolchok-pok (Garo-Meghalaya); Dettaka-rong (Mikir-Assam), FengChuiNan (Mandarin Chinese)

Brief description

A tall perfectly glabrous tree. Leaves 3.8–5.1 cm diameter, coriaceous, pale brown on both surfaces, narrowed into a petiole 1.3 cm long, nerves 8–12 pairs entire, yellow. Male panicles from the axis of fallen leaves, 7.6–12.7 cm long and nearly as broad, branched from the base, quite glabrous. Flowers loosely clustered, pedicels as long as perianth, slender; stamina column globosely trigonous, fleshy, concave; anthers about 8, wholly combined. Fruits shortly peduncled, 3.8 cm long; pericarp rather thin, glabrous, aril yellow, lacerate at the tip only.

6. *K. andamanica* (Warb.) W.J. de Wilde

Vernacular name: Oro, Aurw (Jarawa-Andaman and Nicobar)

Brief description

A moderate sized tree. Young twigs are brown or yellowish brown, sparsely tomentose. Leaves oblong to ovate, alternate, acuminate or sub-acute at apex, 15–18 cm long and 5–8 cm broad, lateral nerves 15–20 pairs, petiole 2 cm long. Fruits 1–2 cm, ellipsoid, base rounded, brown tomentose, stalked 4 mm long. Mace entire red covering the entire seed.

7. *M. beddomei* King

Syn.: *M. laurifolia* Hook. F. et Th., var. lanceolata, Hook.

Vernacular name: Patthapanu (Malayalam-Kerala); Katjathi Kai (Tamil); Jajikai (Kannada), Jayaphal (Marathi).

Brief description

A large tree reaching 27 m in height, with a diameter of about 0.76 m. Bark blackish-green, rather smooth; wood yellowish brown, moderately hard. Leaves coriaceous, glabrous and usually glaucous, smooth above. Leaf nerves and transverse nervules conspicuous, leaves 12.7–25.4 cm long, 6.4–10.2 cm broad. Ovary sessile, globose, narrowed to the apex, appressedly pubescent; stigma sessile, large, slightly oblique, subglobose. Fruit globose, 5.1–6.4 cm in diameter, the pericarp fleshy, the lacinate of the orange–red arillus with their ends separate. Seed globose, smooth, aril red, fleshy extending to the apex.

8. *Myristica* sp.

Brief description

Evergreen, tall and slender straight tree up to 6 m height, stem girth 43 cm. Bark smooth, brownish grey Plagiotropic and orthotropic branching pattern characteristic of Myristicaceae.

Leaf simple, alternate and elliptic lanceolate, petiole grooved, petiole length 0.96 cm. Lamina 20.2 cm × 6.4 cm, apex acuminate, margin entire, glabrous, shining above and glaucous beneath, coriaceous. Lateral nerves 10–13 pairs, pinnate, prominent

and brownish. The voucher specimen of the species is deposited in the herbarium facility of IISR.

9. *Gymnocranthera canarica* (King) Warb.

Syn.: *M. canarica* King; *M. farquhariana* Wall.

Vernacular names: Pintikkaya, Undaipanu, (Malayalam-Kerala); Pindi, Pindikai (Kannada-Karnataka); Undipannu (Tamil-TamilNadu).

Brief description

A very large evergreen tree with oblong leaves up to 25.4 cm long, 10.2 cm broad, glaucous beneath. Leaves alternate, entire, evergreen and pergamaceous. Flowers small, dioecious, perianth-3 to 4 lobed. Androecium sessile, the connectives combined in an oblong thick column; anthers 6–12 elongate. Ovary ovoid; stigmas sessile, connate, scarcely bilobed. Fruit is globose about 2.5 cm diameter. Seed conform to the fruit; testa woody; albumen ruminate; cotyledons divaricate, connate at the base. Bark smooth brown, wood yellowish grey, coarse, moderately hard.

