

DNA fingerprinting and molecular characterization of mango (*Mangifera* spp.) cultivars in Vietnam using ITS DNA barcode

Viet The Ho^{1*}, Ngan Tram Tu² and Thi An Nguyen¹

¹*Ho Chi Minh City University of Food Industry, 140 Le Trong Tan Street, Ho Chi Minh City, Vietnam*

²*Ho Chi Minh City University of Technology, 475A Dien Bien Phu Street, Ho Chi Minh City, Vietnam*

*Corresponding author's e-mail: thehv@hufi.edu.vn

Abstract

Ho, Viet The, Tu, Ngan Tram & Nguyen, Thi An (2021). DNA fingerprinting and molecular characterization of mango (*Mangifera* spp.) cultivars in Vietnam using ITS DNA barcode. *Bulg. J. Agric. Sci.*, 27 (1), 128–134

Mango (*Mangifera* spp.) is an important fruit tree in Vietnam. In this study, internal transcribed spacer (ITS) DNA barcode was used to characterize the genetic richness of 16 mango accessions from different provinces in Vietnam. The results show the discrimination power of ITS region. There are 14 sites of single nucleotide polymorphism detected among analyzed accessions. The dendrograms generated by cluster analysis show that the genetic similarity of samples is not strictly follow the geographical distribution of sampling sites. The obtained results provide molecular biological information for classification, identification plant origins, breeding, and conservation programs; furthermore, utilization of molecular marker analysis could provide new insights to breeders for molecular assisted selection of mango.

Keywords: DNA barcode; diversity; ITS; *Mangifera indica* L.; molecular markers

Introduction

Mango (*Mangifera* spp.) belongs to Anacardiaceae family which consist at least 69 species and distributed in several countries in tropical region. Morphological characteristic is mostly applied for mango classification due to its easiness to perform and carry out on the field with low cost. However; there are several limitations for this method such as low number, complex inheritance pattern and vulnerable to changes of environment (Ahmedand & Mohamed, 2014). The confused identification makes it more difficult for authenticating cultivar correctly and leads to variation in controlling mango quality. For better conservation and breeding programs, several studies have focused of germplasm characterization and identification of genetic relatedness among plant accessions (Tran & Do, 2012; Ahmedand & Mohamed, 2014; Mansour et al., 2014).

To overcome the problem in morphology-based taxonomy, sequencing of genomic DNA has been served as stan-

dardized method for plant identification due to the present of homologous DNA sequences among related species (Laprise & Rodgers, 2010). Numerous studies have shown the efficiency of using DNA barcode for plant identification, it is also considered as a useful tool to study evolution of plant at molecular level (Baldwin et al., 1995; Kress et al., 2005; Hollingsworth et al., 2011; Tehen et al., 2014). DNA barcode is relatively new technique which uses the standardized genomic regions to distinguish among species and has been used intensively for identifying at species-level. Mitochondrial cytochrome oxidase I (COI) gene was generally used for phylogenetic study of animals (Smith et al., 2012; Aziz et al., 2016). However; this gene is not useful in plant since lacking of sufficient variation due to its low mutation rate (Kress et al., 2005; Fazekas et al., 2009). Besides plastid DNA sequence, nuclear ribosomal internal transcribed space (ITS) region is increasingly used in plants (Chase et al., 2007; Giudicelli et al., 2015). Numerous researchers reported that ITS show higher performance in the comparison

to other plastid markers (Muellner et al., 2011; Yang et al., 2012; Zhang et al., 2014).

Mango is an important fruit tree in Vietnam and has been considered and plants for poverty production in rural areas. With total of 74 600 ha and the production of 728 100 tons/year (GSO, 2016), Vietnam stands at 13th place of mango export country (VTO, 2016). Considering the importance of mango, the study to identify the different cultivars is necessary. In spite of that, only a few studies reported genetic composition of mango in the country. Tran and Do analyzed genetic diversity of mango genotypes in Southern region by AFLP and ITS markers (Tran & Do, 2012). More recently, ISSR was utilized to exploit genetic variation of “thanh tra” (*Bouea oppositifolia* Roxb.), a distant member of mango family (Le et al., 2018). However, these two studies were only focused in Mekong delta area in the South of Vietnam. In this study, the genetic diversity of 16 mango accessions collected nationwide were analyzed by using ITS DNA barcode marker. The obtained results could provide scientific information for identification, classification and authentication of mango in Vietnam.

