Indian J. Genet., 79(1): 82-88 (2019)
DOI: https://doi.org/10.31742/IJGPB.79.1.11

Molecular characterization and phylogenetic assessment of a few *Dioscorea* (Dioscoreaceae) species of North-East India

Nilofer Sheikh*, Yogendra Kumar and A. K. Misra

Department of Botany, North Eastern Hill University, Shillong 793 022, Meghalaya

(Received: October 2018; Revised: February 2019; Accepted: February 2019)

Abstract

***Dioscorea* spp. or yam are consumed by the indigenous peoples of North-Eastern region of India as a substitute for potato and most of the species are also used in traditional medicine. North-East region is one of the hotspot for *Dioscorea* species growing in wild habitat which has not been characterized or identified. In the present study, eight morphologically distinct species of *Dioscorea* belonging to section *Enantiophyllum*, *Botryosicyos* and *Opsophyton* were subjected to molecular characterization and phylogenetic assessment using three marker genes (18S rDNA, *matK* and *rbcL*). The results of sequence characterization of the these genes revealed that 18S rRNA gene was highly conserved than *matK* and *rbcL* gene sequences and hence 18Sr RNA gene can be used as better candidates for species delimitation. The phylogenetic analysis of the combined molecular gene sequences also showed that the species belonging to section *Enantiophyllum* were monophyletic in origin.**

Keywords: Molecular characterization, 18S rDNA, *matK*, *rbcL*, maximum parsimony, Bayesian inference

Introduction

The North-eastern region of India forms a distinctive part of the Indo-Burma hotspot which ranks sixth among the 25 biodiversity hotspots of the world and is climatically, ecologically and ethnically very diverse. The region is largely dominated by indigenous groups such as Naga, Khasi, Mizo, Miri, Adi, Aka, Apatan, Maripa, Mushard, Garo, Naga etc., who are socio-culturally different and mainly depend on wild edible fruits, roots and tubers etc. for nourishment. As a wild tuber crop, *Dioscorea* species play a prime role in providing food and medicinal requirements for the local communities in this region. Various species of *Dioscorea* also serve as food security crop for these local communities. (Anon. 1952). About 50 species of *Dioscorea* are distributed in India and approximately 28 species (Sharma and Hore 1995) are distributed in North-East India. The identification of these species has been relied on morphological description. Several workers (Knuth 1924; Burkill 1960; Coursey 1967; Ayensu 1972) have classified this genus based on seed morphology, floral characters, underground organs and anatomical characters, yet the systematics of the genus is not completely resolved due to morphological diversity, dioecy and small flowers. Hence, the study of phylogenetic relationships based on morphological characters among *Dioscorea* species became difficult due to continuous variability which restrict species identification. Phylogenetic assessment based on molecular dataset analysis has provided a clear relationship within the genus. Plastid genome such as *rbcL* and *matK* has provided a rich source of phylogenetic tool to unravel the genetic relationships. Nuclear gene sequences such 18S rDNA in particular have proved a good source of phylogenetic information in Dioscoreales (Caddick et al. 2002a). However, 18SrDNA sequences have proven to contain less phylogenetic signal in comparison to plastid DNA due to low sequence divergence (Soltis et al. 1997). Wilkin et al. (2005) successfully reconstructed phylogenetic relationships of 67 *Dioscorea* taxa of Madagascar based on chloroplast *rbcL* and *matK* sequence data and found out that the main old world groups (such as the left twining *D*. sect. *Stenophora* and the right twining *D*. sect. *Enatiophyllum*) were monophyletic. Hsu et al. (2013) with chloroplast *trnL-F, matK, rbcL* and *atpB-rbcL* sequence data reconstructed phylogenetic relationship

*Corresponding author's e-mail: nilofersheikh83@gmail.com
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

of 48 *Dioscorea* species from East and Southeast Asia. The study reflects on *D*. sect. *Combilium* and sect. *Shannicorea* as closely related sections. In India, Mukherjee and Bhat (2013) studied phylogenetic relationship of wild and cultivated yam species of India and inferred from PCR-RFLP analysis of two cpDNA loci included 25 species of *Dioscorea.* Similarly, Barman et al. (2018) also studied the phylogenetic assessment of 16 Indian *Dioscorea* species based on chloroplast gene sequences. But both these studies on Indian *Dioscorea* species inferred by phylogenetic analysis based on chloroplast sequences were restricted to only half of the species distributed in India. Hence, in the present study an attempt had been made to assess the phylogeny of *Dioscorea* species of NE India which would be an addition to the previous phylogenetic studies on the Indian *Dioscorea* species.