Nuclear DNA analysis in wild and related Myristicaceae

The 18S rDNA sequences of nine species were considered for analyses. The sequences showed variations in both their length and nt compositions for all the specimens. The size of the complete region amplified varied from 1,767 to 1,794 in length. The complete 18S rDNA sequences of higher plants vary in length from 1,800 to 1,813 bp with a mean of 1,807 bp (Nickrent and Soltis 1995). Thus, in this study almost 98.6 % of the total length of the molecule was obtained for all the species studied. The similarity search of 18S rDNA sequences with *Myristica fragrans* (AF206968.1), *Knema latericia* (Sinclair) de Wilde (AF206946.1), *Mauloutchia chapelieri* (Baill.) Warb. (DQ007409.1) showed 96–99 % similarity at a query coverage of 72–100 %, the local blast that resulted in comparison with 18S rDNA gene sequences of Myristicaceae present in NCBI are displayed in Table 2. The GC content (%) varied from 52.77 to 51.04. The frequencies (%) of nt are A (23.31), T (23.82), C (24.48) and G (28.39). The alignment of all the sequences produced 195/1,516 variable sites and 1,257/1,516 conserved sites (online resource 1). Most of the 18S rDNA gene is largely conserved in length across all land plants (Soltis et al. 1999). Some of the regions were more variable than others. The variations were distributed in an uneven manner from 16th to 1,511th nt. This is in agreement with earlier observation in Myristicaceae (Nickrent and Soltis 1995). Total numbers of SNP sites found in the alignment were 146/1,516. *Knema andamanica* was the most distinct and included 18 variable regions and 15 *InDel* with 27 SNP sites specific to this species. The transition/transversion rate ratios were $k_1 = 2.07$ (purines) and $k_2 = 4.216$ (pyrimidines). The overall transition/transversion bias was $R = 1.533$, where $R = [A \times G \times k_1 + T \times C \times k_2]/[(A + G) \times (T + C)]$. The transitions were found to outnumber transversions similar to earlier studies on this family (Nickrent and Soltis 1995). The most frequent single type nt substitution was the T to C transition followed by the C to G transition. The total number of mutations identified was

Table 2 Blast analysis of 18S rDNA sequences of Myristicaceae

Species	TOP HIT plant from GenBank (accession no.)	Query coverage%	% Identity	Alignment length	Mismatch	Gap	E_value
M. amygdalina Wall. ex Hook. f. et Thomson	*M. fragrans* (AF206968)	100	96.89	835	10	14	0
M. andamanica Hook.f.	*M. fragrans* (AF206968)	100	98.17	820	10	5	0
M. beddomei King	*M. fragrans* (AF206968)	100	99.63	819	3	0	0
M. malabarica Lamk.	*M. fragrans* (AF206968)	99	98.56	627	8	1	0
Myristica sp.	*M. fragrans* (AF206968)	100	98.99	819	9	0	0
M. prainii King	*M. fragrans* (AF206968)	100	98.17	763	12	2	0
M. fragrans Houtt.	*M. fragrans* (AF206968)	99	99.04	628	5	1	0
G. canarica (King) Warb.	*M. fragrans* (AF206968)	100	98.56	766	9	2	0
K. andamanica (Warb.) W.J. de Wilde	*K. latericia* (AF206946)	99	99.03	821	2	6	0