Materials and Methods

Plant materials

Leaf samples of 16 mango genotypes were collected from germplasm of research institutes, university nurseries, and seedling centers in different provinces of Vietnam (Figure 1 and Table 1). After sampling, samples were dried in silica gel and stored until use.

DNA extraction

Total DNA was extracted from dried mango leaves using the Cetyltrimethylammonium bromide (CTAB) method described by Doyle & Doyle (1990). DNA quality was then tested by electrophoresis on 1% agarose gel in TAE 1X buffer and stained with Gelred dye (Biotium, USA). The result was observed under ultraviolet light by Quantum - ST4 3000 gel reader (Montreal - Biotech, Canada). DNA concentrations were determined by spectrophotometer (Optima SP 3000 nano UV-VIS, Japan).

ITS amplification

The ITS region was amplified using the composition of PCR reactions as follows: 7.5 µL 2X Mytaq Red Mix (Bioline, UK), 20 ng DNA, 0.2 µM primer (ITS_FW 5' ACGAATTCATGGTCCGGTGAAGTGTTTCG 3' and ITS_RV 5' TAGAATCCCCGGTTCGCTCGCCGTTAC 3') and PCR water for final volume of 15 µL. The PCR reaction conditions were as follows: initial denaturation

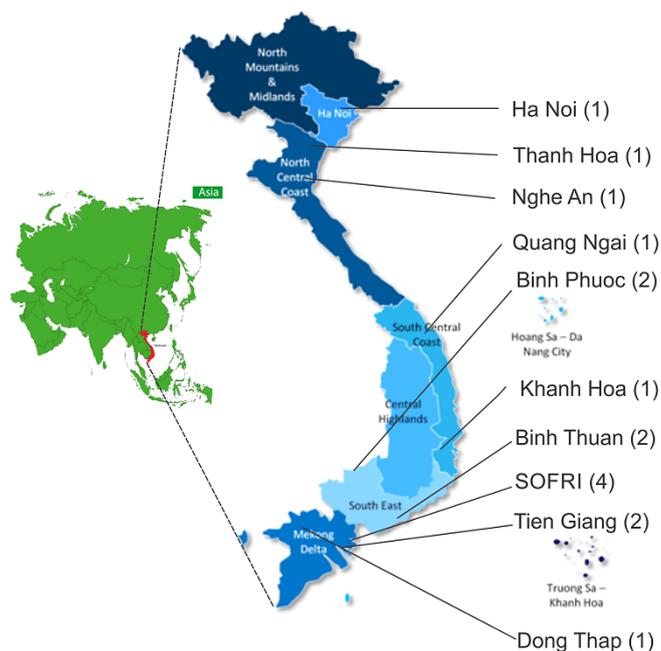


Fig. 1. Targeted areas for collecting mango genotypes (Sample number of each location is shown in parentheses)

at 95°C for 2 minutes; then 35 cycles of 30 seconds at 95°C, 30 seconds at 55°C, and 1 minute at 72°C. Finally, addition 5 minutes was continued at 72 °C to complete the reaction. All reactions were carried out in SureCycler 8800 Thermal Cycler (Agilent, USA). The PCR products were electrophoresed using 1% agarose gel to check the presence or absence of bands. PCR amplification was then purified by ISOLATE II PCR and Gel Kit (Bioline, UK) and then sequenced using the BigDye™ Terminator Cycle Sequencing Kit (Applied Biosystem, USA). The products were next run on ABI 3100 DNA analyzer (Applied Biosystem, USA). The obtained sequences were submitted to GenBank and are publicly accessible under the accession numbers listed in Table 1.

