Research Article **Open Access**

IDENTIFICATION OF ENDEMIC *CURCUMA ALBIFLORA* THW. BY DNA BARCODING METHOD

Wijayasiriwardene T. D. C. M. K.[1*], Herath H. M. I. C.[2] and Premakumara G. A. S.[1]

[1]Industrial Technology Institute, Colombo, Sri Lanka.

[2]Faculty of Graduate Studies, University of Colombo, Sri Lanka.

Abstract

Harankaha is important medicinal plant, which is used in Sri Lankan traditional medicine. However, three plants are reported under the same vernacular name (e.g. *Curcuma albiflora, Curcuma zedoaria*, and *Zingiber zerumbet*) and therefore the raw material may be adulterated or substituted with each other. Phylogenetic analysis of gene sequences and combining with complete genomic sequences helps to identify genetic basis of plants. Standard CTAB method with little modifications was used for the extraction and purification. Extracted DNA was amplified using universal primers for *matK* genes in chloroplast genome *rbcL* gene in chloroplast genome by PCR (polymerase chain reaction). The *rbcL* primers amplified about 600 bp, while *matK* was about 850 bp. Amplified fragments were sequenced and obtained the DNA sequences for the *matK* and *rbcL* genes. Sequenced fragments were analyzed and used for DNA barcoding. DNA barcode was submitted to BOLD (Barcoding of Life Database) online database and then uploaded to the GenBank through the BOLD system (Accession No KF 521885). The distance of interspecific were 0.0010 in *rbcL* and 0.05 in *matK*, which is less than or equal to 0.05 *matK* sequences were selected for identification of *C. albiflora*, *C. zedoaria* and *Z. zerumbet*. *C. albiflora* appeared as a different group as per the Neighbor-Joining method and therefore, it can be identified as a new group. The *matK* gene of *C. albiflora* and other species showed, totally 212 variable sites. However, *rbcL* gene of *C. albiflora* and other species showed only 2 variable sites. Therefore, *matK* gene is suggested to identify *C. albiflora* from other two species claimed as *Harankaha* plants.

Key words: *Curcuma albiflora,* DNA barcoding, MaturaseK

*Corresponding author: Tel: +94 718 609 190; E-mail: drcchandima@iti.lk

1. Introduction

Curcuma is a genus of family *Zingiberaceae*, many species are important in Sri Lankan traditional medicine. By plant morphology, five species of *Curcuma* available in Sri Lanka can be identified. But *C. albiflora*, *C. zedoaria*, and *Z. zerumbet* are named locally under same vernacular name *Harankaha*, which is likely to be adulterated for each other. These plants are used for inflammatory joint disorders, anti-venom for snake bites etc. Comparatively many researches have been conducted on genera *Alpinia*, combined data sets produce a highly resolved tree. *Curcuma* species of *C. albiflora* and *C. oligantha*, which are available in Sri Lanka, should be included in NCBI database for phylogenetic analysis. Phylogenetic analysis of the *matK* coding and noncoding regions is used to derive relationship among genera (Kress *et al.*, 2002). DNA sequencing is one of reliable methods in biological identification which can be performed by many methods such as Maxam-Gilbert sequencing, Sanger's chain-termination methods and new sequencing methods including Lynx therapeutics massively parallel signature sequencing (MPSS), polony-sequencing, pyro-sequencing, solexa-sequencing, solid-sequencing, DNA nanoball sequencing etc. These techniques are applied for reliable identification of microbes, DNA barcode of plant species, and identification of new species of insects etc. (Kumar, 2012). Phylogenetic tree building is a process necessary for an understanding of broad patterns of biological diversity in plant species. MaturaseK (*matK*) gene of chloroplast was used in DNA barcoding of family *Zingiberaceae* (Selvaraj *et al.*, 2008) and studies of angiosperm specie using ribulosebisphosphate carboxylase/oxygenace (*rbcL*) large subunit gene extracted from *Ceratophyllum* genus were reported (Chase *et al.*, 1993; Chase *et al*., 2001).

