Running title: Molecular Authentication of Commercially Cultivated Coffee

**Molecular Authentication of Commercially Cultivated Coffee (*Coffea* spp.) in the Philippines**

**Arleen Panaligan 1,2\*, Miriam Baltazar2,3 and Grecebio Jonathan Alejandro1,4**

*1The Graduate School, University of Santo Tomas, España, 1015 Manila, Philippines*

*2Department of Biological Sciences, Cavite State University, Indang, 4122 Cavite, Philippines*

*3National Coffee Research, Development and Extension Center, Cavite State University, Indang, 4122 Cavite, Philippines*

*4College of Science and Research Centre for the Natural and Applied Sciences, University of Santo Tomas, España, 1015 Manila, Philippines.*

\* For correspondence: acpanaligan@cvsu.edu.ph

*Received \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_; Accepted \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_; Published \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

**Novelty statement**

DNA barcoding using *ITS* and *matK* has not been used to molecularly authenticate commercially cultivated coffee. The results of this study show the potential of DNA barcodes to authenticate coffee species with overlapping characters as well as planting materials even at the seedling stage. This method of authentication will benefit the coffee industry and coffee growers for large-scale plantations.

**Abstract**

Accurate identification of commercially cultivated coffee species is necessary since the cup quality may be attributed to the kind of species. At seedling stage, it is difficult to discriminate *Coffea* species. Even at the matured stage, some morphological characters have been observed to be overlapping. To complement the traditional method of morphology-based identification, DNA barcoding using nuclear ribosomal *internal transcribe spacer* (*ITS*) and *maturase K* (*matK*) regions was performed. Both markers had 100% amplification and sequencing success rates. Although *ITS* had lower resolution in *Coffea* species, it efficiently discriminated *Coffea arabica* and *Coffea liberica.* The *matK* barcode discriminated all the species. Findings revealed that *matK* was an efficient barcode over *ITS* for commercially cultivated *Coffea* species by generating the highest rate of both universality and discrimination power. DNA barcoding as a method of authentication will benefit the coffee industry and coffee growers for large-scale plantations. © 2020 Friends Science Publishers

**Keywords:** *Coffea arabica*; *Coffea canephora; Coffea liberica;* DNA barcode; *ITS; matK*

**Introduction**

Coffee trees belong to the genus *Coffea* L. subgenus *Coffea* under Rubiaceae family and composed of 124 species (Davis *et al*. 2019). In the Philippines, there are three commercially cultivated species, namely *Coffea arabica* L.*, Coffea canephora* Pierre ex A. Froehner andand *Coffea liberica* W. Bull ex Hiern (Figure 1). Excelsa and liberica are varieties of *Coffea liberica*. Generally, the fruits of var. liberica are bigger, more tapered at the base, have thicker and more leathery pericarp than that of var. excelsa (Bridson 1988; Davis *et al*. 2006). The first two species are of high commercial value (Maluf *et al*. 2005; N’diaye *et al*. 2005, Davis *et al*. 2019). *Coffea arabica* is tetraploid (2n=4x=44) and self-fertile while the other two species are diploid (2n=22) and self-incompatible (Davis *et al*. 2006; Clarindo and Carvalho 2008).

Planting materials such as seedlings need accurate identification since the market value of coffee depends on the cup quality which may be attributed to the species. Consumers tend to choose *Coffea arabica* (locally known as Arabica) because of its better flavor and rich aroma. *Coffea canephora* (locally known as Robusta) is more bitter and has higher caffeine content than *Coffea arabica* (Lecolier *et al*. 2009) while *Coffea liberica* (locally known as Kapeng Barako) contains low caffeine (N’diaye *et al*. 2005) and has a strong flavor. Between var. excelsa and var. liberica, the latter is more bitter (Bridson 1988).

Morphology-based identification is the usual method of identifying plants including coffee. However, it is difficult to discriminate *Coffea* species at seedling stage. Generally, distinct characteristics of *Coffea* species can be observed at maturity but they still possess overlapping characters. Since phenotypes are highly affected by the environment, morphological characterization may lead to inconsistent data (Hebert *et al*. 2003). An efficient marker is then necessary to authenticate coffee planting materials. DNA barcoding is a molecular technique of using short, standardized DNA sequences to identify species in which bioinformatics plays an important role (Hebert *et al*. 2003). It can be used to identify plants irrespective to life stages (Ali *et al*. 2014). It also complements traditional taxonomy which is based on morphological characters (Hajibabaei *et al*. 2007).