Data analysis

The obtained electropherograms were edited using FinchTV (Digital World Biology Products, USA). Finally, the sequences were blasted on NCBI BLAST under program BLASTN (National Center for Biotechnology Information, USA). Phylogenetic analysis was conducted according to the neighbor-joining (NJ), pairwise deletion for gaps/missing data, of MEGA 6.0 with 1000 bootstrap replicates. Bootstrap support (BS) was categorized as strong (> 85%), moderate (70%-85%), weak (50%-69%) or poor (< 50%) (Kress

Table 1. Mango samples collected for genetic characterization and corresponding Genband accession number

| No. | Cultivar | Collected location | Sample code | Genbank accession number | Sequence length (bp) |
|-----|-------------|--------------------|-------------|--------------------------|----------------------|
| 1 | Cat Hoa Loc | SOFRI | CHL-SOFRI | MN011936 | 650 |
| 2 | Thai | SOFRI | XTH-SOFRI | MN011937 | 650 |
| 3 | Uc | SOFRI | XU-SOFRI | MN011938 | 550 |
| 4 | Dai Loan | SOFRI | XDL-SOFRI | MN011939 | 650 |
| 5 | Keo | Tien Giang | XK-TG | MN011940 | 450 |
| 6 | Cat chu | Tien Giang | XCC-TG | MN011941 | 650 |
| 7 | Thanh ca | Dong Thap | XTC-DT | MN011942 | 520 |
| 8 | Tu quy | Binh Thuan | XTQ-BT | MN011943 | 400 |
| 9 | Cat trang | Binh Thuan | XCT-BT | MN011944 | 650 |
| 10 | Queo | Quang Ngai | XQ-QN | MN011945 | 410 |
| 11 | Rung | Binh Phuoc | XR-BP | MN011946 | 650 |
| 12 | Tuong | Binh Phuoc | XT-BP | MN011947 | 600 |
| 13 | Cat Hoa Loc | Khanh Hoa | CHL-KH | MN011948 | 650 |
| 14 | Tim | Ha Noi | XTI-HN | MN011949 | 600 |
| 15 | Cat | Thanh Hoa | XC-TH | MN011950 | 520 |
| 16 | Tuong | Nghe An | XT-NA | MN011951 | 540 |

et al., 2002). Two ITS-cashew sequences (KF664192.1 and AB071690.1) of the family Anacardiaceae obtained from NCBI GenBank were included as out-group for phylogenetic analysis.

Results and Discussion

PCR and DNA sequence

There is a concern of the unreliability of PCR on ITS region. In order to find suitable candidate loci for DNA barcode of tropical tree species in India, Tripathi found the relatively low effectiveness of PCR in ITS region with only 74.05% success rate. In 2016, ITS marker was identified as least PCR success with only three of twelve invasive grass species in Australia (Wang et al., 2016). However, in this study, the used PCR protocol is reliable for ITS region with 100% of PCR reactions achieved. Similarly, the high success rate of PCR in ITS gene was published from different research groups. This is also in agreement to Chen and colleagues when they reported ITS is the most suitable region for DNA barcode in medical plants after surveying seven candidate DNA barcodes namely *psbA-trnH*, *matK*, *rbcL*, *rpoC1*, *ycf5*, ITS2, and ITS (Chen et al., 2010). In the comparison with other common DNA barcode markers, ITS was showed higher PCR success rate in the comparison to *matK*, *rbcL*, *ycf5* in study family Araliaceae (Liu et al., 2012). ITS region also exhibited the highest inter-specific divergence which is important for distinguishing different species. In 2016, five species in *Crawfordia* was clearly dis-

criminated by ITS; whereas other markers such as *trnH-psbA* and *rbcL* showed poorer results (Zhang et al., (2016).

The length of ITS sequences showed an average of 571 bp, ranging from 400 to 650 bp. The obtained sequences were then submitted to GenBank and shown in Table 1. The GC content varied from 58 to 62% with average of 60.3%. Basic Local Alignment Tool (BLAST) was used to compare sequence homology of the amplified sequences and sequences from GenBank. The sequence homology of unidentified accessions in this study is shown in Table 2, all cultivars was identified as *M. indica* with BLAST similarity from 99.17% to 100%.