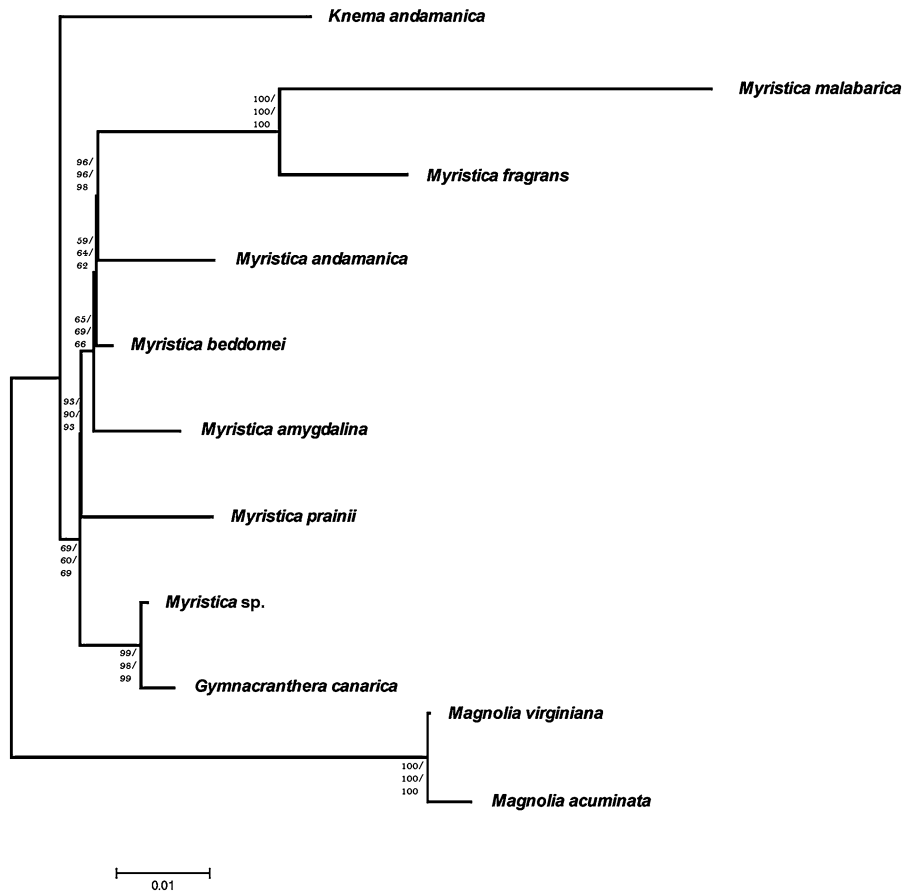


Fig. 1 Consensus phylogenetic tree of the 18S rDNA sequence for inferring relationships among species in Myristicaceae obtained by Bayesian analysis, Maximum Likelihood and Maximum Parsimony methods. Ml branches are shown in the figure, numbers on nodes represent bootstrap support values for ML (*top*), Bayesian posterior probabilities as percentages (*middle*) and MP (*bottom*)

189 and all were found to be silent mutations. The identified regions from nine species of *Myristica* and its wild and closely related genera were deposited in the GenBank Database (Accession numbers JN228257-JN228265).

The consensus phylogenetic tree (Fig. 1) of the nine species based on three methods (online resource 2) grouped the individuals into two separate groups with *K. andamanica* isolated from the rest of the species. In the *Myristica* group, three clusters were evident with *M. malabarica* and *M. fragrans* in cluster 1, *M. andamanica, M. beddomei, M. amygdalina* and *M. prainii* in the second and *Myristica* sp. and *G. canarica* in the third. Within the *Myristica* cluster, *M. malabarica* and *M. fragrans* showed close relationship supported by strong bootstrap values. The closeness between *M. andamanica*, *M. beddomei* and *M. amygdalina* may be explained based on their strong similarity in mace and fruit/seed characters. *M. beddomei* and *M. amygdalina* were almost indistinguishable except for the yellow mace character of *M. amygdalina*. However, *M. prainii* grouped separately from the rest of the *Myristica* species within this cluster, probably due to its distinguishing morphological features like compound fruits, panicles, entire mace, small fruit and seed size. Similar results were observed in the RAPD-ISSR profiles (Sheeja et al. 2013a). The clustering of individuals belonging to same species (*Myristica*) into one group supported by high bootstrap values indicated the utility of this gene region in phylogenetic analysis. Similarly, the two different genera *G. canarica* and *K. andamanica* were found to group seperately from the *Myristica* cluster. *G. canarica* was found to be close to the *Myristica*

Table 3 Comparison between nucleotide polymorphism, divergence and conservation of individual loci in Myristicaceae