The Consortium for the Barcode of Life (CBOL) Plant Working Group (2009) recommended *ribulose-bisphosphate carboxylase* (*rbcL*) and *maturase K* (*matK*) as core barcodes. Although *rbcL* is highly universal, it has a low discriminating power (CBOL Plant Working Group 2009). Between the two core barcodes, *matK* was used in this study because of its greater species discrimination even up to the species level (CBOL Plant Working Group 2009). The nuclear ribosomal *internal transcribe spacer* (*ITS*) was added to *matK* as suggested by the China Plant BOL Group (2011). The *ITS* and *matK,* non-coding and coding regions, respectively, were used to assess possible barcodes to authenticate commercially cultivated *Coffea* species in the Philippines. Hence, this study aimed to evaluate the PCR success rate, sequencing success rate, and discriminatory power of *ITS* and *matK* regions for commercially cultivated *Coffea* species.

 **Materials and methods**

***Sample collection and preservation***

Twenty-four (24) samples representing three commercially cultivated *Coffea* species, namely *Coffea arabica* (6 samples), *Coffea canephora* (9 samples) and *Coffea liberica* (9 samples) were collected from private and government coffee grower institutions in the Philippines. Leaf samples were stored in bags with silica gel. Herbarium vouchers were made and deposited at the University of Santo Tomas Herbarium (USTH). The coffee samples labelled by the coffee growers as ‘arabica’ were coded as A, ‘robusta’ as C, ‘liberica’ as L and ‘excelsa’ as LE.

***Amplification of the candidate DNA barcodes***

Silica gel dried leaves of each *Coffea* samples were crushed separately into fine powder and DNA samples were extracted following the protocol of Dneasy Plant Minikit (Qiagen, Hilden, Germany). Using universal primer pairs (Table 1), the *ITS* and *matK* regions were amplified with a total volume of 25 μL per reaction. The PCR mixture contained 19.45 μL water, 2.5 μL 10x reaction buffer, 0.5 μL 50 mM MgCl2, 0.4 μL 10 mM dNTP, 0.5 μL 10 μM forward and reverse primers, 0.15 μL 5u/ μL Taq DNA polymerase (Vivantis) and 1.0 μL DNA. The PCR amplification was performed using a T100 Thermal Cycler (Bio Rad) as follows: initial denaturation at 97°C for 90 s followed by 35 cycles of 95°C for 30 s, 55°C for 20 s (*ITS*) or 50°C for 20 s (*matK*), 72°C for 1 min, followed by final extension at 72°C for 10 min (Li *et al.* 2012).

One percent (1%) agarose solution containing Gel red stain (Vivantis) in 1x TAE buffer was used for electrophoresis. The sizes of the PCR products were determined using 100 bp plus ladder (Vivantis) and the gel images were obtained using Vilber Lourmat gel documentation system. The PCR products were purified using QIA-quick Purification Kit (Qiagen, Germany) and sent to Macrogen Inc., Seoul, South Korea for bidirectional DNA sequencing.

***Sequence analyses***

The consensus sequences were edited and assembled using Codon Code Aligner v.4.1.1 (Codon Code Co., Centerville, MA, USA). Basic Local Alignment Search Tool (BLAST) was used to gather homologous sequences from the National Center for Biotechnology Information (NCBI) GenBank Database and Multiple Sequence Alignment was performed using MEGA 7 (Kumar *et al*. 2016). Neighbor-Joining (NJ) was constructed in MEGA 7 using 1000 bootstrap replicates.

The discriminatory power of *ITS* and *matK* markers were evaluated using genetic distance-based method. Pairwise distances of these markers were computed using Kimura–2–parameter (K2P) model (Kimura 1980). This was performed using MEGA 7 software (Kumar *et al*. 2016) to determine the genetic variation within species (intraspecific genetic distance) and between species (interspecific genetic distance). The Wilcoxon two–sample test was performed using SPSS software (version 20.0; SPSS Inc., Chicago, USA) to evaluate if there is a significant difference between the interspecific and intraspecific divergences.