Dry rhizomes of *C. zedoaria, C. aromatica,* and *Z. zerumbet* are hard to differentiate because they have very similar external appearance. Main objective of this study was to identify efficient and reliable method to differentiate these herbal materials. Morphological and microscopic identification were carried out in previous studies (Herath *et al.*, 2016). Morphologically, *C. albiflora* can be differentiated by 35 cm plant height, green colour glabrous leaf, absence of coma and fertile bract spreading at lower one third of inflorescence from other two species of *Harankaha*. Microscopically, three sizes of simple starch grains, prismatic crystal available only in leaf of *C. albiflora*. However, the chemical compositions are varying for different geographical areas, harvesting season, way of storage etc (Yu *et al*., 2016). Due to stability of tissue macromolecule against above external factors, authentication by DNA barcode is a reliable method to identify similar plant raw materials for medicinal purposes.

2. Material and Methods

Matured plant samples of *C. albiflora* were collected from natural habitats at Kitulgala of Kegalle and Eratna of Rathnapura districts in the flowering season in 2015. Voucher specimens were authenticated from National Herbarium, Peradeniya. Standard hexadecyltrimethylammonium bromide (CTAB) method with little modification was used for the extraction and

*Corresponding author: Tel: +94 718 609 190; E-mail: drcchandima@iti.lk

purification of DNA (Doyle, 1987; Dhanya *et al*., 2007). Preheated buffer (1 ml) containing 3% CTAB 100 µl of 10% PVP and 3 µl of β-Mercaptoethanol was added to plant material (100 mg) and ground. Proteinase K (10 µl) was also added and the suspension was incubated at 65 ^{o}C water bath for 30 min. Then 1/3 V of 5 M potassium acetate was added and incubated in ice for 1 h. Equal volume of Chloroform: Isoamyl (24:1) was added and mixed by inversion for 15 min, and centrifuged at 10,000 g for 15 min at 4 ^{o}C. The supernatant was transferred and equal volume of 30% PEG was added. The mixture was incubated on ice for 30 min, centrifuged at 12,000 g for 20 min at 4 ^{o}C. The resulting pellet was washed with 70% ethanol, DNA pellet was dissolved in 20 µl of PCR grade water and stored at -20 ^{o}C. Extracted DNA was amplified using universal primers for *matK* genes in chloroplast genome by PCR (polymerase chain reaction) with forward primer sequence (5′-acccagtccatctggaaatcttggtcc-3′) and reverse primer sequence (5′-ctgacagtacttttgtgtttacgag-3′). Moreover, *rbcL* gene in chloroplast genome by PCR with forward primer sequence (5′-atgtcaccacaaacagagactaaag-3′) and reverse primer sequence (5′-gtaaaatcaagtccaccycg-3′). Expected amplicon of *rbcL* primer was about 600 bp, while, *matK* was about 850 bp. The test was triplicated and finally, PCR products were sent to Macrogen Inc. Korea for sequencing and obtained the DNA sequences for the *matK* and *rbcL* genes.

The data analysis was done on three plant species under the common name *Harankaha*, *C. albiflora*, *C. zedoaria* and *Z. zerumbet matK* gene and *rbcL* gene. As *C. albiflora* is endemic to Sri Lanka among other plants claimed as *Harankaha,* DNA extraction and sequencing was performed on *C. albiflora* only. Since *C. zedoaria* and *Z. zerumbet matK* and *rbcL* gene nucleotide base patterns are available in database of National Centre for Biotechnology information (NCBI), they were analyzed with *matK* and *rbcL* gene patterns of *C. albiflora*. Multiple alignments were performed using a built-in CLUSTAL W and aligned sequences were edited by Molecular Evolutionary Genetics Analysis (MEGA) version 7.0. Evolutionary analyses were conducted in MEGA7 (Kumar, 2015). Tajima test statistic (Tajima, 1989) was performed according to Nei *et al*. 1987. The phylogenetic tree was obtained using the Neighbor-Joining method (Saitou *et al.*, 1987). The sum of branch length was computed using the maximum composite likelihood method (Tamura *et al*., 2004). The genetic distance was calculated based on Kimura-2-parameter model (Kimura, 1980). BioRad MyCycler thermal cycler was used for current study.

