Molecular evaluation using two Chloroplast genes of South Indian *Curcuma* species: Insight in to Phylogenetic relationship

Santhoshkumar R. and Yusuf A. *
Interuniversity centre for Plant Biotechnology, Department of Botany, University of Calicut, Kerala-673635, India.

Received: May 08, 2018 Accepted: June 24, 2018

ABSTRACT *Curcuma is one of the largest genera in the family Zingiberaceae. Identification of the genus is difficult and phylogenetic problems due to the morphological variation in the interspecific level, high phenotypic plasticity and many overlapping characters. The present study is focused to evaluate the efficiency of plastid markers rbcL and matK for species discrimination among the genus Curcuma. Intraspecific and interspecific divergence and reconstruction of phylogenetic trees were conducted for species identification. Maximum parsimony trees were generated for rbcL and matK loci. The phylogenetic trees constructed based on the individual matK and rbcL sequences showed slight variation. Phylogenetic analysis using individual data matK tree showed C. aromatica and C. raktakanta with a 100% bootstrap support in a single clade. The matK sequences showed considerable variations between Curcuma species, which can be used as a promising candidate barcode for species identification and also for inferring relationship within the Curcuma species.*

Keywords: *Curcuma*, *matK- rbcL,* Phylogeny

Introduction

Molecular methods are used in plants and other organisms for species identification and to resolve the phylogenetic relationships. In plants, the chloroplast genes, maturase K (*matK*), the larger subunit of the ribulose 1, 5-bisphosphate carboxylase (*rbcL*), and *trnH-psbA* intergenic spacer are widely used for molecular phylogenetic studies (Hilu and Liang, 1997). Barcoding based on the concept of species identification using conserved DNA sequences coded by the universal primers (Daniel et al. 2006). Molecular sequences for species level identification are selected from the chloroplast, mitochondrion, and nucleus (Savolainen and Chase, 2003). Combined sequence data from two or more loci have more discriminatory power for assessing the phylogenetic relationships than single gene locus analysis (Soltis and Soltis, 1998). The combinations of *rbcL* and *matK* genes are recommended by the Consortium for the Barcode of Life (CBOL) plant working group (2009) for species identification. Seven chloroplast genomic regions were identified by the CBOL and suggested a combination of *matK+rbcL* for the identification and authentication of angiosperms.

Chloroplast DNA sequences are utilised in barcoding due to the availability of good-quality sequences and species discrimination capabilities (Burgess et al. 2011). *rbcL* and *matK* sequences are used in barcode based species identification (Asahina et al. 2010) and interspecific phylogenetic analysis (Hilu et al. 1999). The main concept in species identification is to extract the homology of the cloned sequences of the organisms to a reference sequence, either through DNA sequence similarity searches or by phylogeny reconstruction (Altschul et al. 1997).

The family Zingiberaceae is a group of rhizomatous herbs producing delicate fleshy inflorescences. *Curcuma* L. is one of the largest genera among the Zingiberaceae family comprising of 120 species, widely used as spices, medicines, dyes and as ornamental plants (Skornickova et al. 2007). It is distributed in Asia, Australia and South Pacific. Twenty species and one variety have been reported from South India (Sabu, 2006). Species identification through morphological characters is difficult due to great intraspecific variation arising due to interspecific hybridization and polyploidization (Záveská et al. 2012). Pollen morphology is not an adequate feature to classify this genus into separate taxonomic entities. All the *Curcuma* species have a short flowering period and the flowers differ in their color and position of the inflorescence. The terminal or lateral position of the inflorescence and the presence of coma bract and bract color are the major discriminatory traits of the species. However, the position and the color of spikes has been a subject of controversy (Larsen and Smith, 1978), and one of the challenging tasks in plant DNA barcoding is to resolve closely related species (Hollingsworth et al. 2009). The species level DNA barcoding studies showed low resolution for most of the species. However, for *Curcuma* species which has maximum phenotypic plasticity, it is suggested using molecular markers for species discrimination (Kress and Erickson, 2007).