Marker and data Information	Nucleotide polymorphism and divergence	Conservation
18S	Variable sites	Sequence conservation, C: 0.946
Data Information	Invariable (monomorphic) sites: 1,233	Conservation threshold, CT: 1
No. of sequences used: 9	Variable (polymorphic) sites: 182	Region, Start-End & sequence are displayed below:
Selected region: 1–1,516 Number of sites: 1,516	Total no. of mutations: 189	Region_1: 79–179
Total no. of sites (excluding gaps/missing data): 1,415	Singleton variable sites: 135	SCGTGCCGGCGACGCATCATTCAAAT TTCTGCCCTATCAAC TTTCGATGGTAGGATAGTGGCCTA CTAATTGGTGGTGACGGGTGACGR AGAATTAGGGT T
	Parsimony informative sites: 47	
	Nucleotide polymorphism	
	Pi (silent): 0.03596	
	Pi (JC-silent): 0.03685	
	Theta (silent): 0.04914	Region_2: 189–235
	No. of silent substitutions: 189	AGAGGGAGCCTGAGAAACG GCTWCCACATCCAAGG AAGGCAGCAGGC
	Nucleotide divergence	
	Ks (silent): 0.03197	Region_3: 238-303
	Ks (JC-silent): 0.03267	GCAAAYTACCCAATCCTG ACACGGGGAGGTAGTGACAATAA ATAACAATACCGRGCCTC TTCGAGT
		Region_4: 315–416
		GTGAATGAGTACAA ATCTAAATCCCYTAAC GAGGATCCCATTGGAGGGCAT AGTCTGGTGCY AGCAGCCGCGGTAATTT CCAGCTCCAAATAGCTGTATA TT
		Region_5: 1013–1200
		CGTTCTTAGTTG GTGGAGCGRTTTGTCTGGT TAATTCCGTTAACGAACGA GACCTCAGCCTGCTAACTAGSTATGCG GAGGCGACCCTCS GCGGCCAGCTTCTTAGAGGG ACTATGGCCGTTCAGGCCACGGAA GTTTGAGGCAATAACAGGTCTGTG ATRCCCTTAGATGTTCTGGGC CGCACGCGC
		Region_6: 1347-1412
		TGACTACGTCCCTT GCCCTTTGTACACACCG CCCGTCGYTCCTACCG ATTGAATGGT CCGGTGAAG

Table 3 continued

Marker and data Information	Nucleotide polymorphism and divergence	Conservation
matK	Variable sites	Sequence conservation, C: 0.867
Data information	Invariable (monomorphic) sites: 579	Conservation threshold, CT: 0.96
No. of sequences used: 73	Variable (polymorphic) sites: 33	Region, Start-End and sequence are displayed below:
Selected region: 1–613, No. of sites: 613	Total no. of mutations: 35	Region_1: 1–60 AAAAGAGAATCAAAGATTATT CGTATTCTTATATAACTCTCATGTATATGAA TGCGAATC
Total no. of sites (excluding gaps/missing data): 612	Singleton variable sites: 12	
	Parsimony informative sites: 21	
	Nucleotide polymorphism	
	Pi (silent): 0.00706	
	Pi (JC-silent): 0.00710	
	Theta (silent): 0.01177	
	No. of silent substitutions: 35	
	Nucleotide divergence	
	Ks (silent): 0.00697	
	Ks (JC-silent): 0.00700	
rbcL	Variable sites	Sequence conservation, C: 0.868
Data information	Invariable (monomorphic) sites: 362	Conservation threshold, CT: 0.96
No. of sequences used: 86	Variable (polymorphic) sites: 43	Region, Start-End dn sequence are displayed below:
Selected region: 1–456 No. of sites: 456	Total no. of mutations: 45	Region_1: 222-296
Total no. of sites (excluding gaps/ missing data): 405	Singleton variable sites: 30	TTATGTAGC TTACCCY TTAGANCTTTTTGAAG AAGGTTCT GTTACTAACATGYTT ACTTCCATT GNGGGTAATGT
	Parsimony informative sites: 13	
	Nucleotide polymorphism	
	Pi (silent): 0.00467	
	Pi (JC-silent): 0.00469	
	Theta (silent): 0.02211	
	No. of silent substitutions: 45	
	Nucleotide divergence	
	Ks (silent): 0.00462	
	Ks (JC-silent): 0.00463	

cluster. One of the unidentified species of *Myristica* grouped with *G. canarica.* We got similar results in RAPD-ISSR profiling also (Sheeja et al. 2013a). Thus molecular assignment was possible for an unknown species based on the 18S rDNA loci.