3. Results

DNA barcode was submitted to BOLD (Barcoding of Life Database) online database and then uploaded to the GenBank through the BOLD system. The interspecific distance is 0.0010 in *rbcL* and 0.05 in *matK*, which is less than or equal to 0.05. Multiple sequence alignment shows that, there are variable numbers of *InDels* in the gene *matK*. The alignment of *matK* gene of combined nucleotide sequence shows 212 variable sites, zero parsim-info sites, 212 singleton sides, and 571 conserved sites. Nucletide compositions of uncoded, 1st , 2nd , and 3rd codon positions were

*Corresponding author: Tel: +94 718 609 190; E-mail: drcchandima@iti.lk

reported in Table 1. Pairwise distances of *matK* gene analysis were mentioned in Table 2. There were a total of 777 positions in the final dataset. Phylogenetic tree of *matK* gene was shown in Figure 1. The optimal tree with the sum of branch length was 0.4738.

Table 1. Nucleotide composition of *matK* gene of *C. albiflora*, *C. zedoaria*, and *Z. zerumbet*

Uncoded	***C. albiflora***	***C. zedoaria***	***Z. zerumbet***	**1st Codon positions**	***C. albiflora***	***C. zedoaria***	***Z. zerumbet***
T%	36.9	36.9	38.7	**T%**	33.9	34.9	35.1
C%	18.0	14.8	14.9	**C%**	19.9	20.3	20.5
A%	29.2	31.5	31.3	**A%**	30.4	31.0	30.9
G%	15.8	14.8	15.1	**G%**	15.7	13.8	13.5
2nd Codon positions	***C. albiflora***	***C. zedoaria***	***Z. zerumbet***	**3rd Codon positions**	***C. albiflora***	***C. zedoaria***	***Z. zerumbet***
T%	33.6	35.6	35.9	**T%**	43.2	46.0	45.2
C%	18.5	15.3	15.4	**C%**	15.7	8.8	8.9
A%	30.1	33.3	33.2	**A%**	27.2	30.3	29.7
G%	17.8	15.7	15.4	**G%**	13.9	14.9	16.2

Maximum-likelihood estimate of gamma parameter for site rates of the shape parameter was 3.8. The nucleotide frequencies are A (30.63%), T/U (38.35%), C (15.87%), and G (15.14%). The interspecies distance (d_t) between *C. albiflora* and group 2 was 0.4857 and the intra-species distance (d_i) was 0.0043 on *matK* gene. Tajima's neutrality test on *matK* genes were reported in Table 3.

The alignment of *rbcL* gene of combined nucleotide sequence shows 2 variable sites and 1 parsim-info sites, 1 singleton sides, and 191 conserved sites. Nucleotide composition of *rbcL* of *Harankaha* species were reported in Table 4. Pairwise distances of *rbcL* gene analysis of *C. albiflora*, *C. zedoaria*, and *Z. zerumbet* were mentioned in Table 5.