Several studies been conducted using molecular data to decipher the phylogenetic relationships within the family Zingiberaceae. These analyses revealed the pattern of evolutionary relationships to varying degrees in the genus *Curcuma*, but in general, limited to the breadth of taxon sampling as well as resolution. The present study was to find out the suitability of the chloroplast DNA markers, *rbcL* and *matK*, to identify the interspecific relationship between 20 *Curcuma* species and one variety and also to decipher the phylogenetic relationship and compare the same with morphological classification.

Materials and Method

Plant Material

Twenty species and one variety of *Curcuma* were collected from different regions of South India, planted and maintained in the Department of Botany, University of Calicut. Young leaf samples were used for DNA extraction.

DNA Isolation and Purification

Genomic DNA was isolated from the young *Curcuma* leaves using modified CTAB method (Doyle and Doyle, 1987). The extraction buffer contained 2 % (w/v) CTAB, 1M Tris–HCl (pH 8), 0.5M EDTA (pH 8) and 5 M NaCl. The leaf samples were powdered in liquid N_2 and the powder was transferred to preheated (65^{o}C) CTAB buffer. 2% (w/v) PVP was added at the time of homogenization. To the homogenate Proteinase K (10 mg/ml) and 2 % (w/v) β-mercaptoethanol were added and mixed. The homogenate was incubated at 60^{o}C for 1hr with intermittent mixing. The extract was centrifuged at 4^{o}C for 12 min at 12,000 rpm and the supernatant was collected in a new Eppendorf tube. An equal volume of chloroform: isoamyl alcohol (24:1) was added to the supernatant and centrifuged at 12,000 rpm for 12 min at 4^{o}C. The supernatant was extracted twice with equal volume of phenol: chloroform: isoamyl alcohol (25:24:1), followed by extraction with chloroform: isoamyl alcohol (24:1), and centrifuged at 12,000 rpm for 12 min at 4^{o}C. The supernatant was transferred to a fresh Eppendorf tube and one by third volume of chilled isopropyl alcohol was added, and incubated for 1hour at -80^{o}C. The sample was centrifuged at 12,000 rpm for 12 min at 4^{o}C and the pellet was dissolved in minimum quantity TE buffer. 10µl of RNaseA (10mg/ml) was added and incubated at 40^{o}C for 30 min and an equal volume of chloroform: isoamyl alcohol (24:1) were added, mixed well and centrifuged at 12,000 rpm at 4^{o}C. The upper layer was transferred to a fresh Eppendorf tube, chilled isopropyl alcohol was added and mixed well. The samples were incubated at -80^{O}C for one and half hours and centrifuged at 12,000 rpm for 12 min at 4^{o}C to pellet the DNA. DNA pellet was washed with 70 % (v/v) chilled ethanol and air dried at room temperature and re-dissolved in 20 µl TE buffer and stored at -80^{o}C. The quality and quantity of DNA samples were assessed using a Nanodrop spectrophotometer (Thermo, USA) and also by running the DNA in 1% (w/v) agarose gel.

PCR amplification

The Polymerase chain reactions (25 µl) were conducted in a reaction mixture containing 2.5µl 10 X buffer, 1U *Taq* DNA polymerase, 10mmol dNTP mix, 50mmol $MgCl_2$, 10 ng template DNA, 90nmol forward primer, 85 nmol reverse primer for *rbcL* and 88.4nmol forward primer and 72.4nmol reverse primer for *matK*. PCR conditions for *rbcL* (initial denaturation at 94°C for 2 min, 30 cycles comprising a denaturation at 94°C for 15 sec, annealing temperature in a gradient of 53 to 57°C for 30 sec, extension at 72°C for 1 min and final extension at 72°C for 10 min) and for *matK* (initial denaturation at 94°C for 2 min, 30 cycles comprising a denaturation at 94°C for 15 sec, annealing temperature in a gradient of 55 to 60°C for 30 sec, extension at 72°C for 1 min and the final extension at 72°C for 10 min) The amplified PCR product was electrophoresed using 1% (w/v) agarose gel and the band was eluted and purified using Minigel purification kit (Takara, Japan). Sequencing was done from Scigenom Lab Pvt, Ltd (Cochin, Kerala), on a charge basis.

rbcL and *mat*K data analysis

DNA sequences obtained using the primers from all the species were minimally edited and manually aligned using BioEdit software. Maximum parsimony (MP) analysis were performed using heuristic search with MULPARS, tree-bisection-reconnection (TBR) branch swapping, and RANDOM stepwise addition with 1000 replicates. The MP tree was constructed using MEGA 6.0 software. The species identification and homology between the sequences were carried out using BLAST method.