K. andamanica was clearly distinct from the rest of the samples, with the maximum number of SNP sites. A sister-group relationship between the Asian genera of *Knema* and *Myristica* was reported based on molecular data (Sauquet et al. 2003). In our earlier study using ISSR and RAPD, we observed maximum unique bands in *K. andamanica*. A diagnostic SCAR marker was developed from *K. andamanica* for species authentication (Sheeja et al. 2013b). The sister species *Magnolia* grouped separately and could be easily distinguished based on the 18S sequences. The genetic distance matrix using MEGA5 varied from 0.004 (*G. canarica* and *Myristica* sp.) to 0.081 (*K. andamanica* and *M. malabarica*). This indicated that the genetic divergence at the 18S rDNA sequences is low. None of the genera of Myristicaceae have widely disjunct distributions being endemic to a few areas (Sauquet et al. 2003). Habitat restricted plants might have lower genetic variation (Gray 1996). Based on the comparison between 18S and combined RAPD/ISSR analysis (Sheeja et al. 2013a), it was observed that 18S analysis is more informative and resolved the species in a logical manner than the RAPD and ISSR markers. Even though bootstrap values of spines in the present study were good, we still suggest that any further analysis of phylogeny of Myristicaceae may take into account an approach combining data from multiple genes (e.g. *rbcL*) and morphology (Soltis et al. 1999; Newmaster et al. 2008). In many plants, the primary challenge is to identify suitable DNA region/regions to discriminate closely related or sister species (Newmaster et al. 2008). Several studies have utilized partial sequences of 18S for phylogenetic analysis (Nickrent et al. 1995). Despite the phylogenetic promise of the 18S rDNA region, relatively few sequences were determined and the potential remains unexplored. 18S rDNA regions have been already identified as potential barcodes for authentication of *Curcuma* species (Cao et al. 2010; Sasaki et al. 2002). Our study supported this observation in Myristicaceae.

Attempts made for identification and comparison of polymorphism, divergence and conservation threshold using intraspecific data information of nine 18S, 73 *matK* and 86 *rbcL* sequences using DNAsp.5 revealed that polymorphism, divergence and conservation is higher in 18S than in *matK* and *rbcL* markers in Myristicaceae (Table 3). The overlap SNP sites for 18S, *matK* and *rbcL* in Myristicaceae are displayed in online resource 4 of Tables 4–6. The phylogenetic tree based on *rbcL* and *matK* sequences revealed lower genetic divergence (online resource 3) than 18S. Even though there is not much species overlap between 18S and that based on *rbcL* and *matK* sequences from GenBank, definitely the results provide additional information on species discrimination within Myristicaceae.

The present study suggested that comparative sequencing of 18S loci is better than other conventional markers in addressing phylogenetic relationships at the species level and above in Myristicaceae. The SNPs and variable sites identified may be exploited for developing diagnostic markers and for identifying query sequences/barcodes for accurate identification of these endangered species. This is of much advantage in case of Myristicaceae, for identification purposes since the species are morphologically similar and primarily separated by androecium characters in small, short lived flowers (Newmaster et al. 2008). Complete sequencing of entire 18S region is more critical for phylogenetic inference, since earlier studies suggest that number of variable sites in 18S region is lower and more uneven than other barcoding loci (Nickrent and Soltis 1995). This will help to maximize the number of variable sites and also facilitate proper alignment of sequences. Also large scale sampling and evaluation involving other related members and outgroup taxon is equally important.

Acknowledgments Director, IISR for providing necessary facilities and Distributed Information Sub Centre (IISR) for computational analysis of data

References

Arunachalam KD, Subhashini S (2011) Preliminary phytochemical investigation wound healing activity of *Myristica andamanica* leaves in Swiss Albino mice albino mice. J Med Plants Res 5:1095–1106

Atschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Cao H, Sasaki Y, Fushimi H, Komatsu K (2010) Authentication of *Curcuma* species (Zingiberaceae) based on nuclear 18S

rDNA and plastid trnK sequences. Yao Xue Xue Bao 45:926–933

Chandran MDS, Mesta DK (2001) On the conservation of the *Myristica* swamps of the Western Ghats. In: Uma Shaanker R, Ganeshaiah KN, Kamaljit SB (eds) Forest genetic resources: status, threats and conservation strategies. Oxford and IBH Publishing Co Pvt. Ltd, New Delhi, pp 1–19

Conley J (2002) Nutmeg only a spice? In: Whitelaw WA (ed) Proceedings of the 11th annual history of medicine days, Health Science Centre, Calgary, pp 21–26

Dong W, Liu J, Yu J, Wang L, Zhou S (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS ONE 7:e35071

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle

Gray A (1996) Genetic diversity and its conservation in natural populations of plants. Biodiversity Lett 3:71–80

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

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321

Iyer IR, Jayaraman G, Ramesh A (2009) Direct somatic embryogenesis in *Myristica malabarica* Lam., an endemic, threatened medicinal species of Southern India and detection of phytochemicals of potential medicinal value. Indian J Sci Technol 2:11–17