**Results**

***PCR success rate, sequencing success rate and discriminatory power***

Table 2 summarizes the *Coffea* samples used in this study and their accessions. A total of 48 sequences were newly generated in this study from *ITS* and *matK* markers (Table 2). The non-coding *ITS* and coding *matK* regions were 100% amplified and sequenced (Table 3). The sequence characteristics from multiple sequence alignment of the two candidate barcodes are presented in Table 3. Interspecific distances of *ITS* and *matK* were higher than their intraspecific distances. Barcode *matK* had higher discriminatory power than *ITS* (Figures 2 and 3).

**Discussion**

The universality of the DNA barcodes was assessed by getting the percentage of PCR success and sequencing success. A single dark band on agarose gel and obtaining a consensus sequence indicated successful PCR amplification and DNA sequencing.

Amplification and sequencing of *ITS* became a problem in studies on medicinal plants and Arecaceae (Chen *et al*. 2010; Yang *et al*. 2012). Difficulty in amplification and sequencing of *matK* was also observed in some studies (Chen *et al*. 2010; Dong *et al*. 2012). In this study, the non-coding *ITS* and coding *matK* regions were 100% amplified and sequenced, which means that these two barcodes are universal for *Coffea* species. The high universality of *ITS* was found congruent with the findings in Selaginellaceae and *Salvia* (Gu *et al*. 2013; Wang *et al*. 2013) and *matK* in Arecaceae, *Ardisia* and Philippine *Leea* (Yang *et al*. 2012; Liu *et al*. 2013; Cabelin and Alejandro, 2015).

The discriminatory power of *ITS* and *matK* was evaluated using Kimura two-parameter distance and tree-building methods. If the interspecific distance is significantly higher than its intraspecific distance, the discrimination is considered successful (CBOL Plant Working Group 2009). Using Wilcoxon two-sample test, the interspecific divergences of the candidate DNA barcodes were significantly higher than their intraspecific distances (*p* ˂ 0.001). This indicates that these candidate barcodes have a potential to discriminate *Coffea* at the species level.

 A species is considered resolved if all individuals of that species form a monophyletic clade (Kuzmina *et al*. 2012). The *ITS* region was considered efficient in discriminating species in Selaginellaceae and Gentianaceae (Gu *et al*. 2013; Zhang *et al*. 2016). The low discriminatory power of *ITS* has previously been reported in *Ardisia* (Liu *et al*. 2013). Out of the twenty-four coffee samples belonging to the three different species, two members of the *Coffea canephora* interclustered with *Coffea arabica* (Figure 2). However, not all species have to be distinguished for a DNA barcode to be considered effective (Lahaye *et al*. 2008). The ambiguity of the taxon will necessitate further analysis and checking of its identity. Some species would be better resolved by other DNA regions (Fazekas *et al*. 2008). In the present study, the *maturase K* region had higher percentage of resolved monophyletic taxa than *ITS* (Figure 3). The result was in agreement with other studies in sedges and fig cultivars (Starr *et al*. 2009; Castro *et al*. 2015). This result contradicts the report that *matK* had low species resolution in Gentianaceae (Zhang *et al*. 2016).

In the Philippines, there are four known coffee varieties and these are ‘arabica’, ‘robusta’, ‘liberica’, and ‘excelsa’ (Figure 1) corresponding to *Coffea arabica, Coffea canephora, Coffea liberica* and *Coffea excelsa,* respectively. However, there were reports that ‘liberica’ and ‘excelsa’ belong to a single species that is *Coffea liberica*. ‘Liberica’ refers to *Coffea liberica* var. liberica from West Africa and ‘excelsa’ to *Coffea liberica* var. dewevrei from Central Africa (Charrier and Berthaud 1985; Davis *et al*. 2006). The NJ trees for *ITS* and *matK* showed that the locally known Kapeng Barako which refers to ‘liberica’ and ‘excelsa’ grouped in a single clade (Figures 2 and 3), that may indicate that they are of the same species.

**Conclusion**

Although *ITS* had lower discriminatory power in *Coffea*, it efficiently discriminated *Coffea arabica* and *Coffea liberica*. Considering the ease of amplification and sequencing, *ITS* is a satisfactory DNA barcode for these two species. The *matK* region was able to discriminate the three *Coffea* species, indicating that this DNA barcode is efficient for authentication of commercially cultivated coffee at the species level.