The alignment of DNA sequences of 16 mango samples is showed in Table 3. There are 14 variable sites, in which there were unique sited that found specifically for each mango cultivar. MN011951 show high number of variation, the next are MN01140 and MN01148. These unique nucleotide sites can be used as diagnosis character for discriminating different mango cultivars. Such characters have been widely applied in molecular identification studies of different plant species such as *Taxus* L. (Liu et al., 2011), onion (Ipek et al., 2014), *Terminalia* sp. (Mishra et al., 2017); *Melilotus* sp. (Wu et al., 2017). By comparing several DNA barcode markers namely *rbcL*, *matK*, ITS and *psbA-trnH*, Mishra and colleagues reported that ITS marker has highest efficiency for identifying *Decalepsis* at species level (Mishra et al., 2017). A study on *Passiflora* in Brazil also reported that ITS showed higher distinguishing capacity in the comparison to either ITS1 or ITS2 alone (Giudicelli et al., 2015).

Table 2. Statistical simulation of BLAST sequence homology of unidentified specimens with ITS region

| Cultivar name | Accession number | Targeting species | BLAST similarity, % | Sequence cover, % | E-value |
|---------------|------------------|-------------------------|---------------------|-------------------|---------|
| Xoai Uc | MN011938 | <i>Mangifera indica</i> | 100 | 100 | 0 |
| Xoai Dai Loan | MN011939 | <i>Mangifera indica</i> | 99.38 | 100 | 0 |
| Xoai keo | MN011940 | <i>Mangifera indica</i> | 99.33 | 100 | 0 |
| Xoai tim | MN011949 | <i>Mangifera indica</i> | 99.17 | 100 | 0 |

Table 3. Variable sites in 16 mango accessions based on ITS DNA barcode

| Accession number | Variable sites | | | | | | | | | | | | | |
|------------------|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 142 | 153 | 160 | 180 | 184 | 190 | 222 | 235 | 265 | 271 | 307 | 348 | 364 | 413 |
| MN011936 | C | T | T | C | T | T | C | C | C | G | T | C | A | C |
| MN011937 | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011938 | . | C | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011939 | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011940 | . | C | . | . | . | . | . | . | . | . | . | . | C | G |
| MN011941 | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011942 | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011943 | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011944 | . | C | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011945 | . | C | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011946 | . | C | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011947 | . | C | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011948 | . | C | . | T | . | . | A | . | . | . | . | T | . | . |
| MN011949 | . | C | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011950 | . | C | . | . | . | . | . | . | . | . | . | . | . | . |
| MN011951 | T | C | C | . | C | C | . | T | T | A | G | . | . | . |

Phylogenetic trees

The genetic distance among mango accessions based on the Kimura-2 parameter (K2P) is shown in Table 4. The lowest distance was 0.000, while the highest was 0.033. The low distance between sequence pairs such as MN011937 and MN011941; MN011947 and MN011949 (*M. indica*) has raised a question either individuals in these two pairs is actually the sample species, but has reported as a different species, or misidentified due to morphological complexity. This phenomenon was previously reported by Malaysian research group when using mitochondrial DNA cytochrome oxidase subunit to identify different catfish isolation (Abdullah et al., 2017). Large genetic distance between MN011951 and

MN011940 suggesting that the different in genetic composition of these cultivar and may have been even two different species. Another advantages of ITS which is high evolutionary rate, high copy number of rRNA genes make its more power to discriminate at low taxonomic levels (Baldwin et al., 1995). ITS region was also supported as strong candidate for DNA barcode because it might evolve 3-4 times faster than the plastic regions and also be used successfully in studying genetic relatedness of different plants (Chase et al., 2007).

Whereas two pairs of sequences consisting of MN011946 (*M. minutiflora*) and MN011938 (*M. indica*); MN011949 (*M. indica*) và MN011945 (*M. repa*) belonging to different