Janovec JP, Harrison JS (2002) A morphological analysis of the *Compsoneura sprucei* complex (Myristicaceae), with a new combination for the Central American species *Compsoneura mexicana*. Syst Bot 27:662–673

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Krishnamoorthy B, Rema J (2001) Nutmeg and mace. In: Peter KV (ed) Hand book of herbs and spices. Wood head Publishing, Cambridge, pp 238–248

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Mol Biol Evol 7:377–394

Manjunatha BK, Mankani KL, Mukunda SKR, Divakara BK, Divakara BK, Sridar PK (2011) Antioxidant and hepato protective effect of *Myristica malabarica* seed aril extracts on carbon tetrachloride induced hepatic damage. Global J Biotechnol Biochem 6:25–30

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Miller JC (1990) A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. Genetics 125:873–879

Newmaster SG, Fazekas AJR, Steeves AD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour 8:480–490

Nickrent DL, Soltis DE (1995) A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbcL* sequences. Ann Mo Bot Gard 82:208–234

Patro BS, Bauri AK, Mishra S, Chatterjee S (2005) Antioxidant activity of *Myristica malabarica* extracts and their constituents. J Agric Food Chem 53:6912–6918

Porwal MC, Padalia H, Roy PS (2012) Impact of tsunami on the forest and biodiversity richness in Nicobar Islands (Andaman and Nicobar Islands), India. Biodivers Conserv 21:1267–1287

Rambaut A, Drummond AJ (2009) Tracer version 1.5. Available: http://tree.bio.ed.ac.uk/software/tracer/ Accessed 2012

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

Sasaki Y, Fushimi H, Cao H, Cai SQ, Komatsu K (2002) Sequence analysis of Chinese and Japanese *Curcuma* drugs on the 18S rRNA gene and trnK gene and the application of amplification-refractory mutation system analysis for their authentication. Biol Pharm Bull 25:1593–1599

Sauquet H, Doyle JA, Scharaschkin T et al (2003) Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. Bot J Linn Soc 142:125–186

Sharief MU (2007) Plants folk medicine of Negrito tribes of Bay Islands. Indian J Tradit Knowl 6:468–476

Sheeja TE, Rajesh Y, Krishnamoorthy B, Parthasarathy VA (2008) Optimization of DNA isolation and PCR parameters in *Myristica* sp. and related genera for RAPD and ISSR analysis. J Spices Aromat Crops 17:91–97

Sheeja TE, Sabeesh C, Shabna OV, Shalini RS, Krishnamoorthy B (2013 a) A comparative analysis of genetic diversity of wild and related genera of *Myristica* using RAPD and ISSR markers. J Spices Aromat Crops 22 (in press)

Sheeja TE, Anju PR, Shalini RS, Siju S, Dhanya K, Krishnamoorthy B (2013b) RAPD, SCAR and conserved 18S rDNA markers for a red-listed and endemic medicinal plant species, *Knema andamanica* (Myristicaceae). Physiol Mol Biol Plant. doi: 10.1007/s12298-013-0166-6

Soltis DE, Soltis PS, Nickrent DL et al (1997) Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann Mo Bot Gard 84:1–49

Soltis PS, Soltis DE, Wolf PG, Nickrent DL, Chaw SM, Chapman RL (1999) The phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal? Mol Biol Evol 16:1774–1784

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460

Tambat B, Vishwanath K, Chaithra GN, Hareesh TS (2006) Improved germination of *Gymnacranthera canarica* Warb, an endangered, endemic tree species of *Myristica* swamps, Western Ghats. Seed Sci Technol 34:625–630

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Wiart CR (2006) Ethnopharmacology of medicinal plants: Asia and the Pacific. In: Jennifer Hackworth (ed) Humana Press Inc, New Jersey

Zachariah TJ, Leela NK, Maya KM, Rema J, Mathew PA, Vipin TM, Krishnamoorthy B (2008) Chemical composition of leaf oils of *Myristica beddomeii* (King), *Myristica fragrans* (Houtt.) and *Myristica malabarica* (Lamk.). J Spices Aromat Crops 17:10–15

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas, Austin http://zipcodezoo.com/Key/Plantae/Myristica_Genus.asp