Results

The individual and combined *matK* and *rbcL* nucleotide sequence for each species were used to determine the phylogenetic relationship between *Curcuma* spp. PCR-amplified fragments of both *matK* and *rbcL* were

sequenced and the sequences were submitted to the NCBI Genbank and accession numbers were provided. The cloned sequences of *matK* gene ranged between 730 bp to 790 bp. *Zingiber officinale* sequences were used as the outgroup (KY 448306, KY226001). The accession numbers provided by NCBI were listed below.

Species	*rbcL*	*matK*
C. coriacea	KX608612	KU934093
C. karnatakensis	KX608613	KU736742
C. bhatii	KU697332	KX170829
C. raktakanta	KX608610	KX455852
C. oligantha var *oligantha*	KX650824	KX455853
C. aeruginosa	KX608611	KX455854
C. vamana	KX608615	KX455855
C. haritha	KX608606	KX148521
C. oligantha var *lutea*	KX608609	KX418654
C. amada	KX608605	KX650813
C. decipiens	KX608618	KX650814
C. kudagensis	KU886554	KX650815
C. aromatica	KX650825	KX650816
C. mutabilis	KX608607	KX650817
C. pseudomontana	KX608616	KX650818
C. neilgherrensis	KX608608	KX650819
C. longa	KX608614	KX650820
C. xanthorrhiza	KX650821	KX650811
C. inodora	KX650822	KX650809
C. montana	KX650823	KX650810
C. ecalcarata	KX608617	KX650812

Phylogenetic analysis based on *matK* gene sequences

The phylogenetic tree constructed using the *matK* sequences for all the 21 samples showed that each species has clearly distinctive clades. In the *matK* phylogram, the 20 *Curcuma* species and one variety were distinctively assorted into two Groups (Fig. 1) Group I has two clades A, B and two monoclades. Clade A has two subclades C and D. *C. montana ,C. ecalcarata , C. amada ,C. decipiens , C. longa, C. pseudomontana* were grouped under the sub clade C and *C. vamana ,C. xanthorrhiza ,C.aeruginosa ,C. haritha* were clustered in sub clade D. Clade B comprised of 7 species *C. oligantha var. oligantha, C. oligantha* var*. lutea, C. mutabilis, C. karnatakensis, C. coriacea, C. neilgherrensis and C. kudagensis. C. inodora* and *C. bhatii* existed as two monoclades. *C. raktakanta* and *C. aromatica* were grouped together in Group II with 100% bootstrap similarity. The overall mean distance between all sequences was 0.108 and transition/transversion range between the species was 1.27.

Phylogenetic analysis based on *rbcL* gene sequences

The Phylogenetic tree generated based on the *rbc*L genes sequences derived two groups in MP-analysis (Fig. 2). Group 1 comprised of two clades and two monoclades. Clade A has 7 species *C.bhatii* , *C. vamana*, *C. kudagensis*, *C. oligantha* var. *lutea*, *C. coriacea, C. pseudomontana* and *C.aromatica* . Clade B comprised of 9 species with 7 species forming a monoclade having *C. aeruginosa ,C. haritha*, *C. ecalcarata* , *C. amada ,C. decipiens , C. raktakanta* and *C. inodora. C. longa* and *C. karnatakensis* exist as two separate clades. Group II had two clusters of 5 species; *C. xanthorrhiza, C. montana, C. mutabilis, C. neilgherrensis* and *C. oligantha var. oligantha.* Overall mean distance between the sequences was 0.228 and transition/transversion between the species was 1.42.