*Corresponding author: Tel: +94 718 609 190; E-mail: drcchandima@iti.lk

Table 2. Estimates of pairwise distances between *matK* sequences of *C. albiflora*, *C. zedoaria*, and *Z. zerumbet*

Species	1	2
C. albiflora		
C. zedoaria	0.4733	
Z. zerumbet	0.4677	0.0065

Table 3. Results of Tajima's Neutrality Test on *matK* of *C. albiflora*, *C. zedoaria*, and *Z. zerumbet*

Description	Value
Number of sequences	3
Number of segregating sites	212
Nucleotide diversity	0.1823

Table 4. Nucleotide composition of *rbcL* gene of *C. albiflora*, *C. zedoaria*, and *Z. zerumbet*

Uncoded	***C. albiflora***	***C. zedoaria***	***Z. zerumbet***	**1st Codon positions**	***C. albiflora***	***C. zedoaria***	***Z. zerumbet***
T%	30.2	29.6	29.6	**T%**	44	43	43
C%	19.4	19.6	20.2	**C%**	15.8	15.9	16.7
A%	27.6	27.9	27.7	**A%**	27.0	27.5	27.0
G%	22.8	22.9	22.5	**G%**	13.3	13.8	13.8
2nd Codon positions	***C. albiflora***	***C. zedoaria***	***Z. zerumbet***	**3rd Codon positions**	***C. albiflora***	***C. zedoaria***	***Z. zerumbet***
T%	22	22	21	**T%**	25	24	25
C%	16.3	16.9	17.1	**C%**	26.2	25.9	26.9
A%	25.0	24.9	25.1	**A%**	30.8	31.2	30.9
G%	36.7	36.5	36.6	**G%**	18.5	18.5	17.1