Material and methods

Plant material and taxon sampling

Plant sample of eight species of *Dioscorea* (Fig. 1) *viz*., *Dioscorea pentaphylla* L., *Dioscorea alata* L., *Dioscorea belophylla* (Prain) Haines, *Dioscorea glabra* Roxb, *Dioscorea pubera* Bl., *Dioscorea oppositifolia* L., *Dioscorea lepcharum* Prain *et* Burk., *Dioscorea bulbifera* L. were collected from North-eastern region of India. The eight species were identified at Botanical Survey of India, Eastern Regional Centre, Shillong. Specimen vouchers (appendix) were deposited at the Herbarium, Botany department, North Eastern Hill University, Shilllong.

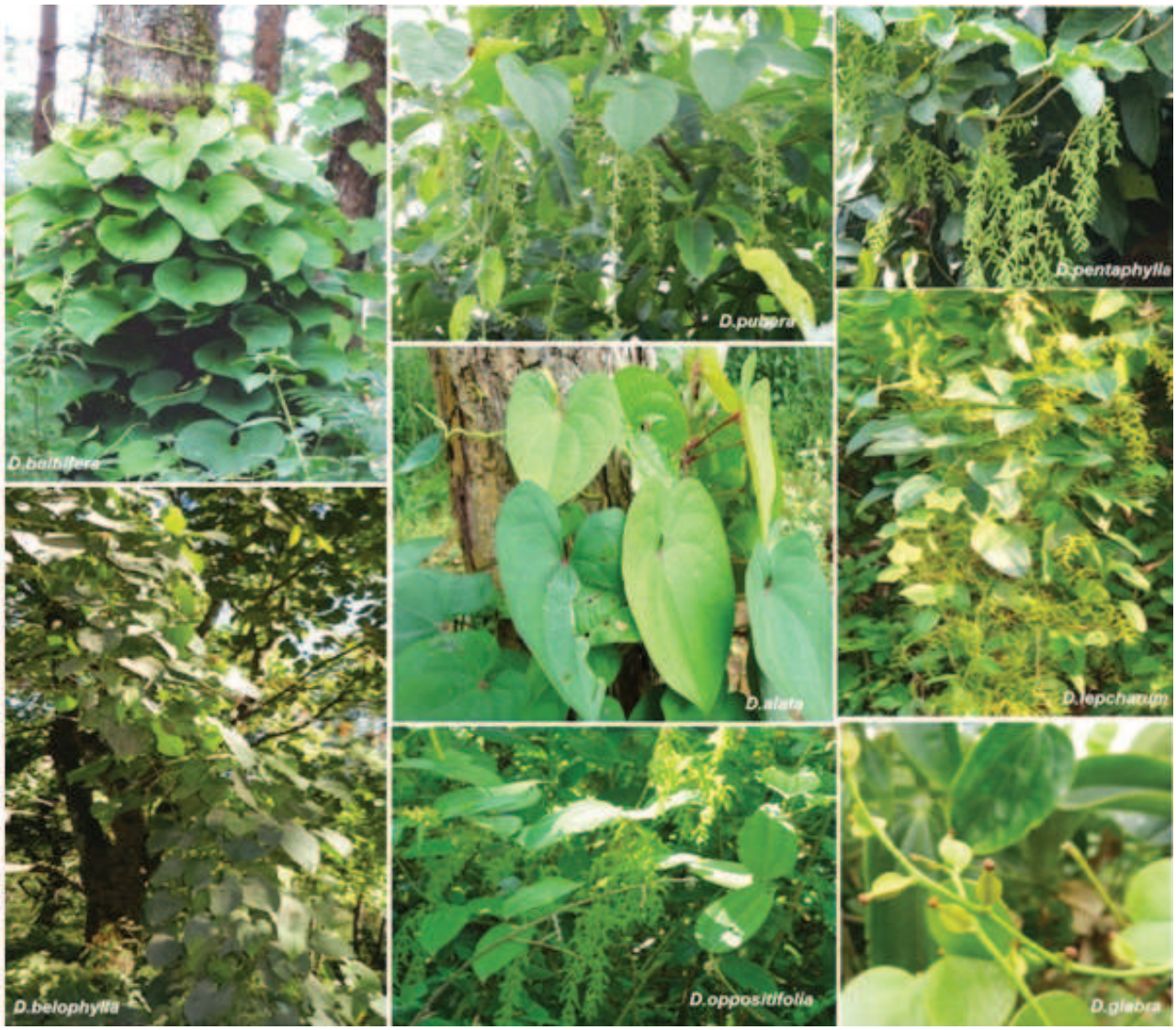

Fig. 1. *Dioscorea* species of North east India

DNA extraction, amplification and sequencing

Total genomic DNA was isolated from fresh leaves following Doyle and Doyle (1987) method of DNA extraction with the addition of the saturated phenol extraction step prior to ethanol precipitation. Polymerase chain reaction was used to amplify 18S rDNA, *matK* and *rbcL* regions. PCR primers were designed using DS GENE version 1.1 software to amplify these regions. Sequencing was done utilizing these primers (Ns18S-F5'- GCAAGTATGAAC TAATTCGG-3´; Ns18S-R5'-TGATCTATCCCCA TCACG-3'; NSrbcL-F5´GATACTGATATCTTA GCGGCAT-3´; NSrbcL-R5´-CCCGATTAGCTAC TGCACC-3´; NSmatK-F5´-CAGGAGTATATTTA CACAC-3´; NSmatK-R5´-GCAGGTCATTGATACA GAT-3´). DNA amplification was performed in *Applied Biosystems® GeneAmp® PCR System 9700*. Amplified PCR products were puried and sequence at Bangalore Genei, India and Axygen Scientific Pvt. Ltd. India. The nucleotide sequences for all the genomic regions have been submitted to the Genbank databases (www.ncbi.nlm.nih.gov) (Table 1).