Table 4. Genetic distance among 16 mango accession based on K2P method, corresponding sample codes are shown in parentheses

| | Accession | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|----|----------------------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | MN011936 <i>M.indica</i> (CHL-SOFRI) | | | | | | | | | | | | | | | |
| 2 | MN011937 <i>M.indica</i> (XTH-SOFRI) | 0.003 | | | | | | | | | | | | | | |
| 3 | MN011938 <i>M.indica</i> (XU-SOFRI) | 0.003 | 0.007 | | | | | | | | | | | | | |
| 4 | MN011939 <i>M.indica</i> (XDL-SOFRI) | 0.003 | 0.003 | 0.007 | | | | | | | | | | | | |
| 5 | MN011940 <i>M.indica</i> (XK-TG) | 0.014 | 0.014 | 0.010 | 0.014 | | | | | | | | | | | |
| 6 | MN011941 <i>M.indica</i> (XCC-TG) | 0.003 | 0.000 | 0.007 | 0.003 | 0.014 | | | | | | | | | | |
| 7 | MN011942 <i>M.me-</i> <i>kongensis</i> (XTC-DT) | 0.003 | 0.003 | 0.007 | 0.003 | 0.010 | 0.003 | | | | | | | | | |
| 8 | MN011943 <i>M.indica</i> (XTQ-BT) | 0.003 | 0.003 | 0.007 | 0.003 | 0.010 | 0.003 | 0.000 | | | | | | | | |
| 9 | MN011944 <i>M.indica</i> (XCT-BT) | 0.007 | 0.007 | 0.003 | 0.003 | 0.010 | 0.007 | 0.007 | 0.007 | | | | | | | |
| 10 | MN011945 <i>M.repa</i> (XQ-QN) | 0.007 | 0.003 | 0.003 | 0.007 | 0.010 | 0.003 | 0.007 | 0.007 | 0.003 | | | | | | |
| 11 | MN011946 <i>M.</i> <i>minutifolia</i> (XR-BP) | 0.003 | 0.007 | 0.000 | 0.007 | 0.010 | 0.007 | 0.007 | 0.007 | 0.003 | 0.003 | | | | | |
| 12 | MN011947 <i>M.indica</i> (XT-BP) | 0.007 | 0.003 | 0.003 | 0.007 | 0.010 | 0.003 | 0.007 | 0.007 | 0.003 | 0.000 | 0.003 | | | | |
| 13 | MN011948 <i>M.indica</i> (CHL-KH) | 0.014 | 0.014 | 0.010 | 0.014 | 0.014 | 0.014 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | | | |
| 14 | MN011949 <i>M.indica</i> (XTI-HN) | 0.007 | 0.003 | 0.003 | 0.007 | 0.010 | 0.003 | 0.007 | 0.007 | 0.003 | 0.000 | 0.003 | 0.000 | 0.010 | | |
| 15 | MN011950 <i>M.indica</i> (XC-TH) | 0.007 | 0.007 | 0.003 | 0.007 | 0.007 | 0.007 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.007 | 0.003 | |
| 16 | MN011951 <i>M.indica</i> (XT-NA) | 0.031 | 0.031 | 0.027 | 0.028 | 0.034 | 0.031 | 0.031 | 0.031 | 0.024 | 0.028 | 0.027 | 0.028 | 0.035 | 0.028 | 0.027 |

species show high genetic similar. Thus, single ITS for DNA barcode could not always be perfect for mango classification, suggesting that the combination of ITS with other DNA barcode markers distributed in different genome regions could lead to more successful result in classification of mango taxonomy (Ipek et al., 2014).

The phylogenetic tree constructed by using MEGA 6.0 with the neighbor-joining (NJ) method is shown in Figure 2. The bootstrap value showing the separation support of the in-group and out-group is up to 99%. Two ITS sequences of cashew obtained from Genbank (accession number KF664192.1 and AB07.1690.1) formed a clear and ob-

servable clade to other mango accessions indicating the reliability of phylogenetic result. The studied mango accessions were divided into two main groups based on NJ analysis. Only accessions MN011951 collected from Nghe An is separated from others. Generally, studied accessions are distributed in groups responding to geographical locations. Though, some accessions are grouped together but they were collected in different distant locations such as MN011949, MN011945 and NM011947 collected from North (Hanoi), Central region (Quang Nam), and South region (Binh Phuoc), respectively. This result suggests there is the mixture among accessions.