*Corresponding author: Tel: +94 718 609 190; E-mail: drcchandima@iti.lk

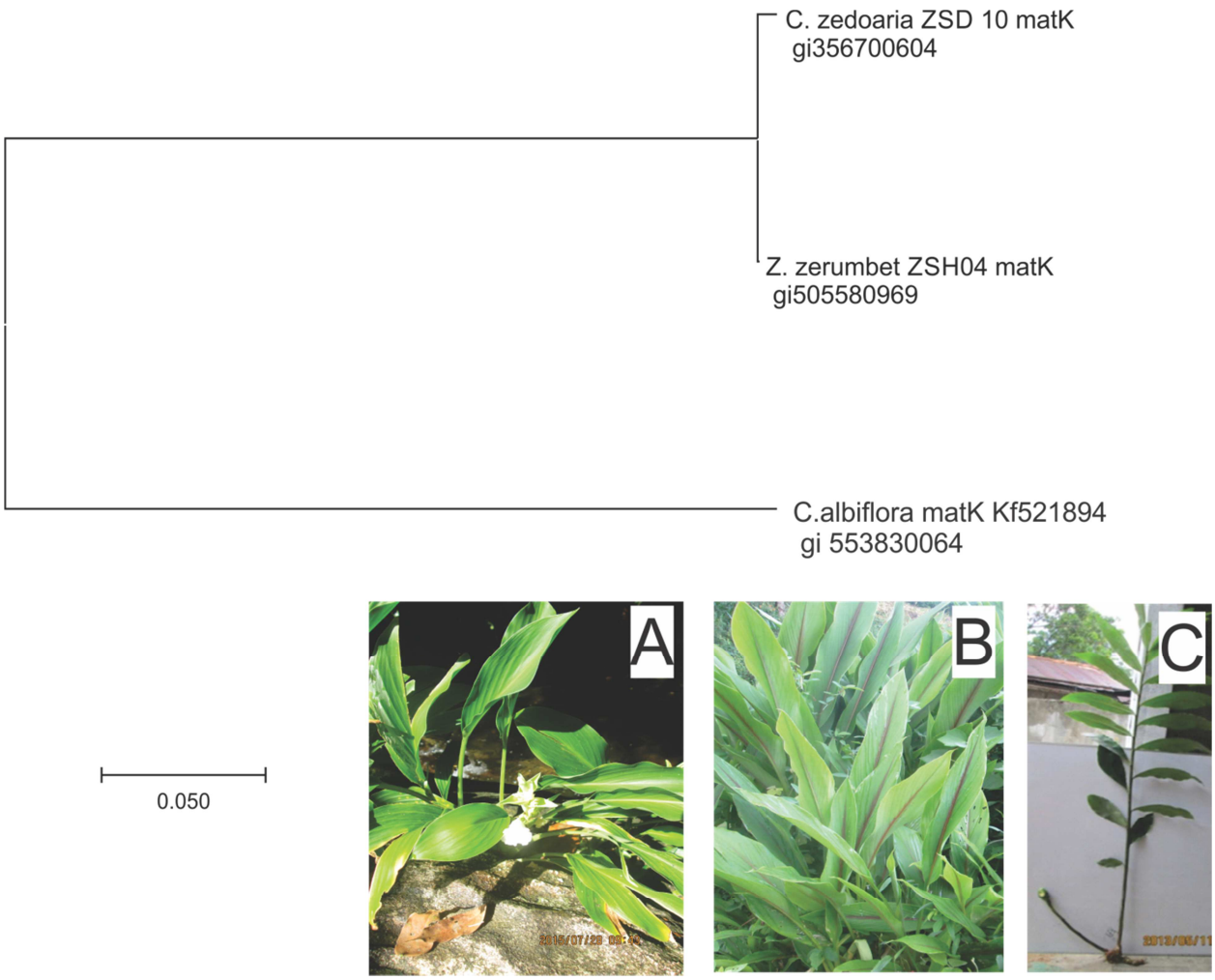

Figure 1. Evolutionary relationships on *matK* of *C. albiflora*, *C. zedoaria*, *Z. zerumbet* and their phenotypic appearance; A: *C. albiflora*, B: *C. zedoaria*, C: *Z. zerumbet*

Phylogenetic tree of *rbcL* gene of *C. albiflora*, *C. zedoaria*, and *Z. zerumbet* was shown in Figure 2. The optimal tree with the sum of branch length was 0.0135. There were a total of 522 positions in the final dataset. The maximum likelihood estimated value of the shape parameter for the discrete gamma distribution is 200. The nucleotide frequencies were 27.39 % (A), 30.08 % (T/U), 19.99 % (C), and 22.54 % (G). The interspecies distance (d_t) between *C. albiflora* and group 2 was 0.0087 and the intra-species distance (d_i) was 0.009 on *rbcL* gene. Results of Tajima's Neutrality Test were mentioned in Table 6.

*Corresponding author: Tel: +94 718 609 190; E-mail: drcchandima@iti.lk

Table 5. Estimates of pairwise distances between *rbcL* sequences of *C. albiflora*, *C. zedoaria*, and *Z. zerumbet*

Specie	1	2
C. albiflora		
C. zedoaria 765 *rbcL* gi751582784	0.0077	
C. zerumbet ZSH04 *rbcL* gi5055800901	0.0096	0.0096

Table 6. Results of Tajima's Neutrality Test on *rbcL*

Description	Value
Number of sequences	3
Number of segregating sites	2
Nucleotide diversity	0.005

4. Discussion

Chloroplast DNA is less in rhizome and also interferences are high due to polysaccharide content, when used rhizome for DNA extraction. Polysaccharides, polyphenols, and resin may co- precipitate with DNA. The *matK* gene of *C. albiflora* and other species showed, totally 212 variable sites. However, *rbcL* gene of *C. albiflora* and other species showed only 2 variable sites. Therefore, *matK* gene is suggested to identify *C. albiflora* from other *Harankaha* plants. *C. albiflora* was separately clustered into one group, while *C. zedoaria* and *Z. zerumbet* were clustered into another group (group 2). The interspecies distance (d_t) between *C. albiflora* and group 2 was 0.4857 and the intra-species distance (d_i) was 0.0043 on *matK* gene. Chen (2012) has reported that higher interspecific divergences and lower intraspecific variations are desirable for DNA barcode. In this study, d_i/d_t was 112.9. Therefore, interspecific divergence is significantly large and suitable for DNA barcoding (Yu *et al*., 2016). Moreover, the neighbor-joining tree also showed that *C. albiflora* can cluster into new group. *C. albiflora matK* gene showed total of 859 nucleotide positions, but other two species showed less than 800 nucletide positions(*C. zedoaria*-783 positions and *Z. zerumbet*-777 positions). While, pairwise distances of *matK* between *C. albiflora* and other species was around 0.46, pairwise distance of *rbcL* was about 0.008. Cytosine nucleotide percentage of *C. albiflora* of *matK* was significantly higher (15.7%) at 3rd codon position than other two species. However, *rbcL* was not showed significant difference of nucleotide composition. Therefore,