Discussion

The morphology based classifications of *Curcuma* species are confusing, which is based on the position of the inflorescence and flower colour (Škorničková, 2007). *Curcuma* species have been classified based on highly specialized floral morphologies, unusual pollination and unusual ecological habitat adaptations despite their morphology being well within the range of variation. The systematic studies conducted so far have been confusing and incomplete due to inherent taxonomic and biological problems of the genus. Original descriptions of many *Curcuma* species are unclear and erroneous and type specimens are often lacking or incomplete. In addition, high intra- and inter population variation has led to disagreement concerning species concepts and boundaries. Chromosome numbers have been widely investigated in

Curcuma species, however, high variability 2n=20-105 was observed in the somatic chromosome numbers of this genus (Škorničková et al. 2007).

Universal barcode based phylogeny utilize universal primers (*rbcL* and *matK,* individuality or in combination) for plant DNA barcoding, identification and discrimination. The amplification success rate of the chloroplast sequences *rbcL* and *matK* was100% and 99% for all the twenty species and one variety, suggesting to deduce the evolutionary relationship among the *Curcuma* species. The results showed that the sequence generated phylogenetic trees are incongruent and incompatible. Comparison of the two barcode markers showed that *matK* can serve as a potential barcode region for the identification of the genus *Curcuma,* because this region showed less divergence and exhibited sufficient genetic distance between the species, consequently this gene showed better performance than the *rbcL* region, with the better variability identified in the *matK* sequences. Studies showed that *matK* gene in angiosperms evolved much faster (2–3 times) than *rbcL* (Soltis et al. 1996). *matK* locus has been used as better candidate than other loci and proposed as a candidate locus for barcoding in angiosperm families like Myristicaceae, Podostemaceae, Lamiaceae and Fabaceae (Newmaster et al. 2008; Kelly et al. 2010; Theodoridis et al. 2012; Gao et al. 2010).

Based on the *matK* phylogram, *C. aromatica* and *C. raktakanta* from different geographical locations clustered together in a single clade with 100% bootstrap similarity. However, these two species were clustered together in a single clade in *rbcL* and *matK* combined phylogenetic tree in spite of the morphology based close relationship between *C. aromatica* with *C. haritha* and both the species have a few morphological characters congruent with *C. raktakanta,* but *C. raktakanta* closely resembles with *C.aeruginosa* morphologically (Sabu, 2006). Barcode based phylogenetic analysis results were not in agreement with the morphological classification of these two *Curcuma* species. *C. aromatica* is a seed setting species (George, 1981) and the seedling variation due to out-crossing may be the reason for the morphological variation as observed with *C. raktakanta*. Barcode based phylogenetic analysis using *matK* alone and *rbcL -matK* combination, *C. aromatica* and *C. raktakanta* showed close proximity, *C. haritha* clustered along with *C. aeruginosa* existed as a single clade. The phylogenetic approach affirms the paraphyletic origin of these two species with (100 %) bootstrap value; however, *C. raktakanta* did not show any relationship with *C. aeruginosa.* Morphological character based classification showed dissimilarities between *C. inodora* and *C.decipiens* (Sabu, 2006), barcode based phylogeny did not show any conclusive evidence for similarities.

The cohabiting species, *C. decipiens* and *C. aurantiaca* showed high similarity in the floral, vegetative and rhizome characters and showed discrimination in the anther spur length between these two species. However, *C. decipiens* and *C. ecalcarata* share many common vegetative, rhizome and floral traits and also cohabit at high altitudes in rocky patches (Sabu, 1991). But our results showed that these two species are clustered in different clades. *C. xanthorrhiza* and *C. montana* were morphologically distant primitive species clustered together in group II with lesser morphological differences. *C. pseudomontana* closely resembles *C. montana*, having a confused taxonomy except for the development of side corms. These two species share many common floral and vegetative characters and occur in similar habitat (Sabu, 1991) with *C. pseudomontana* exhibiting both central and lateral spikes (Santapau, 1945).

The molecular systematic study using the individual as well as combinational *rbcL* and *matK* barcode genes indicated that genotypic characters are not in agreement with the morphological features based phenotypic characters particularly for closely related species within the genus *Curcuma*, the *matK* gene sequence could provide a partial evidence for the identification of medicinal *Curcuma* species. Indeed, each group of *Curcuma* species has a unique sequence pattern in the *matK* gene region, so that they could be easily distinguished at the DNA level.