**Acknowledgements**

The authors would like to thank Research Centre for Natural & Applied Sciences (RCNAS) of the University of Santo Tomas and Cavite State University for providing the laboratory facilities. The first author is grateful to DOST-PCIEERRD and NMNH, Paris, France for the sponsorship to training; Commission on Higher Education-Faculty Development Program (CHED-FDP) for the study grant; and DOST-PCAARRD for the dissertation grant.

**Author contributions**

AC conceptualized and designed the study, performed the experiment, analyzed and interpreted the results, wrote and revised the manuscript. MB and GJA edited and approved the manuscript.

**References**

Ali MA, G Gyulai, N Hidvegi, B Kerti, FM Al Hemaid, AK Pandey, J Lee (2014). The changing epitome of species identification–DNA barcoding. *Saudi J Biol Sci* 21: 204–231

Anthony F, MC Combes, C Astorga, B Bertrand (2002). The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theor Appl Genet* 104: 894–900

Bridson DM (1988). *Coffea*. Flora of tropical East Africa, Rubiaceae, pp:706–709. Rotterdam: Balkema

Cabelin VL, PJ Santor, GJD Alejandro (2015). Evaluation of DNA barcoding efficiency of cpDNA barcodes in selected Philippine *Leea* L. (Vitaceae).  *Acta Bot Gall: Bot Lett* 1–8

Castro C, A Hernandez, L Alvarado, D Flores (2015). DNA barcodes in fig cultivars (*Ficus carica L*.) using *ITS* regions of ribosomal DNA, the *psbA-trnH* spacer and the *matK* coding sequences. *Am J Plant Sci* 6: 95–102

CBOL Plant Working Group (2009). A DNA barcode for land plants. *PNAS USA* 6 (31): 12974–12797

Charrier A, J Berthaud (1985). Coffee: botany, biochemistry and production of beans and beverage, pp:13–47. In: Clifford MN, Willson KC, editors. London & Sydney: Croom Helm

Chen S, H Yao, J Han, C Liu, J Song, Y Shi Zhu, X Ma, T Gao, X Pang, K Luo, Y Li, X Li, X Jia, Y Lin, C Leon (2010). Validation of the *ITS2* region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5(1): 1–8

China Plant BOL Group (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (*ITS*) should be incorporated into the core barcode for seed plants. *PNAS* 108(49): 19641–19646

Clarindo WR, CR Carvalho (2008). First *Coffea arabica* karyogram showing that this species is a true allotetraploid. *Plant Syst Evol* 274: 237–241

Davis AP, R Govaerts, DM Bridson, P Stoffelen (2006). An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Bot J Linn Soc* 152: 465–512

Davis AP, H Chadburn, J Moat, R O’Sullivan, S Hargreaves, EN Lughadha (2019). High extinction risk for wild coffee species and implications for coffee sector sustainability. *Sci Adv* 5: 1–9

Dong W, J Liu, J Yu, L Wang, S Zhou (2012). Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One* 7 (4): 1–9

Fazekas AJ, KS Burgess, PR Kesanakurti, SW Graham, SG Newsmaster, BC Husband, DM Percy, M Hajibabaei, SC Barett (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* 3 (7): 1–12

Gu W, J Song , Y Cao Y, Q Sun Q, H Yao, Q Wu, J Chao, J Zhou, W Xue, J Duan (2013). Application of the *ITS2* region for barcoding medical plants of Selaginellaceae in Pteridophyta. *PLoS One* 8 (6): 1–8

Hajibabaei M, GA Singer, PD Hebert, DA Hickey (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet* 23 (4): 167–172

Hebert PD, A Cywinska, S Ball, JR De Waard (2003). Biological identifications through DNA barcodes. *R Soc* 313–321

Kathurima WC, GM Jenji, SM Muhoho, R Bouglanger, BM Gichimu, EK Gichuru (2012). Genetic diversity among commercial coffee varieties, advanced selections and museum collections in Kenya using molecular markers. *Int J Biodivers Conserv* 4 (2): 39–46

Kimura M (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16(2): 111–120

Korir NK, J Han, L Shangguan, C Wang, E Kayesh, Y Zhang, J Fang (2012). Plant variety and cultivar identification: advances and prospects. *Crit Rev Biotechnol* 1–15