Sequence alignments

Sequences, obtained were subjected to multiple sequence alignment using clustal x program (Thompson et al. 1997) with default setting. A separate alignment matrix for each genomic region was produced. Alignments of all genomic regions were combined to a single nexus file comprising several data partition. Sequences characters such as number of conserved sites, variable sites, parsimony informative sites, indels, GC%, transition and transversion of the three genomic regions (18SrDNA, *matK* and *rbcL*) were calculated (Table 2) by using both MEGA7 (Kumar et al. 2016) and Seqstate v.1.21 (Muller, 2005).

Phylogeny based on molecular analysis

A phylogenetic tree combining three genomic regions were constructed using MEGA7 (Kumar et al. 2016). Maximum Parsimony (MP) methods were used to analyse the aligned sequence data matrix of the three genes. In the MP analysis, characters were equally weighted and a heuristic search option with Tree Bisection Regrafting (TBR) branch-swapping and 10 random stepwise additions was used. Bootstrap values for MP analyses were estimated from 1000 replicates with the above heuristic settings. The best fitting substitution model of evolution was selected by the BIC criterion and maximum likelihood ratio test for each region and the combined matrix using jModeltest 0.1 (Posada 2008). The substitution model selected for nuclear 18SrDNA gene is JC+G and for chloroplast *matK* and *rbcL* is HKY+G whereas the model selected for combined nuclear and chloroplast dataset (18SrDNA+*matK* +*rbcL*) is F18+G. The parameters

Table 1. Details of species, specimen voucher and their gene bank accession number