- phological and genetic characteristics of twelve marian plum (*Bouea oppositifolia* (Roxb.) Meisne.) varieties using ISSR markers in Binh Minh district, Vinh Long Province- Viet Nam. *Can Tho University Journal of Science*, 54(1b), 50-60.
- Liu, J., Moller, M., Ga, L. M., Zhang, D. Q. & Li, D. Z.** (2011). DNA barcoding for the discrimination of Eurasian Yews (*Taxus L.*, Taxaceae) and the discovery of cryptic species. *Molecular Ecology Resources*, 11, 89-100.
- Liu, Z., Zeng, Z., Yang, D., Chu, G., Yuan, Z. and Chen, S.** (2012). Applying DNA barcodes for identification of plant species in the family Araliaceae. *Gene*, 499(1), 76-80.
- Mansour, H., Mekki, L. E. & Hussein, M. A.** (2014). Assessment of genetic diversity and relationships among Egyptian mango (*Mangifera indica* L.) cultivars grown in Suez Canal and Sinai region using RAPD markers. *Pakistan Journal of Biological Sciences*, 17(1), 56-61.
- Mishra, P., Kumar, A., Sivaraman, G., Shukla, A.K., Kaliamoorthy, R., Slater, A. & Velusamy, S.** (2017). Character-based DNA barcoding for authentication and conservation of IUCN Red listed threatened species of genus *Decalepis* (Apocynaceae). *Scientific Reports*, 7:14910.
- Muellner, A. N., Schaefer, H. & Lahaye, R.** (2011). Evaluation of candidate DNA barcoding loci for economically important timber species of the Mahogany family (Meliaceae). *Molecular Ecology Resources*, 11, 450-60.
- Smith, K. F., Thia, J., Gemmill, C. E. C., Cary, S. G. & Filder, A. E.** (2012). Barcoding of the cytochrome oxidase I (COI) indicates a recent introduction of *Ciona savignyi* into New Zealand and provides a rapid method for *Ciona* species discrimination. *Aquatic Invasions*, 7(3), 305-313.
- Techen, N., Parveen, I., Pan, Z. & Khan, I. A.** (2014). DNA barcoding of medicinal plant material for identification. *Current Opinion in Biotechnology*, 25, 103-110.
- Tran, N. & Do, T.** (2012). Genetic diversity of mango (*Mangifera* sp.) cultivars using molecular biological techniques. *Can Tho University Journal of Science*, 22(a), 175-185.
- VTO** (2016). Overview and situation of Vietnamese mango import and export (In Vietnamese). Available at: <http://vietnamtradeoffice.net/tong-quan-va-tinh-hinh-xuat-khau-xoai-viet-nam/>
- Wang, A., Gopurenko, D., Wu, H. & Lepschi, B.** (2016). Evaluation of six candidate DNA barcode loci for identification of five important invasive grasses in eastern Australia. *Plos One*, 12(4), e0175338.
- Yang, J. B., Wang, Y. P., Möller, M., Gao, L. M., & Wu, D.** (2012). Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae). *Molecular Ecology Resources*, 12, 267-75.
- Yao, H., Song, J. Y., Ma, X. Y., Liu, C., Li, Y., Xu, H. X., Han, J. P., Duan, L. S. & Chen, S. L.** (2009) Identification of *Dendrobium* species by a candidate DNA barcode sequence: the chloroplast psbA-trnH intergenic region. *Planta Medica*, 75, 667-669.
- Yu, H., Wu, K., Song, J., Zhu, Y., Yao, H., Luo, K., Dai, Y., Xu, S. & Lin, Y.** (2014). Expedient identification of Magnoliaceae species of DNA barcoding. *Plant Ocmics*, 7(1), 47-53.
- Zhang, D., Duan, L. & Zhou, N.** (2014). Application of DNA barcoding in *Roscoea* (Zingiberaceae) and a primary discussion on taxonomic status of *Roscoea cautleoides* var. *Pubescens*. *Biochemical Systematics and Ecology*, 52, 14-9.
- Zhang, D., Jiang, B., Duan, L. & Zhou, N.** (2016). Internal transcribed spacer (ITS), an ideal DNA barcode for species discrimination in *Crawfordia wall* (Gentianaceae). *African Journal of Traditional, Complementary and Alternative Medicines*, 13(6), 1010-1106.

Received: November, 22, 2019; Accepted: December, 30, 2019; Published: February, 28, 2021