The *matK* and *rbcL* matrices better supported the phylogenetic analysis of *Curcuma* species. MP is one of the correct methods that can easily determine the insertions/deletions of nucleotides and provide important phylogenetic information. Although MP is quite competent in obtaining the correct topology (Nei, 1991, 1996), it may also give incorrect trees when the rate of nucleotide substitution is fairly constant among the taxa (Takezaki and M. Nei, 1994; Zharkikh and W.H. Li, 1993) hence selected MP tree method for discrimination of closely related *Curcuma* species.

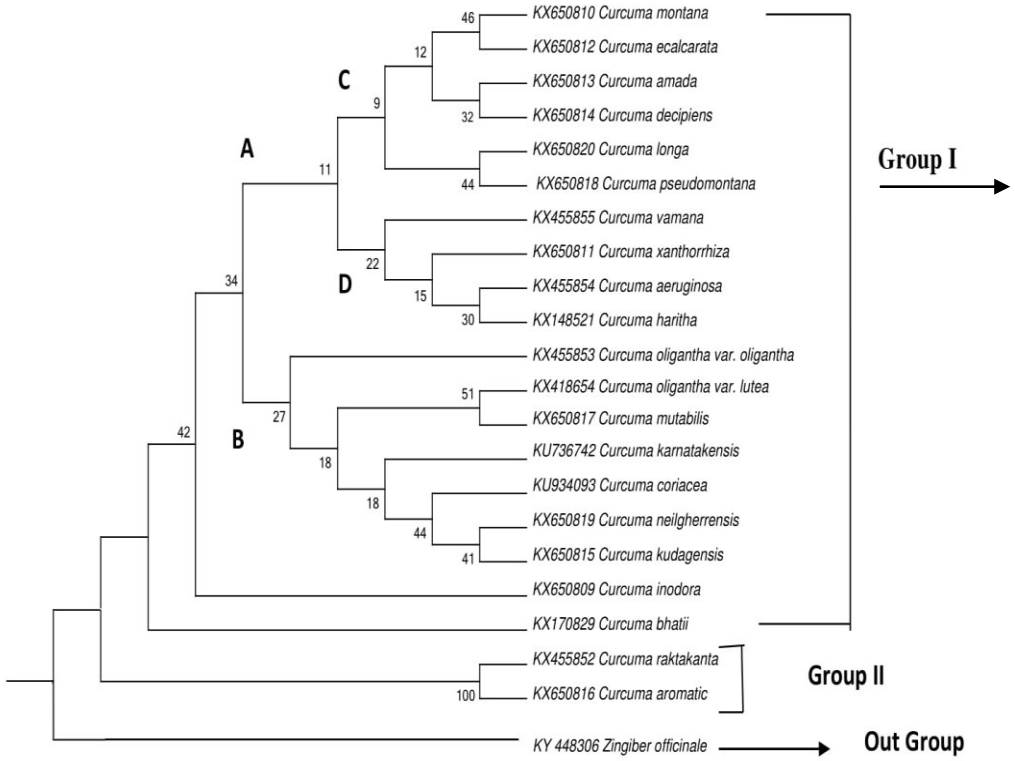

Fig 1: Phylogenetic tree of *matK* showing Evolutionary relationship between 20 curcuma species and on variety. Bootstrap values from 1000 replications were employed for MP method