Section	Species	Specimen voucher*	Gen Bank accessions		
			18SrRNA	*matK*	*rbcL*
Botryosicyos (Hochst.) Uline	*Dioscorea pentaphylla* L.	NEHU-11946	KC921380	KF372562	KF372564
EnantiophyllumUline	*Dioscorea alata* L.	NEHU-11944	KC921381	KF372557	KF372566
	Dioscorea belophylla (Prain) Haines	NEHU-11950	KC921384	KF372555	KF372568
	Dioscorea giabra Roxb.	NEHU-11937	KC921388	KF372558	KF372567
	Dioscorea pubera Bl.	NEHU-11948	KC921383	KF372556	KF372563
	Dioscorea oppositifolia L.	NEHU-11940	KC921387	KF372559	KF3 72569
	Dioscorea lepcharmn Prainet Burk.	NEHU-11942	KC921385	KF372560	KF372570
Opsophyton Uline	*Dioscorea bulbifera* L.	NEHU-11935	KC921382	KF372561	KF372565
Outgroup	*Trichopus zeylanicus* L.		AF309394	AY973845	AF3 07477

Specimen vouchers deposited at the herbarium, Department of Botany, North Eastern Hill University, Shillong

Table 2. Comparative sequence characterization of three genomic regions

Genomicregion	Charac-ters	Conserved sites	Variable sites	Parsimony informative sites	Transi-tion	Trans-version	Ts/Tv rate(R)	Indel	GC%
18S rDNA	1414	1305(**92.2%**)	103(**7.2%**)	17(**1.20%**)	43	57	0.79	10	52.4%
rbcL	1229	1007(**81.93%**)	64(**5.21%**)	48(**3.91%**)	74	26	1.01	5	43.9%
matK	1256	1001(**79.6%**)	91(**7.24%**)	41(**3.26%**)	54	46	1.08	12	31.9%

of substitution model for individual and combined dataset such as Bayesian Information Criterion (BIC), Maximum Likelihood value (lnL), nucleotide frequencies, substitution rate matrix and gamma rate distribution were also recorded (Table 3). Bayesian inference (BI) of phylogeny analysis was used since the parsimony trees are less resolved and have weaker overall support. Bayesian inference (BI) analysis were performed using MRBAYES v.3.1.2 (Ronquist and Huelsenbeck 2003) for 1000000 generations applying the default setting {Markov chain Monte Carlo (MCMC), two runs with four chains each, heating temperature 0.2, saving one tree every 100generations}. The posterior probabilities (calculated with MrBayes) were recorded to represent the support for nodes. All the trees were viewed with program Figtree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Outgroup selection

Trichopus zeylanicus, a species of the genus *Trichopus* of Dioscoreaceae was selected as the outgroup (Wilkin and Caddick 2000; Wilkin et al. 2005) for the present study. Gene sequences for the three genomes of *T. zeylanicus* was retrieved from Genbank (www.ncbi.nlm.nih.gov).

Results and discussion

Molecular characterization

Sequence alignment produced 1414, 1229 and 1256 characters for 18S rDNA, *rbcL* and *matK* respectively. Among all the molecular markers, 18S rDNA is the most conserved as evident from number of conserved sites (1305 out of 1414, 92.2%). The number of parsimony informative sites in 18S rDNA region is about 1.12% which suggested that this region is less evolved. 18SrDNA typically evolves at one-third to one-half the rate of *rbcL*. The *rbcL* region was recorded to have low variable sites (5.21%) and low indels. Higher number of indels was recorded in *matK* followed by 18SrDNA. All markers recorded frequent transition rather than transversions. The percentage of GC content (52.4%) was higher in 18S rDNA region followed

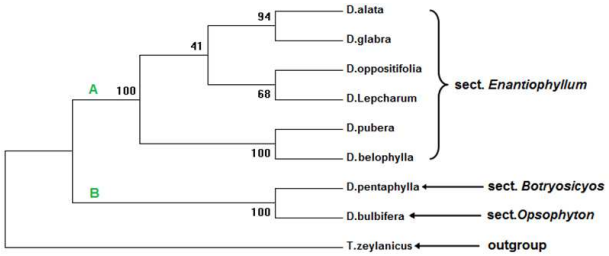

Fig. 2. Single most parsimonious tree obtained from maximum parsimony analysis of combined dataset (18SrDNA, *matK* and *rbcL*). Numbers at the nodes indicates bootstrap values