Kumar S, G Stecher, K Tamura (2016). MEGA 7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *BMC Ecol* 33(7): 1870–1874

Kuzmina ML, KL Johnson, HR Barron, PD Hebert (2012). Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library. *BMC Ecol* 12(25): 1–11

Lahaye R, M Van Der Bank, D Bogarin, J Warner, F Pupulin, G Gigot, O Maurin, S Duthoit, TG Barraclough, V Savolainen (2008). DNA barcoding the floras of biodiversity hotspots. *PNAS* 105(8): 2923–2928

Lecolier A, P Besse, A Charrier, TN Tchakaloff, M Noirot (2009). Unraveling the origin of *Coffea arabica* ‘Buorbon pointy’ from La Reunion: a historical and scientific perspective. *Euphytica* 1–11

Li M, K Au, H Lam, L Cheng, R Jiang, P But, P Shaw (2012). Identification of Baiying (Herba Solani Lyrati) commodity and its toxic substitute Xungufeng (Herba Aristolochiae Mollissimae) using DNA barcoding and chemical profiling techniques. *Food Chem* 135: 1653–1658

Liu Y, K Wang, Z Liu, K Lou, S Chen, K Chen (2013). Identification of medical plants of 24 *Ardisia* species from China using the *matK* genetic marker. *Pharmacogn Mag* 9 (36): 331–337

Maluf MP, M Silvestrini, L Ruggiero, OG Filho, CA Colombo (2005). Genetic diversity of cultivated *Coffea arabica* inbred lines assesed by RAPD, AFLP, and SSR marker systems. *Sci Agric* 62(4): 366–373

N’diaye A, V Poncet, J Louarn, S Hamon, M Noirot (2005). Genetic differentiation between *Coffea liberica* var. liberica and *C. liberica* var. dewevrei and comparison with *C. canephora*. *Plant Syst Evol* 253: 95–104

Starr JR, RF Naczi, BN Chouinard (2009). Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Mol Ecol Resour* 9(1): 151–163

Wang M, H Zhao, L Wang, T Wang, R Yang, X Wang, Y Zhou, C Ding, L Zhang (2013). Potential use of DNA barcoding for the identification of *Salvia* based on cpDNA and nrDNA sequences. *Gene* 528(2): 206–215

White TJ, T Bruns, T, S Lee, J Taylor (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications, pp: 315–322. New York: Academic Press, Inc

Yang HQ, YR Dong, ZJ Gu, N Liang, JB Yang (2012). A preliminary assessment of *matK*, *rbcL* and *trnH-psbA* as DNA barcodes for *Calamus* (Arecaceae) species in China with a note on *ITS*. *Ann Bot Fenn* 49: 319–330

Zhang D, B Jiang, L Duan, N Zhou (2016). Internal transcribed spacer (ITS), an ideal DNA barcode for species discrimination in Crawfurdia Wall (Gentianaceae). *Afr J Tradit Complement Altern Med*13(16): 101–106



**Fig. 1:** Field photographs of *Coffea*: (A) *Coffea arabica* (arabica); (B) *Coffea canephora* (robusta);

(C) *Coffea liberica* (excelsa); (D) C*offea liberica* (liberica) (Photograph: A.C. Panaligan)



**Fig. 2:** Neighbor-joining tree inferred using Kimura two-parameter distances of *ITS.* Numbers on nodes are bootstrap support (BS) values. A = *Coffea arabica*; C = *Coffea canephora*; L = *Coffea liberica* (liberica); LE = Coffea liberica (excelsa)



**Fig. 3:** Neighbor-joining tree inferred using Kimura two-parameter distance of *matK.* Numbers on nodes are bootstrap support (BS) values. A = *Coffea arabica*; C = *Coffea canephora*; L = *Coffea liberica* (liberica); LE = Coffea liberica (excelsa)

**Table 1:** Universal primers of the two candidate barcodes

|  |  |  |  |
| --- | --- | --- | --- |
| Barcode | Primer  | Primer sequence (5’–3’) | Reference |
| *ITS* | *ITS5* | GGAAGTAAAAGTCGTAACAAGG | White *et al*. 1990 |
|  | *ITS4* | TCCTCCGCTTATTGATATGC |
| *matK* | *3F\_Kim f* | CGTACAGTACTTTTGTGTTTACGAG | CBOL-PWG 2009 |
|  | *1R\_Kim r* | ACCCAGTCCATCTGGAAATCTTGGTTC |