*Corresponding author: Tel: +94 718 609 190; E-mail: drcchandima@iti.lk

matK region was an appropriate DNA barcode for identifying *C. albiflora* from other *Harankaha* plants.

5. Conclusions

Present study was the first report on DNA bar-coding of *C. albiflora.* Phylogenetically, it appears as a different group as per the Neighbor-Joining method and therefore, it can be identified as a new group. The *matK* gene of *C. albiflora* showed significant 212 variable sites. Therefore, DNA barcoding study provided a reliable proof to identify *C. albiflora* from other *Harankaha* species and helps in differentiation of substitutes and adulterants.

6. Acknowledgement

National Sciences Foundation, Sri Lanka (Grant No: RG/2011/NRB/04).

References

Chase M.W., Fay, M.F. (2001) Ancient flowering plants: DNA sequences and angiosperm classification, *Journal of Genome Biology* **2(4)**: 1-4.

Chase M.W., Soltis, D.E, Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.L. (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*, *Annals of the Missouri Botanical Garden* **80(3)**: 528–580.

Chen, S.L. (2012) *DNA Barcoding Molecular Markers in Chinese Medicinal Materials*, People's Medical Publishing House, Beijing, 18 .

Dhanya, K., Kizhakkayil, J., Syamkumar, S., Sasikumar, B. (2007). Isolation and Amplification of Genomic DNA from Recalcitrant Dried erries of Black Pepper (Piper nigrum L.)- A medicinal Spice. *Molecular Biotechnology* **37(2)**: 165-168.

Doyle, J.J., Doyle, J.L. (1987) A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, **19**: 1-15.

Herath, H.M.I.C., Wijayasiriwardene, T.D.M.C.K., Premakumara, G.A.S. (2016) Morphological & microscopic identification of *Curcuma albiflora* Thw. *Journal of Ayurveda and Herbal Medicine* **2(1)**: 15-19.

Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences, *Journal of Molecular Evolution* **16**:111-120.

Kress, W.J., Linda, M., Prince, K. J. (2002) The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data, *American Journal of Botany* **89(10)**: 1682-1696.

Kumar, B. R. (2012) DNA representation: In: A. Munshi (ed.), *DNA Sequencing – Methods and Applications*. pp. 3-14, InTech Janeza Trdine 9, 51000 Rijeka, Croatia.

Kumar, S., Stecher, G., and Tamura, K. (2015) MEGA7: Molecular Evolutionary Genetics Analysis version 7 for bigger datasets, *Molecular Biology and Evolution* (submitted).

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

Selvaraj, S., Sarma, R.K., Sathishkumar, R. (2008) Phylogenetic analysis of chloroplast *matK* gene from Zingiberaceae for plant DNA barcoding, *Bioinformation* **3(1)**: 24-27.

Tajima, F. (1989) Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism, *Genetics* **123**: 585-595.

Tamura, K., Nei, M., Kumar S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences* (USA), 11030-11035.

Yu, X., Xie, Z., Wu, J., Tao, J., Xu, X. (2016) DNA barcoding identification of *kadsurae caulis* and *spatholobi caulis* based on internal transcribed spacer 2 region and secondary structure prediction, *Pharmacognosy Magazine*, **12(2)**: 165-9.

*Corresponding author: Tel: +94 718 609 190; E-mail: drcchandima@iti.lk