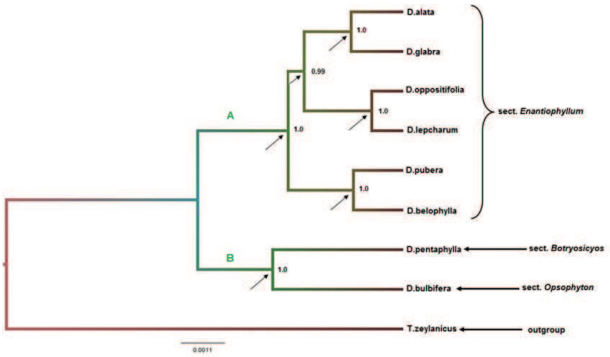

Fig. 3. 50% majority-rule consensus tree of Bayesian inference (BI) inferred from a combined dataset (18SrDNA, *matK* and *rbcL*). Posterior probablilities are given at the nodes

by GC percentage of 43.9 and 31.9 was recorded for *rbcL* and *matk*, respectively. Sequence characterization of the three genes (18S rRNA, *matK* and *rbcL*) showed that 18S rRNA gene was highly conserved than *matK* and *rbcL* gene sequences. 18Sr RNA gene is proved to be better candidates in Dioscoreales (Caddick et al. 2002a; Merckx et al. 2006) and is mostly used at an interfamilial rather than intergeneric or interspecific level (Hamby and Zimmer 1988; Soltis and Soltis 1998). Sun et al. (2012) regarded *matK* as the best regions for use as DNA barcodes for phylogenetic reconstruction, whereas *rbcL* was the

Table 3. Nucleotide substitution models for respective datasets for Bayesian analysis

Dataset	Model	BIC	lnL	Gamma shape	Frequency				Substituition matrix					
					A	C	G	T	AC	AG	AT	CG	CT	GT
18srDNA	JC+G	5375.91	–2633.38	0.2200	0.2441	0.2385	0.2840	0.2334	0.6797	1.5450	1.0725	0.6366	1.2315	1.000
matK	HKY+G	11549.73	–5707.13	0.6060	0.3086	0.1639	0.1585	0.3690	0.9938	1.5132	0.7983	0.8468	1.495	1.000
rbcL	HKY+G	4152.19	–2009.80	0.6150	0.2678	0.1969	0.2453	0.2900	0.1055	0.1718	0.1015	0.0549	0.1718	0.1718
18srDNA+ matK+ rbcL	F18+G	35384.5	–17618.3	7.3980	0.2723	0.2036	0.2299	0.2942	1.0278	1.1552	1.0644	1.0769	1.1676	1.000

BIC = Bayesian Information Criterion; lnL = Maximum Likelihood value; +G = Gamma distribution; JC = Jukes-Cantor; GTR = General Time Reversible; HKY = Hasegawa-Kishino-Yano; F81 = Felsenstein 1981

least suitable marker for *Dioscorea.* But the present study reveals that all the three genes can be used as candidates for species discrimination in *Dioscorea*.

Phylogenetic analysis

The combined Maximum Parsimony (MP) analysis of the three genomic regions resulted in a parsimonious tree (length = 405) with consistency index (CI) = 0.80 and retention index (RI) = 0.78. The most likely combined Bayesian inference trees had a log score of -17618.3. Both the consensus trees obtained by Maximum Parsimony and Bayesian inference analysis forms two major clades (Figs. 2 and 3). CladeA which is subdivided further into 3 subclades, represented by clustering of species such as *D. alata, D. glabra, D. oppositifolia, D. lepcharum, D. pubera* and *D. bellophylla* belonging to the *D*. sect. *Enantiophyllum*. CladeB is represented by two species, *D. bulbifera* (sect. *Opsophyton*) and *D. Pentaphylla* (sect. *Botryosicyos*). CladeB shows close association with the outgroup, *T. zeylanicus*. The combined tree topology inferred by Bayesian analysis shows 100% posterior probability at each node in all the clades. CladeA strongly supports monophyly of *D*. sect. *Enantiophyllum.*