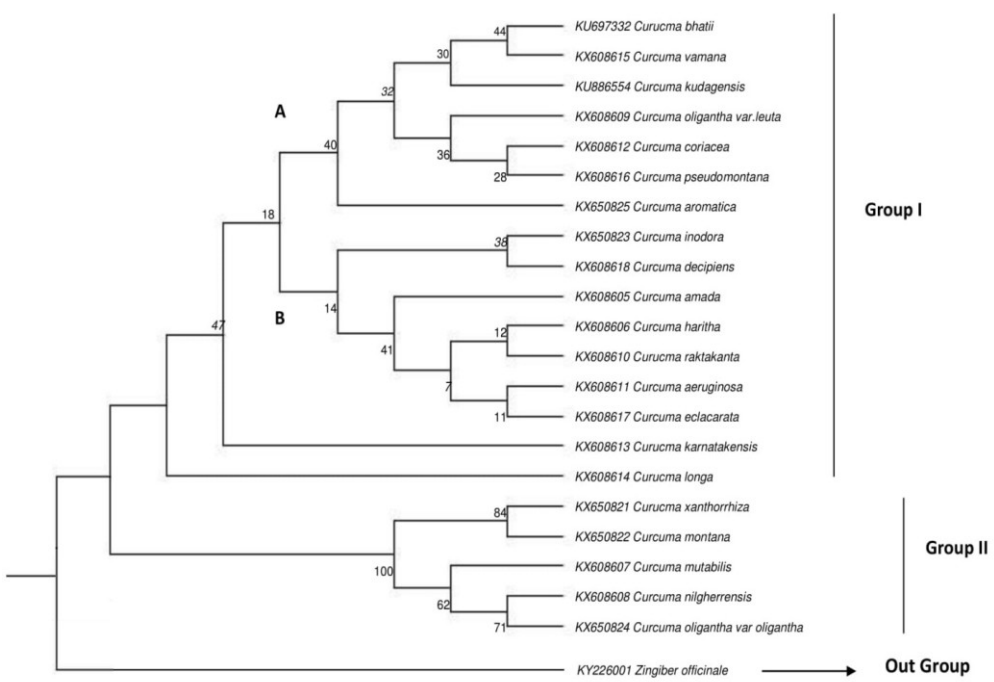

Fig 2: Phylogenetic tree of *rbcL* showing Evolutionary relationship between 20 curcuma species and on variety. Bootstrap values from 1000 replications were employed for MP method

Conclusions

Phylogenetic analysis is a useful and effective tool for the evaluation of species relationship in genus *Curcuma*. The current results showed the higher discriminatory power of markers *rbc*L and *mat*K for *Curcuma* species identification. Comparison of the two barcode markers proved that *mat*K is a better candidate for resolving the interspecific relationship among *Curcuma* species.

Acknowledgements

The authors as thankful to the Director, Interuniversity Center for Plant Biotechnology, University of Calicut for providing the facilities and SR is grateful to University Grant Commission for providing RGNF for conducting this study.

Compliance with ethical standards

Conflict of interest

The authors declared that they have no conflict of interest.