 *ITS = internal transcribe spacer; matK = maturase K*

**Table 2:** List of *Coffea* species used in the study and their accessions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Species | Place of origin/ collection place | Code | USTHaccession | GenBank accession  |
|  *ITS* |  *matK* |
| *C. arabica*  | Indang, Cavite | A1 | 014856 | MK611791 | MK722268 |
| *C. arabica*  | Indang, Cavite | A3 | 014857 | MK611792 | MK722267 |
| *C. arabica*  | Ampasit, La Trinidad, Benguet | A31-1 | 014858 | MK615726 | MK722270 |
| *C. arabica*  | Ampasit, La Trinidad, Benguet | A31-2 | 014859 | MK615727 | MK722269 |
| *C. arabica*  | BSU, La Trinidad, Benguet | A311-1 | 014860 | MK615728 | MK722266 |
| *C. arabica*  | BSU, La Trinidad, Benguet | A311-2 | 014861 | MK615729 | MK722265 |
| *C. arabica* | Mascarenes / Nicaragua – | – | DQ153609 | AB973195 |
| *C. canephora*  | Indang, Cavite | C1 | 014862 | MK615730 | MK722259 |
| *C. canephora*  | Indang, Cavite | C2 | 014863 | MK615731 | MK855097 |
| *C. canephora*  | Alfonso, Cavite | C6 | 014864 | MK615732 | MK722261 |
| *C. canephora*  | Alfonso, Cavite | C7 | 014865 | MK615733 | MK722260 |
| *C. canephora*  | Indang, Cavite | C8 | 014866 | MK615734 | MK722264 |
| *C. canephora*  | Indang, Cavite | C9 | 014867 | MK615735 | MK722263 |
| *C. canephora*  | Indang, Cavite | C10 | 014868 | MK615736 | MK722262 |
| *C. canephora*  | NOMIARC, Bukidnon | C512-2 | 014869 | MK615737 | MK722258 |
| *C. canephora*  | NOMIARC, Bukidnon | C513-2 | 014870 | MK615738 | MK855098 |
| C. canephora | Cameroon/ Vietnam |  – | – | DQ153593 | AB973198 |
| *C. canephora* | Mexico/ Indonesia |  – | – | MF417755 | AB973197 |
| *C. liberica*  | Indang, Cavite | L1 | 014871 | MK615739 | MK722250 |
| *C. liberica*  | Indang, Cavite | L3 | 014872 | MK615740 | MK722249 |
| *C. liberica*  | Alfonso, Cavite | L6 | 014873 | MK615741 | MK722253 |
| *C. liberica*  | Silang, Cavite | L7 | 014874 | MK615742 | MK722252 |
| *C. liberica*  | Alfonso, Cavite | L8 | 014875 | MK615743 | MK722251 |
| *C. liberica*  | Indang, Cavite | LE1 | 014876 | MK615744 | MK722255 |
| *C. liberica*  | Indang, Cavite | LE3 | 014877 | MK615745 | MK722254 |
| *C. liberica*  | Alfonso, Cavite | LE7 | 014878 | MK615746 | MK722257 |
| *C. liberica*  | Silang, Cavite | LE8 | 014879 | MK615747 | MK722256 |
| *C. liberica* var. dewevrei | Central African Republic – | – | DQ153603 |  |
|  |  |  |  |  |
| *C. liberica* var. liberica | Congo/ not indicated – | – | DQ153610 | AM412465 |

USTH = University of Santo Tomas Herbarium; *ITS = internal transcribe spacer; matK = maturase K*

Table 3: Sequence characteristics from multiple sequence alignment of the two candidate barcodes

|  |  |  |
| --- | --- | --- |
| Characteristics | *ITS* | *matK* |
| Number of samples | 24 | 24 |
| PCR success (%) | 100 | 100 |
| Sequencing success (%) | 100 | 100 |
| Aligned sequence length (bp)  | 711 | 846 |
| Mean Interspecific K2P distances | 0.024 ± 0.006 | 0.005 ± 0.002 |
| Mean Intraspecific K2P distances | 0.007 ± 0.002 | 0.002 ± 0.001 |

K2P = Kimura–2–parameter; *ITS = internal transcribe spacer; matK = maturase K*