The analyses clearly support a strong relationship between *D. bulbifera* and *D. pentaphylla* and monophyly of *D.* sect. *Enantiophyllum*. The overall tree topology is similar to previously reported phylogenies on *Dioscorea* species (Wilkin et al. 2005; Hsu et al. 2013; Mukherjee and Bhat, 2013; Barman et al. 2018). Based on their twining stems, compound leaves, underground organ morphology and seed characters, Prain and Burkill (1936, 1938) divided the Asian *Dioscorea* species into nine sections (sect. *Botryosicyos*, *Combilium*, *Enantiophyllum*, *Lasiophyton*, *Opsophyton*, *Paramecocarpa*, *Shannicorea*, *Stenocorea* and *Stenophora*) out of which three sections (sect. *Enantiophyllum*, *Botryosicyos* and *Opsophyton*) were included in present analysis. The monophyly of the sect. *Enantiophyllum* (*D. alata, D. glabra, D. pubera, D. belophylla, D. lepcharum* and *D. oppositifolia*) supports the hypothesis of Wilkin et al. (2005) that the main Old World lineages of *Dioscorea*, such as the right-twining section *Enantiophyllum,* are monophyletic. *D. bulbifera* is the main species of *D.* sect. *Opsophyton* distributed in the wild state in both Asia and Africa. The formation of many axillary tubers (bulbils) is the distinct characteristic of *D. bulbifera* but intraspecific classifications are still diverse. Prain and Burkill (1936) have treated the African form with angular bulbils as a single variety, *D. Bulbifera* var. *anthropophagorum*, and the Asian form with globular bulbils has been divided into nine varieties according to highly variable characters such as the colour, shape, and dimension of bulbils and leaves. *D. pentaphylla*, a representative of sect. *Botryosicyos* of this section show many morphological characters including perennial crown with annual tubers, left twining, usually pubescent and spiny, compound leaves and capsules that are longer than their wide. It is also characterized by one main vein per leaflet. Wilkin et al. (2005) in his treatment considered *D. bulbifera*is unresolved at the base of a combined compound-leafed clade in the strict consensus tree obtained from *matK* alone. In the present study both the species are sole representative of each of the single sections and hence in the analysis were grouped in one clade. According to Prain and Burkill (1936), *D.*

pentaphylla and *D. bulbifera* have few characters common such as left twining direction of stem and seeds winged at the base only. Hence there must be possibility for these two species to form group close to each other. To the best of our knowledge, the present study is a first attempt on the use of 18S rDNA along with *Matk* and *rbcL* gene sequences for inferring phylogeny on Indian *Dioscorea* species and also the result of sequence characterization reveals that 18S rDNA gene can be used as a better candidate for species discrimination in the genus *Dioscorea*.

Authors' contribution

Conceptualization of research (NS, YK, AKM); Designing of the experiments (AKM); Contribution of experimental materials (NS); Execution of field/lab experiments and data collection (NS, YK); Analysis of data and interpretation (NS, AKM); Preparation of the manuscript (NS).

Declaration

The authors do not have any conflict of interest.

Acknowledgements

The authors are thankful to the Head of Botany Department, NEHU, Shillong for availing necessary facilities and also the Joint Director, BSI, Eastern Regional Centre, Shillong for help and support during the course of the research work.

References

Anonymous. 1952. The Wealth of India. CSIR, New Delhi, 3: D-E.

Ayensu E. S. 1972. Dioscoreales, in: Metcalfe C, (eds.), Anatomy of the monocotyledons, vol.6. Oxford, Clarendon Press 1-182.

Burkill I. H. 1960.The organography and the evolution of the Dioscoreaceae, the family of the yams. Bot. J. Linn. Soc., **56**: 319-412.

Barman P., Bhat K. V. and Geeta R. 2018. Phylogenetic analysis of Indian Dioscorea and comparison of secondary metabolite content with sampling across the tree. Genet Resour Crop Evol., **65**: 1003-1012.

Caddick L. R., Rudall P. J., Wilkin P., Hedderson T. A. J. and Chase M. W. 2002a. Phylogenetics of Dioscoreales based on combined analyses of morphological and molecular data. Bot. J. Linn. Soc., **138**: 123-144.