References:-

1. Altschul, S., Madden, T., Scha¨ffer, A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D.J. (1997). Gapped BLAST and PSIBLAST: a new generation of protein database search programs. Nucleic Acids Research, 25, 3389–3402.
2. Asahina, H., Shinozaki, J., Masuda, K., Morimitsu, Y., and Satake, M. (2010). Identification of medicinal Dendrobium species by phylogenetic analysis using matK and rbcL sequences. Journal of Natural Medicines, 64, 133-138.
3. Burgess, K.S., Fazekas, A.J., Kesanakurti, P.R., Graham, S.W., Husband, B.C., Newmaster, S.G., Percy, D.M., Hajibabaei, M., & Barrett, S.C. (2011). Discriminating plant species in a local temperate flora using the rbcL+ matK DNA barcode, Methods in Ecology and Evolution, 1-8.
4. CBOL Plant Working Group. (2009). A DNA barcode for land plants. Proceedings USA.- 106, 12794-12797.
5. Doyle, J.J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bulletin Botanical Society of America, 19,11-5.
6. Daniel, R., Cameron, S., & Will, K. (2006). Are Plant DNA Barcodes a Search for the Holy Grail? Trends in Ecology & Evolution, 21(1):1-2.
7. Gao, T., Yao, H., Song, J., Zhu, Y., Liu, C., & Shilin, C. (2010). Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. BMC Evolutionary Biology, 10, 324.
8. George, H. (1981). Variability in the open pollinated progenies of turmeric C. aromatic Salisb. M. Sc (Hort.) Thesis, Kerala Agricultural University, Thrissur.
9. Hilu, K., & Liang, H. (1997). The matK gene: sequence variation and application in plant systematics. American Journal of Botany, 84, 830–839.
10. Hilu, K.W., Alice, L.A., Jacobs, S.W.L., & Everett, J. (2000). Phylogenetic relationships in subfamily Chloridoideaea (Poaceae) based on matK sequences: A preliminary assessments, Grasses, systematics and evolution, Melbourne CSIRO Australia, 173-179.
11. Hollingsworth, M.L. Andra Clark, A.L.E.X., Forrest, L.L., Richardson, J., Pennington, R.T., Long, D.G., Cowan, R., Chase, M.W., Gaudeul, M., & Hollingsworth, P.M. (2009).Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Molecular Ecology Resources, 9, 439-457.
12. Kelly, L., Ameka, G., & Chase, M. (2010). DNA barcoding of African Podostemaceae (river-weeds): A test of proposed barcode regions. Taxon, 59: 251–260.
13. Kress, W.J., & Erickson, D.L. (2007). A two-locus global DNA barcode for land plants: the coding rbcL gene complements the noncoding trnH-psbA spacer region. PLoS ONE. 2:e508.
14. Larsen, K., & Smith, R.M. (1978). A new species of Curcuma from Thailand. Notes from the Royal Botanic Garden Edinburgh. 36(2), 269-272.
15. Leong- Skornickova, J., Sıda, O., Jarolımova, V., Sabu, M., Fér, T., Trávnícek, P., & Suda, J. (2007). Chromosome numbers and genome size variation in Indian species of Curcuma (Zingiberaceae). Annals of Botany, 100,505–526.
16. Newmaster, S.G., Fazekas, A. J., Steeves, R. A. D., & Janovec, J. (2008). Testing candidate plant barcode regions in the Myristicaceae. Molecular Ecology Resources, 8, 480–490.
17. Sabu, M. (1991). A taxonomic and phylogenetic study of South Indian Zingiberaceae Ph.D Thesis, Department of Botany, University of Calicut. 201–243.
18. Sabu, M. (2006). Zingiberaceae and Costaceae of South India. Association for Angiosperm Taxonomy. -126-186.
19. Santapau, H. (1945). Curcuma pseudomontana Grah. Journal of the Bombay Natural History Society, 45, 618-623.
20. Savolainen, V., & Chase, M.W. (2003). A decade of progress in plant molecular phylogenetics. Trends in Genetics, 19, 717 – 724.

21. Škorničková, J. (2007). Taxonomic studies in Indian Curcuma L. PhD. Thesis Charles University, Prague, Czech Republic.
22. Soltis, D.E., Kuzoff, R.K., Conti, E., Gornall, R., & Ferguson, K. (1996). matK and rbcL gene sequence data indicate that Saxifraga (Saxifragaceae) is polyphyletic. American Journal of Botany, 83, 371-382.
23. Soltis, D.E., & Soltis, P.S. (1998). Choosing an approach and an appropriate gene for phylogeneticanalysis. In: Soltis, D.E. Soltis, P.S. and Doyle, J.J. (eds) Molecular systematics of plants II: DNAsequencing. Kluwer, Dordrecht, 21–24.
24. Theodoratos, S., Stefanaki, A., Tezcan, M., Aki, C., Kokkini, S., & Vlachonasios, K.E. (2012). DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Cesme-Karaburun Peninsula (Turkey). Molecular Ecology Resources, 12(4), 620–633.
25. Záveská, E., Fér, T., Šída, O., Krak, K., Marhold, K., & Leong-Škorničková, J. (2012). Phylogeny of Curcuma (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus Ecomata.- Taxon. 61, 747–763.
26. Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. Annual Review of Genetics, 30: 371-403.
27. Nei, M. (1991). Relative efficiencies of different tree making methods for molecular data. In Recent Advances in Phylogenetic Studies of DNA Sequences. Oxford University Press., 133-47.
28. Takezaki, N., & Nei, M. (1994). Inconsistency of the maximum parsimony method when the rate of nucleotide substitution is constant. Journal of Molecular Evolution, 39: 210-18.
29. Zharkikh, A., & Li, W.H. (1993). Inconsistency of the maximum parsimony method: the case of five taxa with a molecular clock. Systematic Biology, 42: 113-25