Coursey D. G. 1967. Yams: An account of the nature, origins, cultivation and utilization of the useful members of the Dioscoreaceae: Tropical Agricultural series. Longmans, green and Co. Ltd., London, 230pp.

Doyle J. J. and Doyle J. L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, **19**: 11-15.

Hamby K. R. and Zimmer A. A. 1988. Ribosomal RNA as a phylogenetic for inferring phylogeny within the grass family (Poaceae). Plant Syst. Evol., **160**: 29-37.

Hsu K. M., Tsai J. L., Chen M. Y., Ku H. M. and Liu S. C. 2013. Molecular phylogeny of Dioscorea (Dioscoreaceae) in East and Southeast Asia. Blumea, **58**: 21-27.

Knuth R. 1924.Dioscoreaceae. In: Engler A (Ed.), Das Pflanzenreich, IV.43: 1-387pp. Engelmann, Leipzig.

Kumar S., Stecher G. and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetic Analysis Version 7.0 for bigger datasets. Mol. Bio. Evol., **33**(7): 1870-1874.

Mukherjee P. and Bhat K. V. 2013. Phylogenetic relationship of wild and cultivated yam species (*Dioscorea* spp.) of Indian inferred from PCR-RFLP analysis of two cpDNA loci. Plant Syst. Evol., **299**: 1587-1597.

Merckx V., Schols P., De Kamer H. M., Maas P., Huysmans S. and Smet E. 2006. Phylogeny and Evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondria data. American J. Bot., **93**(11): 1684-1698.

Muller K. 2005.SeqState: primer design and sequence statistics for phylogenetic DNA data sets. Applied Bioinformatics, **4**: 65-69.

Prain D. andBurkill I. H. 1936.An account of the genus Dioscorea in the East, Part 1: The species which twine to the left. Annals of the Royal Botanic Gardens, Calcutta 14, 1-210pp. Bengal Government Press, Alipore.

Prain D. and Burkill I. H. 1938. An account of the genus Dioscorea in the East, Part 2: The species which twine to the right. Annals of the Royal Botanic Gardens, Calcutta 14, 211-528pp. Bengal Government Press, Alipore.

Posada D. and Crandall K. A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics, **14**: 817-818.

Ronquist F. and Huelsenbeck J. P. 2003.MrBayes 3: Baysian phylogenetic inference under mixed models. Bioinformatics, **19**: 1572-1574.

Sharma B. D. and Hore D. K. 1995. Genetic Resource of Yams in N. E. India with Special Reference to Garo Hills, Meghalaya. India J. Hill Farm, **8**: 145-151.

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. and Chase M. W. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Annals of the Missouri Botanical Garden,

84: 1-49.

Soltis D. E. and Soltis P. S. 1998.Choosing an approach and an appropriate gene for phylogenetic analyses. In: D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.] Molecular systematics of plants II, DNA sequencing, 1-42pp. Kluwer, New York, USA.

Sun X. Q., Zhu Y. J., Guo J. L., Peng B., Bai M. M. and Yang Y. Y. 2012. DNA Barcoding the Dioscorea in China, a Vital Group in the Evolution of Monocotyledon: Use of matK Gene for Species Discrimination PLoS ONE, **7**(2): e32057. doi:10.1371/journal.pone.0032057.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F. and Higgins D. G. 1997. The CLUSTALX window interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res., **25**: 4876-4882.

Wilkin P. and Caddick L. R. 2000. Palaeotropical compound-leaved yams (Dioscorea; Dioscoreaceae): monophyly and relationships. In: Wilson KL, Morrison DA (eds), Monocots: systematics and evolution, 497-504pp. CSIRO, Melbourne.

Wilkin P., Schols P., Chase M. W., Chayamarit K., Furness C. A., Huysmans S., Rakotonasolo F., Smets E. and Thapyai C. 2005.A plastid gene phylogeny of the yam genus, Dioscorea: roots, fruits and Madagascar. Syst. Bot., **30**: 736-749.